THE UNIVERSITY OF CHICAGO

QUANTITATIVE APPROACHES TO STUDY RNA-MEDIATED GENE REGULATION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
AND
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOPHYSICAL SCIENCES

BY
MATTHEW ALAN REYER

CHICAGO, ILLINOIS
JUNE 2021

Dedicated to the homie Big J. Sorry I wasn’t there.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to acknowledge my two advisors, Jingyi Fei and Aaron Dinner. I spent most of my time in Jingyi's lab, but I valued the time with Aaron that I did have. Thank you Jingyi and Aaron for all the work and time you invested into me to help make me the scientist I am today.

I would also like to acknowledge members of the Fei and Dinner labs, both past and present, including Jeff Zhang, Seongjin Park, Emily Heideman, Shriram Chennakesavalu, Magda Bujnowska, Shafi Azam, Amine Driouchi, and Jinjun Wu in the Fei Lab, and Lu Hong in the Dinner Lab. Thank you all.

Next, I thank the Biophysics program, and in particular Michele Wittels, Julie Feder, and Adam Hammond. I do not know what I would have done without Michele Wittels and Julie Feder. They made my life so much easier by allowing me to just focus on my research. Adam, thank you for always having your door open, I always appreciate our conversations.

Thank you to my friends, both Chicago- and Oregon-type. You always had my back.

And finally, thank you to my family, the most important people in my life, for affording me the luxury of pursuing my interests.

ABSTRACT

Bacteria exist in stressful and constantly changing environments. In order to to sense, respond to, and survive in various stress conditions, microbes have evolved diverse mechanisms to regulate their gene expression. RNAs, and sRNAs in particular, are prevalent mediators of gene regulation in bacteria. The identities of many sRNAs and the stresses they are associated with have often been revealed by traditional biochemical means, but a deeper understanding of the underlying physics that governs how RNAs efficiently regulate their targets could allow us to manipulate and leverage sRNAs for the sake of synthetic biology and human health.

Here, I present my contributions to the quantitative study of RNA-mediated gene regulation. First, I present the general model that describes how individual kinetic parameters contribute to the overall efficacy of sRNA-mediated regulation and the hierarchy of the sRNA-regulon. Additionally, I found that certain sRNAs that have canonically been described as post-transcriptional regulators can regulate their targets co-transcriptionally. Next, I describe our recent work regarding interactions between the ribosome and the sRNA accessory protein, Hfq, and how such interactions might contribute to the sRNA target search process. Finally, I offer a set of the image analysis tools I have developed that can be utilized to study RNA-mediated regulation. In all, these tools and findings can help develop quantitative descriptions of the physics of RNA-mediated gene regulation.

CHAPTER 1

INTRODUCTION

1.1 Gene regulation in bacteria

Bacteria must adapt to stressful and changing environments in order to survive. Bacteria also have DNA. The specific ways that bacteria convey the genetic information contained in DNA in response to the environment dictates whether or not they survive.

The genetic information encoded in DNA is converted to a functional gene product via a process broadly referred to as ”gene expression.” Gene expression canonically follows the blueprint defined by the ”central dogma of molecular biology;” DNA is copied into messenger RNA (mRNA) by the RNA Polymerase (RNAP), in a process called transcription, and mRNA is decoded by the ribosome to produce protein in a process called translation. Both protein and certain types of RNA can serve as functional gene products.

Gene expression is one of the most, if not the most, studied processes in biology. And though the general scheme of gene expression is well-understood, it is the specifics of the process that determine how a cell fares in its environment. In particular, the quantity of gene products, and the timing of their production, can determine the fate of the cell. A given number of functional gene products may be appropriate for one environment, but incompatible with another. Because cells exist in changing, and potentially hostile, environments, they must be able to rapidly respond to those changes by adjusting their gene expression accordingly. The process by which cells make these adjustments is called ”gene regulation.”

Every step of the gene expression process can be regulated, and bacteria effect these regulatory actions through myriad means. For example, bacteria can utilize proteins called transcription factors (TFs)[1–3] to either repress or activate transcription. Or, bacteria can

use RNA-binding proteins (RBPs)[4–6] to affect the stability or accessibility of mRNAs. The number of proteins involved in bacterial gene regulation is extensive, but this thesis focuses elsewhere.

In this thesis, we investigate the role of RNAs in gene regulation. We are interested, though, not just in the identities of the RNAs involved, but the fine details of RNA-mediated regulation. In particular, we are interested in mathematical and physical descriptions of RNA-mediated regulation, and quantitative approaches to generate those descriptions. We believe that these mathematical and quantitative studies can serve as a complement to traditional molecular biology approaches, that they can shed additional light on the mechanisms and dynamics of RNA-mediated gene regulation, and they can potentially be leveraged for synthetic biology purposes to positively impact human health[7–9].

1.2 Early Studies of RNA-mediated Gene Regulation in Bacteria

In the beginning, there were mRNAs, transfer RNAs (tRNA), and ribosomal RNAs (rRNA). These RNAs comprised the middle step in the central dogma and the mechanistic arrow pointing to the final step, and for ages we believed that RNAs were confined to this role as intermediaries. Then, in the 1980s, a new sort of RNA was discovered. An independent *E. coli* gene, *micF* was found to have its own promoter and to generate a small, non-coding RNA. The RNA transcribed from *micF* was later found to bind to the *ompF* RNA, which generates the OmpF outer membrane protein, and reduce the protein levels in response to high temperatures[10–13]. The discovery of *micF* and its roles in temperature regulation represented the first known case of a non-coding RNA playing a regulatory role in gene expression, rather than just serving as an intermediate step to translation[14]. Thus, a new class of RNAs was born.

Since the discovery of *micF* there have been many more regulatory RNAs discovered in both prokaryotes and eukaryotes[15–18]. There are regulatory RNAs that impact every

step of gene expression[19]. Eukaryotic regulatory RNAs fall under many categories, including long non-coding RNAs (lncRNA[20]), microRNAs (miRNA[21]), small nucleolar RNAs (snoRNA)[22], etc. In bacteria, though, the predominant regulatory RNA is referred to simply as the small RNA (sRNA).

1.3 Current methods to discover bacterial small RNAs and their roles

Starting from humble beginnings, hundreds of sRNAs have now been discovered in bacteria. Before the 2000s, we knew of only a few sRNAs. The explosion in the discovery of sRNAs was sparked by computational, genome wide searches for sRNA-encoding genes[23–25]. Even today, more sRNAs are being discovered and the functions of many of them are still being determined. Nonetheless, it is already clear that sRNAs are important and prevalent vehicles of gene regulation in bacteria.

Often, but not strictly, genes encoding sRNAs are found in intergenic regions (IGRs). In fact, initial searches for sRNA-encoding genes focused mostly on IGRs, seeking promoters and terminators within them. Later on, deep-sequencing-based methods[26,27,27–30] greatly increased the pool of potential sRNAs. In addition to those found in IGRs, sRNA-encoding genes have been found in the 5'- and 3'-regions of mRNA-encoding genes[31], within protein-coding genes[26], as *cis*-encoded antisense transcripts[32–34], and more. It seems that the origins of sRNAs are indeed plentiful.

sRNAs mostly act (with some exceptions) by Watson-Crick base-pairing interactions with target nucleic acids. *Cis*-encoded sRNAs are transcribed from the opposite strand of their target RNA and therefore have perfect complementarity with those targets. *Trans*-encoded sRNAs, in contrast, are transcribed in locations distant from their target RNAs, and do not have perfect complementarity; they often bind their targets with incomplete base-pairing. Though it might seem intuitive that *cis*-encoded sRNAs with perfect complementarity may be more effective and prevalent regulators, in fact imperfect comple-

mentarity *trans*-encoded sRNAs are far more prevalent in bacteria. For the most part, we will focus on *trans*-encoded sRNAs in this thesis.

Beyond identifying sRNAs, defining their role is another challenge altogether. sRNA gene deletions typically cause only subtle, if any, phenotypic changes[35]. Because of this, finding the role of an sRNA is often reduced to finding its target mRNAs. However, because sRNAs utilize imperfect base-pairing, and therefore it is insufficient to simply search for sRNA-complementary sequences. Furthermore, sRNAs typically regulate not just one, but multiple targets[36–39], making the search for true positive targets even more difficult.

