

---

Wayne State University Dissertations


---

January 2021

## Global Coordination Between Gene Expression Processes In *C. Crescentus*

James Raymond Aretakis  
*Wayne State University*

Follow this and additional works at: [https://digitalcommons.wayne.edu/oa\\_dissertations](https://digitalcommons.wayne.edu/oa_dissertations)

 Part of the [Microbiology Commons](#), [Molecular Biology Commons](#), and the [Systems Biology Commons](#)

---

### Recommended Citation

Aretakis, James Raymond, "Global Coordination Between Gene Expression Processes In *C. Crescentus*" (2021). *Wayne State University Dissertations*. 3437.  
[https://digitalcommons.wayne.edu/oa\\_dissertations/3437](https://digitalcommons.wayne.edu/oa_dissertations/3437)

This Open Access Dissertation is brought to you for free and open access by DigitalCommons@WayneState. It has been accepted for inclusion in Wayne State University Dissertations by an authorized administrator of DigitalCommons@WayneState.

**GLOBAL COORDINATION BETWEEN GENE EXPRESSION PROCESSES IN  
*C. CRESCENTUS***

by

**JAMES RAYMOND ARETAKIS**

**DISSERTATION**

Submitted to Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

**DOCTOR OF PHILOSOPHY**

2021

MAJOR: BIOLOGICAL SCIENCES

Approved by:

---

Advisor

Date

---

---

---

## **DEDICATION**

To the special people in my life, my family and friends.

Figuratively, and literally, I wouldn't be here without each, and every one of you.

## ACKNOWLEDGEMENTS

The race that is getting your PhD is a marathon, not a sprint. When I was rotating through different research labs, I received the best advice of my PhD, “Make sure you and your mentor get along, you will be working with them for a long, long time”. Thankfully I can say that I found a mentor who I really enjoy and appreciate and which fostered a great working relationship. Dr. Jared Schrader was a great P.I. to work for during the first three years of my PhD, and absolutely amazing for the last two. There were many struggles, global and personal, during those final years and I always felt that Jared put what was best for me and the other lab members first, always... which is saying a lot during a global pandemic.

I want to thank my PhD committee; Dr. Penelope Higgs, Dr. Athar Ansari, and Dr. Tamara Hendrickson. I had the pleasure of getting to know Penelope during our group lab meetings, and not only did she provide fresh points of view on my research and my PhD career, I always enjoyed our conversations together. Dr. Ansari is one of the professors I casually interacted with the most at the department. I almost felt like a pseudo lab member, even without the background in yeast. Waiting for him to walk by, made eating lunch outside of his lab worth the walk down the hall. Finally, Tamara is someone that I usually met with annually, but her enthusiasm for science, and for providing feedback on my project, was something I looked forward to each and every committee meeting. Finally, I want to thank all the Higgs lab members, past and present. At first, we were the only bacteriologists in the building, and the myxo/caulobacter bond runs deep.

Coming into lab every day can either be a trial or a joy, and the Schrader lab became a special place because of the people in it. Each and every one added so much to my PhD, outside of just the science, that it can't even begin to put into words how each and every experience got me through to where I am today. I do want to especially thank those who were in the lab when I started. Dr. Nadra Al-Husini taught me as much as I could absorb and made sure that us new graduate students didn't get lost in that big, new, *Caulobacter* lab. Mohammed Husain Bharmal,

was the other graduate student who started at the same time in Schrader Lab as I, and we weathered many a conference, and meeting together. I have spent more time with you these last five years, than anyone. I am eternally grateful.

Lastly, I want to thank all of my friends and family. I wouldn't still be talking in sentences, let alone typing my PhD thesis if it wasn't for the late-night phone calls, visits, 6<sup>th</sup> floor "get togethers", and being there when I needed to vent about my (frequent) terrible, horrible, no good, very bad days. Life is about finding what is important, and at the end of the day it is you guys. Thank you, Thank you, Thank you.

## TABLE OF CONTENTS

Dedication .....	ii
Acknowledgements .....	iii
List of Figures .....	vii
List of Tables .....	viii
Chapter 1 INTRODUCTION .....	1
Global coordination between gene expression processes in <i>C. Crescentus</i> .....	1
Chapter 2 Absolute Measurements Of mRNA Translation In <i>C. Crescentus</i> Reveal Important Fitness Costs Of Vitamin B12 Scavenging .....	4
Summary .....	4
Introduction .....	4
Results .....	6
Absolute quantitation of mRNA translation rates .....	6
Global analysis of <i>C. crescentus</i> absolute mRNA translation levels .....	9
Analysis of vitamin B12 and methionine metabolism.....	13
Discussion .....	18
Materials and Methods .....	21
Acknowledgments .....	29
Supporting information .....	29
Chapter 3 Bacterial mRNA Decay Is Globally Coordinated with Transcription And Translation. 37	
Introduction .....	37
Results .....	39
Global measurement of <i>C. crescentus</i> mRNA half-lives by Rif-Seq .....	39
Processed noncoding RNAs are generated from diverse locations.....	42
Co-transcriptional mRNA decay occurs on long mRNAs .....	44

Steady state mRNA levels are determined by a combination of transcription and decay.....	42
5' P sites are likely generated by RNase E and occur predominantly in the CDS	47
Translation affects mRNA decay across the transcriptome.....	50
Discussion .....	53
Materials and Methods .....	56
Appendix 4 Methodology for ribosome profiling of key stages of the <i>Caulobacter crescentus</i> cell cycle .	62
Introduction .....	62
Caulobacter Cell Cycle Ribosome Profiling Methodology.....	63
Conclusion .....	86
Acknowledgments .....	87
Supporting information .....	87
Appendix 5 Rif-Correct – Bacterial mRNA Half-Life Estimation .....	89
Summary .....	89
Introduction .....	89
Implementation.....	91
Inputs.....	91
Data Processing.....	93
Output.....	94
Results/Discussion.....	94
Conclusion .....	96
References .....	97
Abstract .....	110
Autobiographical statement .....	112

## LIST OF FIGURES

Figure 2.1: Absolute quantitation of <i>C. crescentus</i> protein synthesis by ribosome profiling.....	8
Figure 2.2: Global analysis of <i>C. crescentus</i> protein synthesis.....	12
Figure 2.3: <i>C. crescentus</i> cells are starved for B12 in laboratory growth media.....	14
Figure 2.4: Excess protein synthesis rates for methionine biosynthetic genes correlates with fitness cost .....	17
Figure S2.1: Absolute protein levels of unsynchronized cells in PYE medium.....	34
Figure S2.2: PYE fractional protein synthesis Proteomap .....	35
Figure S2.3: Non-linear curve fit for B <sub>12</sub> dependent riboswitch repression .....	36
Figure 3.1: Global measurement of <i>C. crescentus</i> mRNA half-lives by Rif-Seq .....	40
Figure 3.2: Processed noncoding RNAs are generated from diverse locations .....	43
Figure 3.3: Co-transcriptional mRNA decay occurs on long mRNAs .....	45
Figure 3.4: Steady state mRNA levels are determined by a combination of transcription and decay .....	46
Figure 3.5: 5' P sites are likely generated by RNase E and occur predominantly in the CDS. ...	49
Figure 3.6: Translation affects mRNA decay across the transcriptome .....	51
Figure 4.1: Schematic of <i>Caulobacter</i> ribosome profiling procedure .....	64
Figure 4.2: Example of <i>Caulobacter</i> cell cycle polysomes.....	70
Figure 4.3: rRNA depletion of total RNA samples for total RNA-seq .....	77
Figure 5.1: Using Rif-Seq to determine mRNA half-lives .....	90
Figure 5.2: Rif-Correct software workflow .....	92
Figure 5.3: <i>Rif-Correct Software Fit</i> .....	95



## LIST OF TABLES

Table S2.1: Absolute measurements of mRNA translation for M2G and PYE .....	29
Table S2.2: Average molecules of protein translated per cell and average proteins per cell.....	30
Table S2.3: Fraction of total protein synthesis for each KEGG category.....	31
Table S2.4: Comparison of absolute translation level to DNA binding sites.....	31
Table S2.5: Doubling time measurements at different concentrations of cyanocobalamin.....	32
Table S2.6: List of bacterial strains .....	32
Table S2.7: Table of accession numbers and organism names from figure 4.....	33
Table 4.1: Expected completion time for each section .....	87
Table 4.2: Oligos for library prep generation.....	87

## CHAPTER 1 – INTRODUCTION

### 1.1 – Global coordination between gene expression processes in *C. Crescentus*

The rise of biotechnology is one of the defining advances of our time by transforming the way we address public health and environmental concerns (Sahin, Kariko, and Tureci 2014). Advances in health and medicine rely on a deeper understanding of the most fundamental biological components and processes, such as cells, mRNA molecules, and gene expression. An ultimate goal is to understand how human cells work, but starting in simpler systems, like bacteria, hastens progress and furthers our overall knowledge. Additionally, understanding how these bacterial systems are organized and regulated is necessary for understanding how they may be controlled for advances in areas like synthetic biology (Chen et al. 2020), and for discovery of new antibiotics (Martens and Demain 2017).

The overall goal of my thesis is to investigate/understand how bacterial systems coordinate the steps of gene expression. Gene expression is an essential process for all organisms, which requires coordination between different sub-steps. These sub-steps include mRNA transcription, translation, and mRNA decay. In order for bacterial cells to grow and divide successfully, these processes need to be rapid and efficient (Maitra and Dill 2015). For bacteria to have high fitness, thousands of different proteins must be optimally synthesized, at the correct time, and in the correct amounts. Although transcription has been the most studied step in gene expression thus far, post-transcriptional processes are necessary for fine-tuning the amount and type of proteins produced. A systems biology approach can therefore be used to understand the coordinated processes from a large set of dynamic interactions (Bruggeman and Westerhoff 2007). Also, these sub-steps are interconnected, making them difficult to study individually (Deana and Belasco 2005). To understand the regulation of translation and mRNA decay, we produced large genome-wide RNA-sequencing datasets in *Caulobacter crescentus*, to help us investigate these processes in a systematic and holistic way.

To begin investigating genome-wide translation levels in *C. crescentus*, I detailed the methods for performing ribosome profiling, which measures gene by gene, global translation rates (Ingolia et al. 2009). This methodology article is published in *Methods in Enzymology* (Aretakis, Al-Husini, and Schrader 2018) and is found in the first appendix. To investigate mRNA decay, genome-wide mRNA half-lives can be determined using a modified RNA-seq with Rifampicin assay (Rif-Seq) (Bernstein et al. 2002; Chen et al. 2015). Here, I wrote a freely available software to correct and calculate Rif-Seq mRNA half-lives in *C. crescentus*. The second appendix contains the detailed methods article for this software that has been submitted for publication. Collectively, the large datasets for these linked post-transcriptional processes (translation and mRNA decay) allows us to simultaneously investigate their individual rates, as well as the relationship between them.

In my first thesis chapter, I investigate the rate and energy expenditure of *C. crescentus* translation, the most energy expensive step in gene expression. Cells in the log phase of growth use half of their energy on actively translating ribosomes (Russell and Cook 1995). Since gene expression is so energetically costly, the correct number of ribosomes must be allocated to the correct mRNAs to maximize the efficiency of this precious resource in the bacteria (Molenaar et al. 2009). An example of this can be seen when proteins with different protein complex stoichiometries are produced at different ratios from the same operon (Li et al. 2014). From one polycistronic mRNA, ribosomes must make individual proteins in differing amounts. Quantifying translation rates allows for a direct investigation of how the cell is spending its energy, and therefore what is necessary for high cell fitness. Ribosome profiling allows for this genome-wide direct reporting of translation rates that allows for quantification and investigation into this process.

In the second chapter, I examine the post-transcriptional process of mRNA decay from a systematic perspective, using Rif-Seq. Rif-Seq is a technique that allows for transcription to be shut-off, and mRNA decay to be measured. This has been used by other groups to study mRNA

half-lives in organisms such as *Escherichia coli* (Bernstein et al. 2002; Chen et al. 2015). This technique also allows for the discovery of stable non-coding RNA, quantification of co-transcriptional degradation, and the different post-transcriptional expression strategies that prioritize efficiency or regulation. Also, I analyzed how translation rates influence mRNA degradation and vice versa. It has been observed that there is an inverse relationship between mRNA degradation and translation; for example, when mRNA half-lives are short there is less translation of that corresponding gene (Deana and Belasco 2005). In eukaryotic systems, translation has also been found to influence mRNA half-lives across the transcriptome. Some groups have found that translation elongation is the influencing step (Presnyak et al. 2015), while others have found translation initiation to be the key step (Chan et al. 2018). Although not all of these questions are answered in this thesis, our understanding of the interplay between translation and mRNA decay has expanded, and the ground work is set for future investigations into the importance of specific translation sub-steps.

This thesis comprises the beginning research into the quantification and exploration of not only mRNA decay and translation, but also the feedback and interplay between these two important gene expression sub-steps. Understanding the basic properties of how gene expression is regulated sets the foundation for greater insights, and biotechnology utility. There are as many bacterial cells in the human body as “human” cells (Sender, Fuchs, and Milo 2016), and the bacterial cells in the environment all around us hold potential tools waiting to be discovered (Deveau et al. 2008). If we lift our perspective to more complicated systems, our ability to understand the complexities rests, in part, with starting to dissect simpler systems that share their attributes (Al-Husini et al. 2018). Although this thesis does not completely explain mRNA decay and translation in *C. crescentus*, it sets the ground work for further illuminating these sub-steps and gene expression as a whole.

## CHAPTER 2 – ABSOLUTE MEASUREMENTS OF MRNA TRANSLATION IN *C. CRESCENTUS* REVEAL IMPORTANT FITNESS COSTS OF VITAMIN B12 SCAVENGING

This chapter has been published:

Aretakis, J. R., A. Gega, and J. M. Schrader. 2019. 'Absolute Measurements of mRNA Translation in *Caulobacter crescentus* Reveal Important Fitness Costs of Vitamin B12 Scavenging', *mSystems*, 4.

Jared M. Schrader and I wrote the paper, and gathered the data. Alisa Gega cloned strains from gBocks.

### 2.1 – Summary

*Caulobacter crescentus* is a model for the bacterial cell cycle which culminates in asymmetric cell division, yet little is known about the absolute levels of protein synthesis of the cellular parts needed to complete the cell cycle. Here we utilize ribosome profiling to provide absolute measurements of mRNA translation in *C. crescentus*, providing an important resource with quantitative genome-wide measurements of protein output across individual genes. Analysis of protein synthesis rates revealed ~4.5% of cellular protein synthesis are for genes related to vitamin B<sub>12</sub> import (*btuB*) and B<sub>12</sub> independent methionine biosynthesis (*metE*) when grown in common growth media lacking B<sub>12</sub>. While its facultative B<sub>12</sub> lifestyle provides a fitness advantage in the absence of B<sub>12</sub>, we find that it provides lower fitness of the cells in the presence of B<sub>12</sub>, potentially explaining why many *Caulobacter* species have lost the *metE* gene and become obligates for B<sub>12</sub>.

### 2.2 – Introduction

In bacterial systems biology, global mRNA translation measurements are critical for understanding how cells utilize their resources to achieve their evolutionarily selected cell growth and division cycles. To complete the bacterial cell cycle, the protein parts encoded within the

genome must be transcribed into mRNAs that are translated into the appropriate number of proteins for the daughter cells to be generated. Genome-wide absolute quantitation of protein level measurements have allowed the monitoring of protein resource allocation (Hui et al. 2015; Li et al. 2014) revealing that these cells allocate resources for optimal growth. As the ribosome content is positively correlated with the growth rate (Bremer and Dennis 2008), cells must optimize the fraction of protein synthesis needed to make new ribosomes (enzymes that make proteins) vs the protein synthesis needed to produce the proteomes of the daughter cells to achieve fast generation times (Li et al. 2014; Scott et al. 2014). Optimality has also been observed at the protein-complex level, as translation of a stoichiometric amount of protein subunits to the overall multi-protein complexes has been observed (Li et al. 2014), with different post-transcriptional strategies across species utilized to achieve the optimal protein concentration (Lalanne et al. 2018). Therefore, to understand the mechanisms controlling the growth and division cycles of diverse bacteria, we must understand how bacteria are able to optimize their protein synthesis resources for maximal fitness.

*Caulobacter crescentus* is an oligotrophic  $\alpha$ -proteobacteria with a carefully orchestrated cell cycle yielding asymmetric cell division and a model organism for the study of the bacterial cell cycle (Lasker, Mann, and Shapiro 2016; Collier 2016). In *C. crescentus*, cells undergo changes in gene expression of ~20% of their entire genome during the process of the cell cycle (Schrader et al. 2016; Laub et al. 2000). Timing of 57% of the cell cycle-regulated mRNAs are controlled at the transcription level by a master regulatory circuit that is composed of 4 transcription factors (DnaA, GcrA, CtrA, and SciP) and a DNA methylase (CcrM) (Lasker, Mann, and Shapiro 2016; Zhou et al. 2015) and 49% of those cell cycle-regulated mRNAs are additionally regulated at the level of mRNA translation (Schrader et al. 2016). Importantly, global *C. crescentus* studies have focused solely on the control of the timing of gene expression in the cell cycle and thus little is

known about the absolute levels of protein synthesis, or how the protein synthesis resources are allocated across the proteome.

Here, we utilize ribosome profiling to achieve a quantitative genome-wide absolute measure of protein synthesis in *C. crescentus*. This resource provides the absolute protein synthesis rate of each protein expressed from the *C. crescentus* genome and a global map of protein synthesis resource allocation. Absolute levels of mRNA translation of cell cycle master-regulators showed higher levels of mRNA translation compared to their known DNA binding sites for all but CcrM, and a relatively low level of mRNA translation of CtrA regulatory proteins relative to the CtrA master regulator itself. PopZ, a polar protein scaffold that recruits asymmetric cell fate specification proteins (Berge and Viollier 2018), is at a limiting concentration compared to its client proteins suggesting that these clients compete for access to the cell pole. Surprisingly, we discovered that the *btuB* vitamin B<sub>12</sub> importer gene and the *metE* methionine biosynthetic gene were among the most highly translated genes in the absence of B<sub>12</sub>, showing that *C. crescentus*' B<sub>12</sub> scavenging pathway requires a surprisingly large amount of the cell's protein synthesis resources. The high cost of protein synthesis of the B<sub>12</sub> scavenging pathway is reduced in the presence of B<sub>12</sub> by riboswitches in the 5' UTR of these two genes. The widely utilized lab strain, NA1000, is a facultative B<sub>12</sub> scavenger due to the *metE* gene which produces methionine in the absence of B<sub>12</sub>, yet many natural *Caulobacter* isolates are obligate B<sub>12</sub> scavengers (Poindexter 1964). We show that while the facultative B<sub>12</sub> scavenging lifestyle generates a fitness tradeoff, where in the absence of B<sub>12</sub> there is a positive fitness advantage from MetE's B<sub>12</sub>-independent methionine production, while in B<sub>12</sub>'s presence there is a fitness disadvantage due to the wasted cost of MetE's protein synthesis, providing an explanation for why many isolates have lost the *metE* gene to become obligates for B<sub>12</sub>.

## 2.3 – Results

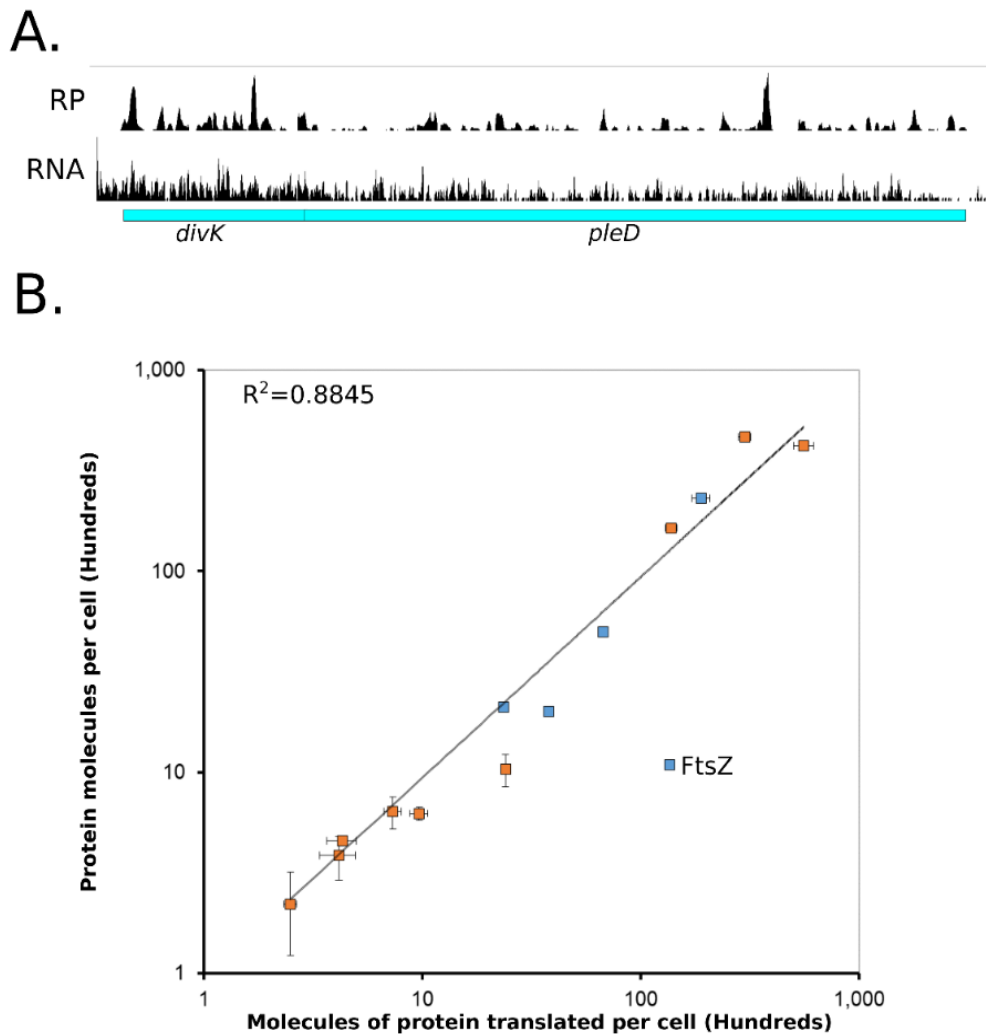
### 2.3.1 – Absolute quantitation of mRNA translation rates

Ribosome profiling provides a global direct measure of the protein synthesis rate by sequencing ribosome protected mRNA footprints (Li et al. 2014; Oh et al. 2011; Ingolia et al. 2009). To determine absolute rates of translation in *C. crescentus*, ribosome profiling was performed in unsynchronized *C. crescentus* NA1000 cells grown in M2G minimal medium. In fast growing bacteria where the rate of translation is the main driving force of protein levels, protein degradation can be negligible and therefore the main driving force of protein levels is mRNA translation (Li et al. 2014). This is largely true in *C. crescentus*, as >95% of proteins were found to have half-lives longer than the cell cycle (Grunenfelder et al. 2001). First, we examined the ribosome footprint density along each ORF on mRNAs as a relative measure for translation (Fig 1A). For example, in the *divK/pleD* polycistronic operon, we find that *divK* has 2.0 fold higher ribosome density as compared to *pleD* (Fig 1A). For absolute quantitation, it is assumed that the average elongation rate is constant for each mRNA which would allow the average ribosome density to be directly proportional to the rate of protein synthesis of each ORF (Li et al. 2014). It is also assumed that all ribosomes will finish translation and make the full-length protein (Li et al. 2014). Next, to reduce the impact of fast and slow-moving ribosomes in the ribosome occupancy profiles along ORFs on the quantitative level of translation, we used winsorization to correct the average ribosome footprint density of each ORF (Table S1). Start codon and stop codon regions were omitted from the analysis to avoid biases in slow moving ribosomes that are initiating or terminating (Oh et al. 2011; Aretakis, Al-Husini, and Schrader 2018).

To convert ribosome density to absolute mRNA translation rates we measured the average protein mass of *C. crescentus* cells which was multiplied by the fractional ribosome density measurement of each gene and divided by the molecular weight (equation 1.  $k_i = \frac{\phi_i P}{mW_i}$ ) (Li et al. 2014). This measure of the average number of proteins translated per cell correlated



well between the protein concentrations reported in the literature as well as protein concentration measurements reported here using YFP intensity of C-terminally tagged gene-fusions (Fig 1B, Fig S1, Table S2). As expected, FtsZ, the key cell division protein which is known to be a substrate of cell cycle dependent proteolysis, (Kelly et al. 1998) has a 13-fold higher amount of translated protein than protein observed in the cell, while stable proteins ranged between 0.65 to



**Figure 2.1 – Absolute quantitation of *C. crescentus* protein synthesis by ribosome profiling.** A.) Ribosome profiling data for cells grown in M2G media of the *divK/pleD* operon. Average ribosome density of *divK* is 2.0 times higher than for *pleD*. mRNA data from (Schrader et al. 2014). B.) Absolute protein levels of unsynchronized cells measured by western blot (blue) or YFP fusions (orange) compared to the absolute molecules of protein translated per cell calculated by ribosome profiling. Vertical error bars indicate the standard deviation in YFP intensity or standard deviation for the western blots, while horizontal error bars indicate the standard deviation from ribosome profiling replicates ( $n=3$ ). FtsZ is expected to deviate from the line since its protein levels are under proteolytic control (Kelly et al. 1998). Data in Tables S2.1 and S2.2.

2.3-fold. The same phenomenon of higher translation levels than protein levels was also observed for the proteolyzed cell cycle regulators DnaA and CcrM, and upon deletion of the Lon protease which is known to be responsible for their proteolysis, the correlation was restored (Peter Chien personal communication, Fig S1) (Wright et al. 1996; Jonas et al. 2013). These data suggest that the absolute measures of mRNA translation are reflecting the absolute protein synthesis rate for each ORF in *C. crescentus* and provide a reasonable measure of steady state protein levels for stable proteins (data can be found in Table S1).

### **2.3.2 – Global analysis of *C. crescentus* absolute mRNA translation levels**

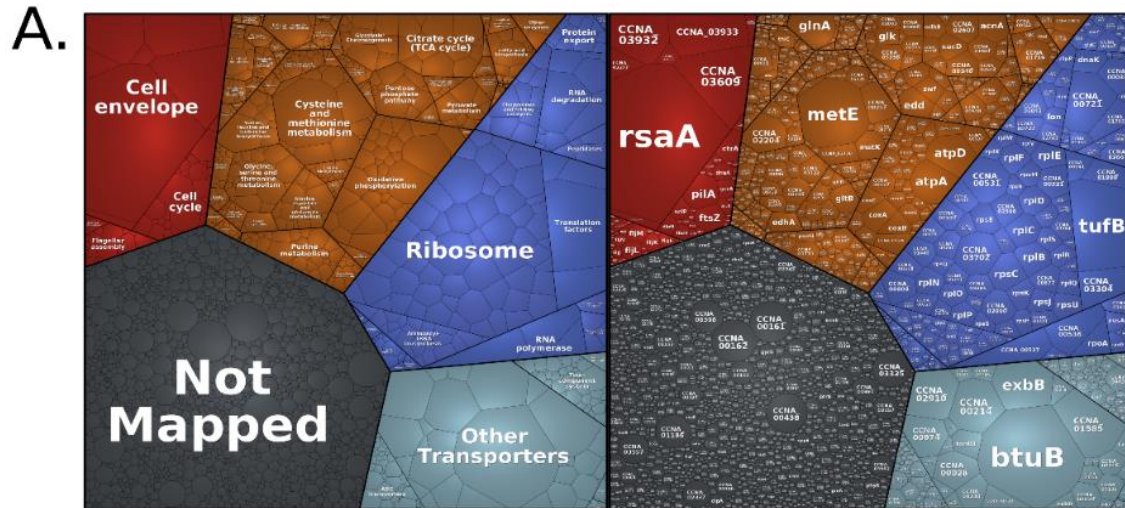
*C. crescentus* cells dedicate a significant percentage of their protein synthesis to several major cellular processes associated with cell growth (Fig 2A, Fig S2). Across major “KEGG categories” we analyzed the percentage of ribosome footprints to understand their allocation of protein synthesis capacity (Table S3). Nutrient transporters (11.8%), the ribosome (11.2%), and the cell envelope (9.8%) represent the largest classes of protein production for *C. crescentus*. A significant fraction of protein synthesis capacity (25.5%) is allocated to produce proteins of unknown function, showing a significant fraction of the cell’s protein synthesis capacity is not understood. By comparing the fraction of the translation KEGG category across minimal media (M2G) (15.9%) and a richer complex media (PYE) (22.8%), we find the cells dedicate a larger amount of protein synthesis capacity to making translational machinery in rich media similar to *E. coli* (Bremer and Dennis 2008). Additionally, we see that a larger amount of protein synthesis capacity in “cell growth and death” is observed in M2G (11.3%) compared to PYE (6.45%), owing largely to increased protein synthesis capacity of an operon of cell-contact dependent toxins and immunity proteins that are known to be expressed in stationary phase (Fig S2) (Garcia-Bayona, Guo, and Laub 2017).

Interestingly, while the cell cycle is a major area of study in *C. crescentus*, the cell cycle genes only represent a small fraction of protein synthesis capacity (1.78%), yet these genes play

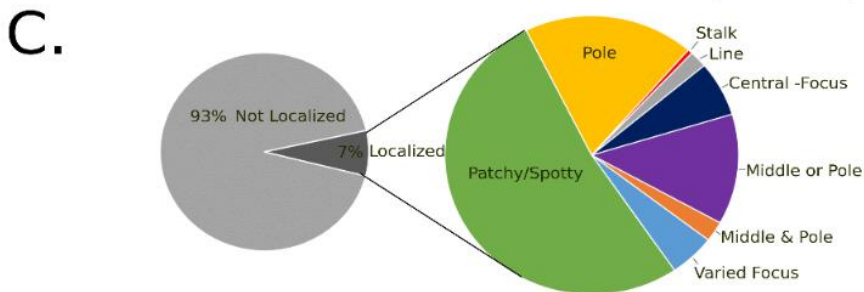
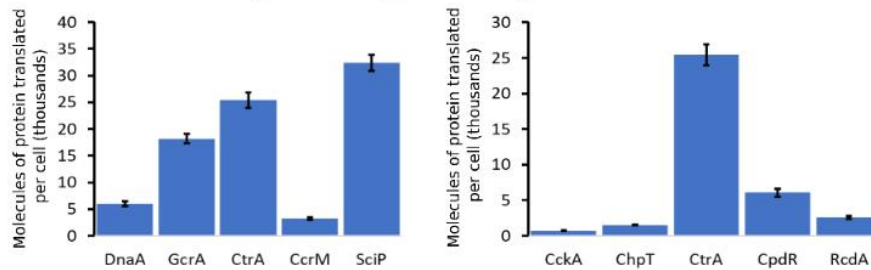
a critical role in shaping the growth and division cycles. The cell cycle-regulatory circuit itself is composed of four transcription factors (DnaA, GcrA, CtrA, and SciP) and a DNA methylase (CcrM) whose spatiotemporal activation facilitate cell cycle progression (Lasker, Mann, and Shapiro 2016). For master-regulator proteins, the cell produces between ~3,000-30,000 copies of each protein which corresponds to between 71-648 fold more proteins than the number of known DNA binding sites that they control (Zhou et al. 2015), with the exception of CcrM (Table S4). 3280 CcrM proteins are translated to methylate the 4542 GANTC sites per chromosome (Fig 2B). While the number of CcrM proteins are approximately 1/3 the number of GANTC sites present after DNA replication, CcrM is a processive enzyme (Berdis et al. 1998), suggesting that each CcrM may on average methylate ~3 GANTC sites. While the number of CtrA proteins translated (25,400 proteins) corresponds closely with the amount measured in predivisional cells (18,000-22,000 (Spencer et al. 2009; Judd et al. 2003)), CtrA is produced at a significantly higher level than its collection of regulatory kinases, phosphotransferases, and proteolytic adapters that control its cell cycle-dependent activity (Fig 2B). GcrA interacts with the RNA-polymerase/ $\sigma^{70}$  complex to activate transcription of target promoters (Haakonsen, Yuan, and Laub 2015) where a ~4-fold excess of GcrA is produced over  $\sigma^{70}$ , suggesting that excess GcrA may accelerate binding to the RNA-polymerase holoenzyme (Wu et al. 2018) to facilitate subsequent recruitment of  $\sigma^{70}$ .

As many proteins were found to have distinct subcellular patterns of protein accumulation in *C. crescentus* as determined in (Werner et al. 2009), we compared protein synthesis capacity to the localization patterns of proteins observed in this dataset (Fig 2C). 7% of protein synthesis occurs for “localized proteins” in *C. crescentus*. Of those localized proteins, most are “patchy/spotty”, while a significant fraction has a subcellular address where the protein accumulates (pole, stalk, or center) (Fig 2C). Many proteins are specifically required to form asymmetric polar protein complexes that function to determine cell fate upon division (Lasker, Mann, and Shapiro 2016). Many of these polarly localized proteins are recruited to the cell pole

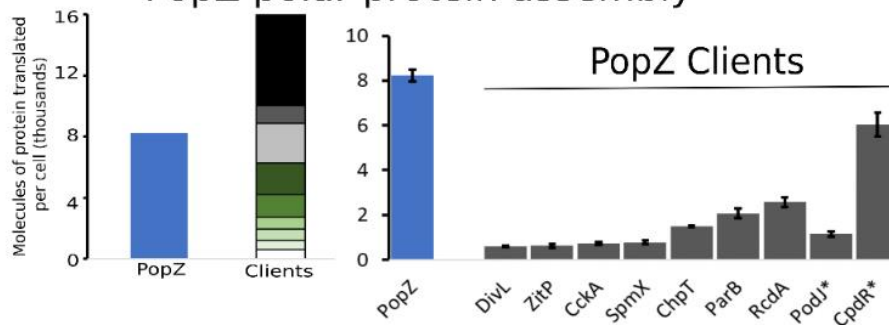
through the multimeric hub protein PopZ (Holmes et al. 2016; Berge et al. 2016; Zhao et al. 2018). Interestingly, by examining PopZ and its known client proteins, we find that PopZ is made in limiting amounts (Fig 2D), suggesting that the clients compete for PopZ binding *in vivo*.