Instead, successful computational algorithms have incorporated information about sRNA ”seed” sequences and sRNA-mRNA co-conservation[40–42]. The seed sequence of an sRNA initiates base-pairing or stabilizes the sRNA-mRNA interaction[43,44]. Several important features shared among sRNA seed sequences have been considered in successful search algorithms: first, seed sequences are highly phylogenetically conserved relative to the rest of the sRNA sequence; second, the seed sequence most be structurally available for binding; and third, since seed sequences often initiate mRNA binding, there is at least imperfect complementarity with target mRNAs. Alorithms such as CopraRNA[40] that incorporate these commonalities have shed light on the regulatory networks and, consequently, functional implications of sRNAs.

A common *in vivo* experimental approach to finding sRNA targets has been to track changes in mRNA levels after sRNA induction. This is made possible by the fact that many sRNAs promote target mRNA degradation[35,45,46]. Microarray analyses are one such method that exploit this[47,48], but are only effective when sRNAs impact mRNA levels. When mRNA levels are unaffected, proteomics-based approaches have also been used to identify targets[49]. Recently, RIL-seq (RNA interaction by ligation and sequencing) has been used for transcriptome-wide identification of sRNA targets without the need for any downstream changes in the target gene expression[50]. In all, there have been a number of

successful and innovative strategies developed in the past twenty years for *in vivo* sRNA-target discovery.

Beyond target discovery, various *in vitro* methods have been developed to study the kinetics, thermodynamics, and mechanics of sRNA-mediated regulation[39,46,51]. However, predictions based on thermodynamic measurements and *in vitro* kinetic analyses have failed to explain various features of sRNA regulatory networks, such as their hierarchical nature[39], perhaps due to their failure to incorporate the roles of regulatory proteins or the simultaneous presence of multiple sRNA binding partners. This is an oversight and challenge we seek to address in Chapter 2.

1.3.1 Chaperones and cellular machineries involved in sRNA regulation

sRNAs do not work alone. Intimately interwoven into the sRNA-mediated regulation process are a number of accessory proteins and cellular machines. Three are of principal interest in this thesis: Hfq, RNase E, and the ribosome.

RNA-binding proteins coordinate with regulatory RNAs to effect gene regulation. Hfq is perhaps the most important and prevalent RBP in bacteria, predicted to exist in more than half of all bacterial species[52,53]. The majority of *trans*-encoded sRNAs in *E. coli* require Hfq for regulation. Hfq is a homo-hexameric, Sm-like protein that performs numerous roles in *E. coli*[54]. Chief among those roles are promoting RNA-RNA interactions and protecting sRNAs from degradation[15,55]. This is performed by three different binding faces that allow Hfq to bind to multiple mRNAs, sRNAs, and rRNAs simultaneously. Additionally, Hfq can regulate mRNAs in an sRNA-independent manner[56,57], though these alternative routes are of lesser interest in this thesis. In general, the ways in which Hfq promotes the sRNA target search process are of great interest.

A third major component of the sRNA-regulation machinery, in addition to Hfq and the sRNA itself, is RNase E. RNase E is an endoribonuclease and the principal component of

the *E. coli* degradosome, which also consists of polynucleotide phosphorylase (PNPase, a 3' exoribonuclease), RNA helicase B (RhlB), and enolase (a glycolytic enzyme)[58–61]. RNase E consists of a catalytic N-terminal, a scaffolding C-terminal, and an RNA binding domain. Through its C-terminal scaffolding region, RNase E forms a ternary complex with Hfq and sRNA to initiate targeted mRNA decay. Full-length RNase E, but no other component of the degradosome, was shown to copurify with Hfq and sRNAs simultaneously, while a C-terminal truncation mutant did not[58].

Finally, while not part of the regulation machinery itself, the ribosome can not be ignored in the sRNA-mediated regulation process. Many inhibitory sRNAs operate by binding their target mRNAs in a region either directly overlapping the RBS or a site close to it. In this way, the sRNA, or Hfq in some instances[62], blocks the ribosome from binding and subsequently prevents translation and promotes degradation. In a sense, sRNA-mediated regulation can be thought of as an interplay between the Hfq-sRNA-RNase E complex and the ribosome, as both attempt to bind the same sites on the mRNA, yet each has a different goal. One might infer that there exists a competition between the two players because of the overlapping binding sites; however, there is now some evidence that suggests this might not be the case[63]. We will explore this subject further in Chapter 3.

1.3.2 Function of sRNAs in stress response

One common trait shared by many sRNAs, outside of some housekeeping sRNAs, is that they are expressed only in response to specific conditions, evincing one of the main roles of sRNAs: defense mechanisms used during stress response.

The definition of what constitutes a stress response to a bacteria is somewhat circuitous, but a broad, general definition is that a bacterial stress response is some change in gene expression, relative to gene expression in normal growth conditions, in response to a change in the environment. Examples of bacterial stresses include Iron-deprivation

stress[64], glucose-phosphate metabolism stress[65], and temperature stress[12]. Under normal growth conditions, sRNAs associated with stress responses are found in very low, even single-digit, copy numbers; but, under specific stress conditions they can increase 100-fold. It is helpful to consider a few examples of sRNAs utilized in stress responses in order to understand the general mechanisms underlying sRNA-mediated gene regulation.

The sRNA RyhB responds to iron deprivation stress[64]. Under normal growth conditions when iron is widely available for the cell, the ferric uptake regulator protein, Fur, represses expression of RyhB. Proteins that utilize iron are synthesized regularly, and proteins responsible for the uptake of more iron are repressed. When iron is deprived, the Fur-repression of RyhB ceases, thereby activating the its expression, which subsequently regulates a collection of genes encoding iron-utilizing proteins including the mRNA *sodB*, which encodes a Fe-superoxidase dismutase[66], via incomplete base-pairing and induced degradation. The logic of RyhB-mediated regulation is straightforward and illustrative: under normal conditions, iron is abundant, proteins that utilize iron are allowed to operate normally, and RyhB is unnecessary; when iron is deprived, RyhB is expressed and down-regulates genes that encode iron-utilizing proteins. In addition to *sodB*, RyhB also regulates mRNAs in the *sdhCDAB* operon, which encodes succinate dehyrogenase, *acnA*, which encodes tricarboxylic acid cycle enzymes, and *ftnA*, which encodes ferritin, among others[64].

Glucose-phosphate stress is characterized by a cytosolic accumulation of phosphorylated sugars. Such an accumulation can be the result of ingesting non-metabolizable sugars, such as the phosphorylated glucose analog α-methyl glucoside-6-phosphate (αMG6P), or by some mutation affecting sugar metabolism. In normal growth conditions, glucose is transported into *E. coli* and phosphorylated by the primary glucose transporter of the phosphoenolpyruvate phosphotransferase system (PTS), EIICB^{glc}[65]. EIICB^{glc} is encoded by the mRNA *ptsG*. An accumulation of phosphosugars, which can lead to growth defects, activates expression of the sRNA, SgrS, which binds *ptsG* via imperfect base pairing

in a region overlapping the ribosome binding site (RBS), thus blocking translation and simultaneously inducing degradation. In addition to regulation of other genes such as *manXYZ*[67,67], which encodes a mannose transporter, *purR*, which encodes a purine biosynthesis operon repressor[68,69], and *yigL*, which encodes a phosphatase that can dephosphorylate non-metabolizable sugars[69,70], this downregulation of *ptsG* by SgrS helps counter the deleterious accumulation of phosphosugars.

In response to high temperature and high osmolatiry conditions, the sRNA MicF targets and represses the translation of the OmpF protein, an outer membrane protein, helping to regulate pore size in a manner compatible with the environment[12]. Additionally, MicF regulates *lrp*, which encodes the leucine responsive protein, *yahO*, which encodes a periplasmic protein, and others[71,72]. Similar to RyhB and SgrS, MicF regulates some of its targets by binding and subsequently blocking its translation and inducing its degradation. This general scheme - induction of sRNA, followed by imperfect base-pairing of target, followed by translation or stability regulation - is common amongst sRNAs utilized in stress responses. The finer details, of course, differ, but the commonalities are helpful in understanding the role of sRNAs in bacteria.

1.4 Diverse mechanisms of sRNA regulation

Regulatory, *trans*-encoded sRNAs regulate at multiple levels of gene expression. Some work by affecting the stability of their target mRNAs. Others activate or inhibit translation. Still others regulate by atypical means. The following are just a few examples of the many mechanisms utilized by sRNAs to control gene expression.