**B.** Cell cycle-regulatory network



**D.** PopZ polar protein assembly

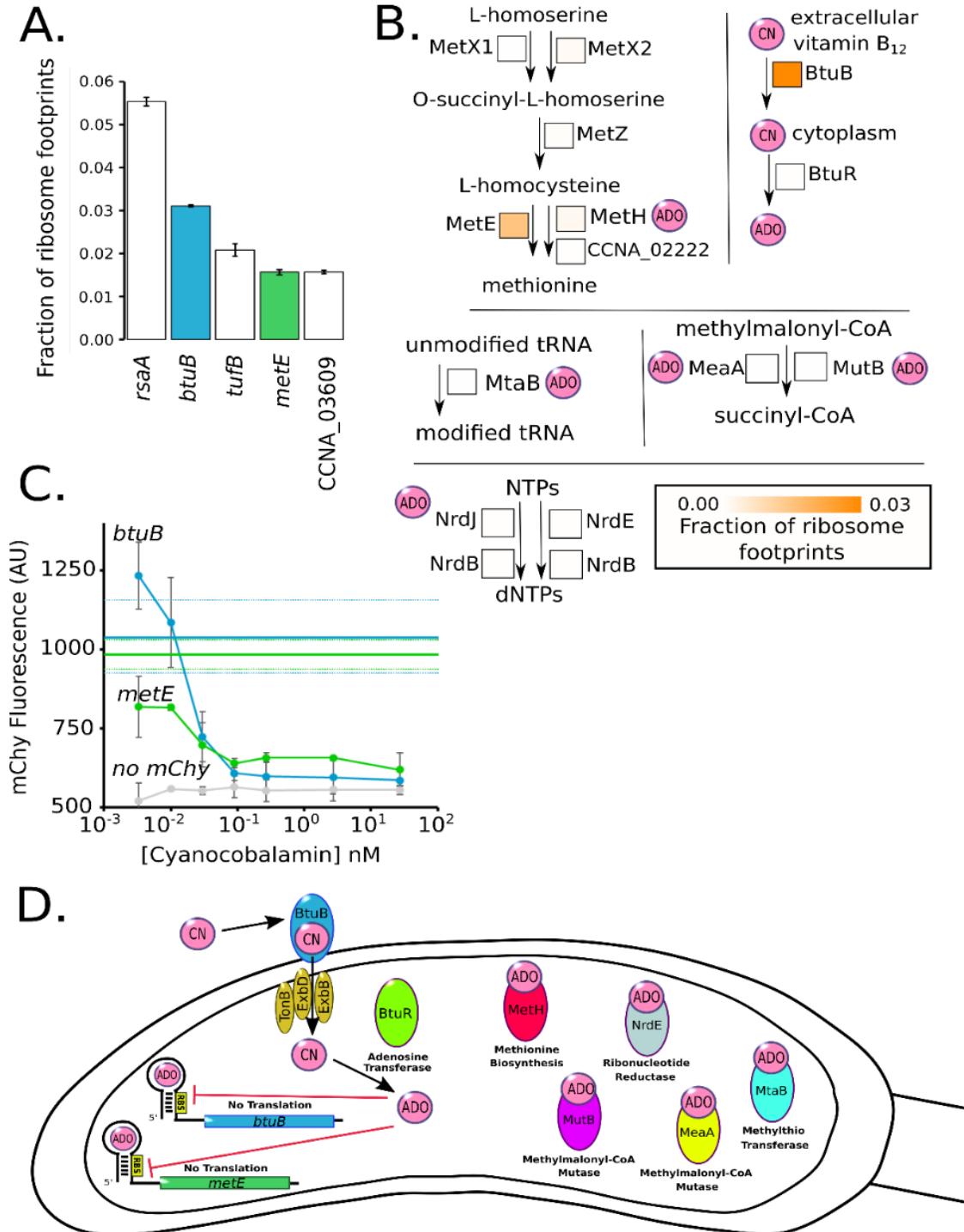


**Figure 2.2 – Global analysis of *C. crescentus* protein synthesis.** A.) Proteomap with each polygon representing a single gene with area scaled to the fraction of ribosome protected mRNA footprints measured. Red is cellular processes, orange is metabolism, blue is genetic information processing, light blue is environmental information processing, and grey are genes of unknown function. B.) Molecules of protein translated per cell for the cell cycle master-regulators (left) and CtrA regulatory network (right). C.) Left, fraction of ribosome protected mRNA footprints encoding localized (dark grey) or non-localized proteins (light grey). Right, zoomed in analysis of the fraction of proteins with different subcellular localization patterns based on (Werner et al. 2009). D.) Polar protein competition. Left, molecules of protein translated per cell for the polar protein scaffold PopZ and its known clients (Holmes et al. 2016; Zhao et al. 2018; Berge et al. 2016). Proteins with known proteolysis are highlighted with asterisks.

### 2.3.3 – Analysis of vitamin B<sub>12</sub> and methionine metabolism

Analysis of the most highly translated proteins found that RsaA, the surface layer protein, was the most highly translated protein in the cell (Fig 3A) (Lau, Nomellini, and Smit 2010). Elongation factor Tu was the 3<sup>rd</sup> most abundant cytoplasmic protein owing to its requirement to deliver aminoacyl-tRNAs to the ribosome during translation (Krab and Parmeggiani 1998). Surprisingly, we also observed that the homolog of the B<sub>12</sub> importer (*btuB* 2<sup>nd</sup> highest) and methionine biosynthetic gene (*metE* 4<sup>th</sup> highest) were among the most highly translated proteins. Vitamin B<sub>12</sub> is an important enzymatic cofactor that in *C. crescentus* is used for the biosynthesis of methionine, dNTP production, tRNA modification, and isomerization of methylmalonyl-CoA to succinyl-CoA (Fig 3) (Zhang et al. 2009). *C. crescentus* cannot synthesize B<sub>12</sub> *de novo* but can import it through the BtuB protein (Menikpurage et al. 2019). In the cytoplasm, both MetE and MetH perform the rate-limiting step of methionine biosynthesis, where MetH requires B<sub>12</sub> but has a higher specific activity than MetE (Fig 3B) (Whitfield, Steers, and Weissbach 1970; Frasca et al. 1988). Of note, both BtuB and MetE are translated at much higher levels than the other components related to methionine biosynthesis (Fig 3B). Both the *btuB* and *metE* genes are the only two genes in *C. crescentus* encoded with B<sub>12</sub> riboswitches in their 5' UTRs (Schrader et al. 2014). We tested the function of these riboswitches by creating 5' UTR fusions to the mCherry

gene driven by the vanillate promoter and subjecting the cells to various concentrations of B<sub>12</sub> in the form of cyanocobalamin (Fig 3C). Both the *metE* and *btuB* 5' UTR reporters showed high translation in the absence of B<sub>12</sub> and exhibited a B<sub>12</sub> concentration dependent translational shutoff (Fig 3C). The *metE* riboswitch appears to be more sensitive to B<sub>12</sub> concentration, with



**Figure 2.3 – *C. crescentus* cells are starved for B<sub>12</sub> in laboratory growth media.** A.) Fraction of ribosome footprints for the most highly translated mRNAs in M2G. B<sub>12</sub> related genes are colored. B.) Pathway of methionine biosynthesis, MetX1 (CCNA\_03309), MetX2 (CCNA\_00559), MetZ, (CCNA\_02321), MetE (CCNA\_00515), MetH (CCNA\_02221), and CCNA\_02222. Pathway for B<sub>12</sub> utilization, BtuB (CCNA\_01826), BtuR (CCNA\_02321). Pathway for tRNA modification, MtaB, (CCNA\_03798). Pathway for nucleotide reduction, NrdJ (CCNA\_01966), NrdE (CCNA\_03607), and NrdB (CCNA\_00261). Pathway for succinyl-CoA biosynthesis MeaA (CCNA\_03177), and MutB (CCNA\_02459). Square boxes next to each enzyme contain an orange heat map which represents the fraction of ribosome footprints. ADO (adenosyl) and CN (cyano) refer to the upper ligand. C.) Negative regulation by B<sub>12</sub> riboswitches on the *btuB* and *metE* genes. Translation reporters for the *btuB* and *metE* genes fused to mCherry assayed in M2G with the indicated concentrations of B<sub>12</sub>. Error bars represent standard deviation of mCherry fluorescence in three biological replicates of the B<sub>12</sub> dilution series (n=3). Solid blue and green horizontal lines indicate the mChy fluorescence without vitamin B<sub>12</sub> for *btuB* and *metE* respectively, and dashed lines indicate the standard deviation. D.) Cartoon of B<sub>12</sub> regulated pathways in *C. crescentus*. *btuB* and *metE* genes contain negative regulatory B<sub>12</sub> riboswitches. BtuB and BtuR are part of the B<sub>12</sub> import and utilization pathway. MetH, NrdE, MeaA, MutB, and MtaB are B<sub>12</sub> dependent enzymes for methionine biosynthesis.

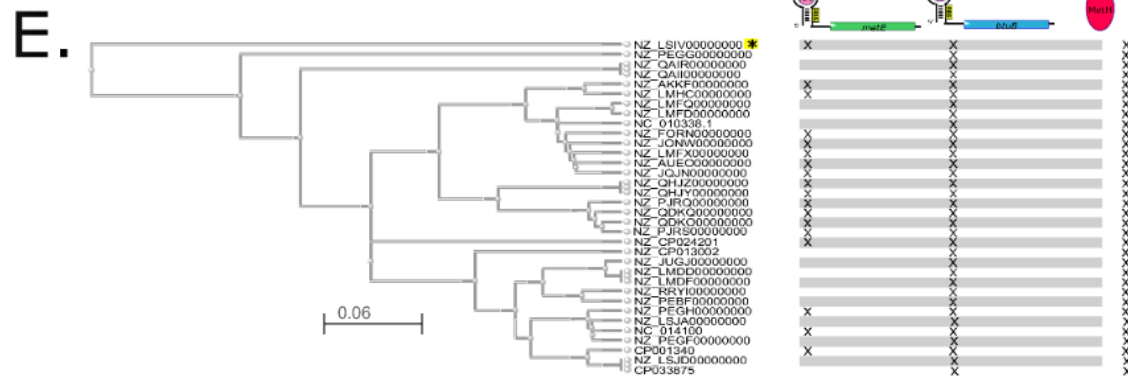
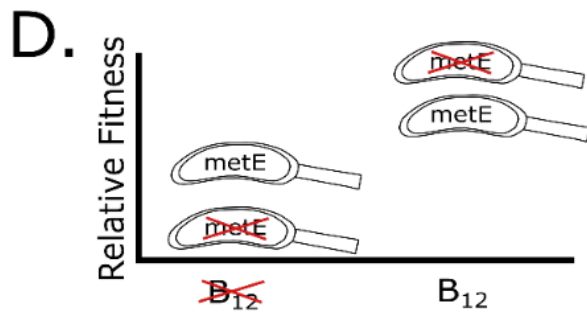
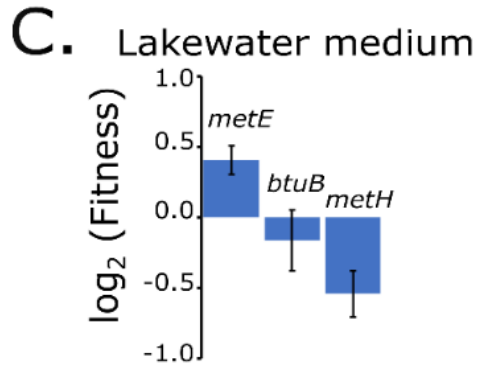
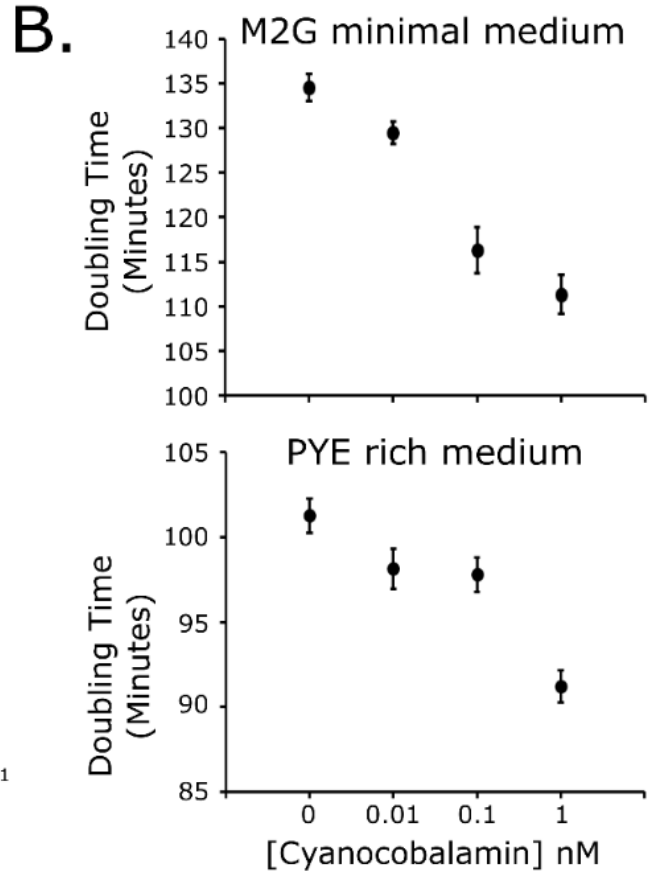
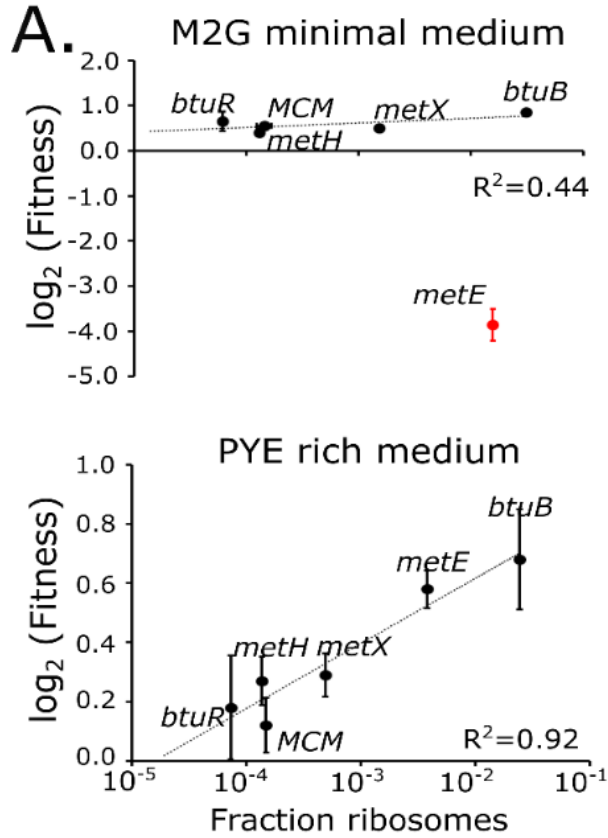
$K_{1/2}=0.062\text{nM}$ , while the *btuB* riboswitch  $K_{1/2}=0.19\text{nM}$ , both in line with the concentrations found in aquatic ecosystems (Fig S3) (Benoit 1957; Pommel 1975). Taken together, these data show that *C. crescentus* cells are investing a large amount of their protein synthesis capacity towards B<sub>12</sub> uptake and the B<sub>12</sub> independent methionine pathway in the absence of B<sub>12</sub>. We therefore hypothesized that the cells are wasting energy in the absence of B<sub>12</sub> by producing these very costly proteins.

To assess if the cells are wasting energy from the B<sub>12</sub> related pathways, we examined the fitness of *C. crescentus* cells with disruptions in the non-essential components of these pathways in the absence of B<sub>12</sub> (Fig 4A). For this we used published Tn-seq datasets (Price et al. 2018) and analyzed B<sub>12</sub> related genes whose disruption would not have polarity effects (single genes or last genes in operons) in M2G minimal media or PYE rich media, neither of which contain B<sub>12</sub> in them (Poindexter 1964; Schrader and Shapiro 2015). In M2G minimal media, cells require the *metE* gene to make methionine (Menikpurage et al. 2019), while the other non-essential B<sub>12</sub>-related components showed increases in fitness when disrupted that were proportional to their protein synthesis costs (Fig 4A). In PYE rich media, which contains methionine in the peptone,

the *metE* gene is no longer essential, but instead its disruption leads to higher fitness (Price et al. 2018). Indeed, all the components of the methionine pathway led to increases in fitness proportional to their protein synthesis cost when disrupted in PYE (Fig 4A) (Price et al. 2018). These data show that in the absence of B<sub>12</sub>, the excessive translation of these proteins leads to unnecessary costs of protein synthesis that limits the fitness of *C. crescentus* cells.

To test the effects of B<sub>12</sub> on *C. crescentus* cell growth we examined the growth rate of cells cultured in M2G media containing B<sub>12</sub> in the form of cyanocobalamin. Here we observe faster growth in a B<sub>12</sub> concentration dependent manner (Fig 4B, Table S5) with up to a 21% faster doubling time observed at 1 nM B<sub>12</sub> in M2G media. 11% acceleration of cell growth rates also occurs in PYE media, which contains methionine in the peptone, suggesting that the growth enhancement of B<sub>12</sub> is likely caused in part from reduced protein synthesis costs from *btuB* and *metE* riboswitches. We attempted to separate B<sub>12</sub>'s effects on methionine synthesis from its effects on other pathways by addition of exogenous methionine in the presence or absence of B<sub>12</sub>, but exogenous methionine leads to a dramatic decrease in growth by an unknown mechanism (Ferber and Ely 1982). Overall, we find that B<sub>12</sub> significantly enhances the growth rate of NA1000 cells.





**Figure 2.4 – Excess protein synthesis rates for methionine biosynthetic genes correlates with fitness cost** A.) Protein synthesis cost measured in the fraction of ribosomes (Table S1) on the X-axis vs the Tn-seq derived fitness values for the for the *btuB* B<sub>12</sub> importer and methionine biosynthetic genes under growth in minimal or rich media as measured previously (Price et al. 2018) (biological replicates, n=2 for M2G, n=10 for PYE.). Black points represent non-essential genes for methionine biosynthesis and red points represent genes required for methionine biosynthesis under the specified growth condition. Error bars represent standard deviation. Curve fits are performed only on non-essential genes. B.) Doubling times of *C. crescentus* cells in M2G and PYE media with indicated concentrations of B<sub>12</sub>. Error bars represent the standard deviation of doubling time measurements (biological replicates, n=3 for each condition). C.) Tn-seq derived fitness values for *metE*, *btuB*, and *metH* under growth in Lake Michigan lake water as measured previously (Hentchel et al. 2019). D.) Fitness tradeoff of facultative vs obligate B<sub>12</sub> scavenging. Relative fitness shown for species with *metE* (facultative) or without *metE* (obligate) in environments lacking or containing sufficient B<sub>12</sub>. E.) Phylogenetic tree of all *Caulobacter* species with completed genomes based on *btuB* and *metH* protein sequences. Each species is labeled by its NCBI Accession Identifier, and the scale represents the Kimura distance. Marks next to species represent the presence of a *metE*, *btuB*, or *metH* gene. All species have a predicted B<sub>12</sub> riboswitch upstream of *btuB* and *metE* genes (not shown) except the species noted with the yellow asterisk. A list of species names can be found in table S7.

## 2.4 – Discussion

### 2.4.1 – Absolute quantitation of protein synthesis in *C. crescentus*

As the cost of protein synthesis is a significant output of the cell's energy, absolute quantitation of translation is a powerful method to measure gene expression and resource allocation. Here we present an absolute protein synthesis resource for *C. crescentus* generated by ribosome profiling which will be vital for systems modeling efforts of the *C. crescentus* cell cycle (Shen et al. 2008; Lin, Crosson, and Scherer 2010; Subramanian and Tyson 2017; Murray et al. 2013) and in the subsequent optimization of synthetic *Caulobacter* genomes (Christen, Deutsch, and Christen 2015; Venetz et al. 2019). Across the proteome the mRNA translation resource allocation showed 1.4% of the translation machinery is dedicated to translation of cell cycle-regulatory genes. We observed that for all the cell cycle-master regulators, except for CcrM, that the number of proteins translated dramatically exceeds the total number of DNA binding sites in the genome (Fig 2B). We hypothesize that the high concentrations of these factors may facilitate rapid activation of target gene transcription during each phase of the cell cycle.

Approximately 7% of protein synthesis capacity is dedicated to genes whose protein products were found to be localized in one of several modes of subcellular organization (Fig 2C) (Werner et al. 2009). Recent reports suggest that some of these foci are formed by liquid-liquid phase separation of the proteins into membraneless organelles (Al-Husini et al. 2018; Banani et al. 2017). For *C. crescentus* BR-bodies, the concentration of the condensate forming protein RNase E (6.3  $\mu$ M) appears to correspond closely to the transition boundary for liquid-liquid phase separation (Al-Husini et al. 2018), potentially allowing control of the assembly of these bodies. Interestingly, we observed that the polar protein scaffold PopZ, which facilitates recruitment of asymmetrically localized signaling proteins to the cell poles (Holmes et al. 2016; Berge and Viollier 2018; Zhao et al. 2018), is present at approximately  $\frac{1}{2}$  the concentration of its client proteins suggesting that clients compete for PopZ access (Fig 2D). Dynamic competition of clients for PopZ may be important for the ordered assembly of unique proteins at each cell pole and may impact the spatial activation of downstream signaling outputs (Lasker et al. 2020; Schrader et al. 2014; Holmes et al. 2016).

#### **2.4.2 – Implications of B<sub>12</sub> scavenging pathway on environmental fitness**

B<sub>12</sub> is an important enzymatic cofactor that is required for the activity of enzymes involved in biosynthesis of methionine, dNTP production, tRNA modification, and isomerization of methylmalonyl-CoA to succinyl-CoA (Fig 3B,D) (Zhang et al. 2009). Like many bacteria, *C. crescentus* cannot produce B<sub>12</sub> but can scavenge it from the environment (Menikpurage et al. 2019), where it can increase the growth rate of *C. crescentus* by up to 21% (Fig 4B). *BtuB*, the B<sub>12</sub> importer, and *metE*, the B<sub>12</sub>-independent methionine synthase, are among the most highly expressed genes accounting for ~4.5% of all protein synthesis capacity (Fig 3A). To counteract the protein synthesis demand, *btuB* and *metE* genes also contain B<sub>12</sub> riboswitches that reduce translation when sufficient B<sub>12</sub> enters the cytoplasm (Fig 3C). Freshwater bodies typically have B<sub>12</sub> concentrations in the range from 0.11 nM to below the level of detection ( <0.1 pM ) (Daisley

1969; Ohwada and Taga 1972; Pommel 1975; Ohwada 1973; Cavari and Grossowicz 1977; Benoit 1957) suggesting that high levels of BtuB may help facilitate import. Importantly, the conserved *metE* and *btuB* riboswitches are sensitive to B<sub>12</sub> in physiologically relevant ranges ( $K_{1/2}$ =0.062 nM, 0.19 nM, respectively) and their different sensitivities suggest that *metE* translation would be downregulated before shutting off the *btuB* importer. At high concentrations of B<sub>12</sub> that saturate the riboswitches (Fig 4B), a significant portion of the increase in growth rate is likely due to the liberation of protein synthesis resources on these two highly expressed genes (Fig 3A,C). When grown in M2G or PYE media lacking B<sub>12</sub>, disrupting the *btuB* gene increases fitness by freeing up wasted protein synthesis resources (Fig 4A) (Price et al. 2018). Similarly, *metE* disruption leads to increased fitness in PYE which contains methionine, but *metE* disruption becomes essential in M2G minimal media as its required to make methionine (Fig 4A) (Price et al. 2018). Why then does *C. crescentus* have a facultative B<sub>12</sub> lifestyle containing both B<sub>12</sub> dependent and independent methionine biosynthesis pathways?

Perhaps B<sub>12</sub> independent and B<sub>12</sub> dependent pathways exist to buffer fluctuations in environmental B<sub>12</sub> concentrations. The concentration of available B<sub>12</sub> in a freshwater body shows variation of up to 40-fold between different sampling locations and at the same sampling location at different times (Ohwada 1973; Ohwada and Taga 1972; Daisley 1969; Cavari and Grossowicz 1977; Pommel 1975). Having both methionine biosynthesis pathways adds flexibility to generate methionine in either high or low B<sub>12</sub> conditions, however, these pathways have different protein synthesis costs. The B<sub>12</sub> independent pathway requires 1.57% of the total protein synthesis capacity to make sufficient MetE, while the B<sub>12</sub>-dependent MetH pathway uses only 0.156% (Fig 3B). Interestingly, disrupting the *metE* gene in lakewater leads to a fitness advantage, however, disrupting *btuB* or *metH* leads to a fitness decrease as determined in (Hentchel et al. 2019) (Fig 4C). Although not measured directly, the gene fitness signature from the experiment leads us to infer that physiologically relevant B<sub>12</sub> concentrations were present in the sampled lake water

(Hentchel et al. 2019). The increased fitness of *metE* disruptions in lakewater suggest that increased biosynthetic flexibility comes with a negative fitness cost from wasted protein synthesis resources on MetE (Fig 4D). 47% of fully sequenced *Caulobacter* species have lost the *metE* gene but not the *btuB* and *metH* genes, suggesting that the observed environmental fluctuations in B<sub>12</sub> concentration (Pommel 1975; Daisley 1969) alter the selective pressure on *metE* (Fig 4E).

Surprisingly, a recent survey of available metagenomic 16S rRNA sequencing data showed that *Caulobacter* are more abundant in soil/compost than in aquatic ecosystems (Wilhelm 2018). Soil has been shown to have B<sub>12</sub> levels that can range from 20 nM to 0.3 nM, correlated with levels of organic matter (Duda, Malinska, and Pedziwilk 1957), while bodies of freshwater typically have B<sub>12</sub> concentrations in the range 0.11 nM to below the level of detection ( 0.1 pM ) (Benoit 1957; Cavari and Grossowicz 1977; Ohwada and Taga 1972; Ohwada 1973; Pommel 1975). The higher B<sub>12</sub> concentration in soil will enhance the growth rate and may explain the increase in relative abundance in this environment.

## **2.5 – Materials and Methods**

### **2.5.1 – Bacterial strains and cell growth**

A list of all bacterial strain used here can be found in table S6. *C. crescentus* cells were grown in M2G or PYE growth media (Schrader and Shapiro 2015) and supplemented with the appropriate antibiotic concentrations (Thanbichler, Iniesta, and Shapiro 2007). *E. coli* cells used for cloning were grown in LB media and supplemented with the appropriate antibiotics.

### **2.5.2 – Ribosome Profiling**

Ribosome profiling was performed similar to (Schrader et al. 2016; Schrader et al. 2014) except contaminating rRNA fragments generated during MNase digestion were depleted to allow deeper quantitation of resulting mRNA translation similar to (Li et al. 2014). For a detailed protocol of the procedure see (Aretakis, Al-Husini, and Schrader 2018). 500mL of NA1000 cells were

grown in M2G media to an OD600 of 0.5 and treated with 100 $\mu$ g/mL Chloramphenicol for 2 min, then harvested by centrifugation and flash frozen in liquid nitrogen. Cells were then lysed on a mixer mill (Retsch mm400) at 6 cycles of 3 min at 15Hz, thawed, membranes pelleted, and the supernatant was footprinted by addition of MNase (Roche). After footprinting, MNase was quenched with EGTA, and samples were separated by sucrose gradient fractionation. 70S peaks were purified, phenol chloroform extracted, and ethanol precipitated (Aretakis, Al-Husini, and Schrader 2018). Resulting mRNA fragments were size selected by 10% acrylamide 1XTBE/7M Urea PAGE, end repaired, 3' adapter ligated, reverse transcribed, circularized, and depleted of rRNA fragments (Aretakis, Al-Husini, and Schrader 2018; Ingolia et al. 2009; Li et al. 2014). Ribosomal RNA cDNA fragments were removed using biotin-linked DNA oligos (Caulo1:5'/5Biosg/CGCTTACGGGGCTATCACCCA,

Caulo2:5'/5Biosg/TGGCAACTAATCACGAGGGTT,

Caulo3:5'/5Biosg/CTCATCTGGTTGCCCAAAGA,

Caulo4:5'/5Biosg/TGGTTCAGGAATATTCACCTG) and MyOne Streptavidin C1 Dynabeads (Invitrogen) as in (Li et al. 2014). Resulting circular cDNAs were amplified by PCR using Phusion DNA polymerase (Fermentas) with indexing primers (Ingolia et al. 2012), pooled together, and sequenced on an Illumina hiseq 2000. Data for three ribosome profiling replicates were deposited in the gene expression omnibus under accession number GSE126485. The three M2G replicates were further analyzed together with a PYE dataset collected previously (Schrader et al. 2014).

Ribosome footprint reads were mapped to the genome as center weighted reads (Oh et al. 2011), and extreme fast and slow codons were corrected for by winsorization of the bottom 5% and top 95% of nucleotides. The resulting fraction of ribosome footprints ( $\phi_i$ ) of each gene ( $i$ ) compared to the total ribosome footprint total were calculated and converted into the number

of molecules protein translated per cell ( $k_i$ ) by (equation 1.  $k_i = \frac{\phi_i P}{mW_i}$ ) where P is the average protein mass per cell and  $mW_i$  is the protein product's molecular weight as originally described in (Li et al. 2014). Average protein mass per cell (P) was measured as follows: 5mL mid log cultures of NA1000 cells were grown in M2G or PYE media overnight in a roller wheel at 28°C. Once the cells reached an OD600nm of 0.3, a 100 $\mu$ L aliquot of the cells were diluted and counted on PYE plates to measure the number of viable cells and another aliquot was saved for protein concentration measurements. To measure protein content, 100 $\mu$ L cells were spun down in a microcentrifuge at 14,000rpm for 30 seconds, the supernatant removed, and the remaining cell pellets were resuspended in 100 $\mu$ L of 1X Laemmli sample buffer lacking any dyes. After resuspension, samples were boiled for 5 min at 95°C then placed on ice. Lysate protein concentrations were measured using the Pierce™ 660nm Protein Assay (Thermo Fisher) and comparing the lysate A600 with a linear curve fit of BSA standards. The mass of protein from the sample was then divided by the number of viable cells to yield the average protein mass/cell. The resulting average protein mass per cell were: 492x10<sup>-15</sup>g/cell  $\pm$  170 in M2G and 519 x10<sup>-15</sup>g/cell  $\pm$  228 in PYE.

### 2.5.3 – Absolute quantitation of C-terminal YFP fusions

5mL mid log cultures of NA1000 cells harboring C-terminal eYFP fusions (gift of the Shapiro Lab, Stanford U.) were grown in M2G or PYE media overnight in a roller wheel at 28°C. Once the cells reached an OD600nm of 0.3, cells were spotted on M2G agarose pads for imaging. Images were collected on Leica DM6000B microscope with a Hamamatsu C9100 EM-CCD camera and a 100X PH3 PlanApo 1.40 NA objective with in a Semrock model 2427A YFP filter cube with 100ms exposure time. Fluorescence intensity was quantified using ImageJ by segmenting the cells and measuring the average pixel intensity of the cell area. Background intensity was subtracted using the NA1000 average YFP pixel intensity. For each fusion strain, a minimum of 50 cells were used for the analysis with a minimum of two technical replicates. As

MipZ molecules per cell had been previously measured (Thanbichler and Shapiro 2006) by quantitative western blot, we converted the MipZ-YFP pixel intensity to the number molecules/cell and multiplied this conversion factor by the YFP intensities of all other C-terminal YFP fusions. The average across replicates and the standard deviation ( $\sigma$ ) are reported in table S2.

#### **2.5.4 – Doubling time measurements**

Treatments were started from log phase cultures grown overnight in the absence of cyanocobalamin (Sigma-Aldrich) and diluted in fresh media to an  $OD_{600}$  of 0.05. Each treatment was split into a separate flask and the correct amount of cyanocobalamin was added to the concentrations of 1 nM, 0.1 nM, 0.01 nM and 0 nM. 50 mL from each treatment was added to three different 250 mL Kimex flasks, for three replicates of each of the four treatments. An initial  $OD_{600}$  measurement was taken of each replicate using a cuvette and a nanodrop spectrophotometer. The 12 – 250 mL flasks were then placed in a 28 °C shaker incubator at 250 RPM.  $OD_{600}$  time points of each flask were taken throughout the logarithmic growth phase. An exponential regression of the log phase time points was used to calculate the doubling time of each replicate.

#### **2.5.5 – Translation reporter assay**

JS417, JS423, and JS440 strains were started from log phase cultures grown overnight in the absence of cyanocobalamin and diluted in media with vanillate and antibiotic to an  $OD_{600}$  of 0.05. A dilution series of each strain was used to fill tubes with 2 mL of culture at each cyanocobalamin concentration; 27 nM, 2.7 nM, 0.9 nM, 0.3 nM, 0.1 nM, 0.033 nM, and 0 nM. The 21 2-mL-cultures were then grown and induced over an eight-hour period by placing the tubes in a 28 °C shaker incubator at 250 RPM for 8 hours. After eight-hours, 2  $\mu$ L from each culture was pipetted onto an M2G-agrose pad on a microscope slide. Each treatment was imaged on a microscope using both phase contrast and a mChy filter cube. Average



fluorescent intensities were calculated using MicrobeJ (Ducret, Quardokus, and Brun 2016) across a minimum of 100 cells.

### **2.5.6 – B12 homolog identification and phylogenetic tree mapping**

Protein sequences for *btuB*, *methH*, and *metE* were determined for each *Caulobacter* species with a complete genome by using the NCBI Basic Local Alignment Search Tool (BLAST) with default settings by searching the protein sequence of each NA1000 gene for homologs with an E-score of  $<10^{-19}$  (Sayers et al. 2019). The *btuB* and *methH* genes were then used to generate the phylogenetic tree using the NCBI Genome Workbench software and the MUSCLE multiple sequence alignment package (Edgar 2004). The tree is a maximum likelihood generated with the default settings from MUSCLE. Riboswitches were identified using rfam (Gardner et al. 2011) and searching the upstream regions of *btuB* and *metE* genes (up to 1000bp upstream of their predicted operons).

### **2.5.6 – Quantitative Real-Time PCR**

*mrpC* mRNA levels in reporterless backgrounds were examined using quantitative real-time reverse transcriptase PCR. Strains were developed for 24 hours under submerged culture in 100 mm Petri dishes as described above. Cells were harvested and total RNA was isolated using the hot phenol method. 10  $\mu$ g RNA was treated with Rnase-free Dnase I (Fermentas), and 1  $\mu$ g RNA was transcribed into cDNA using random hexamer primers (Amersham) and Superscript III reverse transcriptase (Invitrogen) in a 20  $\mu$ l reaction; a detailed protocol has been published previously. Real-time PCR was performed in a 7300 Real Time PCR System (Applied Biosystems) using 2  $\mu$ l of a 1:20 dilution of the cDNA reaction as template and 0.2  $\mu$ M primers specific to the 5' UTR of *mrpC* (oPH655 and oPH656; see Supplementary Table 2) in a 26  $\mu$ l reaction volume using SYBR green PCR master mix (Applied Biosystems). Real-time-PCR cycle conditions were as follows: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 seconds and 60°C for 1 min. Melting and dissociation curves were determined following 95°C for 15

seconds, 60°C for 30 seconds and 95°C for 15 seconds; single peaks were verified for each sample. Genomic DNA was used as a positive control and water and mock cDNA reactions lacking reverse transcriptase were used as negative controls. Ct values for each reaction were assigned automatically by the system software (7300 System SDS software v1.2.3). Average Ct values were calculated from triplicate technical replicates for two independent biological replicates. Fold induction was determined as  $2^{-(Ct_{\text{sample}} - Ct_{\text{wt}})}$  and the Log2 of each fold induction was plotted.

For analysis of *mrpC* mRNA levels in separated populations at 36 hours, cells were harvested, populations separated, RNA isolated, and cDNA generated as described above. Average qPCR Ct values were calculated from duplicate or triplicate technical replicates. Primer efficiencies *E* for *mrpC* and *16S rRNA* primers were determined as 1.98 and 1.94, respectively. The NA/A ratio was determined as  $(E_{\text{mrpC}})^{\Delta Ct_{\text{mrpC}}} / (E_{\text{16S RNA}})^{\Delta Ct_{\text{16S RNA}}}$ , where  $\Delta Ct = Ct_{\text{NA}} - Ct_{\text{A}}$ , respectively. The ratios for two independent biological experiments and the calculated average were plotted as circles and line, respectively.

### 2.5.7 – Proteomap generation

Categories were taken from predefined Kyoto Encyclopedia of Genes and Genomes (KEGG) categories (Kanehisa and Goto 2000). Categories were then ranked based on priority, and any gene that may be present in more than one category were deleted from those with lower priority. The 200 most numerous proteins were then hand checked. Any that had not be automatically assigned to a KEGG category were compared against other organisms to place them in their most appropriate category. The categorized genes along with the ribosome profiling data was used to create the Proteomap (Liebermeister et al. 2014).

### 2.5.7 – Strain Construction

This insert, *btuB\_5'UTR*, was generated by IDT as a gblock construct for the +1 TSS site through the start codon of the *btuB* gene.

*btuB\_5'UTR*

gBlock:

AAGCGTTCAATTGGATCCAATCTTGACGTCCGTTTGATTACGATCAAGATTGGATCCAGCG  
 TCAGGTTCTCCTCGAAAGAGGATGAAAAGGGAACGAGGTTGAAGACCTCGGCTGCCCCCGCA  
 ACTGTAAGCGGCGAGCTTCGCGTCACATGCCACTGGGCCAAAAGGCCTGGGAAGGCGA  
 CGCCCAGAAGCATTGACCCGTGAGCCAGGAGACCTGCCCGGCGCAGTCGTTTCATCGCTC  
 GGCCGGGGTGCGCCGAACGAACGGGATCTCCCGAGAAACGACAGTCAACAGGCCGCGC  
 GACGGCCTGAGCGTCCGCGTCTTCGCGGGCGGTCTGGGAGGTCGCGTGGGTCTTCATAA  
 CGGGAAGACTGTATTATGTTAATTAATATGCATGGTAC

The plasmid pRVChyC-2 (Thanbichler, Iniesta, and Shapiro 2007) was cut with MfeI and PacI, and the gblock segment *btuB\_5'UTR* was inserted into the plasmid by gibson assembly. Next, the resulting plasmid was transformed into *E. coli* DH5 alpha cells and selected on LB-kan plates. The resulting kanR colonies were then screened by PCR for the insert and verified by sanger sequencing (genewiz). The purified plasmid was then transformed into NA1000 cells by electroporation and plated on PYE-kan plates. The resulting colonies were screened for mChy fluorescence after induction with vanillate.

**JS423**

This insert, *metE\_5'UTR*, was generated by IDT as a gblock construct for the +1 TSS site through the start codon of the *metE*.

*metE\_5'UTR*

gBlock:

AAGCGTTCAATTGGATCCAATCTTGACGTCCGTTTGATTACGATCAAGATTGGATCCAGTC  
 GTGGTCTGCGGACGTTTCGCGTCCGGAGCTAAGAGGGAAGTCGGTGAGGGCGTGAAACCC  
 TGAATCCGGCGCTGCCCCGCAACTGTGAGCGGCGAGCCGCTGTCCGTTTCGTGTCACT  
 GACGCGCCGAAGCTGGTTCGGGGATGCGTCCGGAAGGCCAGGGCAGGGGTGACGACCC  
 GTGAGCCAGGAGACCTGCCTCGACAGATAACGTCTCCGGCGGGGTGTCCGGTCTGGCC

GCTTGCTCAGCGCGACCGGACAAAAGCGCCCGTGCGCGCTCGACCGCGCGCGTCCCGAT  
CAGCCTCGCCAAAACACCGGCAGAGGCTTTTCAAAG ATGTTAATTAATATGCATGGTAC

The plasmid pRVChyC-6 (Thanbichler, Iniesta, and Shapiro 2007) was cut with MfeI and PaeI, and the gBlock segment *metE\_5'UTR* was inserted into the plasmid by gibson assembly. Next, the resulting plasmid was transformed into *E. coli* DH5 alpha cells and selected on LB-chlor plates. The resulting chlorR colonies were then screened by PCR for the insert and verified by sanger sequencing (genewiz). The purified plasmid was then transformed into NA1000 cells by electroporation and plated on PYE-chlor plates. The resulting colonies were screened for mChy fluorescence after induction with vanillate.

#### **JS440**

The plasmid pVMCS-2 (Kan<sup>R</sup>) (Thanbichler, Iniesta, and Shapiro 2007) was transformed into NA1000 cells via electroporation and selected for on PYE-Kan plates.

#### **JS441**

This strain was generated by PCR amplifying the last 500bp of the  $\beta'$  RNA polymerase gene into the pYFPC-1 plasmid (Thanbichler, Iniesta, and Shapiro 2007). The insert was PCR amplified from the NA1000 chromosome using Betaprime\_forward and Betaprime\_reverse primers and the plasmid pYFPC-1 (Thanbichler et al., 2007) was PCR amplified by pYFPC\_forward and pYFPC\_reverse primers. The plasmid was then DPN 1 treated, and the insert was placed into the plasmid by gibson assembly. Next, the resulting plasmid was transformed into *E. coli* DH5 alpha cells and selected on LB-spec plates. The resulting specR colonies were then screened by colony PCR for the insert and verified by sanger sequencing (genewiz). The purified plasmid was then transformed into NA1000 cells by electroporation and plated on PYE-spec/strep plates. The resulting colonies were screened for YFP fluorescence.

PCR primers:

Betaprime\_forward 5' TAATATGCATGGTGTCTCGACGAGATCCAGGAGG

Betaprime\_reverse 5' ATCTTAAGGTTTCGGCGTCCGAAAGCGC

pYFPC\_forward 5' GCTTTCGGACGCCGAAACCTTAAGATCTCGAGCTCCG

pYFPC\_reverse 5' GGATCTCGTCTGACACCATGCATATTAATTAAGGCGCC

## 2.6 – Acknowledgements

The authors thank: Peter Chien for sharing absolute protein concentrations of DnaA, Lon, and CcrM, Adam Perez for sharing the tipN-YFP strain, Paola Mera for sharing data on *btuB* and for thoughtful discussion, and members of the Higgs lab for thoughtful discussions. This work was supported by NIH R35 GM124733 to JMS, WSU start-up funds to JMS, and a WSU Chemical Biology Interface research experience award to JRA.

## 2.7 – Supporting information

**Table S2.1 – Absolute measurements of mRNA translation for M2G and PYE.** ([Table S1, XLSX file, 0.7 MB.](#))

**Table S2.2 – Average molecules of protein translated per cell and average proteins per cell.**

<b>M2G</b>	Ribosome profiling		YFP-intensity	
Protein	Average molecules of protein translated per cell	$\sigma$	Average Proteins per cell	$\sigma$
B'-YFP	1.38E+04	8.17E+02	1.64E+04	4.37E+02
CckA-YFP	7.34E+02	6.72E+01	6.37E+02	1.15E+02
DnaB-YFP	4.33E+02	6.61E+01	4.56E+02	1.57E+01
HolC-YFP	4.16E+02	7.74E+01	3.86E+02	9.44E+01
Hu2-YFP	5.58E+04	5.85E+03	4.20E+04	1.98E+03
L1-YFP	2.99E+04	1.90E+03	4.65E+04	1.95E+03
MipZ-YFP	2.41E+03	1.13E+02	1.03E+03	1.91E+02
SMC-YFP	2.50E+02	1.63E+01	2.21E+02	9.85E+01
TipN-YFP	9.70E+02	9.20E+01	6.23E+02	4.80E+01
<b>PYE</b>	Ribosome profiling <sup>a</sup>		YFP-intensity	
Protein	Average molecules of protein translated per cell	$\sigma$	Average Proteins per cell	$\sigma$
L1-YFP	5.36E+04	ND	4.65E+04	1.39E+03
B'-YFP	1.64E+04	ND	1.64E+04	8.22E+01
Ccka-YFP	8.77E+02	ND	6.37E+02	1.11E+01
DnaB-YFP	5.79E+02	ND	4.56E+02	3.62E+01
HolC-YFP	3.17E+02	ND	3.86E+02	1.62E+01
Hu2-YFP	4.45E+04	ND	4.20E+04	9.78E+02
Mipz-YFP	1.03E+03	ND	1.03E+03	1.33E+01
Smc-YFP	2.01E+02	ND	2.21E+02	1.92E+00
TipN-YFP	3.72E+02	ND	6.23E+02	1.18E+01

<sup>a</sup>Ribosome profiling data from (Schrader et al. 2014).

**Table S2.3 – Fraction of total protein synthesis for each KEGG category<sup>a</sup>.**

Category	PYE	M2G
Cell growth and death	6.45%	11.26%
Cell motility	1.19%	1.09%
Membrane transport	9.23%	11.94%
Signal transduction	1.38%	1.30%
Folding, sorting and degradation	5.88%	5.08%
Replication and repair	0.62%	0.46%
Transcription	2.62%	2.37%
Translation	22.82%	15.88%
Amino Acid Metabolism	8.11%	8.54%
Carbohydrate Metabolism	5.51%	6.09%
Energy Metabolism	4.54%	4.47%
Glycan Biosynthesis and Metabolism	0.27%	0.24%
Lipid Metabolism	1.03%	0.93%
Metabolism of Cofactors and Vitamins	1.09%	1.53%
Metabolism of other Amino Acids	0.90%	0.97%
Metabolism of Terpenoids and Polyketides	0.19%	0.15%
Nucleotide Metabolism	1.91%	1.69%
Other Enzymes	0.43%	0.41%
Xenobiotics Biodegradation and Metabolism	0.05%	0.07%
Not Mapped	25.77%	25.52%

<sup>a</sup>KEGG category data from (Kanehisa et al. 2014).

**Table S2.4 – Comparison of absolute translation level to DNA binding sites.**

Gene	Molecules translated per cell	$\sigma$	DNA Sites in cell cycle-regulated promoters	Total DNA binding sites <sup>a</sup>
<i>dnaA</i>	6000	459	77	84
<i>gcrA</i>	18200	844	94	217
<i>ctrA</i>	25400	1480	183	187
<i>ccrM</i>	3280	233	96	4542
<i>sciP</i>	32400	1480	61	61

<sup>a</sup>Number of binding sites were summed together from (Zhou et al. 2015) based on data from (Taylor et al. 2011; Fioravanti et al. 2013; Fumeaux et al. 2014; Haakonsen, Yuan, and Laub 2015; Fiebig et al. 2014; Kozdon et al. 2013).

**Table S2.5 – Doubling time measurements at different concentrations of cyanocobalamin.**

	Cyanocobalamin concentration (nM)			
	0	0.01	0.1	1
M2G (Minutes)	135	129	116	111
$\sigma$	1.55	1.24	2.65	2.24

PYE (Minutes)	101	98.1	97.8	91.2
$\sigma$	1.02	1.18	1.00	0.952

**Table S2.6 – List of bacterial strains.**

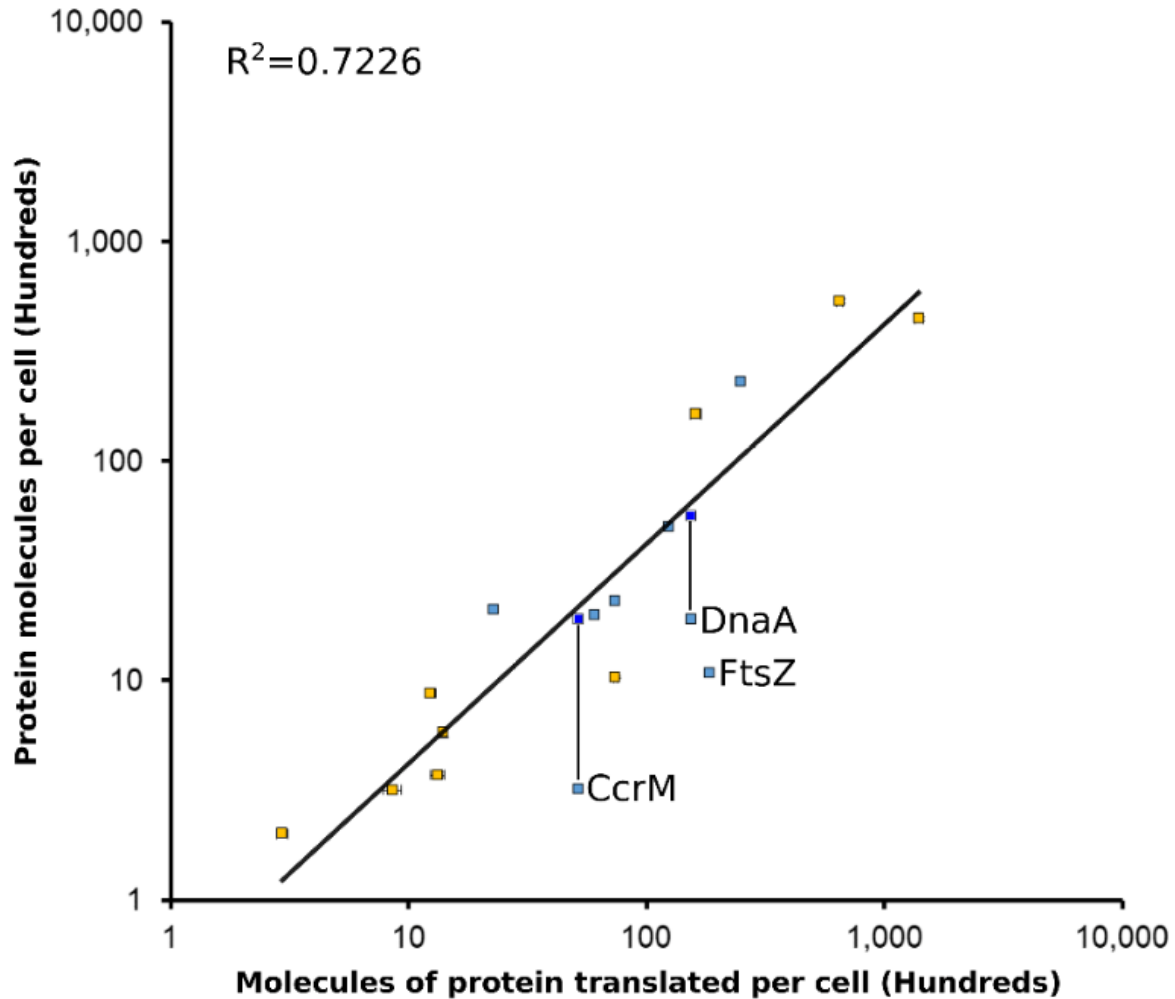
Strain	Genotype	Source
NA1000	Synchronizable variant of CB15	(Evinger and Agabian 1977)
JS417	NA1000 pRV( <i>btuB</i> _5'UTR)CHYC-2 Kan <sup>R</sup>	This Study
JS423	NA1000 pRV( <i>metE</i> _5'UTR)CHYC-6 Chlor <sup>R</sup>	This Study
JS440	NA1000 pRVMCS-2 Kan <sup>R</sup>	This Study
JS290	NA1000 <i>L1::L1-yfp</i> Gent <sup>R</sup>	(Bayas et al. 2018)
JS441	NA1000 $\beta'$ :: $\beta'$ -yfp Spec <sup>R</sup> Strep <sup>R</sup>	This Study
NJH429	NA1000 <i>cckA::cckA-yfp</i> Rif <sup>R</sup>	(Iniesta, Hillson, and Shapiro 2010)
LS3587	NA1000 <i>dnaB::dnaB-yfp</i> Kan <sup>R</sup>	(Jensen, Wang, and Shapiro 2001)
LS3586	NA1000 <i>holC::holC-yfp</i> Kan <sup>R</sup>	(Jensen, Wang, and Shapiro 2001)
MS307	NA1000 <i>hu2::hu2-yfp</i> Kan <sup>R</sup>	(Lee et al. 2011)
MT97	NA1000 <i>mipZ::mipZ-yfp</i>	(Thanbichler and Shapiro 2006)
LS3394	NA1000 <i>smc::smc-yfp</i> Kan <sup>R</sup>	(Jensen and Shapiro 2003)
<i>tipN-YFP</i>	NA1000 <i>tipN::tipN-yfp</i>	(Gift from Adam Perez)



**Table S2.7 – Table of accession numbers and organism names from figure 4**

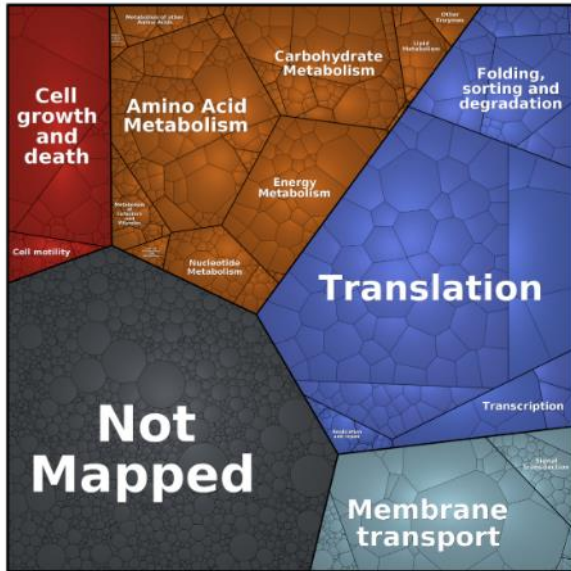
Accession Number	Name <sup>a</sup>
NZ_LSIV00000000	<i>Caulobacter</i> sp. CCH4-E1
NZ_PEGG00000000	<i>Caulobacter</i> sp. B11
NZ_QAIR00000000	<i>Caulobacter</i> sp. HMWF009
NZ_QAII00000000	<i>Caulobacter</i> sp. HMWF025
NZ_AKKF00000000	<i>Caulobacter</i> sp. AP07
NZ_LMHC00000000	<i>Caulobacter</i> sp. Root655
NZ_LMFQ00000000	<i>Caulobacter</i> sp. Root1455
NZ_LMFD00000000	<i>Caulobacter</i> sp. Root487D2Y
NC_010338.1	<i>Caulobacter</i> sp. K31
NZ_FORN00000000	<i>Caulobacter</i> sp. UNC279MFTsu5.1
NZ_JONW00000000	<i>Caulobacter</i> sp. UNC358MFTsu5.1
NZ_LMFX00000000	<i>Caulobacter</i> sp. Root1472
NZ_AUEO00000000	<i>Caulobacter</i> sp. URHA0033
NZ_JQJN00000000	<i>Caulobacter henricii</i> strain CF287
NZ_QHJZ00000000	<i>Caulobacter</i> sp. D4A
NZ_QHJY00000000	<i>Caulobacter</i> sp. D5
NZ_PJRQ00000000	<i>Caulobacter flavus</i> strain CGMCC1 15093
NZ_QDKQ00000000	<i>Caulobacter</i> sp. 774
NZ_QDKO00000000	<i>Caulobacter radialis</i> strain 695
NZ_PJRS00000000	<i>Caulobacter zae</i> strain 410
NZ_CP024201	<i>Caulobacter mirabilis</i> strain FWC 38
NZ_CP013002	<i>Caulobacter henricii</i> strain CB4
NZ_JUGJ00000000	<i>Caulobacter</i> sp. OV484
NZ_LMDD00000000	<i>Caulobacter</i> sp. Root342
NZ_LMDF00000000	<i>Caulobacter</i> sp. Root343
NZ_RRYI00000000	<i>Caulobacter</i> sp. 602-1
NZ_PEBF00000000	<i>Caulobacter</i> sp. FWC2
NZ_PEGH00000000	<i>Caulobacter</i> sp. BP25
NZ_LSJA00000000	<i>Caulobacter</i> sp. CCH9-E1
NC_014100	<i>Caulobacter segnis</i> ATCC 21756
NZ_PEGF00000000	<i>Caulobacter</i> sp. X
CP001340	<i>Caulobacter crescentus</i> NA1000
NZ_LSJD00000000	<i>Caulobacter</i> sp. CCH5-E12
CP033875	<i>Caulobacter</i> sp. FWC26

<sup>a</sup>Accession numbers and names were taken from the NCBI GenBank database.

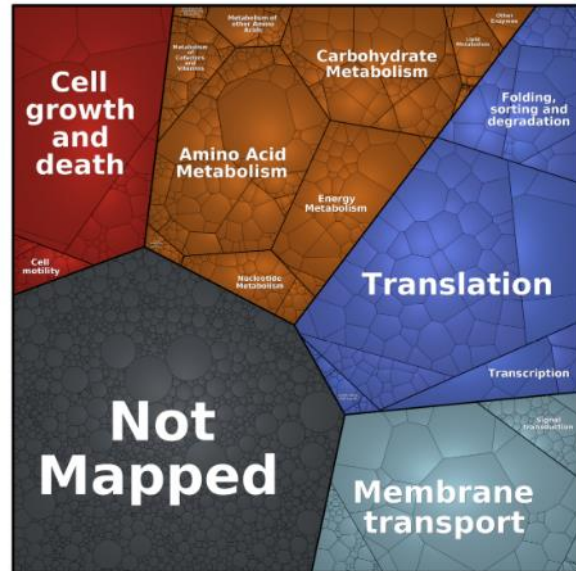


**Figure S2.1 – Absolute protein levels of unsynchronized cells in PYE medium** Absolute protein levels of unsynchronized cells in PYE media measured by western blot (blue) or YFP fusions (orange) compared to the absolute molecules of protein translated per cell calculated by ribosome profiling. PYE ribosome profiling data were collected in (Schrader et al. 2014). Vertical error bars indicate the standard deviation in YFP intensity or standard deviation for the western blots. FtsZ, CcrM, and DnaA are indicated as their protein levels are under proteolytic control by Lon protease (Wright et al. 1996; Jonas et al. 2013). CcrM and DnaA protein levels in a strain lacking their protease, Lon, are indicated in dark blue (Peter Chien personal communication). Data in Tables S1 and S2.

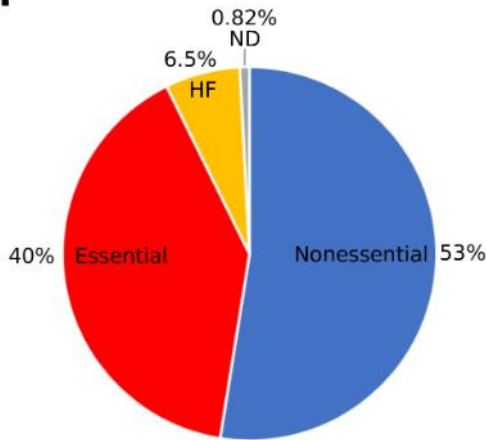
**A.** PYE Media



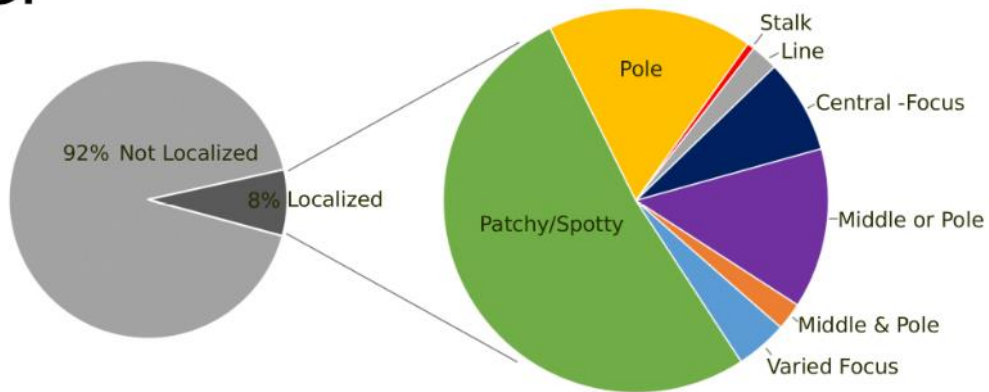
M2G Media



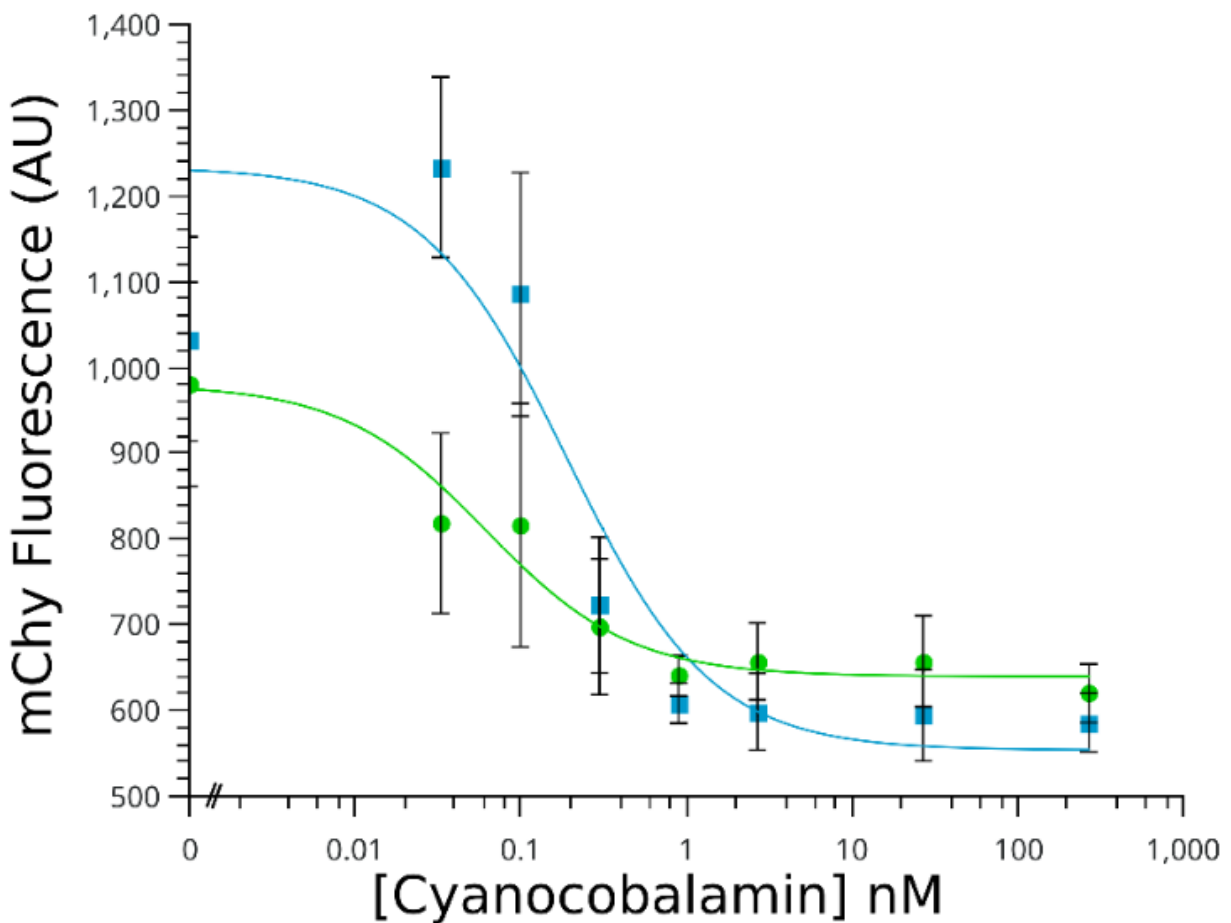
**B.**



**C.**



**Figure S2.2 – PYE fractional protein synthesis Proteomap** Global analysis of *C. crescentus* protein synthesis in PYE and M2G media. A.) Left, Proteomap of cells grown in PYE media with the area scaled to the fraction of ribosome protected mRNA footprints measured. Right, Proteomap of cells grown in M2G media shown at the same level as cells grown in PYE Media with KEGG categories (Kanehisa and Goto 2000) B.) Fraction of ribosome protected mRNA footprints compared to gene essentiality as determined in (Christen et al. 2011). Red is essential genes, blue is nonessential, yellow is high fitness genes, and grey is genes that were not determined (Christen et al. 2011). C.) Left, fraction of ribosome protected mRNA footprints in localized (dark grey) or non-localized mRNAs (light grey) for cells grown in PYE media. Right, zoomed in analysis of the fraction of proteins with different subcellular localization patterns as determined in (Werner et al. 2009).



**Figure S2.3 – Non-linear curve fit for B<sub>12</sub> dependent riboswitch repression.** mCherry intensities at each cyanocobalamin concentration were fit to a Michaelis-Menton equation using QTIplot software to determine the K<sub>1/2</sub> of cyanocobalamin. R<sup>2</sup> values were 0.95 for *metE* (green) and 0.85 for *btuB* (blue) respectively.

## CHAPTER 3 – BACTERIAL MRNA DECAY IS GLOBALLY COORDINATED WITH TRANSCRIPTION AND TRANSLATION

### 3.1 – Introduction

Gene expression is an essential process that occurs in two sub-steps, mRNA transcription, and translation, and is shut off by a mRNA decay. Bacteria must precisely coordinate each of these steps in order to optimally express their genes for maximal fitness (Maitra and Dill 2015). Steady state mRNA levels are commonly used to approximate gene expression; however, these levels are complex arising from many biochemical sub-steps that impact expression and are therefore not well correlated with protein levels (Ingolia 2014). Interactions between transcription, translation, and mRNA decay processes can begin as soon as the nascent mRNA leaves the RNA Polymerase (RNAP) (McGary and Nudler 2013). Additionally, we know that mRNAs in bacteria expressed from T7 RNAP do not couple with ribosomes during transcription, altering both translation and mRNA decay (Yang et al. 2019). In *E. coli* using Rif-seq, it was determined that approximately 30% of mRNAs undergo co-transcriptional degradation. Interestingly, *E. coli* is organized with a vast majority of its ribosomes at the cell poles, with far fewer localized within the nucleoid (Castellana, Hsin-Jung Li, and Wingreen 2016). In contrast, *C. crescentus* and many other bacteria have a far more equal distribution of ribosomes throughout the cytoplasm. In addition, the mRNA decay machinery in *E. coli* is localized to the cell's inner membrane (Chao et al. 2017), whereas in *C. crescentus* it is localized to the nucleoid (Montero Llopis et al. 2010), raising the possibility that there may be more co-transcriptional coordination with mRNA decay.