1.4.1 Translation-level control by sRNAs

Perhaps the most common mechanism of action in sRNA-mediated regulation is by direct occlusion of the RBS binding site. Many sRNAs bind sites in the 5'-untranslated

regions (UTR) of their target mRNAs that either directly overlap the RBS or close enough that the bound Hfq-sRNA complex sterically prevents the ribosome from binding. Apart from the downstream destabilization effects this causes, the inhibition of translation leads to a direct and immediate downregulation of the translated gene product of the target mRNA. SgrS regulation of *ptsG* is an example of an sRNA binding directly over the RBS[73]. SgrS regulation of another mRNA in its regulon, *manXYZ*, represents an example of an sRNA that binds outside of the RBS, but is able to inhibit translation anyway. SgrS binds the polycistronic mRNA *manXYZ* at a site upstream of and completely separate from the RBS. However, in a role reversal, SgrS binding promotes Hfq-mediated occlusion of the RBS[62].

Other sRNAs can activate, rather than inhibit, translation. Such instances can occur when an mRNA has a leader sequence with a structure that occludes the RBS within a stem-loop. An example is the activation of the translation of sigma factor RpoS by the small RNA, DsrA[74–76]. Under normal growth conditions, the translation of RpoS is inhibited because its RBS is buried within a stem-loop in its leader sequence. In low temperatures, DsrA is synthesized, and by imperfect pairing between its first stem-loop and a site within the folded leader region of RpoS, it unwinds the leader sequence of RpoS, thus exposing its RBS and promoting translation.

1.4.2 Degradation-level control

Because of the interaction between RNase E and the Hfq-sRNA complex, sRNA binding also often guides RNase E to cleaveage sites on the target mRNAs, thereby destabilizing the target. This mechanism is entwined with translation inhibition, as ridding the mRNA of ribosomes grants easier access to the cleavage sites by RNase E. Therefore, sRNAs that block translation often also induce degradation, as is the case with SgrS-*ptsG* regulation. Induced mRNA degradation may merely be a secondary effect in some cases, though, as some sRNAs can efficiently silence their targets even without mRNA degrada-

tion[77]. Outside of the direct recruitment of RNase E by the Hfq-sRNA complex, it appears that the downstream effects of translation inhibition alone may passively encourage RNase E-mediated degradation, which we will discuss in Chapter 2.

An example of atypical, positive sRNA-mediated regulation at the degradation level is the stabilization of target mRNA by the *Salmonella* sRNA RydC. RydC-Hfq complex binds the mRNA *cfa* at a site that prohibits RNase E binding, thus preventing cleavage and upregulating *cfa*[78]. SgrS upregulates the mRNA *yigL* in a similar manner[70].

1.4.3 Transcriptional control

The sRNA ChiX is an interesting example of co-transcriptional regulation of a single gene within a polycistronic mRNA. ChiX regulates the *chiPQ* operon by co-transcriptionally binding to a site overlapping the *chiP* RBS, thus preventing further translation. But, the interesting regulation happens downstream, in the *chiQ* region. By preventing ribosomes from binding, ChiX binding leaves the mRNA bare and exposes a Rho binding site. This allows the transcription termination factor Rho to bind and inhibit further transcription of the downstream *chiQ* sequence[79].

1.5 Prominent mathematical and computational studies of bacterial small RNAs

Several general features of sRNA-mediated regulation make it ripe for quantitative study. Because a single sRNA is able to regulate multiple target mRNAS, studies that treat sRNAs and mRNA as nodes in a regulatory network are appealing[80]. The kinetics and robustness of sRNA-mediated regulation have been studied, much in the same light as study of other gene regulatory networks. In this section, we describe a small subset of the important quantitative descriptions of sRNA-mediated regulation.

The earliest quantitative studies of sRNA regulation were conducted by the Hwa

lab[81,82]. Being the first, they were able to start with the simplest mathematical model, which considered only the transcription and turnover rates of sRNAs and their mRNA targets, and assumes coupled degradation of the targets. With this simple model alone, the authors were able to establish the ”linear-threshold” model of sRNA-mediated regulation, a popular model which is still referred to today. The model states that the sRNA transcription rate relative to the mRNA transcription rate alone can explain the target expression behavior. Below a threshold, when the sRNA synthesis rate is much higher than the target mRNA synthesis rate, the expression of the target gene is completely silenced. Beyond that threshold, target gene expression rises linearly with the mRNA transcription rate[81]. The authors also found that sRNAs were effective in repressing fluctuations that arise from stochastic gene expression. This relatively simple model explained fairly well the RyhB-*sodB* regulation network.

Erel Levine explored the effect that the presence of multiple mRNA targets would have on the quantitative characteristics of sRNA regulation. The authors posited that sRNAs could mediate indirect interactions, or cross-talk, between disparate target mRNAs, as the co-degradation of the sRNA with one of the targets could reduce its availability to regulate the other. Using a RyhB-targeted, plasmid-encoded reporter gene and tracking the copy numbers of two different chromosomal RyhB targets (in two separate experiments), the authors showed that regulation of one target by an sRNA could indeed rescue the second target from regulation. Incorporating their experimental findings into a model that accounted for the transcription rates and binding rates of two common targets of a single sRNA, they predicted a ”hierarchical cross-talk effect,” wherein the expression of a weakly-interacting target mRNA is highly effected by the expression of a strongly-interacting target, and a strongly-interacting target mRNA is hardly affected at all by the expression of a weakly-interacting target[82].

In a study from the Margalit Lab, the authors shed light on the advantages of sRNA-mediated regulation over other forms of regulation, such as transcription factor driven

regulation[83]. The authors showed with a mathematical model that sRNA-mediated regulation could be enacted faster than TF-mediated regulation, due to the fact that sRNAs do not need to be translated into proteins before they regulate their targets, a time-consuming process. Similar to the Hwa Lab study, the Margalit Lab model indicated that the expression levels of the target gene were essentially determined by the relative synthesis rates of the sRNA and the target mRNA, as well as the co-degradation rate. They also found that by adjusting the ratio of these rates, they could fine-tune the expression levels of the target.

The Wingreen lab similarly compared the quantitative characteristics of sRNA and TF regulatory networks. In addition to validating the linear-threshold behavior of proteins in sRNA regulatory networks, the authors further explored the output noise behavior, defined as the protein expression variance divided by the mean expression level, in the various protein expression regimes (*i.e.* repressed, expressing, and crossover). They found that even in the face of large intrinsic noise, in the form of transcriptional bursting, sRNA-based networks much more reliably tempered protein output noise due to the buffering effects of sRNAs on mRNA fluctuations. Additionally, sRNA-based networks were found to be more effective filters of high-frequency input noise, defined as rapidly-changing cellular inputs such as fluctuating external concentrations of some molecular signal. This could be functionally important, as sRNAs are often expressed in response to rapidly changing, external signals. TF-based networks, on the other hand, were found to be superior for precise control of protein expression, due to their ability to sensitively respond to small input signals. In all, there were clear distinctions between sRNA- and TF-based networks that point to the physiological contexts in which one is superior to the other[84].

One final study worth mentioning was conducted by the Stavans Lab, which described the effect of a small RNA on cell-to-cell phenotypic variation in *E. coli*[80]. Using RyhB regulation of *sodB* and *fumA* as a model system, the authors experimentally determined that in response to different sRNA levels, not only do the average target gene expression

levels change, but so too do the distributions of expression levels across different cells. Interestingly, while the the cell-to-cell variability in target protein levels remained close to constant across different RyhB levels, phenotypic variability, defined as the cell-to-cell variability in the *ratio* of the two target proteins, increased as a response to stress when RyhB decoupled the expression of the target genes. Thus, total phenotypic noise could increase in response to a sRNA, even when individual protein noise remained steady. Beyond these revelations, this study pointed to the fact that there are interesting dynamics exhibited by sRNA-regulatory networks when we move beyond just average expression levels, and observe distributions of gene expression levels.

1.6 Current challenges

What remains to be done in the study of sRNA-mediated regulation? It is notable that while the models described above do a decent job at describing sRNA-regulatory networks generally, they do sometimes fail in the finer details, or leave details unstudied. For example, the hypothesis of faster and more efficient regulation provided by sRNAs relative to transcription factors, response curves measured in later studies cast doubt on it[85]. The Hwa Lab's linear-threshold model is perhaps overly simplistic by considering only sRNA and mRNA synthesis rates, and is only concerned with mean target protein expression levels. How are the response curves and thresholds affected by binding rates, for example? Or different types of degradation? And how can we reconcile the different predictions made from *in vitro* studies of thermodynamics and kinetics with the true, *in vivo* kinetics? These are the sorts of questions that will be addressed in the remainder of this thesis.

CHAPTER 2

KINETIC MODELING REVEALS ADDITIONAL REGULATION AT CO-TRANSCRIPTIONAL LEVEL BY POST-TRANSCRIPTIONAL SRNA REGULATORS

2.1 Introduction

To cope with changes in both natural and host environments, microbes have evolved diverse mechanisms to sense, respond to, and adjust to stress conditions. Small RNAs (sRNAs) are common mediators of gene regulation in bacteria, especially in stress responses, and have been observed to provide survival benefits during infections, biofilm formation, and exposure to toxins and antibiotics[86–90]. In the canonical scheme of sRNA-mediated gene regulation (Figure 2.1A), sRNAs, often along with a chaperone protein, Hfq, target and bind mRNAs via incomplete Watson-Crick base-pairing[91]. As many sRNA binding sites on target mRNAs partially overlap with the ribosome binding site (RBS), binding of sRNAs can affect mRNA translation. In addition, the stability of the mRNAs can be affected through RNase E-mediated co-degradation of the sRNA-mRNA complex[55,91,92]. Previous biochemical studies suggest two mechanisms for sRNA-mediated degradation: (1) sRNA-mediated reduction of translation leads to a change in degradosome access to the target mRNA, thereby increasing the degradation rate of sRNA-bound mRNA[46,93–95] (here referred to as “passive degradation”, or “translation-coupled degradation”, interchangeably), and (2) modulation of degradation through direct recruitment of the degradosome[45,46,51,58,65] or direct obstruction of RNase E cleavage sites[78] (here referred to as “active degradation”). For a particular target mRNA, distinct sRNAs may regulate at one or more levels of expression – translation or mRNA stability – by different molecular mechanisms[45,55,65,92]. However, how control at each of these levels quantitatively contributes to the overall efficacy of sRNA-mediated regulation is not well characterized.

One characteristic feature of sRNA regulators is their ability to regulate multiple target mRNAs[39,44,96,97]. Previous studies have shown that the regulation of various targets by the same sRNA can exhibit a hierarchical pattern; *i.e.* certain targets are more effectively regulated than others[81,98]. Such prioritization in regulation helps optimize stress responses when sRNA abundance is limited[99]. However, the *in vivo* kinetic determinants that set the regulation hierarchy are largely unclear. A previous *in vivo* kinetic characterization of the sRNA target search and sRNA-mRNA co-degradation processes suggests that the *in vivo* binding affinity between specific sRNA-mRNA pairs can contribute to setting the regulatory hierarchy[38], whereas the *in vitro* binding affinity does not seem to correlate with the regulation hierarchy[39]. In addition, a recent RIL-seq (RNA Interaction by Ligation and sequencing) based study found a positive correlation between the Hfq occupancy of the target mRNA and sRNA-target interaction frequency, indicating that the binding efficiency of Hfq may affect the regulation priority of the target mRNA[100]. These observations suggest that *in vivo* target search and regulation kinetics may be collectively determined by complex molecular interactions and kinetic pathways that are difficult to fully recapitulate *in vitro* and therefore require *in vivo* characterization.

In this work, we sought to provide a comprehensive model of sRNA-mediated regulation at the level of translation and mRNA stability. To achieve this goal, we utilized a genetically and biochemically well-characterized E. coli sRNA, SgrS, as a model. SgrS is the central regulatory effector of the glucose-phosphate stress response. Intracellular accumulation of phosphorylated glycolytic intermediates, such as the phosphorylated glucose analog α-methyl glucoside-6-phosphate (αMG6P), along with depletion of other glycolytic intermediates, launches transcription of SgrS, and subsequent regulation of several mRNA targets[101]. The best characterized targets include negatively regulated *ptsG* mRNA (encoding glucose transporter)[65,102], *manXYZ* mRNA (encoding mannose transporter)[67,103], *purR* mRNA (encoding purine biosynthesis operon repressor)[68,69], as well as positively regulated *yigL* mRNA (encoding a phosphatase that can dephosphorylate non-metabolizable sugars so they can be excreted to relieve stress)[69,70].

By implementing a combined single-cell imaging and mathematical modeling approach, we determined the kinetic parameters of SgrS regulation of a subset of its target mRNAs. Unexpectedly, our data reveal that instead of acting exclusively on fully synthesized transcripts, SgrS is able to regulate some targets co-transcriptionally. We found that another sRNA, RyhB also acts on nascent mRNA co-transcriptionally, suggesting that this may be a general feature for sRNAs previously characterized as post-transcriptional regulators. We found that co-transcriptional regulation is attenuated when Rho factor activity is inhibited, indicating that this co-transcriptional regulation is likely due to Rho-dependent termination following sRNA-mediated repression of translation. Finally, our data suggest several important kinetic steps that may determine the efficiency and differential regulation of multiple mRNA targets by an sRNA. Binding of sRNA to the target mRNA is likely the rate-limiting step and may dictate the regulation hierarchy observed within an sRNA regulon. Our approach may be used as a general platform for dissecting kinetic parameters and providing mechanistic details for sRNA-mediated regulation.

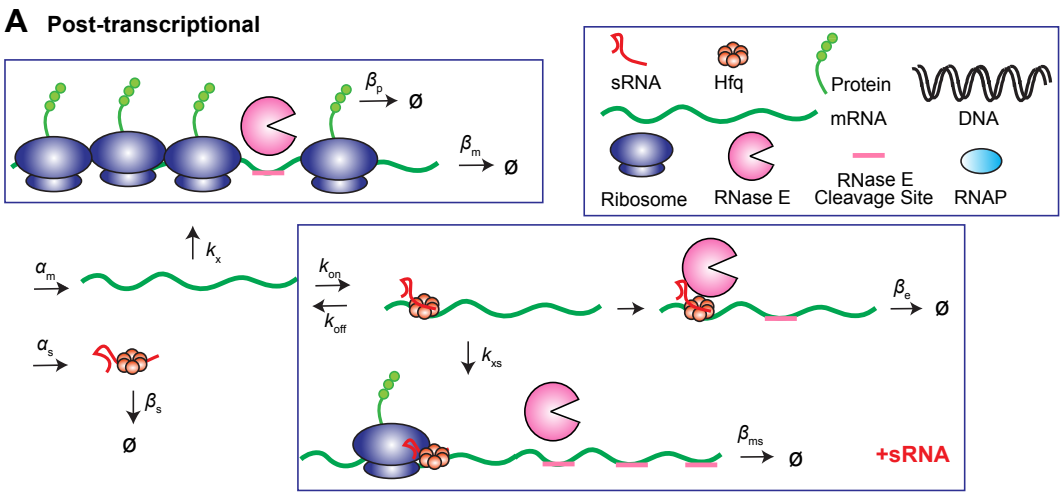
2.2 Results

2.2.1 Kinetic model and experimental measurement of sRNA-mediated regulation

Since SgrS has been biochemically characterized as a post-transcriptional gene regulator, we first set up a post-transcriptional regulation model to describe this process, including regulation at the levels of both translation and degradation (Figure 2.1A). In the absence of the sRNA, this model includes basal levels of mRNA transcription (α_m, as rate constant), translation (k_x), and endogenous mRNA and protein degradation (β_m and β_p respectively). When the sRNA is produced, its transcription rate is defined by α_s and the effective degradation rate by β_s. β_s approximates endogenous degradation as well as target-coupled degradation with all other mRNA targets except for the specific target

mRNA of interest. The sRNA binds to an mRNA target with an on-rate of k_{on} and dissociates with an off-rate of k_{off}. Upon binding, the translation activity of the bound mRNA changes to k_{xs}. The sRNA-mediated degradation is described by β_{ms} for translation-coupled degradation and β_e for active degradation.