In addition to mRNA decay, mRNAs in *E. coli* and *S. enterica* have also been found to be processed into stable ncRNAs from the 5' UTR, CDS, or 3' UTR (Chao et al. 2012), (Dar and Sorek 2018), (Shell et al. 2015) and (Kawano et al. 2005) and that these processed RNAs can have important cellular functions. While *C. crescentus* RNA-seq studies have identified many

ncRNAs, the methods used to detect them were biased for identifying them in intergenic regions and that did not overlap with known mRNAs (Zhou et al. 2015; Schrader et al. 2014).

Coordination between mRNA translation and decay has been well documented. Mutations that lower translation levels of model mRNAs often show faster decay rates (Bicknell and Ricci 2017). There are two main models as to what factors affect mRNA decay, and consequently overall gene expression. The first model is the ribosome protection model. This model posits that that ribosomes translating on mRNA provide a zone of protection from nucleases over sensitive portions of mRNA transcripts. This is thought to be due to ribosomes relatively large size sterically hindering the degradosome complex from degrading the mRNA (Presnyak et al. 2015). However, there is a second “induced decay” model that states that slowly moving or stalled ribosomes trigger mRNA decay. In this model, slow moving or stalled ribosomes coordinate with the degradation machinery to cut the mRNA transcript (Chan et al. 2018). At the transcriptome-wide scale however, such a relationship between translation and mRNA decay has not been clearly established. While codon usage frequency has been shown to correlate strongly with mRNA level and half-life (Boel et al. 2016), and RNA decay sites by the major mRNA decay nuclease RNase E have been mapped (Chao et al. 2017), we lack direct information about how dynamic movement of ribosomes and initiation of mRNA decay by RNase E are coordinated.

Here we used Rif-Seq (a modified RNA-Seq protocol) to measure genome-wide mRNA half-lives and analyze global trends with transcription rates and translation rates in the bacterium *C. crescentus*. Using this approach, we found that the fraction of mRNAs that undergo co-transcriptional mRNA decay is very similar to *E. coli* (~30%), suggesting that the different subcellular localization patterns of ribosomes and RNase E do not lead to a global difference in co-transcriptional mRNA decay. Interestingly, we also observed that many mRNAs are processed into stable ncRNAs, where we identified 45 new stable ncRNAs, the majority of which result from mRNA processing. To better study understand the coordination of mRNA decay and translation, we compared mRNA half-lives and translation efficiency as measured by ribosome profiling (Li

2015) which revealed a strong global coordination. To identify how translation and mRNA decay are coordinated, we compared the location of 5' P sites generated by mRNA decay (Zhou et al. 2015) with the position of ribosomes determined by ribosome profiling (Schrader et al. 2014; Aretakis, Gega, and Schrader 2019). Here we find that 5' P cut sites are more likely to occur in mRNA locations devoid of ribosomes, suggesting that the ribosome protection model may be most likely to explain the coordination of mRNA decay and translation. Rif-seq global mRNA half-life profiling in the presence of translation initiation inhibitor retapamulin led to a mild increase in mRNA half-life, while addition of the elongation inhibitor chloramphenicol led to a dramatic stabilization of mRNA half-life across the transcriptome, providing further support for the ribosome protection model. Altogether, these data provide further evidence that transcription, mRNA decay, and translation are all highly coordinated in bacteria.

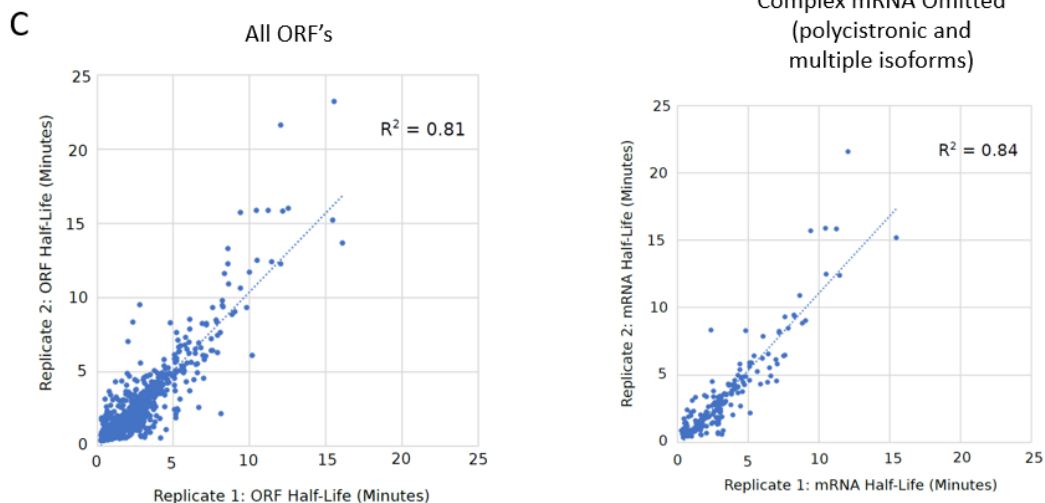
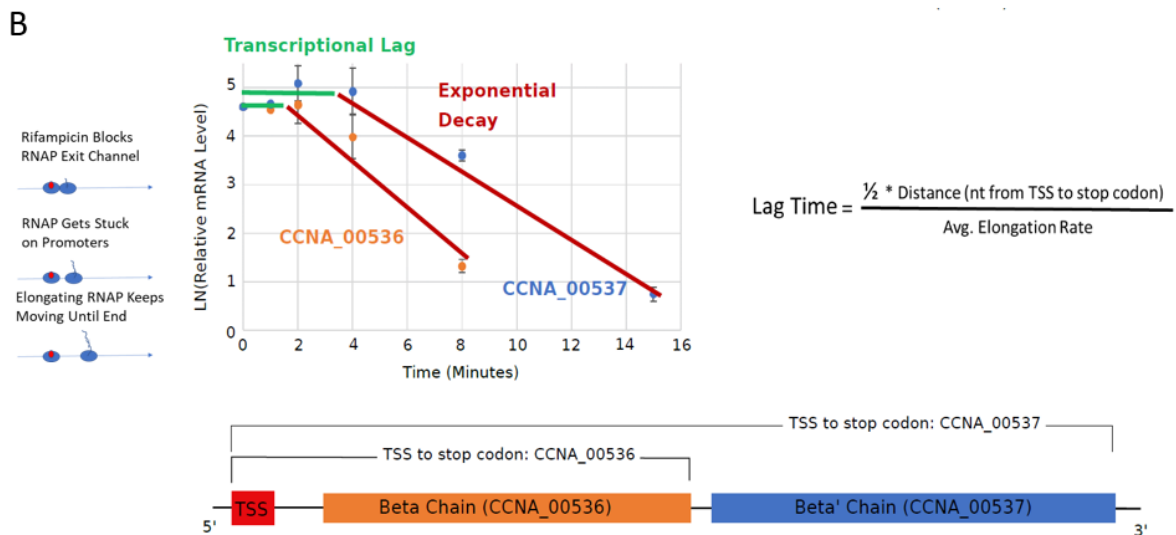
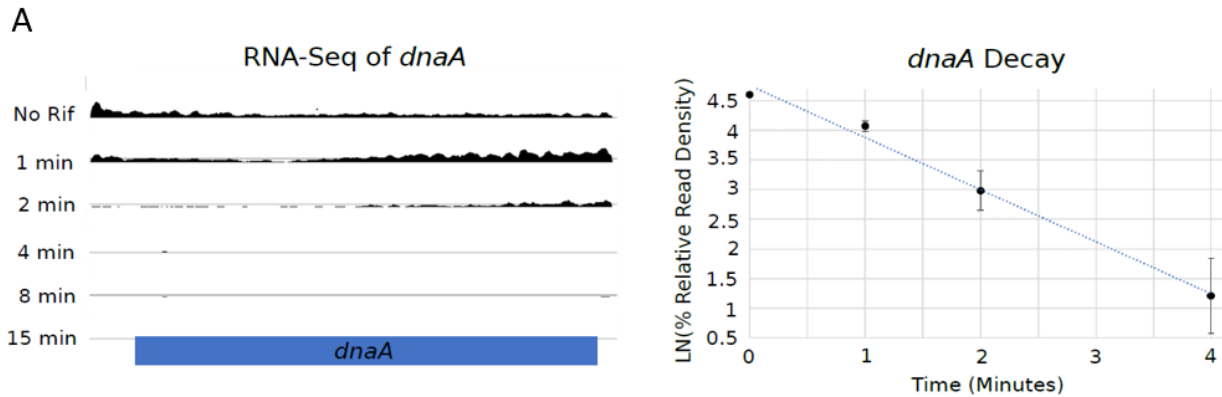
## **3.2 – Results**

### **3.2.1 – Global measurement of *C. crescentus* mRNA half-lives by Rif-Seq**

In order to quantify genome wide mRNA half-lives, Rif-Seq was performed on steady-state *C. crescentus* cells grown overnight to mid-log phase in M2G media (see materials and methods). The results are shown for the *dnaA* gene (figure 1 A, left). A steady state sample was taken, and then transcription was halted with the antibiotic rifampicin and samples were taken at different time points to measure the decay of the mRNA transcript. mRNA half-lives were calculated by natural log transformation of the relative read density of the different time points, and fitting to a linear regression (figure 1 A, right). The slope of this linear regression allows for the calculation of the mRNA transcript's half-life, which for the *dnaA* gene was 1.06 minutes.

A correction needs to be applied to the mRNA decay data to compensate for rifampicin's mechanism of action in which RNAP mRNA transcription initiation is halted, but not elongation. There is “transcriptional lag” associated with using rifampicin, in which it blocks the RNAP exit channel for newly initiating RNAP, but not for currently elongation RNAP, allows transcription to finish and run-off the transcript (figure 1 B, left). We analyzed many different decay graphs,

examined half-lives from genes within the same polycistronic operon, and compared mRNA half-lives for the same gene across our biological replicates, to determine the best correction for this “transcription lag”. Our results agree with a previous group (Chen et al. 2015), and we used a Lag



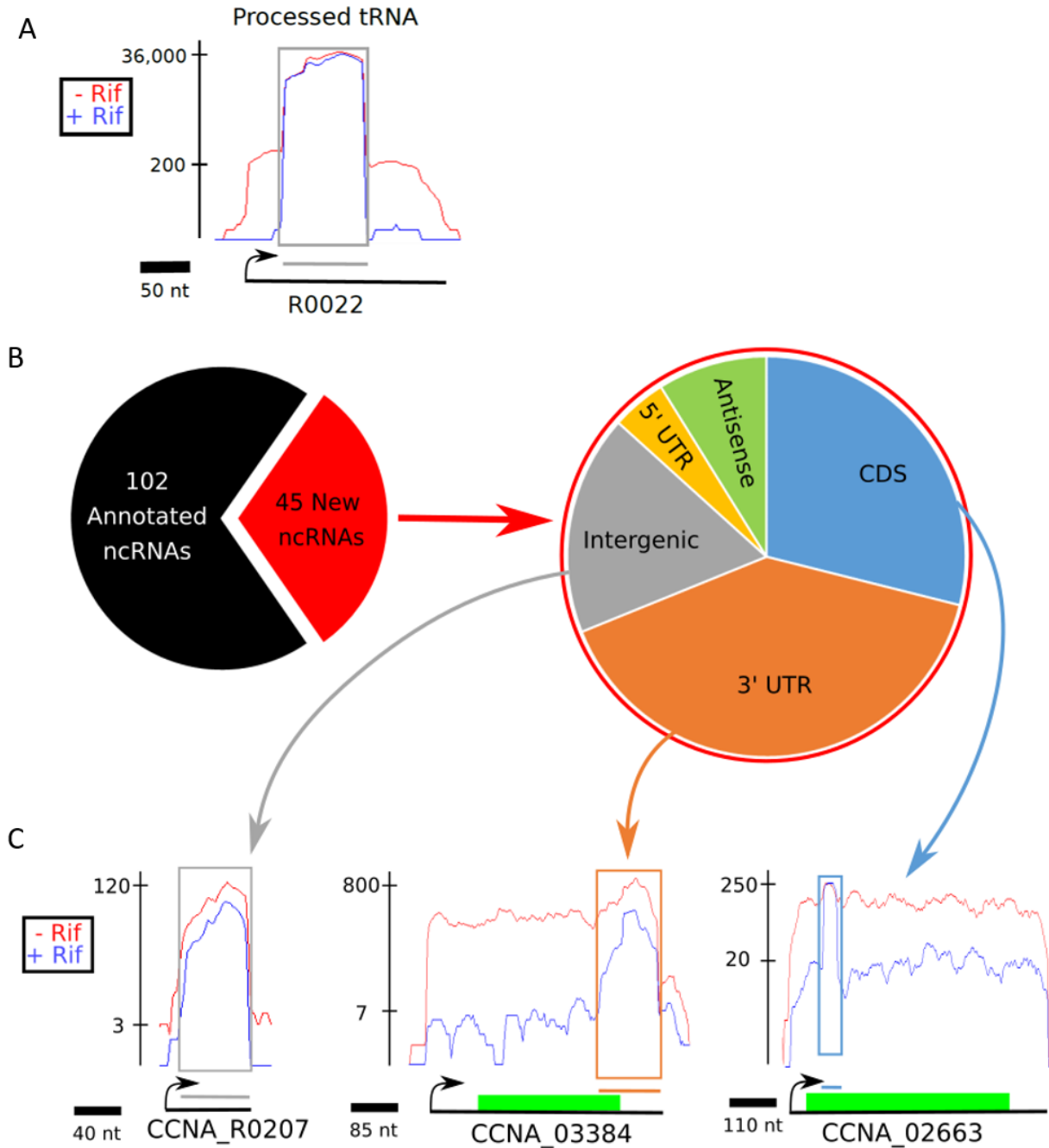


**Figure 3.1 – Global measurement of *C. crescentus* mRNA half-lives by Rif-Seq.** A) (left) Normalized mRNA level (RPKM)s for *dnaA* at each time point. (Right) Natural log transformed RPKM data plotted against time for the *dnaA* mRNA. The half-life was measured to be 1.06 minutes. B) Rifampicin inhibits polymerase transcription initiation, but not elongation. This introduces a potential lag in observed mRNA decay where the elongating RNAP run off. An example of the lag in mRNA decay is shown for a polycistronic operon containing genes CCNA\_00536 and CCNA\_00537. Transcriptional runoff was corrected for by calculating the average time it takes for the RNA polymerase to elongate through a particular gene, calculated as the lag time. Lag time is calculated as half the distance from the TSS to the end of the CDS, divided by the average RNAP mRNA Elongation Rate (C) (left) For all mRNAs whose half-life were measured in both Rif-Seq biological replicates the half-lives were plotted ( $R^2 = 0.81$ ,  $n=1258$  CDSs) and (right) for simple (monocistronic and single isoform) mRNAs ( $R^2 = 0.84$ ,  $n=227$  CDSs).

Time that is half the distance from the TSS to the end of the CDS, divided by the average RNAP mRNA Elongation Rate. (figure 1 B, top, right). This correction calculates when half of the ribosomes have run-off a particular mRNA's section of the transcript. The calculation of the average RNAP elongation rate on mRNA was used in this correction and will be discussed later.

To illustrate this “transcription lag”, real data is shown for two genes, CCNA\_00536 and CCNA\_00537, encoding beta and beta' chains for the RNAP, respectively. There is a “transcriptional lag” that is proportional to the distance from the TSS to the end of the CDS. (figure 1 B, bottom). This “transcriptional lag” can be seen in the natural log transformed RPKM data plotted against time for the two genes in this polycistronic operon (figure 1 B center). There is a clear “transcriptional lag” phase, where mRNA levels do not decrease, and an exponential decay phase where mRNA levels decay. Across the genome, RNA half-lives per gene were calculated by using transcriptome data to calculate and correct for the transcriptional lag. This was accomplished with the Rif-Correct Software package (see materials and methods)

We then compared the calculated mRNA half-lives between our two biological replicates. These two datasets comprised all ORF's that had calculated mRNA half-lives in both replicates. We found relatively strong agreement and correlation between the two replicates (figure 1 C, left). However, there seemed to be some outliers that suggested further issues with our calculated half-lives that went beyond our initial “transcriptional lag” correction. Directly reviewing the sequencing data for many of the outlier genes revealed an interesting discovery. There were fragments of



**Figure 3.2 – Processed noncoding RNAs are generated from diverse locations.** A) The raw read counts of tRNA CCNA\_R0022 at prior to rifampicin treatment (in red), and 15 minutes after the addition of rifampicin (in blue). An algorithm was designed to detect stable RNA fragments that remain 15 minutes after transcription is inhibited with rifampicin. The stable processed region is shown as a gray solid line and is highlighted by a box. B) The stable RNA detection algorithm found 102 previously annotated ncRNAs, and 45 additional stable ncRNAs. The pie chart on the right shows the sources of the new stable ncRNAs. C) Newly identified ncRNAs are generated from multiple regions of the genomes. CCNA\_R0207 is an example of a new intergenic ncRNA. CCNA\_03384 has an example of a new ncRNA that is processed from the 3' region of an mRNA. CCNA\_02663 has a ncRNA that is processed from its CDS region. Graphs show the relative read density for each stable fragment at prior to rifampicin treatment (in red), and at 15 minutes (in blue) after rifampicin addition. The different genome regions had the following number of new ncRNA; 5' UTR, 2; CDS, 13; 3' UTR, 18; Intergenic, 8; Antisense, 4.

some mRNA transcripts that seemed to resist degradation over the time scale of our Rif-Seq experiment (15 minutes total). The RPKM of a specific gene's mRNA levels is calculated by averaging the sequencing reads across the entire length of that gene's transcript. These stable, degradation resistant, partial mRNA fragments were artificially increasing the calculated mRNA level near the end of the time-series, and subsequently the half-life for the entire mRNA. Other groups have found similar degradation resistant fragments in *E. coli* (Dar and Sorek 2018).

### 3.2.2 – Processed noncoding RNAs are generated from diverse locations

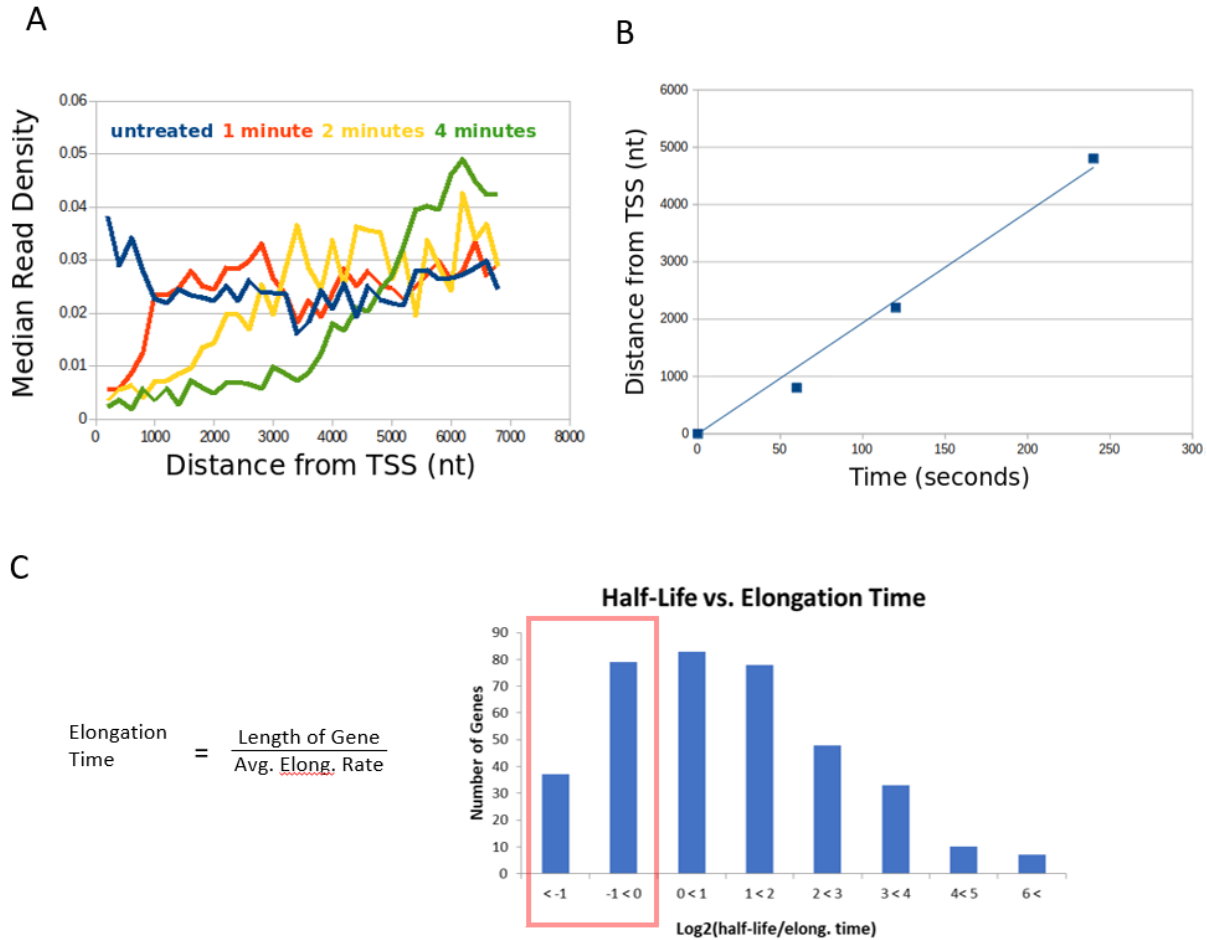
To create a better overall genome-wide half-life dataset, we searched for the stable mRNA fragments in *C. crescentus*. tRNAs are known to be stable in bacteria over the short timescale of our Rif-Seq data points (Neidhardt 1964). An algorithm was designed and trained to detect the stable processed tRNA in *C. crescentus* (see materials and methods). The algorithm was able to successfully identify 40 out of 40 tRNA that were degradation-resistant 15 minutes after transcription was inhibited with rifampicin (figure 2 A). Through this method, 147 ncRNAs were found, 45 of which were newly discovered ncRNAs. These new stable ncRNA occurred in different regions throughout the genome (figure 2 B, right) (Bharmal, Aretakis, and Schrader 2020) (see materials and methods). CCNA\_03384, and CCNA\_02663 contain a new class of ncRNA that are “carved” out of mRNA by degradation around the stable fragment (figure 2 C) (Dar and Sorek 2018). This new class of processed ncRNA are unlike typical intergenic ncRNA, as they do not have their own initiator and terminator sequences. Although a biological function for this new class of ncRNA is currently not known, they are present at significant abundance within the cell.

The detection of these new degradation resistant ncRNA from within known genes allowed for a further refinement of our mRNA half-lives dataset for our subsequent analyses. mRNA that were omitted included complex mRNA such as those that were polycistronic, or had multiple isoforms from multiple known multiple TSSs, internal TSSs, or ncRNA within the boundaries of the corresponding gene. We used this smaller half-life dataset for all subsequent analyses, except where specified. This reduced the number of outliers that did not strongly correlate between our

mRNA half-life biological replicates (figure 1 C, right). Without the complex mRNA, we have more confidence in our overall mRNA half-life dataset, and with the conclusions we draw from this data.

### 3.2.3 – Co-transcriptional mRNA decay occurs on long mRNAs

As had been previously mentioned, the average elongation rate for RNAP on mRNA was needed for the “transcriptional lag” calculation. A runoff experiment was performed in order to determine the RNAP rate of elongation. Read density data was combined for highly expressed operons over 7,000 bases and plotted against distance from TSS (figure 3 A) (see materials and methods). The half-max median read density was then plotted versus the time point in seconds (figure 3 B). The performing a linear regression of the relationship between the two variables gives a slope of 19.36 nt./sec ( $R^2 = 0.994$ ), which represents the average RNAP elongation rate on mRNA (see materials and methods). Calculation of the RNAP elongation rate on mRNA not only allows for the “transcriptional lag” calculation that is needed to correct for rifampicin’s mechanisms of action, but also allows for an exploration of co-transcriptional decay in *C. crescentus*. Co-transcriptional mRNA decay occurs when an mRNA transcript starts to be degraded while it is still being transcribed by the RNAP. In order to determine which mRNA, undergo co-transcriptional decay, RNAP average mRNA elongation rate is used along with the length of each transcript to calculate overall elongation time (figure 3 C, left). This was then compared to the mRNA half-life for each gene. (figure 3 C, right) (see materials and methods). Roughly a third of the genes in our half-life dataset had half-lives shorter than their calculated average elongation time, indicating that these genes should experience co-transcriptional decay.



**Figure 3.3 – Co-transcriptional mRNA decay occurs on long mRNAs.** A) Runoff of RNAP in long mRNAs. Combined read data for highly expressed operons longer than 7,000 nt., at 0 minutes, 1 minute, 2 minutes and 4 minutes after the addition of Rifampicin was plotted as distance from TSS in 300 nucleotide bins versus median read density. B) The half-maximal read density vs distance was plotted for each time a linear regression yielded a slope of 19.36 nt./sec, representing the transcription elongation rate. The nucleotide distance from the TSS was calculated corresponding to the half-max of each time points median read-density. C) Histogram of the number of mRNAs with different ratios of half-life over transcription elongation time. 30.9% of mRNAs have a  $\text{Log}_2(\text{half-life}/\text{elongation time})$  less than 0, representing mRNA half-lives that are shorter than the amount of time it takes to be produced by RNA polymerase, indicating that these mRNAs undergo co-transcriptional degradation.

### 3.2.4 – Steady state mRNA levels are determined by a combination of transcription and decay

mRNA transcription rate in the cell is controlled by two gene-expression sub-step rates, mRNA transcription and mRNA decay. Measuring transcription on a gene by gene basis globally is difficult but if the rates of mRNA transcription and mRNA decay are known for a gene then the transcription rate can be calculated. RNA-Seq data was used to find the RPKM, and relative mRNA copy number across the cell. These relative mRNA levels were then used with mRNA copy number data from *E. coli*, which was found to be ~1400 total mRNA copies per cell, which allowed us to calculate individual mRNA absolute copy number (Neidhardt 1987) (see materials and methods) (figure 4 A). Once absolute mRNA copy numbers for each gene were determined the relationship between single gene mRNA copy number, mRNA decay rate, and transcription rate was used to determine the transcription rate for each gene. (figure 4 B) (see materials and methods). A Proteomap for each gene with both a calculated mRNA transcription rate (figure 4

A

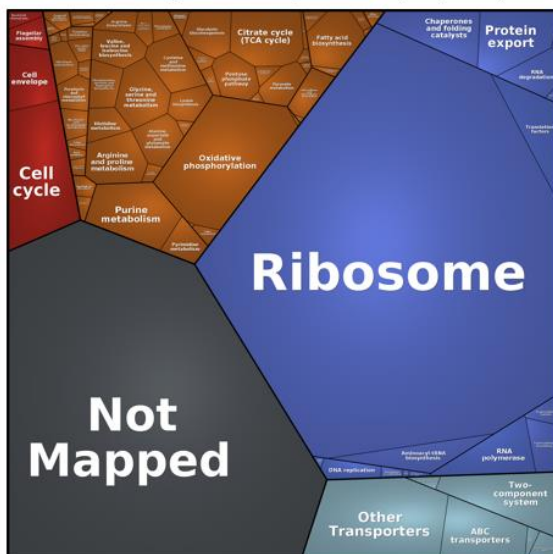
$$\text{mRNA Copy Number per Cell} = \frac{\text{mRNA RPKM}}{\text{Total RPKM}} \times \text{Total mRNA per Cell}$$

B

$$\text{Transcription rate (calculated) (mol/cell/min)} = \frac{\text{mRNA copy number (measured)}}{\text{Decay rate (measured)}}$$

C

Transcription Rates (molecules/cell/min.)



D

Half Life (min.)



**Figure 3.4 – Steady state mRNA levels are determined by a combination of transcription and decay.** A) Equation for the calculation of absolute mRNA copy number. Relative individual mRNA Copy Number of RNA-Seq data was used to calculate the absolute mRNA copy number for all mRNAs. The total mRNA per cell was estimated to be 1400 (Cannistraro and Kennell 1985) which was used to convert read density (RPKM) into absolute mRNA copy number. B) Equation for the calculation of absolute transcription rate (molecules/cell/minute). Absolute transcription rates were calculated using the absolute mRNA copy number per cell and the measured mRNA half-life. C) Proteomap (Liebermeister et al. 2014), with each polygon representing a gene category with area scaled to the calculated mRNA transcription rate in molecules per cell per minute normalized to the summed transcription rate of all mRNAs (represented by the square). D) Proteomap (Liebermeister et al. 2014), with each polygon representing a gene category with area scaled to the calculated mRNA half-life in molecules per cell per minute normalized to the summed half-lives of all mRNAs (represented by the square).

C), and mRNA half-life (figure 4 D) was created (see materials and methods). These size of the area in each polygon corresponds to the relative level of either transcription rate, or mRNA half-life, respectively, for each gene ontology category.

### 3.2.5 – 5' P sites are likely generated by RNase E and occur predominantly in the CDS

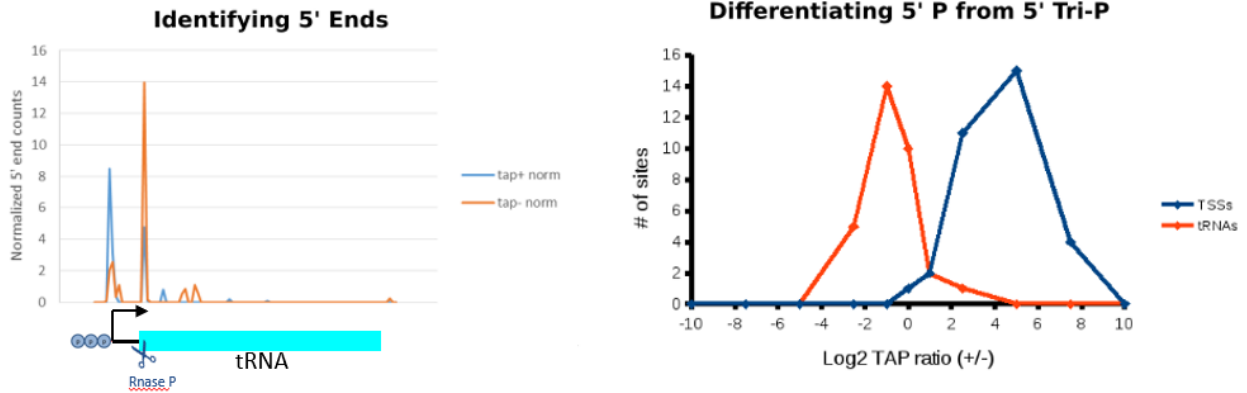
To understand mRNA degradation more deeply an investigation into the specific sites where the RNases cut transcripts for degradation is needed. Monophosphate sites and mRNA tri-phosphate sites correspond to mRNA processing or degradation sites, and mRNA TSS sites, respectively. 5' global RACE experiments take advantage of this difference in 5' end phosphorylation state to identify mRNA processing or degradation sites and TSS sites throughout the genome (see materials and methods). tRNAs can be used to validate and illustrate the power of this technique by showing the connection between sites identified by 5' global RACE, and known processing and TSS sites. 5' global RACE data (Zhou et al. 2015) was overlaid a figure of a tRNA transcript which shows that there is an increase in TAP + library normalized 5' end counts at the TSS site at a tri-phosphate 5' end, and an increase in the TAP - library normalized 5' end counts at the processing site with a mono-phosphate 5' end (figure 5 A , left). A positive Log2 ratio of a TAP + library over a TAP - library indicates a TSS site, and a negative Log2 ratio indicates a processing site (figure 5 B, right).

The sequence motif of 5' monophosphate sites show a nucleotide preference similar to RNAs that were crosslinked to RNase E (Chao et al. 2017). The log sequence of monophosphate

5' ends shows a preference for an adenine or an uracil at the first nucleotide after the cut site, and an uracil or a cytosine two nucleotides after the cut site. This is a similar preference that has been found in other experiments that specifically enriched for RNase E cutting in the bacterium *Salmonella enterica* (Chao et al. 2017) (figure 5 B) (see materials and methods). 5' monophosphates identified by the 5' global RACE experiment, not associated with a known processing site, were identified as degradation cut sites and pooled for further analyses. From the nucleotide information around the identified cut sites the minimum free energy was calculated (see materials and methods). A higher minimum free energy indicates less secondary structure in that area of the mRNA. The highest minimum free energy was found to correspond to the area directly at, or around the cut sites. (figure 5 C). This indicates that cutting takes places in areas calculated to have less mRNA secondary structure. Finally, we explored in what regions along mRNA transcripts did mRNA degradation mRNA take place. mRNA transcripts were divided into different regions; 5' UTR, Start, Coding, Stop and 3' UTR (see materials and methods). For this analysis we only used monocistronic mRNA that had a known TSS site 30 nucleotides or more upstream of the CDS. In these different regions we found the greatest number of cuts occurred in the Coding region of the mRNA (figure 5 D). However, the region with the greatest probability of cutting, based on cut site per nucleotide, was the 5' UTR, while the Coding region had the second lowest probability of cutting. For comparison, we used monocistronic tRNA with 3' UTR's as a control due to their known processing sites, and we found that the small number of these tRNA have a larger probability of cutting than any of the mRNA regions in our analysis. Furthermore, we calculated what the probability of cutting in any of these mRNA regions was if cuts were randomly distributed throughout the genome and found that this probability was lower than any of the observed cutting in mRNA regions.