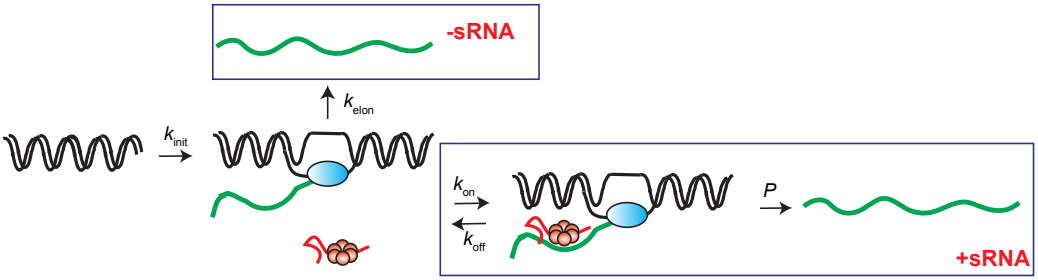
Production of the sRNA, SgrS (from the endogenous chromosomal gene), was induced by exposing E. coli cells to glucose-phosphate stress using α-methylglucoside (αMG)[65]. The target mRNAs (containing SgrS binding sequences) fused to the superfolder GFP (sfGFP) gene[104] were carried on low-copy number plasmids under the control of a tetracycline promoter (P*tet*) (Figure 2.2A and Supplementary Figure 2.7). In contrast to the induction scheme commonly used in previous studies in which the changes in target mRNA abundance or translation were recorded after sRNA induction, we chose to induce SgrS before target mRNA induction and then record the levels of SgrS, target mRNA and protein simultaneously as a function of time. Time t = 0 was defined as the time at which the target mRNA was induced (Figure 2.2B). Fractions of cells were fixed at different time points. SgrS and the target mRNAs were fluorescently labeled with DNA oligo probes through a standard fluorescence in situ hybridization (FISH) method[38]. Translation of the sfGFP fusion produced a direct fluorescent readout for protein levels (Figure 2.2B). The single-cell sRNA, mRNA, and protein levels were characterized by their volume-integrated fluorescent signals[105]. sRNA and mRNA copy numbers were further determined by comparing fluorescent intensities and the Ct values in the reverse transcription and quantitative PCR (RT-qPCR)-based calibration (Figure 2.2C and Supplementary Figure 2.8). The abundance of endogenous *ptsG* mRNA is 1-2 copies per cell under our growth conditions measured by RT-qPCR, <1% of the *ptsG-sfGFP* fusion mRNA expressed from the plasmid, suggesting that the contribution of the endogenous mRNAs to our measurement is negligible. As the copy numbers of sRNA and mRNA were in the range of tens to hundreds per cell, we described the time-dependent changes in sRNA, mRNA and protein deterministically by mass action equations (Figure 2.1B).



B **Post-transcriptional**

$$\frac{d[s]}{dt} = \alpha_S - \beta_s[s] - k_{on}[m][s] + k_{off}[ms] \qquad \frac{d[ms]}{dt} = k_{on}[m][s] - k_{off}[ms] - \beta_{ms}[ms] - \beta_e[ms]$$

$$\frac{d[m]}{dt} = \alpha_m - \beta_m[m] - k_{on}[m][s] + k_{off}[ms] \qquad \frac{d[p]}{dt} = k_x[m] + k_{xs}[ms] - \beta_p[p]$$



D **Co-transcriptional**

$$\frac{d[s]}{dt} = \alpha_S - \beta_s[s] - k_{on}([m][s] + [mi][s]) + k_{off}([ms] + [mis]) \qquad \frac{d[m]}{dt} = k_{elon}[mi] - \beta_m[m] - k_{on}[m][s] + k_{off}[ms]$$

$$\frac{d[mi]}{dt} = k_{init} - k_{elon}[mi] - k_{on}[mi][s] + P\,k_{off}[mis] \qquad \frac{d[ms]}{dt} = k_{on}[m][s] - k_{off}[ms] - \beta_{ms}[ms] - \beta_e[ms]$$

$$\frac{d[mis]}{dt} = k_{on}[mi][s] - k_{off}[mis] \qquad \frac{d[p]}{dt} = k_x[m] + k_{xs}[ms] - \beta_p[p]$$

Figure 2.1 Model for determination of kinetic parameters of sRNA-mediated regulation *in vivo*.
(A) Kinetic model describing sRNA-mediated, post-transcriptional regulation. **(B)** ODE for post-transcriptional regulation model. **(C)** Kinetic model for co-transcriptional regulation. **(D)** ODE for co-transcriptional regulation model. Parameters are described in the text.

We chose to pre-induce the sRNA for two reasons. First, by capturing the sRNA-

mediated changes in the production of new proteins, we can more accurately measure regulation at the translational level. sRNA-mediated regulation generally occurs within minutes[65,81,106]. However, many proteins, including the reporter sfGFP, have long lifetimes in *E.coli*, which are essentially determined by rate of dilution due to cell division[107]. Therefore, the fluorescent signal from already existing proteins in the cell can overwhelm any protein level changes caused by sRNAs. Second, and more importantly, we were interested in the timing of sRNA-mediated regulation of target mRNAs and more specifically, whether sRNAs can act on the newly synthesized mRNAs co-transcriptionally. In the case of pre-induced mRNA, the mature mRNAs outcompete the nascent mRNAs owing to their relative abundances, which may make any effect at the transcriptional level undetectable.

For each sRNA-mRNA pair, we measured the time-course changes of sRNA, mRNA and protein levels in four genetic backgrounds: wild-type (WT), Δ*sgrS*, *rne*701, and *rne*701 Δ*sgrS*. Time-dependent changes in mRNA and protein upon mRNA induction were recorded in the absence of SgrS for the determination of parameters describing basal transcription (α_m) and translation (k_x) activities. By comparing the fusion mRNA and protein levels in the Δ*sgrS* strain in the presence of αMG with the corresponding levels in the WT strain in the absence of αMG, we noticed that the presence of αMG alone (*i.e.*, without ensuing production of SgrS) reduced the efficiency of induction of the mRNA fusion (Figure 2.2C, Supplementary Figure 2.9, and Table 2.1). Therefore, to quantify the regulation by the sRNA specifically, we use the Δ*sgrS* and *rne*701 Δ*sgrS* grown in the presence of αMG as our "-sRNA" condition to quantify the basal transcription and translation activities of the target mRNA in the WT and *rne*701 backgrounds, respectively.

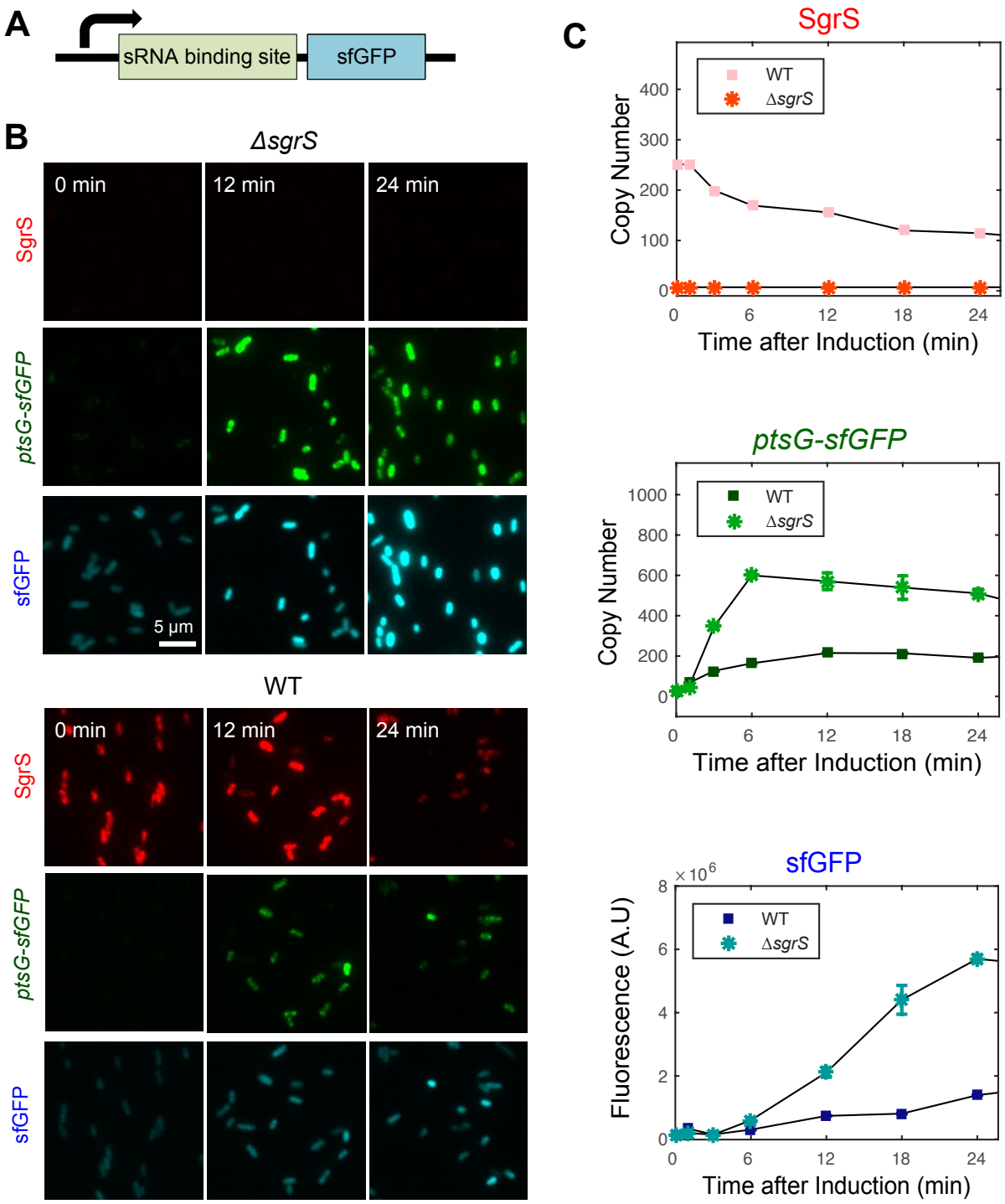


Figure 2.2 Illustration of experimental setup and representative results.
(A) Illustration of the target mRNA, including the 5' UTR and part of the coding region from the endogenous mRNA target containing the SgrS binding site, and a coding region for sfGFP reporter. **(B)** Representative images of SgrS, *ptsG-sfGFP* mRNA, and sfGFP signal in the absence (upper) or presence (lower) of sRNA induction over 24 minutes. **(C)** Measured sRNA, mRNA, and protein levels from images in (B), representing volume-integrated single cell fluorescence values, converted to copy numbers for the case of RNA molecules.

Comparison of the kinetic behaviors in the WT vs. *rne*701 strain allowed us to separate the effect of sRNA-mediated passive and active degradation. The *rne*701 allele encodes a truncated RNase E protein lacking part of the C-terminal unstructured region[108], including RhlB, enolase, PNPase and Hfq binding sites[51,109–111]. The *rne*701 mutant still fully retains its catalytic function but has an impaired ability to interact with Hfq and other degradosome components without the C-terminal unstructured region[58,108,112]. Consistent with previously reported[38,113], β_m of the fusion mRNAs, measured using rifampicin pulse-chase experiments, did not show any difference between the WT and *rne* mutant background (Supplementary Figure 2.10). This result suggests that rne701 has a minor effect on endogenous mRNA degradation, and that accessibility to the translated mRNA is unlikely to be affected by the partial truncation of the C-terminal region of RNase E. Our model therefore assumes that SgrS-mediated mRNA degradation in the *rne*701 background is primarily through translation-coupled degradation.

Finally, to further constrain our model, we experimentally measured a subset of parameters. In addition to β_m, we measured β_s by first inducing SgrS and then washing away the inducer (Supplementary Figure 2.11). β_s was slightly slower in the *rne*701 background (Supplementary Figure 2.11), suggesting that active co-degradation with target mRNAs contributes to the ensemble sRNA turnover, consistent with previous results[38]. We approximated sfGFP protein half-life using the cell doubling time ($\sim$90 min) under our experimental condition (Supplementary Figure 2.12). The six-minute folding time of sfGFP was accounted for in the model by building in a six-minute delay between mRNA and protein production (*i.e.*, the protein present at t = 12 minutes is translated by the mRNA present at t = 6 minutes). α_s was determined by measuring the time-dependent production of SgrS upon induction.

2.2.2 Simulation predicts that SgrS may regulate ptsG co-transcriptionally

Under the assumption that SgrS regulates *ptsG-sfGFP* mRNA post-transcriptionally, we fixed the α_m and k_x values obtained from the Δ*sgrS* and *rne*701 Δ*sgrS* strains and fit the rest of the parameters in the time-dependent levels of SgrS, target mRNA, and sfGFP in the WT and *rne*701 strain in the presence of αMG, including β_e, β_{ms}, k_{xs}, k_{on}, and k_{off}, using Bayesian Markov Chain Monte Carlo (MCMC) modeling (Materials and Methods). However, the optimized parameters of the post-transcriptional regulation model did not accurately describe the experimental data, specifically the amplitude of sRNA-induced repression (Figure 2.13A). We therefore considered an alternative model that included the possibility that SgrS could regulate its targets co-transcriptionally, rather than exclusively post-transcriptionally.

Initially, we modeled co-transcriptional regulation by allowing α_m to change in the presence of sRNA (denoted α_{ms}). This model fit the data well (Figure 2.13B). The resulting α_{ms} (0.87 $\pm$ 0.05 molecules $\cdot s^{-1}$) was smaller than α_{ms} (1.9 $\pm$ 0.3 molecules $\cdot s^{-1}$), *i.e.*, transcription was slower in the presence of the sRNA. Since the FISH probes for the target mRNA specifically bind to the sfGFP coding region in the mRNA fusion downstream of the sRNA binding site, we infer that generation of the full-length mRNA, and therefore the fluorescent signal, was reduced upon sRNA binding, or sRNA-mediated regulation may occur during transcription. In addition, the reduction in α_m, was more pronounced in the WT *rne* background (α_{ms} = 0.46 α_m) compared to in the *rne*701 background (α_{ms} = 0.98 α_m), suggesting that a fully assembled degradosome contributes to the strength of co-transcriptional regulation.

2.2.3 SgrS decreases the abundance ratio of downstream to upstream regions relative to the SgrS binding site on the target mRNA

Since, according to our model, co-transcriptional regulation by SgrS reduces the production of full-length *ptsG-sfGFP* mRNA, we reasoned that this may be reflected by a decrease in the abundance of downstream (from the SgrS binding site) relative to upstream regions on the *ptsG-sfGFP* mRNA (henceforth referred to as the "D/U ratio"). To experimentally measure the D/U ratio, we devised a RT-qPCR assay. In this assay, we utilized two sets of primers: one amplifying the region upstream of the SgrS binding site, and the other amplifying the downstream region, in the coding region (Figure 2.3A). To evaluate the D/U ratio change specifically contributed by the co-transcriptional regulation, we compared RT-qPCR results on extracted RNA from cells at 1 and 15 min after induction (Figure 2.3B). These times were chosen based on the fact that the lifetime of *ptsG-sfGFP* mRNA is around 7-8 min (Supplementary Figure 2.10 and Table 2.1). At 1 min after induction ($D/U_{t=1}$), the contribution by endogenous degradation on the read-though ratio should not dominate. In addition, since the cellular level of mRNA at 1 min post-induction is low (Figure 2.2C), the fraction of nascent mRNAs, *i.e.*, the mRNAs still being transcribed, compared to fully synthesized mRNAs, should be relatively high. We therefore considered this pool of *ptsG-sfGFP* mRNAs as relatively enriched in nascent mRNAs and expected that effects at the co-transcriptional level would be enhanced in this sample. At 15 min after induction ($D/U_{t=15}$), *ptsG-sfGFP* mRNA levels reach steady-state, with a high cellular abundance (Figure 2.2C); thus, the fraction of nascent mRNAs should be minimal compared to fully synthesized mRNAs. We therefore reasoned that effects at the co-transcriptional level are largely buried by effects at the post-transcriptional level at this time point.