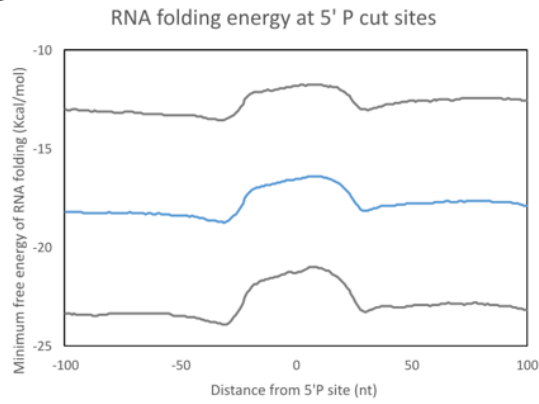
A



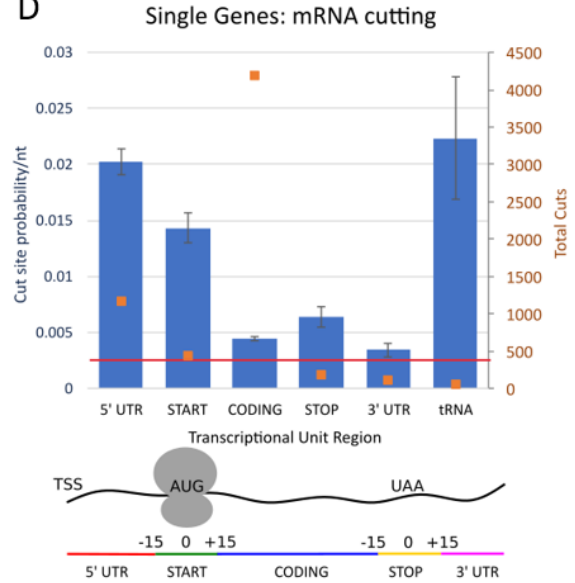
B



C



D

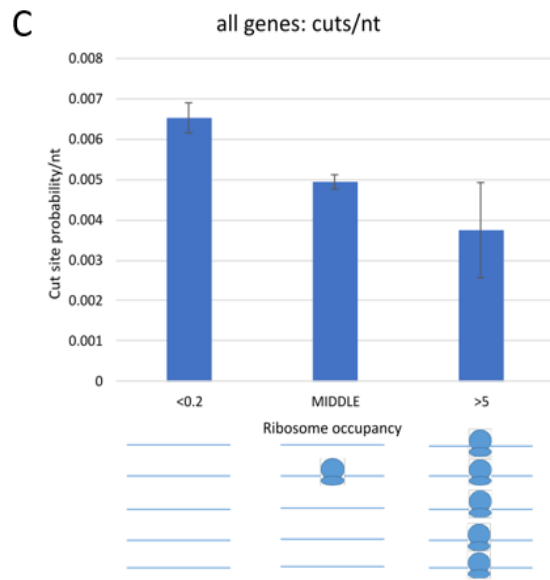
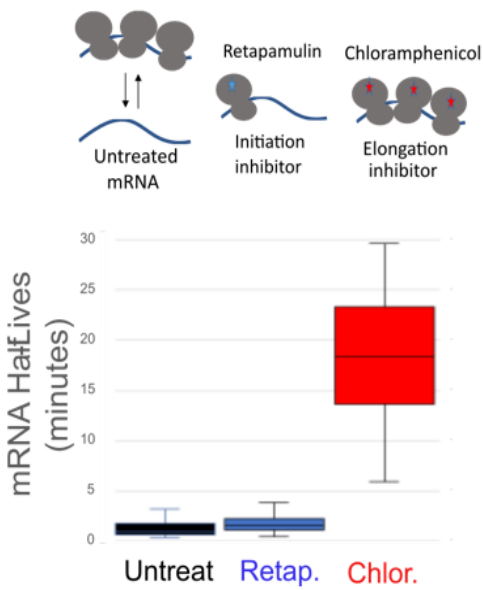
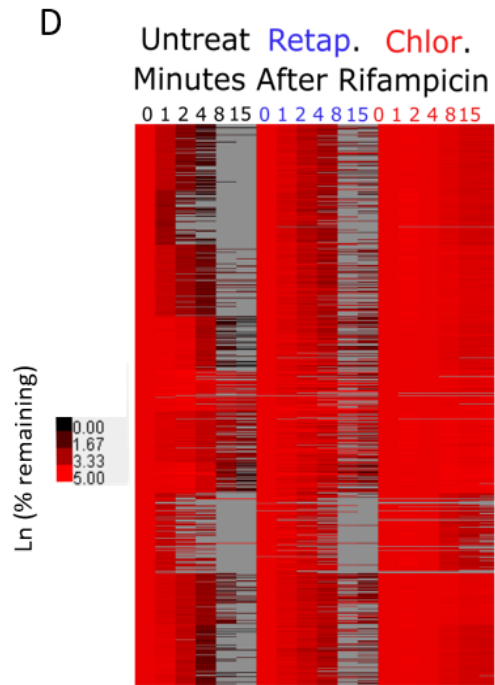
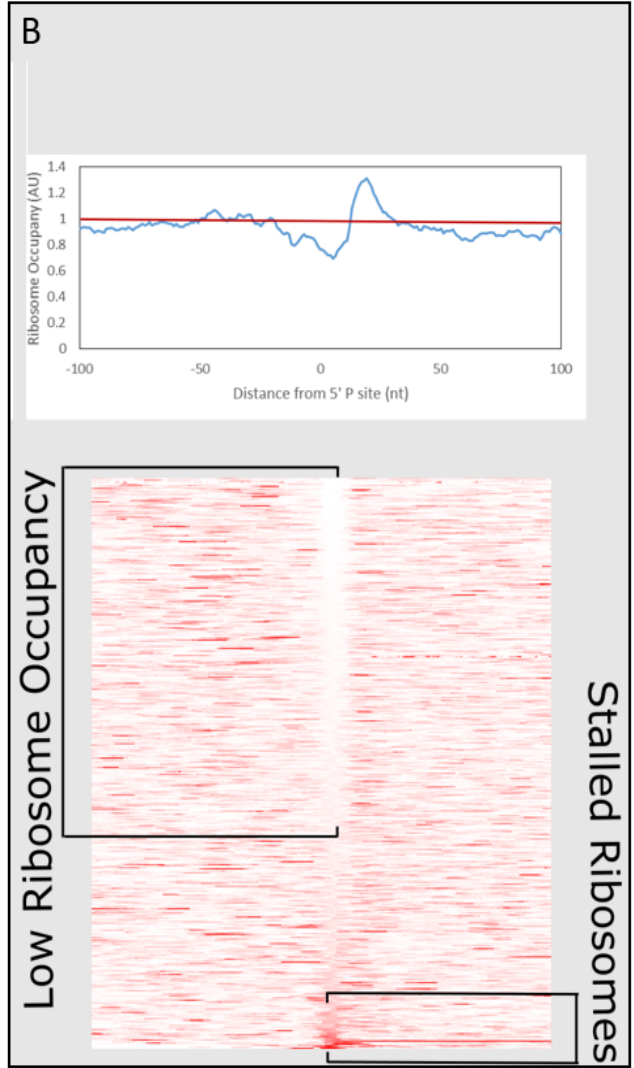
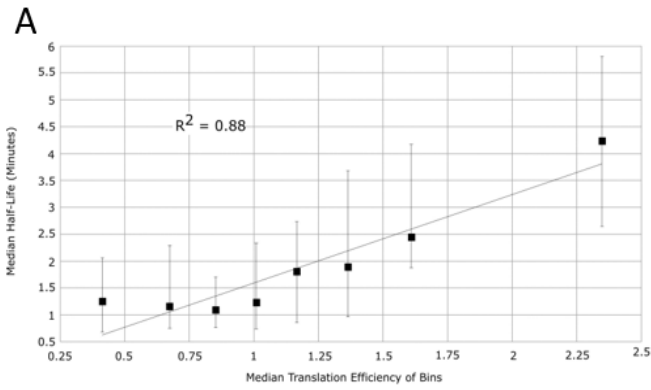


**Figure 3.5 – 5' P sites are likely generated by RNase E and occur predominantly in the CDS.**

A) (Left) 5' global RACE data (Zhou et al. 2015) analyzed previously for 5' PPP sites, were reanalyzed for identification of 5' P sites generated by nucleases. Raw data for a tRNA gene for the TAP + and TAP – negative libraries showing Tri-phosphate native ends at the TSS, and monophosphate processed ends at the mature 5' end of the tRNA. (Right) Log2 Ratio of TAP + library to TAP – library show that processed ends of tRNA have a negative Log2 ratio and primary TSS have a positive Log2 ratio for tRNAs that is distinguishable. B) The sequence motif of 5' P sites show a sequence preference like RNAs that were crosslinked to RNase E (Chao et al. 2017). C) Metagene plot of the minimum free energy of the predicted mRNA secondary structure in the region surrounding the 5' P site. The blue line represents the median free energy, while the lines above and below it indicates the upper and lower quartile of the calculated free energy. D) On the left y-axis cut site probability per nucleotide is shown for different regions for mRNA highlighted in the cartoon below, and for tRNA which are known to be processed (orange squares). On the right y-axis the total cuts identified per region is shown (blue bars). There are more total cuts occurring in the CDS region of mRNA, but there is a higher probability of a cut per nucleotide in the 5' UTR. The horizontal red line represents the probability of finding a cut site if they were distributed randomly across the genome.

**3.2.6 – Translation affects mRNA decay across the transcriptome**

Although, most of the cutting takes place in the CDS region, there is a lower per nucleotide probability of cutting in the CDS region than most of the rest of the mRNA transcript. This led us to analyze the impact of translation and ribosomes on mRNA decay. First, we wanted to investigate if there was a correlation between our calculated mRNA half-lives and translation. mRNA were binned according to their translation efficiency, a measure of how well each mRNA is translated into protein (RNA-Seq RPKM divided by Ribosome Profiling RPKM) and graphed versus half-life (figure 6A) (see materials and methods). We found that there is a relationship between these two measures, and that increased translation efficiency leads to longer mRNA half-lives. Second, we wanted to determine how 5' monophosphate cut sites correlate with ribosomal occupancy. Ribosomal occupancy (Schrader et al. 2014) was normalized and the average set to 1 arbitrary unit. The area around each cut site was combined and graphed (figure 6 B, top) (see materials and methods). The lowest ribosomal occupancy takes place directly at the cut sites. This can be seen in the meta-gene plot, with ribosomal occupancy displayed as a red heat map (figure 6 B, bottom) (see materials and methods). The majority of genes have low ribosomal occupancy at the cut site, while some have middling ribosomal occupancy, and a few highly expressed genes have stalled ribosomes at these cut sites. This shows that for most genes,



**Figure 3.6 – Translation affects mRNA decay across the transcriptome** (Clockwise) A) mRNAs were ranked by translation efficiency and placed into 8 bins with an equal number of mRNAs (overall  $n = 375$ ). mRNAs in polycistronic operons were omitted. The median mRNA half-life of each bin is plotted. The error bars represent the upper and lower quartiles of half-life. A linear regression trend line is shown. B) (Top) Metagene plot centered on mRNA cut sites is plotted versus ribosome occupancy with the red line showing the average occupancy. Actively translating ribosomes were used and initiating and terminating ribosomes were omitted. (bottom) Meta-gene plot of ribosome occupancy as a red heat map (darker red is higher ribosome occupancy) showing that cutting typically takes place where there is low ribosome occupancy. Cut sites were restricted to highly expressed genes with  $>1$  ribosome footprint per codon and were mapped as center weighted footprints (Oh et al. 2011). C) mRNA cut sites were sorted into 3 categories; cut sites that had ribosome occupancy at least five times less than average, cut sites that had ribosome occupancy at least five times more than average, and cut sites in-between. Error bars represent Wald-type confidence interval. D) (top) Transcriptome wide abundances for mRNAs at different points after rifampicin treatment. Cells were either grown normally, pre-treated with Retapamulin (initiation inhibitor), or pre-treated with chloramphenicol (elongation inhibitor). Each line represents a single CDS, and the RNA abundance is colored as a heatmap with the scale plotted to the left. (bottom) For genes found in all three data-sets mRNA half-lives in minutes were plotted as boxplots. Medians are 1.11 minutes for untreated, 1.78 minutes for retap, and 17.85 minutes for chlor.

mRNA cutting takes place at sites of low ribosomal occupancy, and a few highly expressed mRNA skew the average with stalled ribosomes at cut sites. To validate this conclusion, we did an additional analysis starting with ribosomal occupancy instead of 5' monophosphate cut sites. Nucleotides across the genome were put into either a low, middle, or high ribosomal occupancy category. For each category we calculated the cut site probability per nucleotide using the size of the category in nucleotides and 5' P sites (see materials and methods). In agreement with our previous analysis, the greatest probability of cutting is found at low ribosomal occupancy sites, followed by middle, and then the lowest probability of cutting is found at high ribosomal occupancy.

In order to directly test how ribosomes effect mRNA degradation, we used antibiotics to freeze ribosomes along the mRNA transcript. Retapamulin inhibits ribosomal initiation leaving the transcript empty of ribosomes except for a ribosome at the ribosome binding site. While, chloramphenicol inhibits ribosomal elongation freezing ribosomes along the length of the transcript, leaving the ribosomes unable to disassociate. Using a modified Rif-Seq protocol that collected time points after 2 minutes of pretreatment with retapamulin, or chloramphenicol, we

were able to manipulate ribosomes and measure mRNA degradation. There was also a control time course without any antibiotic pre-treatment (see materials and methods). Our untreated library had similar results to previous Rif-Seq experiments, when analyzed alone. Retapamulin Rif-Seq pre-treatment had a marked decrease in bulk mRNA degradation, while pre-treatment with chloramphenicol had a universal, and exceptional, decrease in bulk mRNA degradation (figure 6 D, top). Due to many mRNA having unmeasurable half-lives in the chloramphenicol treatment (mRNA half-lives twice as long as the last timepoint were considered too long to measure accurately), only mRNA with short half-lives appear in all three time courses. Consequently, the median half-life for the untreated dataset decreased by almost a minute, from 1.89 minutes, when restricted to mRNA with measurable half-lives in all three time-courses. Untreated mRNA had a median half-life of 1.11 minutes, while retapamulin treated mRNA had a median half-life of 1.78 minutes, and chloramphenicol treated mRNA had a median half-life of 17.85 minutes. All ORF's were used in the analysis of the untreated, retapamulin, and chlorophenol time courses. These results illustrate how ribosomal placement is a major contributing factor to mRNA half-life.

### 3.3 – Discussion

Overall, *C. crescentus* mRNA lifetimes ( 2.67 minute average) were similar to the mRNA lifetimes in *E. coli* of ( ~2.5 minute average) (Chen et al. 2015), which has a 60 minute shorter cell cycle than *C. crescentus*. Interestingly, despite more homogeneous mixing of ribosomes and major mRNA decay nuclease RNase E with the nucleoid in *C. crescentus*, the fraction of mRNAs undergoing co-transcriptional decay is also nearly the same (30.9% *C. crescentus*, 33.5% *E. coli*) despite large scale differences in the localization of the translation and decay machinery. We also identified 45 stable ncRNAs, most of which were processed from mRNAs (Dar and Sorek 2018). The largest class we identified were in 3' UTRs where many in *S. enterica* were found to be processed from, we also identified processed ncRNAs made from the CDS and 5' UTR where ncRNAs have also been found in *E. coli* (Kawano et al. 2005; Dar and Sorek 2018). While the

function and expression pattern of these ncRNA is uncertain, some are processed from essential genes, such as the essential cell cycle-regulator *gcrA*, suggesting that they may perform important cellular functions.

By comparing the *C. crescentus* mRNA half-lives and estimated transcription rates we can clearly observe differences in mRNA turnover across different gene classes that are likely to make the cells more fit. For instance, ribosomal proteins have a high transcription rate, but have short half-lives, suggesting rapid turnover. This may help the bacteria to adapt to rapid changes in their environment, by allowing them to be rapidly shut off. Other gene categories, such as TonB transporters have lower mRNA transcription rates but significantly longer mRNA half-lives (Figure 4 C, D) suggesting very slow mRNA turnover. In this case, it is possible that cells always need to important nutrients and may still want to express these mRNAs even if environmental conditions change. As bacterial cells must optimize their fitness in unpredictable dynamic environments, these programmed turnover rates of certain cellular mRNA types likely allow the cells to program their transcriptome and prime it for more rapid adaptation to changing environments.

Upon reanalysis of 5' global RACE monophosphate (Zhou et al. 2015) data, we found that most cut sites occur within the coding sequence of mRNAs (figure 5 D) in regions with lower secondary structure content (figure 5 C). Since we see a sequence motif at the cut site that is highly similar to the *S. enterica* RNase E cleavage motif (Chao et al. 2017), it is possible that these represent direct entry sites for RNase E cleavage (Clarke et al. 2014). While we observe fewer cut sites in the 5' UTR, due to the small size of 5' UTRs in *Caulobacter* compared to the size of CDSs, the probability per nt. of finding a cut site in the 5' UTR is actually higher than in the CDS. While RNase E is known to have increased activity when on a 5' P, the enhanced probability of cutting may be related to 5' P accessibility (Deana, Celesnik, and Belasco 2008). Why is it that the highest number of RNA cut sites were identified in the CDS, but the probability per nucleotide of cutting is lowest in this region of the mRNA? Importantly, this is the region that is occupied by ribosomes which likely provide a physical barrier to RNase E cleavage. Indeed, we observed a

strong negative correlation between cut sites and ribosome occupancy (figure 6 B, C]. Indeed, by perturbing ribosome occupancy by addition of retapamulin, which blocks initiation site ribosomes but allows ribosome runoff, we found only a slight stabilization of mRNA lifetimes. Addition of chloramphenicol, which blocks elongation and allows increased ribosome accumulation across the CDS, led to a dramatic stabilization in mRNA lifetime, with many mRNAs showing only minimal decay after 15 minutes of transcription shutoff. These data suggest that the ribosome protection effect provides a major stabilization to mRNA lifetime and may be the main source driving the correlation between translation levels and mRNA half-life (figure 6 A).

When reviewing the ribosome protection model or the factor induced model, the two purposed models that help explain the relationship between mRNA decay and translation, we see our support falls behind the former. Degradosome cut sites occur with lower probability in the coding regions of transcripts (figure 5 D). For the majority of mRNA cutting occurs where there is an absence of ribosomes (figure 6 B), and nucleotides classified as having lower ribosomal occupancy have a higher probability of cutting (figure 6 C), and in only a small population do the cuts occur near ribosome peaks (figure 6 B). The factor induced model predicts that stalled ribosomes would induce mRNA degradation at the site of the stalled ribosome, while the ribosome protection model predicts that stalled ribosomes would sterically hinder the degradosome from attaching to and degrading the mRNA. The next step in understanding the interaction between translation and mRNA decay would be to determine which model is dominant. To differentiate between these two major models a set of synthetic transcripts would be engineered with more or less efficient ribosomal binding sites, in combination with optimal and non-optimal codon usage. By comparing the mRNA abundances and decay rates of these synthetic mRNAs, the occupancy of ribosomes and speed of ribosomes could be varied rather dramatically. This would allow us to more directly identify whether ribosome occupancy is truly dominant in causing stabilization, and whether or not accumulation of stalled ribosomes might also strongly influence the overall decay rate.

### **3.4 – Materials and Methods**

#### **3.4.1 – Rif-Seq**

Cell Harvesting 25 ml of *C. crescentus* cells are grown overnight to mid-log phase in M2G media. They are grown in an incubation shaker, in a 250 mL conical flask. Mid-log phase is confirmed as an Optical Density 600 reading between 0.3 to 0.5. 2 ml of Qiagen RNAprotect bacteria reagent is brought to room temperature in 15 ml falcon tubes. 200ug/ml of Rifampicin dissolved in DMSO is added to the 25 mL culture. 1 ml of culture is pulled out of the conical flask at each time point after the addition of rifampicin and added to 2 mL of RNAprotect and vortexed for 5 seconds (timepoints are 0, 1, 2, 4, 8, and 15 minutes). After the last time point is taken, the cells are incubated at room temperature for at least 5 minutes in the RNAprotect. Immediately after the cells are spun down and pelleted in preparation for RNA Extraction. RNA was extracted according to protocol (Al-Husini et al. 2020) and libraries were constructed according to (Aretakis, Al-Husini, and Schrader 2018).

#### **3.4.2 – mRNA Half-life calculation using Rif-Correct Software**

Half-life for each gene was calculated from Rif-Seq according to the protocol in (Aretakis & Schrader *In Submission* 2021). To calculate lag time, RNA polymerase elongation time is calculated to the middle of each CDS feature. This distance is calculated from half the distance from the farthest upstream TSS site, or first CDS start codon if no TSS is known (for operons this would be the first gene in the operon), to the end of the CDS site for that specific gene. This is used as a correction, and the linear regression of the natural log of the RPKM data gives a slope calculation which relates to the half-life.

#### **3.4.3 – Detection of non-coding RNAs**

non-coding RNAs were discovered according to the protocol in (Bharmal, Aretakis, and Schrader 2020). New ncRNAs were categorized as CDS if they overlapped with the CDS region, as 3' UTR or 5' UTR if they overlapped with a previously known UTR region or overlapped with



the region within 30 nucleotides of an upstream or downstream start or stop codon, antisense if they overlapped with the region on the opposite strand of a CDS, or intergenic if they met none of the previous criteria and where not overlapping a known ncRNA.

#### **3.4.4 – Calculating average RNAP elongation rate on mRNA's**

For each time point (0, 1, 2, and 4 minutes) read data was combined for highly expressed operons greater than or equal to 7,000 bases in length in 300 nucleotide bins. The half-max distance for each timepoint was calculated. Then each time point in seconds was plotted against the half-max, and a linear regression was performed, and the slope of the linear regression corresponded to the average RNAP elongation rate on mRNA in nucleotides per second (19.36 nt./sec).

#### **3.4.5 – Half-Life vs Elongation Time**

Elongation time was only calculated for genes that had non-complex mRNA such as those genes that are monocistronic, and without multiple known isoforms (multiple TSS's or internal TSS's). Elongation time was calculated as the length of the gene in nucleotides (from known TSS or start of CDS to end of CDS) divided by the average RNAP elongation rate on mRNA's (19.36 nt./sec).

#### **3.4.6 – smFISH**

See the Material and Methods of our collaborator Dr. Kim; smFISH mRNA Copy Number Per Cell: See the Material and Methods of our collaborator Dr. Kim; smFISH Global analysis of transcription rates: See the Material and Methods of our collaborator Dr. Kim

#### **3.4.7 – mRNA Copy Number Per Cell from *E. coli***

Using the *E. coli* estimate for mRNA copy number per cell (*E. coli* 1400 mRNA per cell) (Cannistraro and Kennell 1985), *C. crescentus* RNA-Seq relative mRNA copy number an absolute copy number for each gene was calculated. The RPKM for *C. crescentus* RNA-Seq in log phase steady state conditions, grown in M2G media (Schrader et al. 2014) was summed across the genome, the percentage as calculated for each individual gene by dividing each gene's RPKM by

the total RPKM. Then this per gene percentage was multiplied by the total mRNA estimate for *E. coli*, to estimate the absolute mRNA copy number for each mRNA per cell.

#### **3.4.8 – Individual mRNA genome-wide transcription rate**

For this analysis all ORF's half-lives were used to provide a systemic view of mRNA transcription and translation. Individual mRNA copy number was divided by the decay rate to give an individual mRNA copy number in molecules per cell per minute.

#### **3.4.9 – Proteomap generation**

See (Aretakis, Gega, and Schrader 2019) for explanation of how *C. crescentus* Gene Ontology categories were determined. For genes where a mRNA half-life had been calculated, individual gene copy number per cell was divided by the mRNA half-life in minutes to give a mRNA transcription rate in mRNA's transcribed per minute. The categorized genes along with the mRNA transcription rate data were used to create the Proteomap for mRNA transcription rate (Liebermeister et al. 2014), and mRNA half-lives were used to create the other Proteomap for mRNA half-lives (Liebermeister et al. 2014).

#### **3.4.10 – 5' Global RACE experiments**

See (Zhou et al. 2015) for a detailed explanation of the 5' end detection. NA1000 cells were grown to mid-long, pelleted and frozen in liquid nitrogen. Total RNA was extracted with hot Trizol, and rRNA was subtracted from the samples. One library was treated with Tobacco Acid Pyrophosphatase (TAP), and another was not. The libraries were then ligated to an RNA adapter with T4 RNA ligase. mRNA libraries were then converted to cDNA libraries and sequenced.

#### **3.4.11 – Sequence Logo of 5' degradation ends**

The 5' degradation ends identified from the 5' global RACE experiment were fed into WebLogo, see (Crooks et al. 2004). These 5' desegregation ends were multiple sequence aligned. At each position nucleotide data is stacked proportional to the nucleotide conversation at that position and scaled/measured overall in bits.

### **3.4.12 – RNA folding energy at 5' P cut sites**

A 50 nt. Sliding window was used to calculate the minimum free energy and reported for the nucleotide centered in the middle of the window. From 100 nt. upstream to 100 nt. downstream of an identified 5' end degradation site, minimum free energy was reported as a median and graphed along with lines representing upper and lower quartiles.

### **3.4.13 – mRNA cutting in different mRNA regions**

mRNA in this analysis were limited to those mRNA and tRNA that are monocistronic, and without multiple known isoforms (multiple TSS's or internal TSS's). Also, mRNA were limited to those with known UTR's greater than or equal to 30 nucleotides. These mRNA were divided up into different regions. With the 5' UTR region defined as from the TSS site to -16 from CDS start. With the start region defined as -15 from CDS start to +15 from CDS start. With the coding region defined as CDS +15 from CDS start to -15 from CDS end. With the end region from -15 from CDS end to +15 from CDS end. The tRNA region was grouped together as a whole. From our 5' end degradation detection we sorted those cuts that fell into any of the mRNA regions, or for tRNA. This gave us total number of cuts in those regions. Then the total number of cuts was divided by the length of all of those regions across our selected mRNA to give our cut site per nucleotide probability, the error bars show the calculated Wald confidence interval. Finally, the total number of cuts from the 5' end degradation detection for each strand was randomly assigned a position in the genome, on their respective strands, using the Random Python library function "randint()". Then the cuts that fell in each mRNA region across the genome was summed. The red line represents the total cuts in each region if cuts happened randomly across the genome as calculated in this fashion.

### **3.4.14 – Translation efficiency vs. half-life**

mRNA in this analysis were limited to those mRNA that are monocistronic, and without multiple known isoforms (multiple TSS's or internal TSS's). Median translation efficiency was calculated according to (Schrader et al. 2014) mRNA were then separated into eight bins based

on their translation efficiency and then the median translation efficiency of that bin was graphed vs median half-life of mRNA in that bin. Lower and upper error bars represent the 25% and 75% quartile for mRNA half-life, respectively.

#### **3.4.15 – Center weighted ribosomal footprints**

Ribosomal profiling data was used to determine the ribosomal occupancy across the genome, for footprints that had greater than 1 read per nucleotide, and the resulting footprint was center weighted. The region around the 5' end degradation site (100 nucleotides upstream and downstream) was graphed versus the ribosomal occupancy with the average set to 1 and represented by the red line. The metagene plot below shows a heat map of ribosomal occupancy (the darker red the higher ribosomal occupancy). The center represents the 5' end degradation site, and shows where there is low ribosomal occupancy, middle ribosomal occupancy and stalled ribosomes at the degradation site.

#### **3.4.16 – Cut site probability per nucleotide**

Ribosomal profiling data was used to determine the ribosomal occupancy across the genome, for footprints that had greater than 1 read per nucleotide, and the resulting footprint center weighted. From that average ribosomal occupancy was calculated. Nucleotides were then put into 1 of three categories; those nucleotides with five times or less ribosomal occupancy than average, those nucleotides with five times or more ribosomal occupancy than average, and those nucleotides with ribosomal occupancy in between. The total number of cuts in each of those categories was summed and divided by the total number of nucleotides in that category to get cuts per nucleotide probability in each category.

#### **3.4.16 – Rif-Seq with Retapamulin or Chloramphenicol pretreatment**

Cell Harvesting 25 ml of *C. crescentus* cells are grown overnight to mid-log phase in M2G media. They are grown in an incubation shaker, in a 250 mL conical flask. Mid-log phase is confirmed as an Optical Density 600 reading between 0.3 to 0.5. 2 ml of Qiagen RNAprotect bacteria reagent is brought to room temperature in 15 ml falcon tubes. At time = -2 minutes (two

minutes before the addition of rifampicin at time = 0 minutes) the 25 mL were pretreated with either no antibiotics, or chloramphenicol to a final concentration of 100 µg/ml, or retapamulin to a final concentration of 12.5 µg /ml. Then 200ug/ml of rifampicin dissolved in DMSO is added to the 25 mL culture. 1 ml of culture is pulled out of the conical flask at each time point after the addition of rifampicin and added to 2 mL of RNAprotect and vortexed for 5 seconds (timepoints are 0, 1, 2, 4, 8, and 15 minutes). After the last time point is taken, the cells are incubated at room temperature for at least 5 minutes in the RNAprotect. Immediately after the cells are spun down and pelleted in preparation for RNA Extraction. RNA was extracted according to protocol (Al-Husini et al. 2020) and libraries were constructed according to (Aretakis, Al-Husini, and Schrader 2018).

## APPENDIX 4 – METHODOLOGY FOR RIBOSOME PROFILING OF KEY STAGES OF THE CAULOBACTER CRESCENTUS CELL CYCLE

This chapter has been published:

Aretakis, J. R., N. Al-Husini, and J. M. Schrader. 2018. 'Methodology for Ribosome Profiling of Key Stages of the *Caulobacter crescentus* Cell Cycle', *Methods Enzymol*, 612: 443-65.

Jared M. Schrader and I wrote the paper. Nadra Al-Husini, Jared M. Schrader and I created and refined the protocol.

### 4.1 – Introduction

Bacterial cell division is the result of a productive round of the cell cycle to yield two daughter cells. The cell cycle is highly coordinated in *Caulobacter crescentus* where it is driven by a cell cycle gene-regulatory network that coordinates gene expression with the major cell cycle events such as chromosome replication and cell division. Recent ribosomes profiling data showed that 484 genes undergo changes in translation efficiency during the cell cycle, suggesting a broad role for translational control in cell cycle-regulation. In this chapter, we focus on how to perform ribosome profiling to measure the translation efficiency across cellular mRNAs at key stages in the *Caulobacter* cell cycle. This methodology relies on the high-yield Iudox gradient synchronization of *Caulobacter* cells followed by ribosome profiling to measure ribosome density and total-RNA-seq to measure mRNA levels.

In bacteria, cell cycle-regulatory networks control progression of the cell cycle by integrating information about key cell cycle events into their gene expression programs. *Caulobacter crescentus* has proven to be a powerful model of the bacterial cell cycle due to the high-yield and rapid ability to perform cell synchronization (Evinger and Agabian 1977; Schrader and Shapiro 2015). In *Caulobacter*, the well-defined cell cycle-regulatory network is controlled by a set of master regulators that are arranged in a transcriptional regulatory circuit (McAdams and

Shapiro 2011; Lasker, Mann, and Shapiro 2016; Collier 2016). While transcriptional control, DNA methylation, phosphosignaling, protein localization, and regulated proteolysis have well established roles in the cell cycle-regulatory circuit, translational control was recently found to play a significant role where 12.4% its genes have cell cycle-regulated translation (Schrader et al. 2016). Ribosome profiling now allows genome-wide interrogation of mRNA translation, making it a powerful tool towards understanding the gene-regulatory networks controlling complex cellular events (Ingolia et al. 2009; Stumpf et al. 2013; Brar et al. 2012). The methodology described here integrates the high-yield of *Caulobacter* cell synchronization with the power of ribosome profiling to measure translation efficiency at key stages of the cell cycle. This method allows unbiased interrogation of translational control in bacterial cell cycles with genome-wide coverage.

#### **4.2 – *Caulobacter* Cell Cycle Ribosome Profiling Methodology**

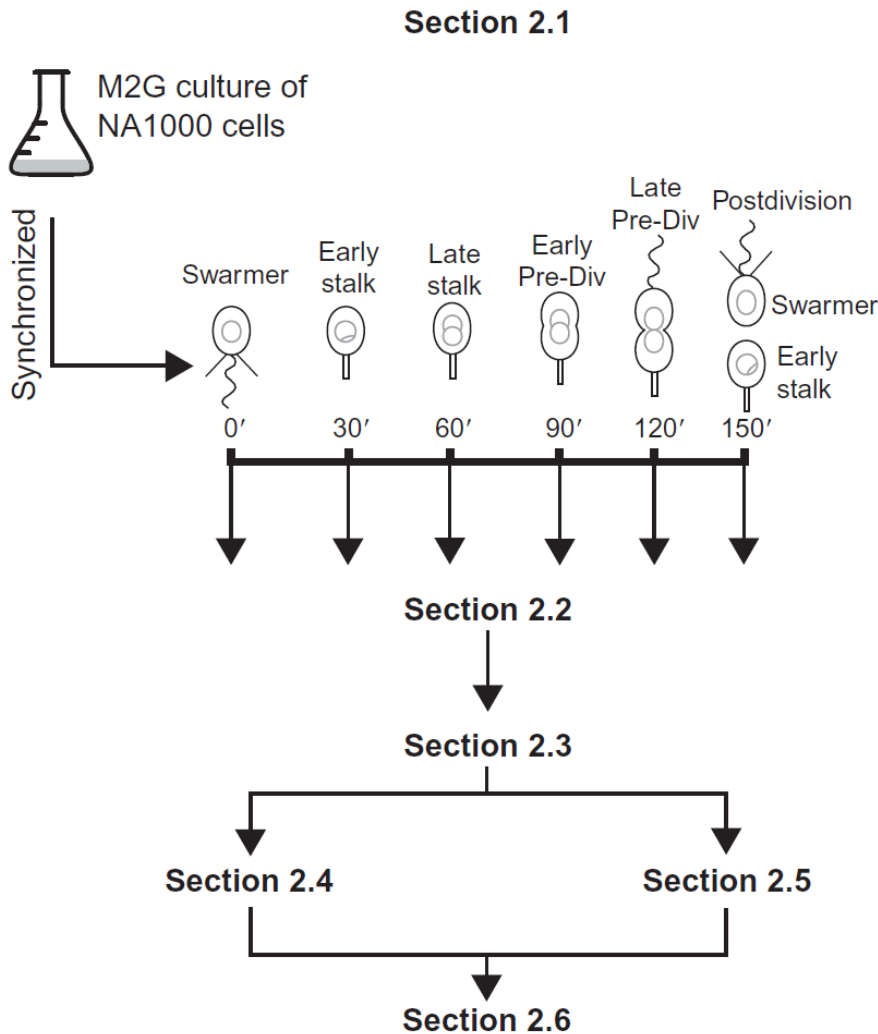
Ribosome profiling described here is performed in multiple phases including 2.1- Cell Growth, Synchronization and Harvesting, 2.2- Harvesting Cells, 2.3- Cell Lysis, 2.4- Footprinting and Footprint Extraction, 2.5- Total RNA-Seq Sample Preparation and PAGE Size Selection, and 2.6- Library Generation (Fig. 1). This protocol takes several days to perform (Table 1) so it is important to include control samples throughout the process. Typically, ribosome profiling in bacteria uses large volumes of cultures but this protocol is adapted to small samples containing only 30 mL of log-phase cells retrieved upon ludox gradient synchronization. To measure translation efficiency total RNA-seq is compared to ribosome footprints to calculate the translation efficiency, which is the metric measured across the cell cycle.