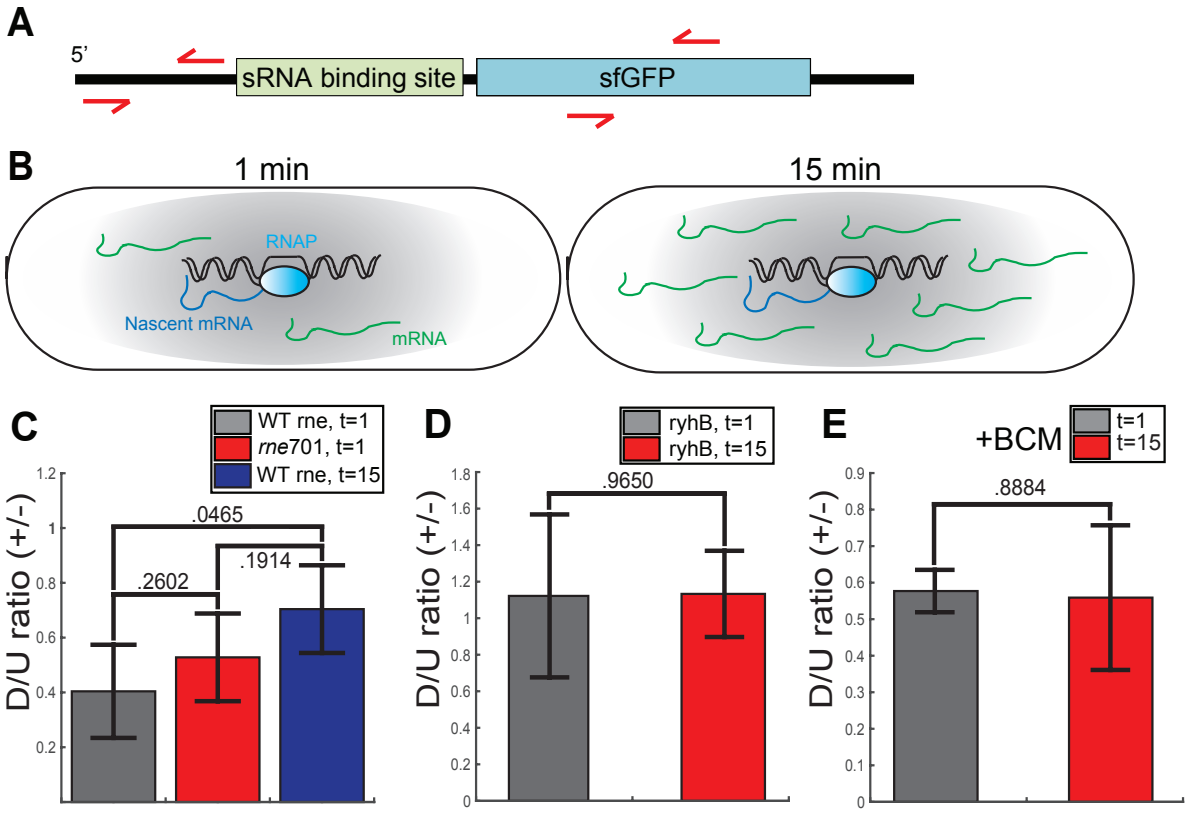


Figure 2.3 RT-qPCR measurement of the D/U ratio.
(A) Schematic illustration of the qPCR primer binding sites relative to SgrS binding site on the mRNA.. **(B)** Schematic illustration of total RNAs extracted at different time points of mRNA induction, which contain different ratios of nascent mRNAs to fully transcribed mRNAs. **(C)** Reduction in the D/U ratio of *ptsG-sfGFP* mRNA affected by SgrS (D/U(+/-), defined by the ratio of D/U in the presence of SgrS to D/U in the absence of the SgrS). **(D)** D/U(+/-) of *ptsG-sfGFP* mRNA affected by RyhB. **(E)** D/U(+/-) of ptsG-sfGFP mRNA affected SgrS with addition of Bicyclomycin (BCM). 50 μg/mL BCM was added 15 minutes before the time of cell collection, i.e. at t = -14 minutes relative to aTc induction of ptsG-sfGFP mRNA for cells collected at t=1 minutes, and t = 0 relative to aTc induction of ptsG-sfGFP mRNA for cells collected at t = 15 minutes.

We measured $D/U_{t=1}$ and $(D/U_{t=15})$ in the WT and Δ*sgrS* cells in the presence of αMG (Figure 2.3C). The comparison between the D/U ratios of WT and ΔsgrS cells (reported as "D/U (+/-)" in Figure 2.3C-E) reflects the change of D/U introduced by SgrS. $(D/U_{t=15})(+/-)$ was about $\sim$0.7, suggesting that the regulation by SgrS caused reduction in the abund-

ance of the downstream region compared to the upstream region of the SgrS binding site on the target mRNA. The reduced ($D/U_{t=15}$) upon SgrS regulation may be explained by the directionality of RNase E activity, *i.e.*, an enhanced RNase E activity on the downstream fragment with 5' monophosphate[114–116]. In comparison, ($D/U_{t=1}$)(+/-) was about 0.4, suggesting that in the nascent-mRNA enriched pool, the regulation by SgrS led to significantly more reduction in the abundance of the downstream region compared to the upstream region, and supporting our prediction that SgrS repressed the generation of the downstream portion co-transcriptionally. As a control, ($D/U_{t=1}$)(+/-) and ($D/U_{t=15}$) remained around 1 when inducing a non-matching sRNA, RyhB, a small RNA that is repressed by Fur (ferric uptake regulator) and produced in response to iron depletion, by adding 2,2'-dipyridyl (referred to as "DIP") into the culture (Figure 2.3D)[64]. In addition, the reduction in $D/U_{t=1}$ was less in the *rne*701 background (($D/U_{t=1}$)(+/-) was $\sim$0.5) (Figure 2.3C), consistent with the predicted trend from the simulation that the co-transcriptional regulation is stronger in the WT background.

2.2.4 Co-transcriptional regulation by SgrS is dependent on Rho activity

Though co-transcriptional regulation was stronger in WT compared to the *rne*701 background, it was nonetheless present in both backgrounds. We inferred that RNase E may play a role in co-transcriptional regulation, however, additional factors may contribute. One mechanism that could underlie co-transcriptional regulation is Rho-dependent termination. As SgrS binding to mRNA targets leads to translational repression, reduction in the transcription-coupled translation could lead to increased Rho access to the mRNA followed by premature termination. To test this possibility, we compared the D/U ratio on extracted RNA from cells at 1 and 15 min after *ptsG* induction in the presence of SgrS and Bicyclomycin (BCM). BCM targets and selectively inhibits Rho-dependent transcription termination[117,118]. In the presence of BCM, the significant disparity between $D/U_{t=1}$(+/-) and $D/U_{t=15}$(+/-) disappeared, indicating that in the absence of Rho-dependent transcription termination, co-transcriptional regulation is insignificant relative to post-transcriptional

regulation (Figure 2.3E). We therefore conclude that SgrS-induced co-transcriptional regulation of *ptsG* is Rho-dependent.

2.2.5 A revised kinetic model containing co-transcriptional regulation module

After experimentally confirming SgrS-dependent co-transcriptional regulation, we then improved the kinetic model by linking sRNA binding directly to the co-transcriptional regulation (Figure 2.1C and D). In this revised model, we assumed that sRNA binds and unbinds nascent and mature mRNAs with the same k_{on} and k_{off} rates. In order to allow for co-transcriptional binding, mRNA transcription is separated into two steps: initiation (k_{ini}) and elongation (k_{elon}). When nascent mRNAs are bound by sRNA during elongation, a free parameter (P) is introduced to the model, representing the probability of generating the full-length, mature mRNA. We allow P to differ between the WT and *rne*701 backgrounds. The revised kinetic model significantly improved the fitting of data for the SgrS regulation of *ptsG* (Figure 2.4A and B). To further validate the improved performance of the co-transcriptional regulation model, we applied the Bayesian information criterion (BIC), where a penalty is applied to the co-transcriptional model for its two added parameters (namely, P, in WT *rne* and *rne*701 background)[119], to select between co-transcriptional and post-transcriptional regulation models. The co-transcriptional model was selected by virtue of having the lower BIC value. Consistent with the qPCR results, P was lower in the WT background than in the *rne*701 background (Figure 2.4A and B, Table 2.1).

To validate the co-transcriptional regulation model, we generated two data sets. In the first, we reduced the induction of SgrS using a lower concentration of αMG and measured α_s experimentally. In the second, we reversed the induction order of SgrS and *ptsG-sfGFP* mRNA, presenting the condition under which newly induced sRNAs regulate pre-existing mRNA targets. We simulated the time courses of SgrS, *ptsG-sfGFP* mRNA, and sfGFP using the best set of parameters obtained from a model with (Figure 2.4A and B) or without

(Supplementary Figure 2.14A) co-transcriptional regulation, respectively. In both cases, the co-transcriptional regulation model predicted the experimental data better (Figure 2.4C and D, Supplementary Figure 2.14B and C).