##### **4.2.1 – Cell Growth, Synchronization and Harvesting**

*Caulobacter* synchronization by ludox gradient synchronization yields a large amount of swarmer cells that can be followed during the cell cycle (Evinger and Agabian 1977). Briefly, 1 L of mid-log phase NA1000 cells are arrested in growth by carbon starvation, then swarmer cells are isolated by ludox gradient centrifugation. This procedure has been covered in depth recently at both large and small scales (Schrader and Shapiro 2015). For each ribosome profiling sample,

1 L of cells grown in M2G should be synchronized, and the 30 mL of resulting swarmer cells at  $OD_{600}=0.5$  should be grown to the desired cell cycle stage and rapidly harvested for ribosome profiling and total-RNA-seq.

#### 4.2.1.1 – Equipment



**Figure 4.1 Schematic of *Caulobacter* ribosome profiling procedure.** – Section 2.1 - Cell Growth, Synchronization and Harvesting: Grow 1 Liter of NA1000 cells, synchronize, and then grow to one of the cell cycle time points. Section 2.2 - Harvesting Cells: To each cell population, harvest by chloramphenicol pre-treatment (2 min at 28 °C) followed by rapid cell collection on ice. Section 2.3 - Cell Lysis: Lysis is performed under liquid nitrogen and a small portion is saved as a frozen aliquot for total RNA-seq. Section 2.4 - Footprinting and Footprint Extraction: Footprint mRNA fragments, select for size, and recover digested mRNA fragments. Section 2.5 - Total RNA-Seq Sample Preparation and PAGE Size Selection: Use the lysate aliquot from section 2.3 and prepare for total RNA-seq. Section 2.6 - Library Generation: Sequencing libraries are generated in the same way from the products of both ribosome footprinting and total RNA-Seq.



- Standard Shaker Incubator
- Spectrophotometer
- Optical Cuvettes
- Corex Tubes
- JA-20 Fixed-Angle Aluminum Rotor (Beckman Coulter)

#### 4.2.1.2 – Buffers and Reagents

- *Caulobacter crescentus* Strain NA1000 (Synchronizable Variant).
- PYE Media: 2 g of Bactopectone, 1 g of Yeast Extract, 2 mL of 0.5 M  $\text{MgSO}_4$ , 5 mL of 0.1 M  $\text{CaCl}_2$ , fill up to 1 L with Milli-Q filtered water (autoclaved solution).
- 20x M2 Salts: 17.4 g of  $\text{Na}_2\text{HPO}_4$ , 10.6 g of  $\text{KH}_2\text{PO}_4$ , 10 g of  $\text{NH}_4\text{Cl}$ , Resuspend in 1 L of Milli-Q filtered water (autoclaved solution).
- M2G Media: 50 mL of 20x M2 Salts, 1 mL of 0.5 M  $\text{MgSO}_4$ , 10 mL of 20% Glucose, 1 mL of Ferrous Sulfate Chelate Solution (Sigma Aldrich), fill to 995 mL with Milli-Q filtered water, and then add 5 mL of 0.1 M  $\text{CaCl}_2$  (add last after dilution to avoid precipitation). Sterilize with a 0.22  $\mu\text{m}$  filter.
- Ludox AS-40 Solution (Sigma).

#### 4.2.1.3 Procedure

- In a shaker incubator at 28 °C and 250 rpm, grow an overnight culture of *Caulobacter* strain NA1000 in 5 mL of PYE media.
- Transfer 200  $\mu\text{L}$  of saturated culture to 25 mL of M2G media and grow overnight in a shaker incubator at 28 °C and 250 rpm.
- Read the  $\text{OD}_{600}$  of the 25 mL culture using spectrophotometer.
- Using a doubling time of 2.33 hours, calculate how much of the 25 mL culture needs to be added to 1 L of culture to give a mid-log ( $\text{OD}_{600}=0.5$ ) 1 L culture in the morning.

- Verify the presence of swarmer cells using a phase contrast microscope and a wet mount of the cells.
- A detailed video protocol of the synchronization procedure can be found in (Schrader and Shapiro 2015).
- Harvest cells by centrifugation and suspend in 180 mL cold M2 salts.
- 60 mL of cold ludox solution should be mixed, and the resulting suspension should be pipetted into eight 30 mL corex tubes and spun for 30 min at 6400 g in a JA-20 rotor.
- The lower band is the swarmer band. The top of the solution can be aspirated off and discarded. The lower swarmer cell band should be collected from each tube and pooled in a new 50 mL conical tube.
- Check again by phase microscopy that the cells are highly enriched for swarmer cells (90-95% motile swarmer cells).
- The swarmer cells should be pelleted and washed twice in 30 mL of cold M2 each. Finally, the cells should be suspended in 30 mL of 28 °C M2G media at an  $OD_{600}=0.5$  and placed into a sterile 250 mL flask and shaken at 250 rpm until the desired cell cycle stage.

#### 4.2.1.4 Notes

- After synchronization, an entire 1 L cell synchrony allows for the isolation of one cell cycle ribosome profiling sample.
- It is important to examine and confirm that the cells are in the proper phase of the cell cycle by imaging with a wet mount on a phase-contrast microscope. Swarmer cells should be small and motile, stalked cells should be small and immotile, and predivisional cells should be larger with a dumbbell shape morphology.

#### 4.2.2 Harvesting cells

*Caulobacter* cells can only be harvested by centrifugation as their cells rapidly clog 0.22  $\mu\text{m}$  filters used for ribosome profiling of *B. subtilis* or *E. coli* cells (Oh et al. 2011). Harvesting

cells by centrifugation is a process that must be performed in a timely matter to minimize artifacts arising from the chloramphenicol addition. After removal from the shaker incubator, cell temperature needs to be rapidly lowered to 4 °C and maintained there until frozen in liquid nitrogen. A quick harvesting time and rapidly lowering cell temperatures by addition of ice cubes are necessary for the efficient harvesting of ribosomes trapped mRNA fragments.

#### **4.2.2.1 Equipment**

- Sorvall Legend X1R Centrifuge (Thermo Scientific)
- Fiberlite F15-8 x 50cy Fixed Angle Rotor (Thermo Scientific)
- 50 mL Conical Tube
- Standard Ice Cube Tray
- Hammer
- Zip-Top Bag
- Standard Syringe Needle

#### **4.2.2.2 Buffers and Reagents**

- 50 mg/mL Chloramphenicol solution in ethanol
- Ice cubes of M2G minimal media with 100 µg/ml of Chloramphenicol, 15 mL each, stored at -80 °C.
- Resuspension Buffer: 20 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 100 mM NH<sub>4</sub>Cl, 1 mM Chloramphenicol
- Lysis Buffer: 20 mM Tris pH 8.0, 10 mM MgCl<sub>2</sub>, 100 mM NH<sub>4</sub>Cl, 5 mM CaCl<sub>2</sub>, 0.4% Triton X100, 0.1% NP-40, 1 mM Chloramphenicol, 100 U/mL RNase free DNase I (Roche)
- Liquid Nitrogen

#### 4.2.2.3 Procedure

- Before harvesting the cells, take an ice cube out of the -80° C freezer, place ice cube in zip-top bag, crush using a hammer and place each crushed cube into a 50 mL conical tube and store the tube on ice. Each tube is sufficient for harvesting 30 mL of cells.
- For each cell cycle time point, 30 mL of cells at  $OD_{600}=0.5$  are treated with chloramphenicol (addition of 60  $\mu$ L chloramphenicol to a final concentration of 100  $\mu$ g/ml) and incubate in 28 °C shaker for 2 minutes at 250 rpm.
- Immediately add the cell culture to a 50 mL prechilled conical tube (4 °C) containing 15mL of crushed ice (M2G broth with 100  $\mu$ g/mL of chloramphenicol), cap the tube and invert twice.
- Place tubes in the prechilled, -10 °C F15-8 rotor and spin at 14,000 rpm (22,789 g) for 1.5 min to pellet cells.
- Discard supernatant by decanting and suspend cell pellet in 20 mL of prechilled Resuspension Buffer.
- Pellet cells again at 14,000 rpm (22,789 g) for 1.5 min to pellet cells in -10 °C F15-8 rotor.
- Discard supernatant by decanting and suspend cell pellet in 300  $\mu$ L of Lysis Buffer.
- Slowly drip the cell pellet/lysis buffer suspension into a liquid nitrogen filled 50 mL conical tube.
- Poke holes in 50 mL conical lid using a hot needle and pour out remaining liquid nitrogen in the tube. Store the cell pellet in a -80 °C freezer.

#### 4.2.2.4 Notes

- Cell collection should be performed as quickly as possible to avoid artifacts resulting from changes in translation occurring in response to the cell collection procedure.
- Cells should be cooled down rapidly by addition to crushed ice (M2G broth with 100  $\mu$ g/mL of chloramphenicol) and maintained chilled throughout the cell collection procedure.
- After the initial spin, the supernatant should still contain small pieces of ice.

- Frozen cell pellets can be stored at -80 °C for at least a year.

### **4.2.3 Cell Lysis**

Cell lysis is performed in a mixer mill with samples frozen in liquid nitrogen to minimize changes in ribosome position occurring in a liquid lysate (Oh et al. 2011). Save a small frozen aliquot for RNA-seq. Most of the lysate is used for ribosome footprinting.

#### **4.2.3.1 Equipment**

- Mixer Mill MM 400 (Roche)
- Grinding Jar 10 mL (Roche)
- 12 mm Grinding Ball (Roche)
- 1.5 mL RNase Free Microcentrifuge Tube

#### **4.2.2.2 Buffers and Reagents**

Not Applicable

#### **4.2.3.3 Procedure**

- Chill assembled jar and grinding ball in liquid nitrogen.
- Open jar and add frozen cell pellet/lysis buffer suspension to chilled jar with grinding ball and close lid.
- Mill at 15 Hz for 3 min. Chill the sealed jar in liquid nitrogen between each run. Five runs total.
- A small scoop of the frozen pulverized cells should be scraped out of the jars using a pre-chilled scoopula and stored in a pre-chilled 1.5 mL microcentrifuge tube at -80 °C for total RNA seq.
- To thaw the remaining lysate, place each half, chamber side up, in warm 30 °C water.

#### **4.2.3.4 Notes**

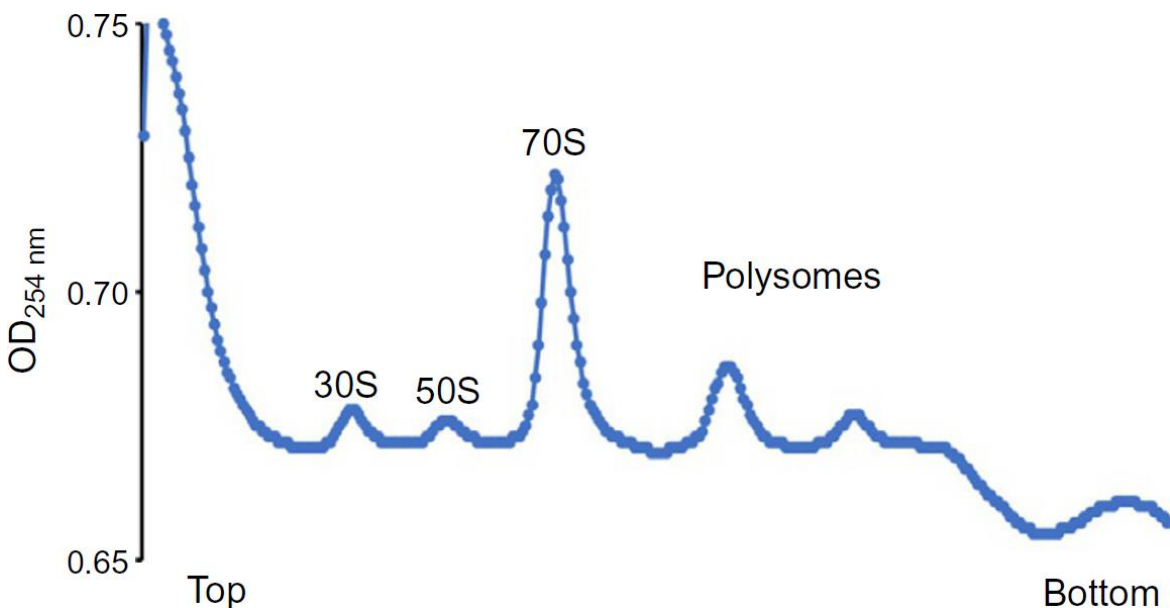
- Use cryogenic gloves when working with liquid nitrogen.

#### 4.2.4 Footprinting and Footprint Extraction

The thawed lysates are then footprinted by addition of micrococcal nuclease (MNase) which generates heterogeneous mRNA footprint sizes unlike the 28 nt footprints generated by eukaryotic ribosomes (Oh et al. 2011; Ingolia et al. 2009; O'Connor et al. 2013). MNase reactions can be quenched by addition of EGTA to chelate  $\text{Ca}^{2+}$  ions needed for nuclease activity. After footprinting with MNase, the 70S ribosomes are purified on a sucrose gradient. It is important to also run an undigested control on the sucrose gradient to ensure that polysomes were preserved upon cell harvesting and lysis (Figure 2).

##### 4.2.4.1 Equipment

- Standard Microcentrifuge
- Nanodrop Spectrophotometer (Thermo Fisher)
- Open-Top Polyclear Tubes (7030 Seton)
- TH-641 Swinging Bucket Rotor Package (Thermo Scientific) (SW 41 Ti equivalent)
- Sorvall WX+ Ultracentrifuge (Thermo Scientific)
- 2 mL Siliconized RNase Free Microcentrifuge Tubes



**Figure 4.2 – Example of *Caulobacter* cell cycle polysomes** Undigested cell lysate derived from a single cell cycle timepoint was separated on a 10-55% sucrose gradient.

- Standard Thermomixer
- Gradient Station (BIO-COMP)

#### 4.2.4.2 Buffers and Reagents

- 10 mM Tris-HCl pH 7.0, from 1 M stock (Ambion)
- Lysis Buffer: 20 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 100 mM NH<sub>4</sub>Cl, 5 mM CaCl<sub>2</sub>, 0.4% Triton X100, 0.1% NP-40, 1 mM Chloramphenicol, 100 U/mL RNase free DNase I (Roche)
- MNase Enzyme 1500 U/μL (Roche)
- SUPERase<sup>•</sup> In RNase Inhibitor Enzyme 20 U/μL (Ambion)
- 0.5 M EGTA pH 8.0
- 10% sucrose solution: 10% sucrose (w/v), 20 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 100 mM NH<sub>4</sub>Cl, 1 mM Chloramphenicol
- 50% sucrose solution: 50% sucrose (w/v), 20 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 100 mM NH<sub>4</sub>Cl, 1mM Chloramphenicol
- Gradient Balance Buffer: 20 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 100 mM NH<sub>4</sub>Cl, 1 mM Chloramphenicol
- 20% SDS
- Acid Phenol Chloroform (Ambion)
- Chloroform
- 3 M NaOAc pH 5.5 (Ambion)
- 100% isopropanol
- 80% ethanol from 100% stock
- Novex 2x TBE-Urea Sample Buffer (Invitrogen)

#### 4.2.4.3 Procedure

- Continue from section 2.3.3 by recovering the pulverized cells from the thawing chamber and placing them into a pre-chilled 1.5mL microcentrifuge tube for footprinting.

- Incubate the lysate for 20 minutes on ice.
- Pellet cell debris at 20,000 g for 10 min in a 4 °C microcentrifuge.
- Transfer supernatant to a pre-chilled siliconized microcentrifuge tube without disturbing pelleted cells.
- Dilute 1 µL of lysate with 99 µL of 10 mM Tris-HCl pH 7.0. Measure  $A_{260}$  by Nanodrop spectrophotometer (blank with 1 µL lysis buffer diluted into 99 µL of 10 mM Tris-HCl pH 7.0).
- Digest 500 mg of lysate RNA with 1.3 µL of MNase and 6 µL of SUPERase•In. Add lysis buffer for a 200 µL reaction. Treat undigested polysome control similarly, except without the addition of MNase.
- Incubate at 25 °C for 1 h while shaking in a thermomixer at 1400 rpm.
- Quench reaction with 2 µL of EGTA and immediately return to ice.
- Chill TH-641 swinging rotors and buckets to 4 °C
- Fill open-top polyclear tube with 55% sucrose solution until you reach the line on the maker block metal stand (~6 mL).
- Carefully layer 10% sucrose solution on top.
- Cap open-top polyclear tubes.
- From gradient using Gradient Master parameters for a SW 41 Ti rotor, short capped tubes, and 7%-47% weight by volume sucrose gradient.
- Remove cap and carefully layer nuclease digested cell lysate and undigested polysome control cell lysate into separate tubes.
- Use gradient buffer to balance tubes.
- Spin at 35,000 rpm (209,627 g) for 2.5 h at 4 °C.
- Fractionate using Gradient Station fractionator, with a piston speed of 0.2 mm/s.
- Manually collect fractionated monosomes, cap 2.0 mL siliconized microcentrifuge tube, flash-freeze in liquid nitrogen and stored at -80 °C indefinitely.



- To extract total RNA from footprinted monosomes, add 40  $\mu\text{L}$  of 20% SDS to 0.7 mL of thawed monosomes fraction solution. Monosomes may need to be split into two tubes of  $\sim 0.7$  mL each.
- Add 0.7 mL of acid phenol pre-warmed to 65  $^{\circ}\text{C}$  to the tube and mix by vortexing.
- Incubate at 65  $^{\circ}\text{C}$  for 5 min while shaking in a thermomixer.
- Chill on ice for 5 min.
- Spin at 20,000 g for 2 min in a microcentrifuge and transfer supernatant to fresh tube.
- Add 0.7 mL of room temperature acid phenol, mix by vortexing, and then incubate at room temperature for 5 min.
- Spin at 20,000 g for 2 min in a microcentrifuge and transfer aqueous layer to fresh tube.
- Add 0.6 mL of chloroform and mix by vortexing.
- Immediately spin at 20,000 g for 1 min in a microcentrifuge and transfer aqueous layer to fresh tube.
- Add 75  $\mu\text{L}$  of 3 M NaOAc pH 5.5 and 0.8 mL of 100% isopropanol. Mix by vortexing.
- Chill at -80  $^{\circ}\text{C}$  for 30 min.
- Pellet RNA at 20,000 g for 1 h in a 4  $^{\circ}\text{C}$  microcentrifuge.
- Wash pellet with 0.8 mL of -20  $^{\circ}\text{C}$  80% ethanol.
- Air dry pellet for 5 min and suspend in 20  $\mu\text{L}$  of 10 mM Tris-HCl pH 7.0.
- Dilute 1  $\mu\text{L}$  of precipitated RNA with 9  $\mu\text{L}$  of 10 mM Tris pH 7.0 and then quantify with Nanodrop spectrophotometer.
- Dilute 50  $\mu\text{g}$  of RNA to 20  $\mu\text{L}$  with 10 mM Tris-HCl pH 7.0 and then add 20  $\mu\text{L}$  of 2x TBE-Urea sample buffer.
- Save samples at -20  $^{\circ}\text{C}$  until section 2.5.3 and run size selection for both samples at once on 1x TBE 7M UREA PAGE gel.

#### 4.2.4.4 Notes

- When layering 10% sucrose on top of 55% sucrose in an open-top polyclear tube care must be taken, especially in the beginning, to ensure that there is no mixing at the boundary between the two layers.
- Carefully transport the sucrose gradient tubes to not disturb the boundary layer. Use smooth, even movements, especially when attaching and removing the buckets from the rotor.
- Surplus lysis buffer can be used to balance rotor buckets instead of gradient buffer.
- All 6 rotor buckets should be filled with sucrose and balanced for the spin even if all buckets are not loaded with samples.
- Fractionated monosomes typically have a final volume of 1.2 mL.
- Use siliconized microcentrifuge tubes to avoid transfer loss

#### 4.2.5 Total RNA-Seq Sample Preparation and PAGE Size Selection

To measure translation efficiency, mRNA levels are required to normalize the ribosome footprint density. To get high-quality total RNA-seq data, rRNA, which is abundant, should be removed from the sample. rRNA depleted total RNA should then be hydrolyzed under denaturing conditions to yield a size distribution that is similar to ribosome footprint fragment length.

##### 4.2.5.1 Equipment

- Standard Thermomixer
- Standard Microcentrifuge
- Standard 1-D Vertical Gel Electrophoresis Equipment
- 2 mL Siliconized Microcentrifuge Tube
- Corning Costar Spin-X Centrifuge Tube Filters (Fisher Scientific)
- 2100 Bioanalyzer (Agilent)

##### 4.2.5.2 Buffers and Reagents

- 20% SDS

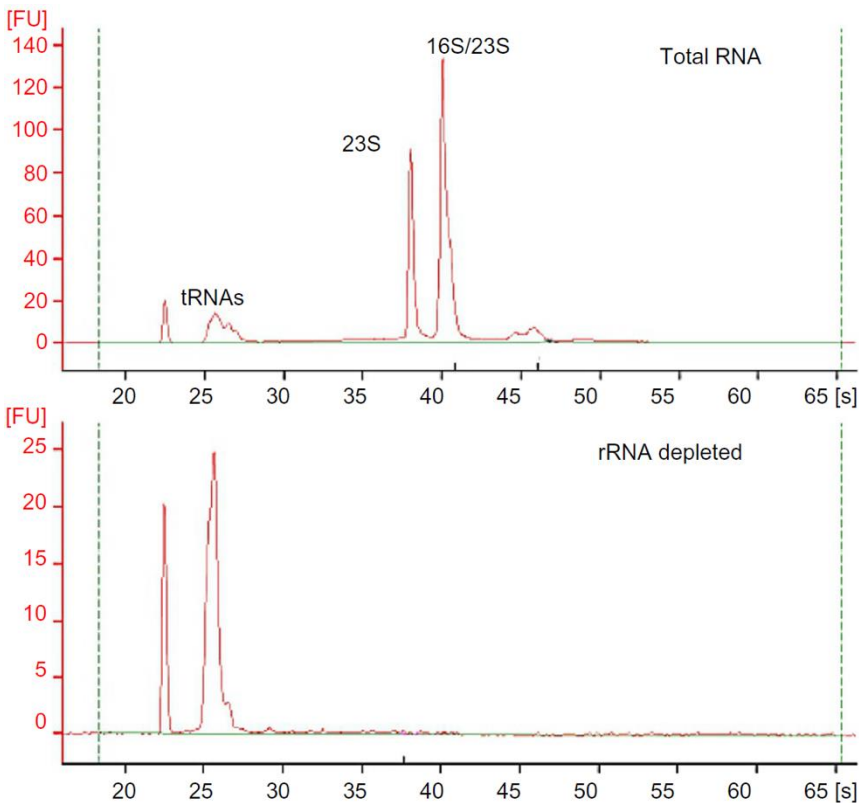
- Acid Phenol Chloroform (Ambion)
- Dry Ice
- 100% ethanol
- Chloroform
- 3 M NaOAc pH 5.5 (Ambion)
- 0.5 M EDTA
- 100% isopropanol
- 80% ethanol from 100% Stock
- 10 mM Tris-HCl pH 7.0, from 1 M stock (Ambion)
- Novex 2x TBE-Urea Sample Buffer (Invitrogen)
- 10 Base Pair Ladder (Invitrogen)
- Novex TBE-Urea Gels, 15%, 10 well (Invitrogen)
- 1x TBE from 10x stock (Ambion)
- SYBR Gold (Invitrogen)
- Elution Buffer: 10 mM Tris-HCl pH 7.0, 1 mM EDTA, 0.5 M NaOAc pH 5.5
- Glycogen (20 mg/mL)
- Bioanalyzer RNA 6000 Nano Kit (Agilent)
- Ribo-Zero rRNA Removal Kit: Gram-Negative Bacteria (Illumina)
- 2x alkaline hydrolysis buffer: 1  $\mu$ L of 0.5 M EDTA, 30  $\mu$ L of 0.1 M  $\text{Na}_2\text{CO}_3$ , 220  $\mu$ L of 0.1 M  $\text{NaHCO}_3$
- o199p RNA oligo (IDT)

#### 4.2.5.3 Procedure

- To frozen pulverized cell powder aliquot from section 2.3.3, add 0.6 ml of RNA lysis buffer including 40  $\mu$ L of 20% SDS.
- Immediately add 0.7 mL of acid phenol pre-warmed to 65  $^\circ\text{C}$  to the tube and mix by vortexing.

- Incubate at 65 °C for 5 min while shaking in a thermomixer.
- Chill in a dry ice - ethanol bath for 5 minutes (or until completely frozen).
- Spin at 20,000 g for 2 min in a microcentrifuge and transfer supernatant to a fresh tube.
- Add 0.7 mL of room temperature acid phenol, mix by vortexing, and then incubate at room temperature for 5 min.
- Spin at 20,000 g for 2 min in a microcentrifuge and transfer aqueous layer to fresh tube.
- Add 0.6 mL of chloroform and mix by vortexing.
- Immediately spin at 20,000 g for 1 min in a microcentrifuge and transfer aqueous layer to fresh tube.
- Add 65 µL of 3 M NaOAc pH 5.5 (or 1/9 volume) and 0.7 mL of 100% isopropanol (or 1.1 volume), and mix by vortexing.
- Chill at -80 °C for 30 minutes.
- Pellet RNA at 20,000 g for 1 h in a 4 °C microcentrifuge.
- Quickly wash pellet with 0.8 mL of -20 °C 80% ethanol.
- Suspend in 50 µL of 10 mM Tris-HCl pH 7.0.
- Assess RNA quality before and after ribosomal removal (Figure 3) using Bioanalyzer and RNA 6000 Nano Kit, following manufactures instructions.
- 5 µg of RNA in a volume of no more than 26 µL can used for the subsequent steps.
- Remove 16S and 23S ribosomal RNAs by subtractive hybridization using Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria). Follow manufactures instructions.
- The protocol requires washing the magnetic beads, treating samples with rRNA removal solution, removing RNA, and purifying rRNA depleted samples.
- Prepare 2x alkaline hydrolysis buffer and add 12.5 µL of 2x alkaline hydrolysis buffer to 12.5 µL of enriched mRNA.
- Incubate at 95 °C for 23 minutes and immediately return to ice.

- Add 560  $\mu\text{L}$  of stop buffer to end hydrolysis.
- Add 650  $\mu\text{L}$  of 100% isopropanol. Mix by vortexing
- Chill at  $-80\text{ }^{\circ}\text{C}$  for 30 min.
- Pellet hydrolyzed RNA at 20,000 g for 1 h in a  $4\text{ }^{\circ}\text{C}$  microcentrifuge.
- Wash pellet with 0.8 mL of  $-20\text{ }^{\circ}\text{C}$  80% ethanol.
- Dry and suspend pellet in 10  $\mu\text{L}$  of 10 mM Tris-HCl pH 7.0.
- Add 10  $\mu\text{L}$  of 2x TBE-Urea sample buffer.
- Prepare 10 base pair ladder in parallel with a solution containing: 1  $\mu\text{L}$  of 10 base pair ladder, 9  $\mu\text{L}$  of 10 mM Tris-HCl pH 7.0, and 10  $\mu\text{L}$  of 2x TBE-Urea sample buffer.



**Figure 4.3 – rRNA depletion of total RNA samples for total RNA-seq.** Bioanalyzer traces before (top) and after rRNA removal (bottom). Stable RNA peaks are labeled. *Caulobacter* 23s rRNA is processed into two fragments, one runs at the same size as the 16S rRNA and the other runs at a shorter length. Y-axis scales are different for the two graphs.

- Prepare o199-P control oligo in parallel with a solution containing: 2  $\mu\text{L}$  of 10 $\mu\text{M}$  o199-P, 8  $\mu\text{L}$  of 10 mM Tris-HCl pH 7.0, and 10  $\mu\text{L}$  of 2x TBE-Urea sample buffer.
- Thaw extracted ribosome footprints sample from section 2.4.3
- Denature ladder and RNA samples at 80  $^{\circ}\text{C}$  for 1 min and return to ice.
- Set up TBE-Urea Gel 15% in 1x TBE and pre-run for 1 hour at 200 volts.
- Wash and then load lanes. Total RNA samples and footprinting samples from section 2.4.3 should be loaded on the same gel. Run for 65 minutes at 200 volts.
- Stain gel with 6  $\mu\text{L}$  of SYBR Gold in 60 mL of 1x TBE for 5 min and excise desired bands from 30-40 nt based on ladder and o199p oligo.
- **RNA recovery from acrylamide gels** will be repeated in this procedure so it will be explained in detail here and only referenced subsequently.
- Pierce a needle through a 0.5 mL tube and insert gel slice in pierced tube and nest pierced tube into a 2 mL siliconized microcentrifuge tube.
- Spin at 20,000 g for 2 min in a microcentrifuge, or until most of the gel has extruded through, and transfer the remaining gel pieces.
- Add 0.5 mL of elution buffer and freeze at -80  $^{\circ}\text{C}$  for 20 minutes.
- Nutate overnight in a cold room or 4  $^{\circ}\text{C}$  refrigerator.
- Incubate at 25  $^{\circ}\text{C}$  for 5 min while shaking in a thermomixer at 1400 rpm.
- Transfer gel slurry to a Spin-X centrifuge tube filter using a wide bore pipette or a cut standard tip.
- Add 2 $\mu\text{L}$  of glycogen and 550 $\mu\text{L}$  of 100% isopropanol (1.1 equivalence volume), then mix by vortexing and chill at -20  $^{\circ}\text{C}$  for 1 hour.
- Pellet RNA at 20,000 g for 1 h in a 4  $^{\circ}\text{C}$  microcentrifuge and aspirate away the supernatant.
- Wash pellet with 0.8 mL of 4  $^{\circ}\text{C}$  80% ethanol.
- Pulse spin to aspirate away remaining ethanol and air dry for 5 min.

- Suspend recovered RNA in 15  $\mu$ L of 10 mM Tris-HCl pH 7.0.

#### 4.2.5.4 Notes

- o199p control and ladder help to identify appropriate RNA fragments.
- Ribosome footprints may need to be spread across up to 4 lanes of the gel due to high concentration to avoid overloading.

#### 4.2.6 Library Generation

Library preparation is performed by 3' end repair, 3' linker ligation, reverse transcription, single strand cDNA circularization, and finally PCR (Ingolia et al. 2012). These steps should be performed on both the footprinted mRNA fragments and the size selected mRNA fragments generated for mRNA-seq. Control o199p RNA oligo can be utilized throughout the library prep to ensure each step has worked efficiently. Quality control steps are noted below to ensure final DNA libraries contain a maximal amount of mRNA fragments. Library generation is performed as described in (Ingolia et al. 2012) but using the T4 RNA ligase KQ mutant (NEB) due to higher efficiency of ligation. DNA oligos utilized here have been adapted to work on the newer illumina sequencers including the Hiseq 4000 and Miniseq, as well as older sequencers (Table 2).

##### 4.2.6.1 Equipment

- Standard Thermomixer
- Standard Microcentrifuge
- Standard 1-D Vertical Gel Electrophoresis Equipment
- 2 mL Siliconized Microcentrifuge Tubes
- 0.2 mL PCR Tube
- Agilent Bioanalyzer
- Qubit Fluorimeter
- Corning Costar Spin-X Centrifuge Tube Filters (Fisher Scientific)
- Standard Nutator

#### 4.2.6.2 Buffers and Reagents

- 10x T4 PNK Buffer supplied with enzyme (New England Biolabs)
- T4 PNK Enzyme (New England Biolabs)
- SUPERase•In RNase Inhibitor Enzyme 20 U/ $\mu$ L (Ambion)
- Dephosphorylation Buffer Master Mix: 2  $\mu$ L of 10x T4 PNK buffer, 1  $\mu$ L SUPERase•In (for each reaction).
- 10 mM Tris-HCl pH 7.0 from 1 M stock (Ambion)
- 0.5 M EDTA
- 3 M NaOAc pH 5.5 (Ambion)
- Precipitation Master Mix: 10 mM Tris pH 7.0, 3 mM EDTA, 0.375M NaOAc pH 5.5.
- Glycogen (20mg/mL)
- 100% Isopropanol
- 100% Ethanol
- 80% Ethanol from 100% stock
- 50% sterile filtered PEG M.W. 8000
- 10x T4 RNA Ligase 2 Buffer, supplied with enzyme (New England Biolabs)
- 100% DMSO
- 1  $\mu$ g/ $\mu$ L of Linker-1 (New England Biolabs)
- T4 RNA Ligase 2, Truncated K227Q Enzyme (New England Biolabs)
- Novex 2x TBE-Urea Sample Buffer (Invitrogen)
- Ligation Master Mix: 8  $\mu$ L of 50% sterile filter PEG M.W., 2  $\mu$ L of 10x T4 Ligase 2 buffer, 2  $\mu$ L of 100% DMSO, 1  $\mu$ L of SUPERase•In, 1  $\mu$ L of 1  $\mu$ g/ $\mu$ L of Linker-1 (for each reaction).
- 10 Base Pair Ladder (Invitrogen)
- Novex TBE-Urea Gels, 15%, 10 well (Invitrogen)
- 1x TBE from 10x stock (Ambion)



- SYBR Gold (Invitrogen)
- Elution Buffer: 10 mM Tris-HCl pH 7.0, 1 mM EDTA, 0.5 M NaOAc pH 5.5
- 10 mM dNTP (New England Biolabs)
- 20  $\mu$ M o225-linker-mini-indexed (IDT)
- 5x FSB Buffer supplied with enzyme (Invitrogen)
- 0.1 M DTT supplied with enzyme (Invitrogen)
- Superscript III (Invitrogen)
- 1 N NaOH
- 10 mM Tris-HCl pH 8.0 from 1 M stock (Ambion)
- Nucleic Acid Master Mix: 1  $\mu$ L of 10 mM dNTP, 1  $\mu$ L of 25  $\mu$ M o225-linker-mini-indexed, 1.5  $\mu$ L of Milli-Q filtered water (for each reaction)
- Reverse Transcription Buffer Master Mix: 4  $\mu$ L of 5x FSB buffer, 1  $\mu$ L of 0.1 M DTT, 1  $\mu$ L of SUPERase•In (for each reaction)
- CircLigase Enzyme Kit (Epicentre)
- Circularization Master Mix: 2  $\mu$ L of 10x CircLigase buffer, 1  $\mu$ L of 1 mM ATP, 1  $\mu$ L of 50 mM  $MnCl_2$  (all components supplied with enzyme)
- High fidelity Phusion Enzyme (New England Biolabs)
- 10 mM dNTP (New England Biolabs)
- 100  $\mu$ M DNA oligos (from Table 2) (IDT)
- Amplification Master Mix: 16.7  $\mu$ L of 5x High Fidelity buffer (supplied with enzyme), 1.7  $\mu$ L of dNTP, 0.4  $\mu$ L of 100  $\mu$ M JS-ind (use separate oligos #1-12 for each biological sample), 0.4  $\mu$ L of 100  $\mu$ M oJS231, 59.2  $\mu$ L of Milli-Q filtered water, 0.8  $\mu$ L of High Fidelity Phusion Enzyme (for each sample)
- 6x DNA loading dye
- 0.05% Tween 20

- EBT: 10 mM Tris-HCl pH 8.0, 0.05% Tween 20
- High Sensitivity DNA kit (Agilent)

#### 4.2.6.3 Procedure

- Place the size selected mRNA footprints, size selected total mRNA fragments, and o199-P control sample on ice.
- To dephosphorylate the samples, prepare dephosphorylation buffer master mix and add to 15  $\mu$ L of the sample.
- Add 2  $\mu$ L of T4 PNK and incubate reaction at 37 °C for 1 h.
- Prepare precipitation master mix and add 90  $\mu$ L to each tube.
- Heat kill enzyme at 75 °C for 10 min. Acid:Phenol chloroform extraction will work as well.
- Add 0.3 mL of 100% ethanol, 2  $\mu$ L glycogen and mix by vortexing
- Chill at -20 °C for 1 h.
- Pellet dephosphorylated RNA at 20,000 g for 1 h in a 4 °C microcentrifuge.
- Remove supernatant, wash pellet with 0.8 mL of 4 °C 80% ethanol, pellet by spinning at 20,000 g for 15 min in a 4 °C microcentrifuge. Repeat washes two additional times.
- Air dry pellet for 5 min.
- Suspend pellet in 5  $\mu$ L of 10 mM Tris pH 7.0.
- Prepare the ligation master mix. The PEG will crash out of solution upon addition of DMSO. Warm the tube in your hands to solubilize.
- Add 2  $\mu$ L of T4 RNA Ligase 2, Truncated K227Q recording lot number, mix by vigorously pipetting and incubate reaction at 25 °C for 2.5 h.
- Add 20  $\mu$ L of 2x TBE-Urea sample buffer to the samples.
- Prepare 10 bp ladder in parallel: 1  $\mu$ L of 10 bp ladder, 5  $\mu$ L of 10 mM Tris-HCl pH 7.0, 6  $\mu$ L of 2x TBE-Urea sample buffer.
- Denature samples at 70 °C for 1 min and return to ice.

- Setup 10% TBE-Urea gel in 1x TBE and pre-run for 1 h at 200 volts. It is useful that o199-P negative control and ligated 0199-P positive control are included as sizing controls and to assess efficiency of ligation.
- Wash wells and load the gel. Every other lane should be skipped as to not cross contaminate samples.
- Run for 50 min at 200 volts.
- Stain gel with 6  $\mu\text{L}$  of SYBR gold in 60 mL of 1x TBE for 5 minutes.
- Excise the desired band and recover ligated products.
- To recover RNA, follow the procedure detailed at the end of section 2.5.3.
- Resuspend in 10  $\mu\text{L}$  of 10 mM Tris-HCl pH 7.0.
- For reverse transcription, prepare nucleic acid master mix and add 3.5  $\mu\text{L}$  to 10  $\mu\text{L}$  of ligated RNA.
- Denature at 60  $^{\circ}\text{C}$  for 5 minutes and return to ice.
- Prepare reverse transcription buffer master mix, add 6  $\mu\text{L}$  to the sample, and then add 1  $\mu\text{L}$  of Superscript III.
- Incubate at 50  $^{\circ}\text{C}$  for 30 minutes and then add 2.3  $\mu\text{L}$  of 1 N NaOH.
- Incubate at 95  $^{\circ}\text{C}$  for 15 min.
- To ethanol precipitate add 23  $\mu\text{L}$  of -20  $^{\circ}\text{C}$  100% ethanol (3 volumes), 2.3  $\mu\text{L}$  of sodium acetate (1/10 volume), and 2  $\mu\text{L}$  of glycogen then mix by vortexing and chill at -20  $^{\circ}\text{C}$  for 1 h.
- Pellet cDNA at 20,000 g for 1 h in a 4  $^{\circ}\text{C}$  microcentrifuge and then aspirate away the supernatant. Wash pellet with 1 mL of -20  $^{\circ}\text{C}$  80% ethanol. Repeat wash 2 more times.
- Pulse spin to aspirate away remaining ethanol and air dry for 5 minutes.
- After the pellet is dry, suspend the sample in 5  $\mu\text{L}$  of 1x Novex TBE-Urea Sample Buffer.
- Prepare 10 bp ladder in parallel: 1  $\mu\text{L}$  of 10 bp ladder, 9  $\mu\text{L}$  of 10 mM Tris-HCl pH 8.0, 10  $\mu\text{L}$  of 2x TBE-Urea sample buffer.

- Denature all samples at 80 °C for 2 min and return to ice.
- Setup 9% TBE-Urea gel in 1x TBE. Pre-run for 1 h at 200 volts.
- Wash lanes and load in samples. Run for 70 min at 200 volts checking predicted size versus dye.
- Stain gel with 6 µL of SYBR gold in 60 mL of 1 x TBE for 5 min.
- Excise desired bands and recover the reverse transcription products.
- To recover DNA pierce a needle through a 0.5 mL tube, insert gel slice in pierced tube, and nest pierced tube into a 2 mL siliconized microcentrifuge tube.
- Spin at 20,000 g for 2 minutes in a microcentrifuge or until most of the gel has extruded through. Transfer the remaining gel pieces.
- Add 0.5 mL of elution buffer (1/9 volume) and freeze at -80 °C for 20 minutes.
- Incubate at 70 °C for 10 min. Shake in thermomixer at 1400 rpm.
- Transfer gel slurry to a Spin-X centrifuge tube filter using a wide bore pipette or a cut standard tip.
- Add 2 µL of glycogen and 550 µL of 100% isopropanol (1.1 volume), then mix by vortexing and chill at -20 °C for 1 h.
- Pellet DNA at 20,000 g for 1 h in a 4 °C microcentrifuge and aspirate away the supernatant.
- Wash pellet with 0.8 mL of 4 °C 80% ethanol.
- Pulse spin to aspirate away remaining ethanol and air dry for 5 minutes.
- After DNA has been recovered, resuspend in 15 µL of 10 mM Tris-HCl pH 8.0.
- Add 4 µL of circularization master mix and 1 µL of CirLigase enzyme and incubate at 60 °C for 1 h. Include o225-linker-mini-indexed alone as negative control.
- Heat kill enzyme at 80 °C for 10 minutes and then place on ice.
- For PCR amplification, prepare amplification master mix and add 79.2 µL to 5µL of circularized ssDNA from proceeding steps. **PCR primers containing DNA barcodes to allow for**

**multiplexed sequencing on illumina sequencers should be used (Table 2) with a different index oligo for each sample.**

- Aliquot 17  $\mu$ L of PCR mix into four separate PCR strip tubes and perform the following PCR reaction:

Initial denaturing	98°C	30 seconds
Denaturing	98°C	10 seconds
Annealing	60°C	10 seconds
Extension	72°C	5 seconds

- Remove PCR tubes after cycles 6, 8, 10 and 12. **The sample with the fewest number of cycles with visible PCR product will be gel extracted.** As a control, include the o225-linker-mini-indexed to mark the size of no-insert products.
- Setup 8% TB polyacrylamide gel in 1x TBE.
- Directly load samples and run at 180 volts for 40 min.
- Stain gel with 6  $\mu$ L of SYBR gold in 60 mL of 1x TBE for 5 min.
- Excise insert DNA band and recover double stranded DNA products by piercing a needle through a 0.5 mL tube.
- Insert gel slice in pierced tube and nest pierced tube into a 2 mL siliconized microcentrifuge tube.
- Spin at 20,000 g for 2 minutes in a microcentrifuge, or until most of the gel has extruded through. Transfer the remaining gel pieces.
- Add 0.5 mL of elution buffer (1/9 volume) and freeze at -80 °C for 20 min.
- Incubate at 70 °C for 10 min in thermomixer at 1400 rpm
- Transfer gel slurry to a Spin-X centrifuge tube filter using a wide bore pipette or a cut standard tip.
- Add 2  $\mu$ L of glycogen and 550  $\mu$ L of 100% isopropanol (1.1 volume), mix by vortexing, and then chill at -20 °C for 1 h.

- Pellet DNA at 20,000 g for 1 h in a 4 °C microcentrifuge and aspirate away the supernatant.
- Wash pellet with 0.8 mL of 4 °C 80% ethanol.
- Pulse spin to aspirate away remaining ethanol and air dry for 5 min.
- Suspend amplified DNA in 10 µL of EBT.
- For quantification of amplified DNA, dilute 1 µL of amplified DNA with 9 µL of 10mM Tris-HCl pH 8.0.
- Assess DNA quality using the bioanalyzer by the steps outlined by Agilent for the High Sensitivity DNA Kit. Quantify DNA concentration using the Qubit fluorimeter.
- Samples should then be pooled in equal amounts for Illumina sequencing (50 bp single end) on an Illumina HiSeq instrument.

#### 4.2.6.4 Notes

- Bring the o199 control oligo through each step-in library prep to assess problematic reactions.
- Ligation efficiency markedly decreases if PEG is old so it is recommended that PEG be made within a month of use.
- T4 RNA Ligase 2, Truncated can be used for ligation instead of T4 RNA Ligase 2, Truncated K227Q mutant, but higher yields have been obtained with the K227Q mutant.
- Addition of 1 N NaOH and heating at 95 °C during the reverse transcription steps quenches the reaction by hydrolyzing RNA template.
- Loading the gel with small volumes avoids formation of “bunny ears” on gels.

#### 4.3 – Conclusion

If followed correctly, this procedure allows the genome-wide measurement of translation at the key stages of the *Caulobacter* cell cycle. For each ORF, the ratio of ribosome footprint density to total RNA-seq density can be used to calculate translation efficiency, which can be assessed across the cell cycle. Currently, *Caulobacter* is the only organism with a high-yield drug free method of cell cycle synchronization making it a powerful organism for the study of the

bacterial cell cycle. Innovative new methodologies in cell synchronization, such as “baby machines” (Bates et al. 2005), should allow for the expansion of ribosome profiling studies into

#### 4.4 – Acknowledgements

Research reported in this publication was supported by NIGMS of the National Institutes of Health under award number R35GM124733. The authors thank Wayne State University startup funds to JMS.

#### 4.5 – Supporting information

**Table 4.1 – Expected completion time for each section.**

Section	Approximate Work Time
2.1- Cell Growth, Synchronization and Harvesting	Three days
2.2- Harvesting Cells	1 hour
2.3- Cell Lysis	1.5 hours
2.4- Footprinting and Footprint Extraction	10 hours
2.5- Total RNA-Seq Sample Preparation and PAGE Size Selection	13 hours
2.6- Library Generation	Three days

**Table 4.2 – Oligos for library prep generation**

o199-P	5' AUGUACACGGAGUCGACCCGCAACGCGA/3phos/ 3'
Linker-1 (NEB)	5' App/CTGTAGGCACCATCAAT/3ddC 3'
o225-linker- mini- indexed	5' /5Phos/AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/iSp18/CACTCA/iSp18/CA AGCAGAAGACGGCATAACGAGATATTGATGGTGCCTACAG 3'
JS-ind1	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACATCACGACACTCTTTCCCTACACGACG 3'
JS-ind2	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCAACGATGTACACTCTTTCCCTACACGACG 3'
JS-ind3	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACTTAGGCACACTCTTTCCCTACACGACG 3'
JS-ind4	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACTGACCAACTCTTTCCCTACACGACG 3'
JS-ind5	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACACAGTGACACTCTTTCCCTACACGACG 3'
JS-ind6	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACGCCAATACACTCTTTCCCTACACGACG 3'
JS-ind7	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCAACGATCACACTCTTTCCCTACACGACG 3'
JS-ind8	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACACTTGAACACTCTTTCCCTACACGACG 3'
JS-ind9	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACGATCAGACACTCTTTCCCTACACGACG 3'
JS-ind10	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACTAGCTTACACTCTTTCCCTACACGACG 3'
JS-ind11	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACGGCTACACACTCTTTCCCTACACGACG 3'
JS-ind12	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACT CCAGTCACCTTGTAACACTCTTTCCCTACACGACG 3'
oJS231	5' CAAGCAGAAGACGGCATAACGAGATATTGATGGTGCC 3'



## **APPENDIX 5 – RIF-CORRECT – BACTERIAL MRNA HALF-LIFE ESTIMATION**

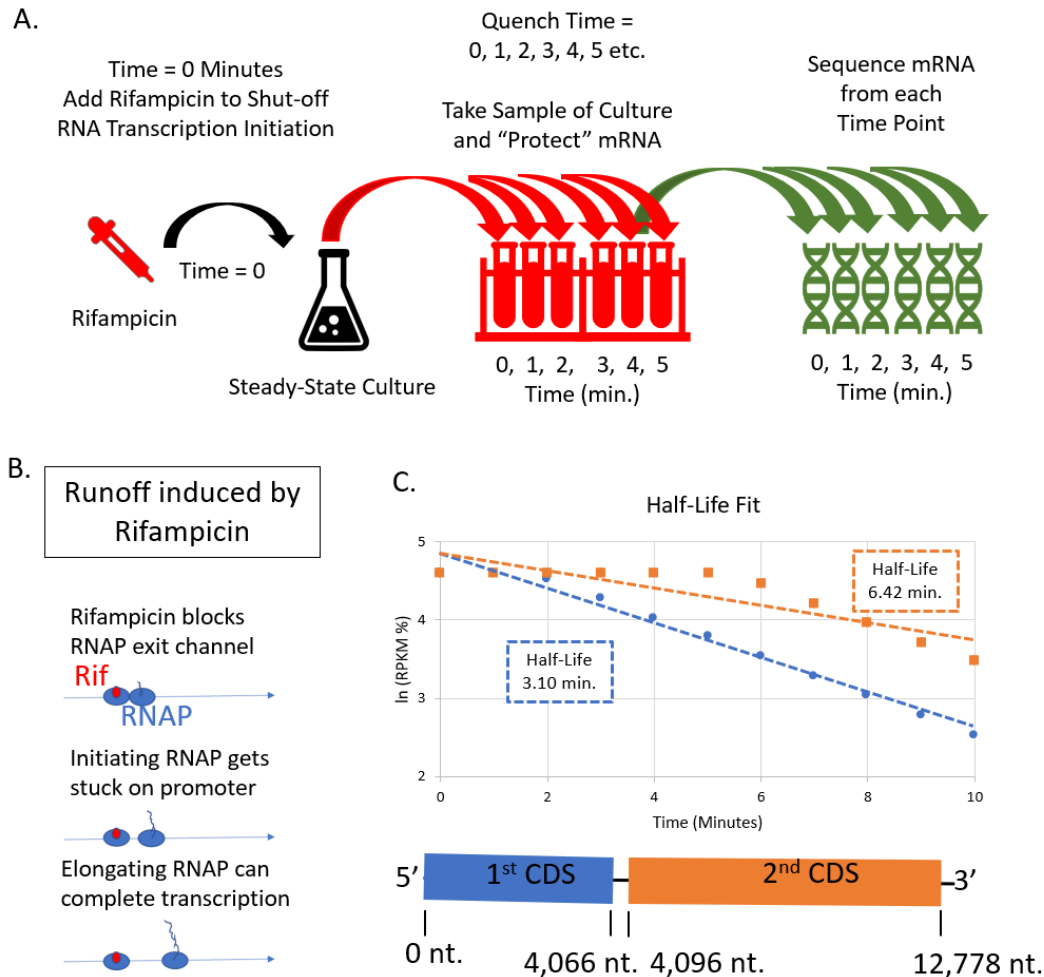
### **5.1 – Summary**

Genome-wide measurement of bacterial mRNA lifetimes using the antibiotic rifampicin has provided new insights into the control of bacterial mRNA decay. However, for long polycistronic mRNAs, the estimation of mRNA half-life can be confounded with transcriptional runoff caused by rifampicin's inhibition of initiating RNA polymerases, and not elongating RNA polymerases. We present the Rif-correct software package, a free open-source tool that uses transcriptome models of transcript architecture to allow for more accurate mRNA half-life estimates that account for the transcriptional runoff. Rif-correct is implemented as a customizable python script that allows for users to control all the analysis parameters to achieve improved mRNA half-life estimates. Rif-correct is the first free open source computational analysis pipeline for Rif-seq dataset mRNA half-life estimation. It is simple to run, fast, and easy to run with a detailed instruction manual and example datasets.

### **5.2 – Introduction**

mRNA half-lives can be measured genome-wide in bacteria using the antibiotic Rifampicin to shut of mRNA transcription and measuring abundance at different time-points by microarray or RNA-seq (Bernstein et al. 2002) (Fig 1 A, Rif-Seq description). While generally successful, complexities arise from the antibiotic rifampicin, which only blocks transcription initiation, and does not inhibit transcription elongation (Alifano et al. 2015) (Fig 1 B). Therefore, some elongating RNA polymerases can continue to transcribe and run off. Therefore, directly fitting a simple half-life equation to the natural log transformed mRNA abundances does not accurately fit the data, especially for long operons, causing an inaccurate determination of apparent half-life from the slope (Fig 1 C). To overcome this some groups have used expensive proprietary software to correct their data for transcriptional runoff (Sunney Xie 2015). Therefore Rif-Correct was developed to be a free open-source software that uses Rif-seq data and RNA polymerase

elongation rate to improve the mRNA half-life estimates by correcting for the mechanism of action issues inherent to rifampicin.



**Figure 5.1 – Using Rif-Seq to determine mRNA half-lives** A) The Methodology for the Rif-Seq experiment is shown. The antibiotic Rifampicin is added to a steady-state log phase bacterial culture at time = 0 to shut-off transcription initiation. Then samples of the culture are taken at different time points after the addition of Rifampicin and the cellular RNAs are quenched in an RNA protection solution. Finally, RNA-Seq libraries are made from each time point which report the relative genome-wide mRNA abundance (RPKM). B) Natural log transformed mRNA abundance (RPKM) for a two gene example operon with the blue circles representing the data for the first gene in the operon and the orange squares representing the data for the second gene in the operon. Linear regressions (shown as the blue and orange dotted line, respectively) does not fit the data very well with R values of -0.98 and -0.87. Importantly, both genes are part of the same mRNA with half-lives of 2.77 minutes, but the estimated half-lives are 3.10 minutes and 6.42 minutes. C) The antibiotic rifampicin’s mechanism of action inhibits RNAP transcription initiation by blocking the exit tunnel (Alifano et al. 2015), but not transcription elongation where the exit channel is already occupied by RNA. Consequently, there is a runoff of elongating RNAP that can cause a delay to the observed decay of the mRNA which is the source of poor curve fitting.

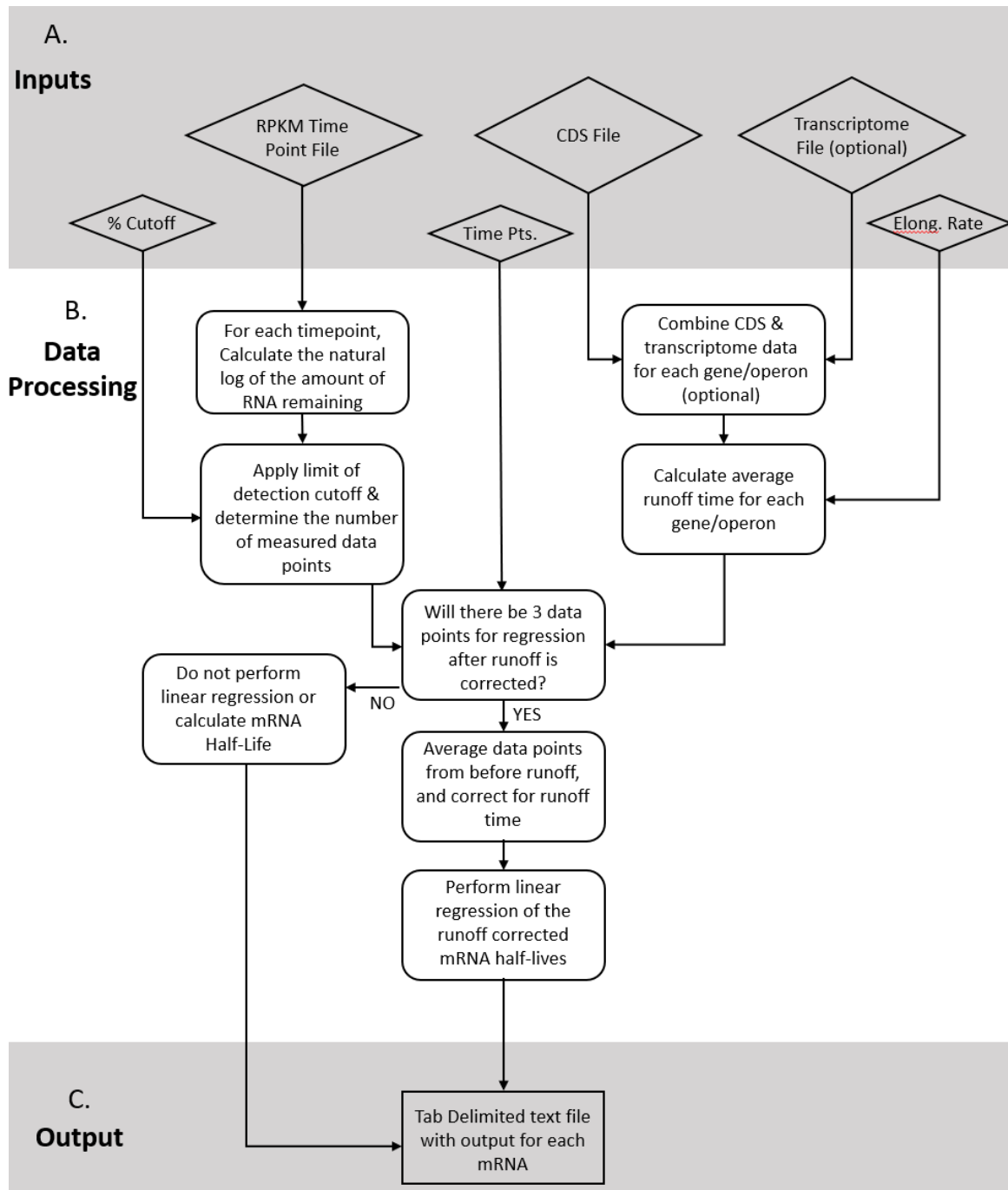
## 5.3 – Implementation

### 5.3.1 – Inputs

To run Rif-correct the user needs to input multiple data files (Fig 2 A). This includes 1) a tab delimited mRNA abundance file with mRNA abundance measurements at each time point after rifampicin collection together with the name for each mRNA. For genome-wide analysis, RNA abundances will be measured by RNA-seq or by microarray. 2) a tab delimited genome annotation file containing the locus tag for each coding sequence with genomic coordinates and strand information, the operon number and list of genes within each operon. This genome annotation file should be ordered with positive stand in genome order from low to high, and then negative strand in transcription order from high to low. To arrange the genome annotation file in this manner, a small standalone python script “CDS order” is provided to perform this step. 3) an optional transcriptome file can be added to take full advantage of the lag correction for operons it is needed. This file is an assembly of the CDS, transcription start site (TSS) data, and operon structure to build probably transcriptional units. This file includes: If CDS with that operon number are in a polycistronic operon, start gene of operon, last gene of operon, strand, farthest upstream promoter site (necessary) and up to 3 other promoter sites (optional), TSS type (optional), number of TSSs (optional), internal TSS (optional), a flag as to whether an operon is simple (single mRNA isoform predicted) or complex ((optional, multiple mRNA isoforms predicted), number and names of internal non-coding RNAs to operon (optional). A complete set of instructions to generate each file and example files are included.

In addition to these data files, the user is also prompted to enter multiple inputs for the data at the command prompt (Fig 2 A). First, the user must input the time points after rifampicin addition in which the RNA abundance measurements were collected. Second, the user must input the elongation rate for RNA polymerase on mRNA in nucleotides per second (An example for *E. coli* is 25 nt/sec in LB media (Chen et al. 2015), and *C. crescentus* is 19.36 nt/sec in M2G

media (Aretakis 2021 “Unpublished Data”). The lower limit of detection to be used in mRNA abundance estimation (optional), expressed in the percent of the original mRNA level remaining



**Figure 5.2 – Rif-Correct software workflow.** A) Inputs for the Rif-Correct Software is shown. Large diamonds denote input files for Rif-correct, smaller diamonds denote inputs typed into the command line by the user. B) Data processing steps and decisions for the Rif-Correct software is shown. C) Output for the Rif-Correct Software is a tab delimited text file. The output file contains information from each inputted mRNA.

(3% is recommended). The lowest acceptable correlation value from linear curve fitting that to be acceptable for reporting of a half-life ( $R > 0.7$  recommended, optional).

### 5.3.2 – Data Processing

The Rif-correct workflow is shown in Figure 2 B. First, the genome annotation file and the optional transcriptome file are combine into one data structure if provided. This creates one master list of all predicted transcriptional units with their corresponding information, but if the transcriptome file is not provided the list is populated with null values and the program will treat each gene as monocistronic and the start codon will be used instead of the TSS for transcription runoff time estimation. The transcription runoff time is calculated for each transcriptional unit based on the average elongation rate of the RNAP on mRNA, if the transcriptome file is not provided then the transcription runoff time is calculated from the start to the end of the CDS. If the elongation rate is set to 0 nt/sec then there will be no runoff correction for any gene, which is not recommended. To calculate runoff time, RNA polymerase elongation time is calculated to the middle of each CDS feature. This distance is calculated from half the distance from the farthest upstream TSS site, or first CDS start codon if no TSS is known (for operons this would be the first gene in the operon), to the end of the CDS site for that specific gene. For polycistronic operons, this runoff time is estimated for each CDS on each transcriptional unit. The mRNA abundance measurement data for each CDS and time point is calculated as a percent based on its level compared to the 0 minute time point, and then the natural log of this ratio is calculated. If the user has placed a limit of detection cutoff, any data points below the cutoff are discarded from further analysis. To account for the transcription runoff time, for each CDS any RNA abundance data points before the runoff time are averaged together, and included as a calculated data point with the estimated runoff time. The number of consecutive time points with RNA abundance measurements are then recorded for each CDS. Next, linear curve fitting is performed to calculate the Rif-corrected mRNA half-life. For curve fitting of mRNA half-lives, there must be at least three

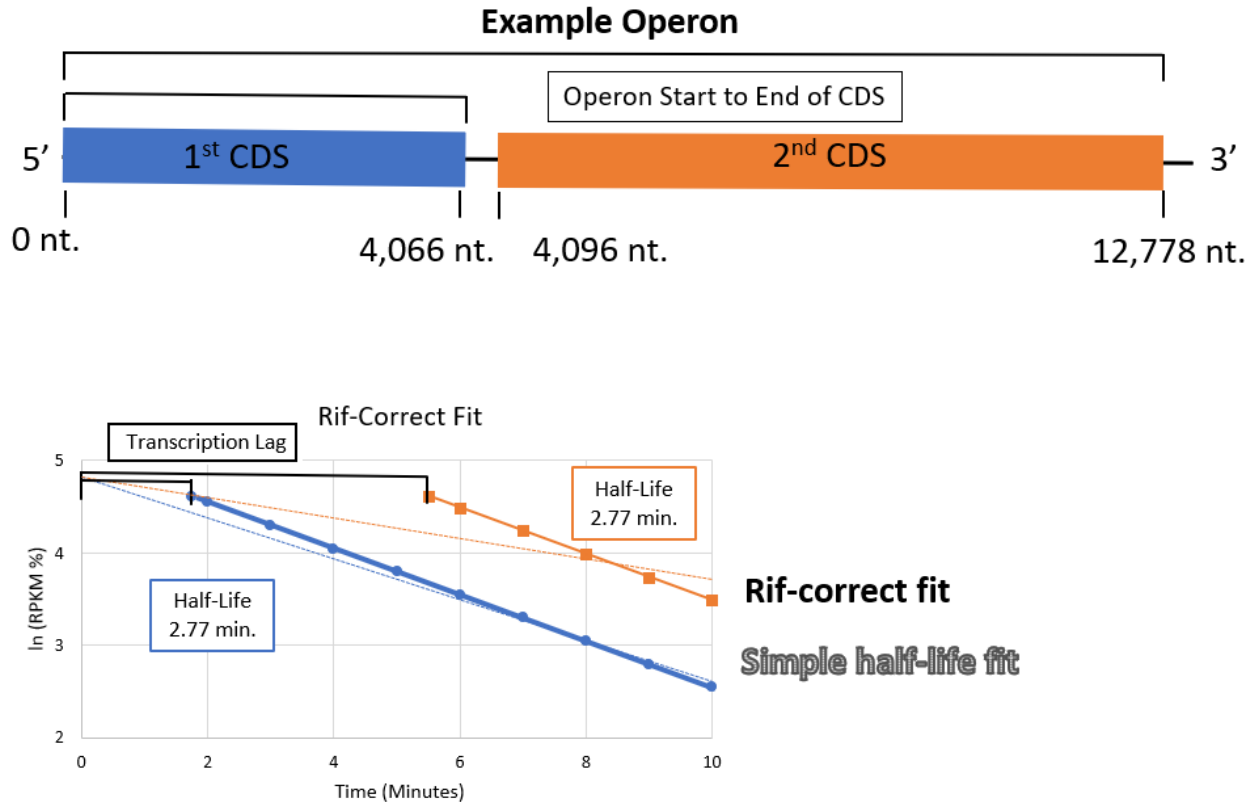
consecutive data points after runoff. If there are not enough data points for a CDS, then no linear regression is performed and no half-life is calculated for that gene and the data for the gene is written to the output file. If there are enough data points for a CDS then no calculation of mRNA half-life will be calculated. If the correlation value of curve fitting is below the optional curve fit threshold for a given CDS, no mRNA half-life will be calculated. Finally, a tab delimited text file is saved (Fig 2 C).

### 5.3.3 – Output

A tab delimited text file with the user inputted name is created in the directory where the program is run. This file contains: Slope of the Linear regression of the natural log transformed and transcription lag corrected data, Y intercept of the Linear regression of the natural log transformed and transcription lag corrected data, R value of the Linear regression of the natural log transformed and transcription lag corrected data, Rif-corrected half-life in minutes based in the slope ( $\ln(2)/\text{slope}$ ), number of datapoints used in linear regression after the low abundance threshold cutoff was applied, Number of data points averaged due to the runoff correction, R value flag if the linear regression R value is above the user inputted R value cutoff ( $R = -1$  is perfect natural log decay,  $r = 0$  is a flat line, and  $r = 1$  is an perfect increasing half-life), Too slow to measure mRNA half-Life flag - if the calculated half-life is negative or twice as long as the last time point used in the linear regression. There are also quality control outputs that match the gene with the user inputted gene information, to help trouble-shot any irrational results, these include user inputted time points and their associated natural log transformed data.

## 5.4 – Results/Discussion

Rif-Correct allows the use of transcriptome data for models of mRNA architecture (or from the CDS +1 site if the TSS is not known) to be used to correct for transcription runoff induced by rifampicin. While many short mRNAs do not require any correction for half-life estimation, long



**Figure 5.3 – Rif-Correct Software Fit.** The example operon data from figure 1 is shown with the 1<sup>st</sup> CDS region in blue, and the 2<sup>nd</sup> CDS region in orange. Bracketed lines represent the operon start to end of CDS used in the transcription runoff correction. Rif-Correct software improves curve fitting and corrected for the transcription runoff. Natural log transformed relative RPKM levels are shown for the two gene example operon with the blue circles representing the data for the first gene in the operon and the orange squares representing the data for the second gene in the operon. Linear regressions are shown for the simple half-life fits (shown as the blue and orange dotted line, respectively  $R^2 = -0.98$  and  $-0.87$ ), and for the Rif-Correct fit (shown as the blue and orange solid line, respectively). The difference in slope between the Rif-Correct fit (2.77 min, and 2.77 min) and the simple fit (3.10 min, 6.42 min) show how using the Rif-Correct Fit, allows for a more accurate half-life calculation (half-lives shown in the blue and orange solid boxes, respectively).

polycistronic operons are aided significantly by the Rif-correct algorithm (Fig 3). For example, hypothetical mRNA abundance data were generated for a hypothetical simple 2-gene operon and half-lives of each ORF region were fit to both a simple half-life equation (Fig 1C) or by the Rif-Correct package (Fig 3). Despite being part of the same mRNA with the same half-life, simple half-life fitting yields a half-life of 3.10 minutes for the first CDS ( $R=-0.98$ ), and 6.42 minutes for the second CDS ( $R=-0.87$ ), while Rif-Correct runoff correction is able to yield 2.77 minutes for the first CDS and 2.77 minutes for the second CDS. Overall, these longer mRNAs lead to artificially

extended mRNA lifetimes. When applied transcriptome wide, such analysis can help to ensure better mRNA half-lives are reported.

## **5.5 – Conclusion**

Rif-Correct is a free open source software that uses a transcriptome-based model to correct for transcriptional runoff to improve mRNA half-life estimates in bacteria. It is highly customizable with parameters to allow researchers to perform the analysis in the best way for their applications and is fast and easy to run. With a growing number of Rif-seq datasets (Vargas-Blanco and Shell 2020), Rif-correct will be broadly useful to bacterial mRNA decay researchers to help them to more easily calculate mRNA half-lives across the transcriptome.



## REFERENCES

- Al-Husini, N., D. T. Tomares, O. Bitar, W. S. Childers, and J. M. Schrader. 2018. 'alpha-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies', *Mol Cell*, 71: 1027-39 e14.
- Al-Husini, N., D. T. Tomares, Z. J. Pfaffenberger, N. S. Muthunayake, M. A. Samad, T. Zuo, O. Bitar, J. R. Aretakis, M. M. Bharmal, A. Gega, J. S. Biteen, W. S. Childers, and J. M. Schrader. 2020. 'BR-Bodies Provide Selectively Permeable Condensates that Stimulate mRNA Decay and Prevent Release of Decay Intermediates', *Mol Cell*, 78: 670-82 e8.
- Alifano, Pietro, Carla Palumbo, Daniela Pasanisi, and Adelfia Talà. 2015. 'Rifampicin-resistance, rpoB polymorphism and RNA polymerase genetic engineering', *Journal of biotechnology*, 202: 60-77.
- Aretakis, J. R., N. Al-Husini, and J. M. Schrader. 2018. 'Methodology for Ribosome Profiling of Key Stages of the *Caulobacter crescentus* Cell Cycle', *Methods Enzymol*, 612: 443-65.
- Aretakis, J. R., A. Gega, and J. M. Schrader. 2019. 'Absolute Measurements of mRNA Translation in *Caulobacter crescentus* Reveal Important Fitness Costs of Vitamin B12 Scavenging', *mSystems*, 4.
- Banani, S. F., H. O. Lee, A. A. Hyman, and M. K. Rosen. 2017. 'Biomolecular condensates: organizers of cellular biochemistry', *Nat Rev Mol Cell Biol*, 18: 285-98.
- Bates, D., J. Epstein, E. Boye, K. Fahrner, H. Berg, and N. Kleckner. 2005. 'The *Escherichia coli* baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle', *Mol Microbiol*, 57: 380-91.
- Bayas, Camille A, Jiarui Wang, Marissa K Lee, Jared M Schrader, Lucy Shapiro, and W.E. Moerner. 2018. 'Spatial organization and dynamics of RNase E and ribosomes in *Caulobacter crescentus*', *Proceedings of the National Academy of Sciences*, In Press.

- Benoit, Richard J. 1957. 'Preliminary Observations on Cobalt and Vitamin B12 in Fresh Water 1', *Limnology and Oceanography*, 2: 233-40.
- Berdis, A. J., I. Lee, J. K. Coward, C. Stephens, R. Wright, L. Shapiro, and S. J. Benkovic. 1998. 'A cell cycle-regulated adenine DNA methyltransferase from *Caulobacter crescentus* processively methylates GANTC sites on hemimethylated DNA', *Proc Natl Acad Sci U S A*, 95: 2874-9.
- Berge, M., S. Campagne, J. Mignolet, S. Holden, L. Theraulaz, S. Manley, F. H. Allain, and P. H. Viollier. 2016. 'Modularity and determinants of a (bi-)polarization control system from free-living and obligate intracellular bacteria', *Elife*, 5.
- Berge, M., and P. H. Viollier. 2018. 'End-in-Sight: Cell Polarization by the Polygamic Organizer PopZ', *Trends Microbiol*, 26: 363-75.
- Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen. 2002. 'Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays', *Proc Natl Acad Sci U S A*, 99: 9697-702.
- Bharmal, M. H., J. R. Aretakis, and J. M. Schrader. 2020. 'An Improved *Caulobacter crescentus* Operon Annotation Based on Transcriptome Data', *Microbiol Resour Announc*, 9.
- Bicknell, A. A., and E. P. Ricci. 2017. 'When mRNA translation meets decay', *Biochem Soc Trans*, 45: 339-51.
- Boel, G., R. Letso, H. Neely, W. N. Price, K. H. Wong, M. Su, J. D. Luff, M. Valecha, J. K. Everett, T. B. Acton, R. Xiao, G. T. Montelione, D. P. Aalberts, and J. F. Hunt. 2016. 'Codon influence on protein expression in *E. coli* correlates with mRNA levels', *Nature*, 529: 358-+.
- Brar, G. A., M. Yassour, N. Friedman, A. Regev, N. T. Ingolia, and J. S. Weissman. 2012. 'High-resolution view of the yeast meiotic program revealed by ribosome profiling', *Science*, 335: 552-7.

- Bremer, H., and P. P. Dennis. 2008. 'Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates', *EcoSal Plus*, 3.
- Bruggeman, F. J., and H. V. Westerhoff. 2007. 'The nature of systems biology', *Trends Microbiol*, 15: 45-50.
- Cannistraro, V. J., and D. Kennell. 1985. 'Evidence that the 5' end of lac mRNA starts to decay as soon as it is synthesized', *J Bacteriol*, 161: 820-2.
- Castellana, M., S. Hsin-Jung Li, and N. S. Wingreen. 2016. 'Spatial organization of bacterial transcription and translation', *Proc Natl Acad Sci U S A*, 113: 9286-91.
- Cavari, BENZION, and NATHAN Grossowicz. 1977. 'Seasonal distribution of vitamin B12 in Lake Kinneret', *Applied and environmental microbiology*, 34: 120-24.
- Chan, L. Y., C. F. Mugler, S. Heinrich, P. Vallotton, and K. Weis. 2018. 'Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability', *Elife*, 7.
- Chao, Y., L. Li, D. Girodat, K. U. Forstner, N. Said, C. Corcoran, M. Smiga, K. Papenfort, R. Reinhardt, H. J. Wieden, B. F. Luisi, and J. Vogel. 2017. 'In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways', *Mol Cell*, 65: 39-51.
- Chao, Y., K. Papenfort, R. Reinhardt, C. M. Sharma, and J. Vogel. 2012. 'An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs', *Embo j*, 31: 4005-19.
- Chen, H., K. Shiroguchi, H. Ge, and X. S. Xie. 2015. 'Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli', *Mol Syst Biol*, 11: 808.
- Chen, Y., D. Banerjee, A. Mukhopadhyay, and C. J. Petzold. 2020. 'Systems and synthetic biology tools for advanced bioproduction hosts', *Curr Opin Biotechnol*, 64: 101-09.

- Christen, B., E. Abeliuk, J. M. Collier, V. S. Kalogeraki, B. Passarelli, J. A. Coller, M. J. Fero, H. H. McAdams, and L. Shapiro. 2011. 'The essential genome of a bacterium', *Mol Syst Biol*, 7: 528.
- Christen, M., S. Deutsch, and B. Christen. 2015. 'Genome Calligrapher: A Web Tool for Refactoring Bacterial Genome Sequences for de Novo DNA Synthesis', *ACS Synth Biol*, 4: 927-34.
- Clarke, J. E., L. Kime, A. D. Romero, and K. J. McDowall. 2014. 'Direct entry by RNase E is a major pathway for the degradation and processing of RNA in Escherichia coli', *Nucleic Acids Res*, 42: 11733-51.
- Collier, J. 2016. 'Cell cycle control in Alphaproteobacteria', *Curr Opin Microbiol*, 30: 107-13.
- Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. 'WebLogo: a sequence logo generator', *Genome Res*, 14: 1188-90.
- Daisley, KW. 1969. 'Monthly survey of vitamin B12 concentrations in some waters of the English Lake District', *Limnology and Oceanography*, 14: 224-28.
- Dar, D., and R. Sorek. 2018. 'Bacterial Noncoding RNAs Excised from within Protein-Coding Transcripts', *mBio*, 9.
- Deana, A., and J. G. Belasco. 2005. 'Lost in translation: the influence of ribosomes on bacterial mRNA decay', *Genes Dev*, 19: 2526-33.
- Deana, A., H. Celesnik, and J. G. Belasco. 2008. 'The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal', *Nature*, 451: 355-8.
- Deveau, H., R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath, and S. Moineau. 2008. 'Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*', *J Bacteriol*, 190: 1390-400.
- Ducret, A., E. M. Quardokus, and Y. V. Brun. 2016. 'MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis', *Nat Microbiol*, 1: 16077.

- Duda, J, E Malinska, and Z Pedziwilk. 1957. 'Relation between the vitamin B12 content and the microorganism count in soil', *Acta microbiologica Polonica (1952)*, 6: 355-65.
- Edgar, R. C. 2004. 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Res*, 32: 1792-7.
- Evinger, M., and N. Agabian. 1977. 'Envelope-associated nucleoid from *Caulobacter crescentus* stalked and swarmer cells', *Journal of Bacteriology*, 132: 294-301.
- Ferber, Donna M, and Bert Ely. 1982. 'Resistance to amino acid inhibition in *Caulobacter crescentus*', *Molecular and General Genetics MGG*, 187: 446-52.
- Fiebig, A., J. Herrou, C. Fumeaux, S. K. Radhakrishnan, P. H. Viollier, and S. Crosson. 2014. 'A cell cycle and nutritional checkpoint controlling bacterial surface adhesion', *PLoS Genet*, 10: e1004101.
- Fioravanti, A., C. Fumeaux, S. S. Mohapatra, C. Bompard, M. Brilli, A. Frandi, V. Castric, V. Villeret, P. H. Viollier, and E. G. Biondi. 2013. 'DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other Alphaproteobacteria', *PLoS Genet*, 9: e1003541.
- Frasca, V., R. V. Banerjee, W. R. Dunham, R. H. Sands, and R. G. Matthews. 1988. 'Cobalamin-dependent methionine synthase from *Escherichia coli* B: electron paramagnetic resonance spectra of the inactive form and the active methylated form of the enzyme', *Biochemistry*, 27: 8458-65.
- Fumeaux, C., S. K. Radhakrishnan, S. Ardisson, L. Theraulaz, A. Frandi, D. Martins, J. Nesper, S. Abel, U. Jenal, and P. H. Viollier. 2014. 'Cell cycle transition from S-phase to G1 in *Caulobacter* is mediated by ancestral virulence regulators', *Nat Commun*, 5: 4081.
- Garcia-Bayona, L., M. S. Guo, and M. T. Laub. 2017. 'Contact-dependent killing by *Caulobacter crescentus* via cell surface-associated, glycine zipper proteins', *Elife*, 6.

- Gardner, P. P., J. Daub, J. Tate, B. L. Moore, I. H. Osuch, S. Griffiths-Jones, R. D. Finn, E. P. Nawrocki, D. L. Kolbe, S. R. Eddy, and A. Bateman. 2011. 'Rfam: Wikipedia, clans and the "decimal" release', *Nucleic Acids Res*, 39: D141-5.
- Grunenfelder, B., G. Rummel, J. Vohradsky, D. Roder, H. Langen, and U. Jenal. 2001. 'Proteomic analysis of the bacterial cell cycle', *Proc Natl Acad Sci U S A*, 98: 4681-6.
- Haakonsen, D. L., A. H. Yuan, and M. T. Laub. 2015. 'The bacterial cell cycle regulator GcrA is a sigma70 cofactor that drives gene expression from a subset of methylated promoters', *Genes Dev*, 29: 2272-86.
- Hentchel, K. L., L. M. Reyes Ruiz, P. D. Curtis, A. Fiebig, M. L. Coleman, and S. Crosson. 2019. 'Genome-scale fitness profile of *Caulobacter crescentus* grown in natural freshwater', *ISME J*, 13: 523-36.
- Holmes, J. A., S. E. Follett, H. Wang, C. P. Meadows, K. Varga, and G. R. Bowman. 2016. 'Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles', *Proc Natl Acad Sci U S A*, 113: 12490-95.
- Hui, S., J. M. Silverman, S. S. Chen, D. W. Erickson, M. Basan, J. Wang, T. Hwa, and J. R. Williamson. 2015. 'Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria', *Mol Syst Biol*, 11: 784.
- Ingolia, N. T. 2014. 'Ribosome profiling: new views of translation, from single codons to genome scale', *Nat Rev Genet*, 15: 205-13.
- Ingolia, N. T., G. A. Brar, S. Rouskin, A. M. McGeachy, and J. S. Weissman. 2012. 'The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments', *Nat Protoc*, 7: 1534-50.
- Ingolia, N. T., S. Ghaemmaghami, J. R. S. Newman, and J. S. Weissman. 2009. 'Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling', *Science*, 324: 218-23.

- Iniesta, A. A., N. J. Hillson, and L. Shapiro. 2010. 'Polar remodeling and histidine kinase activation, which is essential for *Caulobacter* cell cycle progression, are dependent on DNA replication initiation', *J Bacteriol*, 192: 3893-902.
- Jensen, R. B., and L. Shapiro. 2003. 'Cell-cycle-regulated expression and subcellular localization of the *Caulobacter crescentus* SMC chromosome structural protein', *Journal of Bacteriology*, 185: 3068-75.
- Jensen, R. B., S. C. Wang, and L. Shapiro. 2001. 'A moving DNA replication factory in *Caulobacter crescentus*', *Embo j*, 20: 4952-63.
- Jonas, K., J. Liu, P. Chien, and M. T. Laub. 2013. 'Proteotoxic stress induces a cell-cycle arrest by stimulating Lon to degrade the replication initiator DnaA', *Cell*, 154: 623-36.
- Judd, E. M., K. R. Ryan, W. E. Moerner, L. Shapiro, and H. H. McAdams. 2003. 'Fluorescence bleaching reveals asymmetric compartment formation prior to cell division in *Caulobacter*', *Proc Natl Acad Sci U S A*, 100: 8235-40.
- Kanehisa, M., and S. Goto. 2000. 'KEGG: kyoto encyclopedia of genes and genomes', *Nucleic Acids Res*, 28: 27-30.
- Kanehisa, M., S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. 2014. 'Data, information, knowledge and principle: back to metabolism in KEGG', *Nucleic Acids Res*, 42: D199-205.
- Kawano, M., A. A. Reynolds, J. Miranda-Rios, and G. Storz. 2005. 'Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*', *Nucleic Acids Res*, 33: 1040-50.
- Kelly, A. J., M. J. Sackett, N. Din, E. Quardokus, and Y. V. Brun. 1998. 'Cell cycle-dependent transcriptional and proteolytic regulation of FtsZ in *Caulobacter*', *Genes Dev*, 12: 880-93.
- Kozdon, J. B., M. D. Melfi, K. Luong, T. A. Clark, M. Boitano, S. Wang, B. Zhou, D. Gonzalez, J. Collier, S. W. Turner, J. Korlach, L. Shapiro, and H. H. McAdams. 2013. 'Global

- methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle', *Proc Natl Acad Sci U S A*, 110: E4658-67.
- Krab, I. M., and A. Parmeggiani. 1998. 'EF-Tu, a GTPase odyssey', *Biochim Biophys Acta*, 1443: 1-22.
- Lalanne, J. B., J. C. Taggart, M. S. Guo, L. Herzel, A. Schieler, and G. W. Li. 2018. 'Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry', *Cell*, 173: 749-61.e38.
- Lasker, K., T. H. Mann, and L. Shapiro. 2016. 'An intracellular compass spatially coordinates cell cycle modules in *Caulobacter crescentus*', *Curr Opin Microbiol*, 33: 131-39.
- Lasker, K., L. von Diezmann, X. Zhou, D. G. Ahrens, T. H. Mann, W. E. Moerner, and L. Shapiro. 2020. 'Selective sequestration of signalling proteins in a membraneless organelle reinforces the spatial regulation of asymmetry in *Caulobacter crescentus*', *Nat Microbiol*, 5: 418-29.
- Lau, J. H., J. F. Nomellini, and J. Smit. 2010. 'Analysis of high-level S-layer protein secretion in *Caulobacter crescentus*', *Can J Microbiol*, 56: 501-14.
- Laub, M. T., H. H. McAdams, T. Feldblyum, C. M. Fraser, and L. Shapiro. 2000. 'Global analysis of the genetic network controlling a bacterial cell cycle', *Science*, 290: 2144-8.
- Lee, S. F., M. A. Thompson, M. A. Schwartz, L. Shapiro, and W. E. Moerner. 2011. 'Super-resolution imaging of the nucleoid-associated protein HU in *Caulobacter crescentus*', *Biophys J*, 100: L31-3.
- Li, G. W. 2015. 'How do bacteria tune translation efficiency?', *Curr Opin Microbiol*, 24: 66-71.
- Li, G. W., D. Burkhardt, C. Gross, and J. S. Weissman. 2014. 'Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources', *Cell*, 157: 624-35.



- Liebermeister, W., E. Noor, A. Flamholz, D. Davidi, J. Bernhardt, and R. Milo. 2014. 'Visual account of protein investment in cellular functions', *Proc Natl Acad Sci U S A*, 111: 8488-93.
- Lin, Y., S. Crosson, and N. F. Scherer. 2010. 'Single-gene tuning of *Caulobacter* cell cycle period and noise, swarming motility, and surface adhesion', *Mol Syst Biol*, 6: 445.
- Maitra, A., and K. A. Dill. 2015. 'Bacterial growth laws reflect the evolutionary importance of energy efficiency', *Proc Natl Acad Sci U S A*, 112: 406-11.
- Martens, E., and A. L. Demain. 2017. 'The antibiotic resistance crisis, with a focus on the United States', *J Antibiot (Tokyo)*, 70: 520-26.
- McAdams, H. H., and L. Shapiro. 2011. 'The architecture and conservation pattern of whole-cell control circuitry', *Journal of Molecular Biology*, 409: 28-35.
- McGary, K., and E. Nudler. 2013. 'RNA polymerase and the ribosome: the close relationship', *Curr Opin Microbiol*, 16: 112-7.
- Menikpurage, I. P., D. Barraza, A. B. Melendez, S. Strebe, and P. E. Mera. 2019. 'The B12 receptor BtuB alters the membrane integrity of *Caulobacter crescentus*', *Microbiology (Reading)*, 165: 311-23.
- Molenaar, D., R. van Berlo, D. de Ridder, and B. Teusink. 2009. 'Shifts in growth strategies reflect tradeoffs in cellular economics', *Mol Syst Biol*, 5: 323.
- Montero Llopis, P., A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-Wagner. 2010. 'Spatial organization of the flow of genetic information in bacteria', *Nature*, 466: 77-81.
- Murray, Seán M, Gaël Panis, Coralie Fumeaux, Patrick H Viollier, and Martin Howard. 2013. 'Computational and genetic reduction of a cell cycle to its simplest, primordial components', *PLoS Biol*, 11: e1001749.
- Neidhardt, F. C. 1964. 'The regulation RNA synthesis in bacteria', *Prog Nucleic Acid Res Mol Biol*, 3: 145-81.

- Neidhardt, Frederick C. 1987. 'Escherichia coli and Salmonella', *Typhimurium Cellular and Molecular Biology*.
- O'Connor, P. B., G. W. Li, J. S. Weissman, J. F. Atkins, and P. V. Baranov. 2013. 'rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments', *Bioinformatics*, 29: 1488-91.
- Oh, E., A. H. Becker, A. Sandikci, D. Huber, R. Chaba, F. Gloge, R. J. Nichols, A. Typas, C. A. Gross, G. Kramer, J. S. Weissman, and B. Bukau. 2011. 'Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo', *Cell*, 147: 1295-308.
- Ohwada, Kouichi. 1973. 'Seasonal cycles of vitamin B12, thiamine and biotin in Lake Sagami. Patterns of their distribution and ecological significance', *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, 58: 851-71.
- Ohwada, Kouichi, and Nobuo Taga. 1972. 'VITAMIN B12, THIAMINE, AND BIOTIN IN LAKE SAGAMI 1', *Limnology and Oceanography*, 17: 315-20.
- Poindexter, J. S. 1964. 'Biological Properties and Classification of the Caulobacter Group', *Bacteriol Rev*, 28: 231-95.
- Pommel, Bernard. 1975. 'Distribution et signification écologique de la vitamine B12 et de la thiamine dans trois lacs subalpins et jurassien'.
- Presnyak, V., N. Alhusaini, Y. H. Chen, S. Martin, N. Morris, N. Kline, S. Olson, D. Weinberg, K. E. Baker, B. R. Graveley, and J. Collier. 2015. 'Codon optimality is a major determinant of mRNA stability', *Cell*, 160: 1111-24.
- Price, M. N., K. M. Wetmore, R. J. Waters, M. Callaghan, J. Ray, H. Liu, J. V. Kuehl, R. A. Melnyk, J. S. Lamson, Y. Suh, H. K. Carlson, Z. Esquivel, H. Sadeeshkumar, R. Chakraborty, G. M. Zane, B. E. Rubin, J. D. Wall, A. Visel, J. Bristow, M. J. Blow, A. P. Arkin, and A. M. Deutschbauer. 2018. 'Mutant phenotypes for thousands of bacterial genes of unknown function', *Nature*, 557: 503-09.

- Russell, J. B., and G. M. Cook. 1995. 'Energetics of bacterial growth: balance of anabolic and catabolic reactions', *Microbiol Rev*, 59: 48-62.
- Sahin, U., K. Kariko, and O. Tureci. 2014. 'mRNA-based therapeutics--developing a new class of drugs', *Nat Rev Drug Discov*, 13: 759-80.
- Sayers, E. W., R. Agarwala, E. E. Bolton, J. R. Brister, K. Canese, K. Clark, R. Connor, N. Fiorini, K. Funk, T. Hefferon, J. B. Holmes, S. Kim, A. Kimchi, P. A. Kitts, S. Lathrop, Z. Lu, T. L. Madden, A. Marchler-Bauer, L. Phan, V. A. Schneider, C. L. Schoch, K. D. Pruitt, and J. Ostell. 2019. 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Res*, 47: D23-D28.
- Schrader, J. M., G. W. Li, W. S. Childers, A. M. Perez, J. S. Weissman, L. Shapiro, and H. H. McAdams. 2016. 'Dynamic translation regulation in *Caulobacter* cell cycle control', *Proc Natl Acad Sci U S A*, 113: E6859-E67.
- Schrader, J. M., and L. Shapiro. 2015. 'Synchronization of *Caulobacter crescentus* for investigation of the bacterial cell cycle', *J Vis Exp*.
- Schrader, J. M., B. Zhou, G. W. Li, K. Lasker, W. S. Childers, B. Williams, T. Long, S. Crosson, H. H. McAdams, J. S. Weissman, and L. Shapiro. 2014. 'The coding and noncoding architecture of the *Caulobacter crescentus* genome', *PLoS Genet*, 10: e1004463.
- Scott, M., S. Klumpp, E. M. Mateescu, and T. Hwa. 2014. 'Emergence of robust growth laws from optimal regulation of ribosome synthesis', *Mol Syst Biol*, 10: 747.
- Sender, R., S. Fuchs, and R. Milo. 2016. 'Revised Estimates for the Number of Human and Bacteria Cells in the Body', *PLoS Biol*, 14: e1002533.
- Shell, S. S., J. Wang, P. Lapierre, M. Mir, M. R. Chase, M. M. Pyle, R. Gawande, R. Ahmad, D. A. Sarracino, T. R. Ioerger, S. M. Fortune, K. M. Derbyshire, J. T. Wade, and T. A. Gray. 2015. 'Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape', *PLoS Genet*, 11: e1005641.

- Shen, X., J. Collier, D. Dill, L. Shapiro, M. Horowitz, and H. H. McAdams. 2008. 'Architecture and inherent robustness of a bacterial cell-cycle control system', *Proc Natl Acad Sci U S A*, 105: 11340-5.
- Spencer, W., R. Siam, M. C. Ouimet, D. P. Bastedo, and G. T. Marczyński. 2009. 'CtrA, a global response regulator, uses a distinct second category of weak DNA binding sites for cell cycle transcription control in *Caulobacter crescentus*', *J Bacteriol*, 191: 5458-70.
- Stumpf, C. R., M. V. Moreno, A. B. Olshen, B. S. Taylor, and D. Ruggero. 2013. 'The translational landscape of the mammalian cell cycle', *Mol Cell*, 52: 574-82.
- Subramanian, K., and J. J. Tyson. 2017. 'Spatiotemporal Models of the Asymmetric Division Cycle of *Caulobacter crescentus*', *Results Probl Cell Differ*, 61: 23-48.
- Taylor, J. A., M. C. Ouimet, R. Wargachuk, and G. T. Marczyński. 2011. 'The *Caulobacter crescentus* chromosome replication origin evolved two classes of weak DnaA binding sites', *Mol Microbiol*, 82: 312-26.
- Thanbichler, M., A. A. Iniesta, and L. Shapiro. 2007. 'A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in *Caulobacter crescentus*', *Nucleic Acids Res*, 35: e137.
- Thanbichler, M., and L. Shapiro. 2006. 'MipZ, a spatial regulator coordinating chromosome segregation with cell division in *Caulobacter*', *Cell*, 126: 147-62.
- Vargas-Blanco, D. A., and S. S. Shell. 2020. 'Regulation of mRNA Stability During Bacterial Stress Responses', *Front Microbiol*, 11: 2111.
- Venetz, Jonathan E, Luca Del Medico, Alexander Wölfle, Philipp Schächle, Yves Bucher, Donat Appert, Flavia Tschan, Carlos E Flores-Tinoco, Mariëlle Van Kooten, and Rym Guennoun. 2019. 'Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality', *Proceedings of the National Academy of Sciences*, 116: 8070-79.

- Werner, J. N., E. Y. Chen, J. M. Guberman, A. R. Zippilli, J. J. Irgon, and Z. Gitai. 2009. 'Quantitative genome-scale analysis of protein localization in an asymmetric bacterium', *Proc Natl Acad Sci U S A*, 106: 7858-63.
- Whitfield, C. D., E. J. Steers, Jr., and H. Weissbach. 1970. 'Purification and properties of 5-methyltetrahydropteroyltriglutamate-homocysteine transmethylase', *J Biol Chem*, 245: 390-401.
- Wilhelm, R. C. 2018. 'Following the terrestrial tracks of *Caulobacter* - redefining the ecology of a reputed aquatic oligotroph', *ISME J*, 12: 3025-37.
- Wright, R., C. Stephens, G. Zweiger, L. Shapiro, and M. R. Alley. 1996. 'Caulobacter Lon protease has a critical role in cell-cycle control of DNA methylation', *Genes Dev*, 10: 1532-42.
- Wu, X., D. L. Haakonsen, A. G. Sanderlin, Y. J. Liu, L. Shen, N. Zhuang, M. T. Laub, and Y. Zhang. 2018. 'Structural insights into the unique mechanism of transcription activation by *Caulobacter crescentus* GcrA', *Nucleic Acids Res*, 46: 3245-56.
- Yang, S., S. Kim, D. K. Kim, H. Jeon An, J. Bae Son, A. Heden Gynna, and N. Ki Lee. 2019. 'Transcription and translation contribute to gene locus relocation to the nucleoid periphery in *E. coli*', *Nat Commun*, 10: 5131.
- Zhang, Y., D. A. Rodionov, M. S. Gelfand, and V. N. Gladyshev. 2009. 'Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization', *BMC Genomics*, 10: 78.
- Zhao, Wei, Samuel W Duvall, Kimberly A Kowallis, Dylan T Tomares, Haley N Petitjean, and W Seth Childers. 2018. 'A circuit of protein-protein regulatory interactions enables polarity establishment in a bacterium', *BioRxiv*: 503250.
- Zhou, B., J. M. Schrader, V. S. Kalogeraki, E. Abeliuk, C. B. Dinh, J. Q. Pham, Z. Z. Cui, D. L. Dill, H. H. McAdams, and L. Shapiro. 2015. 'The global regulatory architecture of transcription during the *Caulobacter* cell cycle', *PLoS Genet*, 11: e1004831.

**ABSTRACT****GLOBAL COORDINATION BETWEEN GENE EXPRESSION PROCESSES IN *C. CRESCENTUS***

by

**JAMES RAYMOND ARETAKIS****August 2021****Advisor:** Dr. Jared Schrader**Major:** Biological Sciences**Degree:** Doctor of Philosophy

Gene expression is an essential process of all organisms which occurs in two steps, mRNA transcription, and translation and which is shut off by mRNA decay. For a bacterium to grow and divide, these steps need to be tuned to optimize the synthesis rates of thousands of proteins from their genes. Genome-wide experiments were used to determine the absolute transcription rates, mRNA degradation rates, and translation rates on a gene by gene basis in the model bacterium *Caulobacter crescentus*. This allows illumination of the interconnections between the sub-steps of gene expression. Co-transcriptional mRNA decay was determined to affect nearly a third of mRNAs in the bacterium, which likely has regulatory consequences for these genes. Different strategies for producing mRNA at the correct level were found that either optimize for efficiency or for regulatory control. We observed that there is strong evidence that higher translation efficiency and slows mRNA decay, and that this is likely due to protection provided by ribosomes sterically hindering the RNA degradosome from cutting the mRNA. Finally, absolute translation rates revealed the fitness costs of Vitamin B<sub>12</sub> scavenging from the highly expressed Vitamin B<sub>12</sub> importer protein and B<sub>12</sub> dependent and independent methionine biosynthesis pathways. We found that across *Caulobacter* species there was a tradeoff where obligate scavengers had higher fitness when growing in a vitamin B<sub>12</sub> rich environment, but lower fitness when growing in vitamin B<sub>12</sub> deficient environment due to the high cost of protein synthesis.

**AUTOBIOGRAPHICAL STATEMENT****JAMES RAYMOND ARETAKIS****EDUCATION:**

2016 – 2019 Ph.D. (Biology), Wayne State University, Detroit, MI

2005 – 2009 B.Sc. (Biology), University of Michigan, Ann Arbor, MI

**PUBLICATIONS (Orcid ID: <https://orcid.org/0000-0002-0978-6658>):**

Bharmal, Mohammed-Husain\*, **James R. Aretakis\***, and Jared M. Schrader. "An Improved *Caulobacter crescentus* Operon Annotation Based on Transcriptome Data." *Microbiology Resource Announcements* 9.44 (2020).

**\*Co-First Author**

Al-Husini, N., Tomares, D.T., Pfaffenberger, Z.J., Muthunayake, N.S., Samad, M.A., Zuo, T., Bitar, O., **Aretakis, J.R.**, Bharmal, M.H.M., Gega, A. and Biteen, J.S. (2020). BR-bodies provide selectively permeable condensates that stimulate mRNA decay and prevent release of decay intermediates. *Molecular cell*, 78(4), 670-682.

**Aretakis, J.**, Alisa Gega, and Jared M. Schrader. (2019) Absolute measurements of mRNA translation in *C. crescentus* reveal important fitness costs of vitamin B12 scavenging. *mSystems*

**Aretakis, J.**, Al-Husini, N., Schrader, J.M. (2018) Methodology for ribosome profiling of key stages of the *Caulobacter crescentus* cell cycle. *Methods in Enzymology*.

**AWARDS:**

2020, P. Dennis Smith Award for Research in Genetics (Wayne State University, Detroit, MI)

2019, Graduate Research Award (Wayne State University, Detroit, MI)

2019-2020, Thomas C. Rumble Fellowship (Wayne State University, Detroit, MI)

2018, Graduate Teaching Award (Wayne State University, Detroit, MI)

2018, B.E.S.T. Program Phase III "Exploration Grant" (Wayne State University, Detroit, MI)

2017, Best Poster Presentation: Bio. Department Retreat (Wayne State University, Detroit, MI)

2017, Chemistry Biology Initiative Fellowship (Wayne State University, Detroit, MI)

2017, Best Poster Presentation (American Society of Microbiology Michigan Branch Meeting)