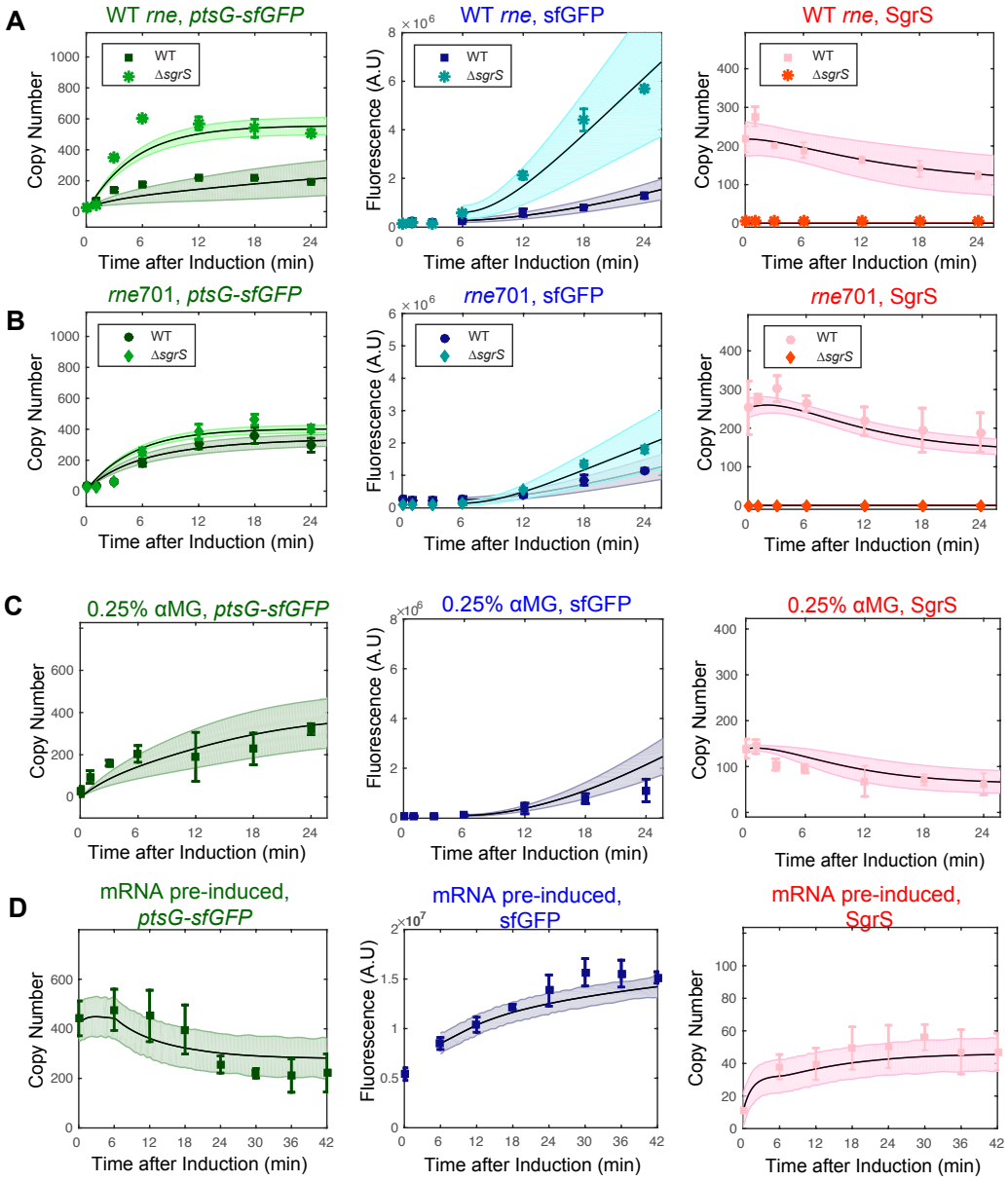


Figure 2.4 Fitting of SgrS regulation of *ptsG* expression with co-transcriptional regulation model.
Time-dependent changes of SgrS, ptsG-sfGFP mRNA, and sfGFP levels in the presence or absence of SgrS, in the **(A)** WT rne background and **(B)** *rne*701 background. Points with error bars represent experimental data from 2-3 biological replicates.

Figure 2.4 *(continued)* Fitting of SgrS regulation of *ptsG* expression with co-transcriptional regulation model.
Each biological replicate contains ∼500-1000 cells. Black lines represent fitting with the best set of parameters using co-transcriptional regulation model with two-step transcription module. Shaded, colored regions represent predicted error of the fitting, calculated by sampling from the means and errors of individual kinetic parameters. **(C)** Simulated prediction (black curve with shaded, colored region) using co-transcriptional regulation model for validation dataset with reduced αMG concentration for SgrS induction, overlaid with experimental data (points with error bars). **(D)** Simulated prediction using co-transcriptional regulation model and experimental data for validation dataset of pre-induced mRNA.

2.2.6 Co-transcriptional regulation may be a widespread mechanism utilized by sRNAs

We next asked if co-transcriptional regulation might be a general feature shared by other previously characterized post-transcriptional sRNA regulators. We applied the same imaging and modeling scheme to RyhB and one of its targets, sodB. We generated two fusion mRNAs, $sodB_{130}$ and $sodB_{130+30}$, containing the RyhB binding site and sfGFP gene (Supplementary Figure 2.7). $sodB_{130+30}$ contains an additional 30 nucleotides which include a RNase E cleavage site and is more sensitive to RyhB regulation at the degradation level[46].

The responses of $sodB_{130}$ and $sodB_{130+30}$ to RyhB regulation were again best captured by the co-transcriptional regulation model as suggested by BIC (Supplementary Figures 2.15 and 2.16), suggesting that co-transcriptional regulation may be a general mechanism of sRNA-mediated regulation. Consistent with the SgrS regulation, co-transcriptional regulation for RyhB was also more efficient in the WT background compared to *rne*701. In addition, the β_e value of the $sodB_{130+30}$ was ∼4.5 fold higher than that of the $sodB_{130}$, in line with the addition of the RNase E cleavage site in $sodB_{130+30}$, serving as a validation of our model.

2.2.7 *Parameters that contribute to regulation efficiency of sRNA over different targets*

We next fit the models to two other SgrS targets, *manX* and purR. It has been established that *ptsG* is the primary target of SgrS, *manX* is a secondary target, and *purR* is a lower-priority target[39]. Consistently, we observed 78%, 53% and 18% repression respectively for ptsG, *manX* and *purR* at the protein level at 24 minutes under the same SgrS induction condition (Table 2.2). At steady state, the model predicted the regulation efficiency to be 57%, 43% and 5% at the protein level, and 48%, 33% and 11% at the mRNA level for ptsG, *manX* and purR, respectively. BIC suggested that the co-transcriptional regulation model better fit *manX* (Supplementary Figure 2.17), but the post-transcriptional model better fit *purR* (Supplementary Figure 2.18), indicating that the contribution of co-transcriptional regulation for *purR* is negligible.

Comparison of the parameters for the three mRNA targets for SgrS and two targets for RyhB (described above) suggests features that contribute to the overall regulation efficiency (Figure 2.5A, Table 2.2).

1. Within the same sRNA regulon, a faster binding rate led to more efficient regulation. We found that k_{on} changed more dramatically than k_{off} among different targets. For SgrS, the difference in k_{off} was within $\sim$2-2.5 fold among the three targets, whereas the change in the k_{on} values was up to $\sim$40 fold between *ptsG* and purR, suggesting that the binding kinetics is dominated by k_{on}. Interestingly, although $sodB_{130}$ and $sodB_{130+30}$ had the same RyhB target site, which led to the similar k_{off}, $sodB_{130+30}$ showed a higher k_{on} than $sodB_{130}$(see Discussion). In addition, RyhB had a much higher k_{off} for the *sodB* constructs compared to SgrS.

2. The repression at the translation level ($k_{xs}/k]_x$) contributed positively to the regulation efficiency among the SgrS targets. The SgrS binding site is located in the 5' UTR of *ptsG* mRNA, partially overlapping the RBS, and within the first 10 codons

and 34 codons in CDS of *manX* and *purR* respectively[62,69]. SgrS inhibits translation initiation on these mRNAs through different mechanisms. On *ptsG* mRNA, base pairing of SgrS directly blocks ribosome binding, while on *manX* and *purR* mRNAs, binding of SgrS guides Hfq to bind at a site close to the RBS to block ribosome binding[62,69]. Our results indicate that direct binding of SgrS at the RBS may be more efficient in repressing translation, and that the efficiency of translational regulation may decrease as the sRNA binding site moves further into the CDS. k_{xs} /$k]_x$ was similar among the two *sodB* constructs upon RyhB regulation, consistent with the fact that they share the same RyhB binding site. However, even though RyhB also regulates *sodB* through directly blocking ribosome binding at the RBS[120,121], the repression of translation was less efficient than for SgrS regulation of *ptsG*, suggesting that the different structures of sRNA-mRNA duplexes may affect translation to different extents.

3. For all target mRNAs, β_{ms} was larger than the corresponding β_m, supporting the translation-coupled degradation model in which reduced translation activity upon sRNA binding leads to faster degradation of the sRNA-bound mRNA. For the three SgrS targets, there was no correlation between β_e and regulation efficiency. Although a higher β_e was observed for *purR*, the most weakly regulated target of SgrS, a much smaller k_{on} value for *purR* limited the regulation efficiency. The impact of active degradation became more evident when comparing the two RyhB targets, in which most other parameters were similar. The higher β_e value of the $sodB_{130+30}$ contributed to a higher regulation efficiency of $sodB_{130+30}$ (67% and 67% at protein and mRNA levels for $sodB_{130+30}$ respectively compared to 48% and 37% for sodB130).

4. We observed a positive correlation between the strength of co-transcriptional regulation and the overall regulation efficiency. Co-transcriptionally bound *ptsG* had a lower probability of generating a full-length mRNA compared to *manX*, while *purR*

was insignificantly affected by co-transcriptional regulation. Similarly, co-transcriptionally bound $sodB_{130+30}$ had a lower probability of generating a full-length mRNA compared to $sodB_{130}$.

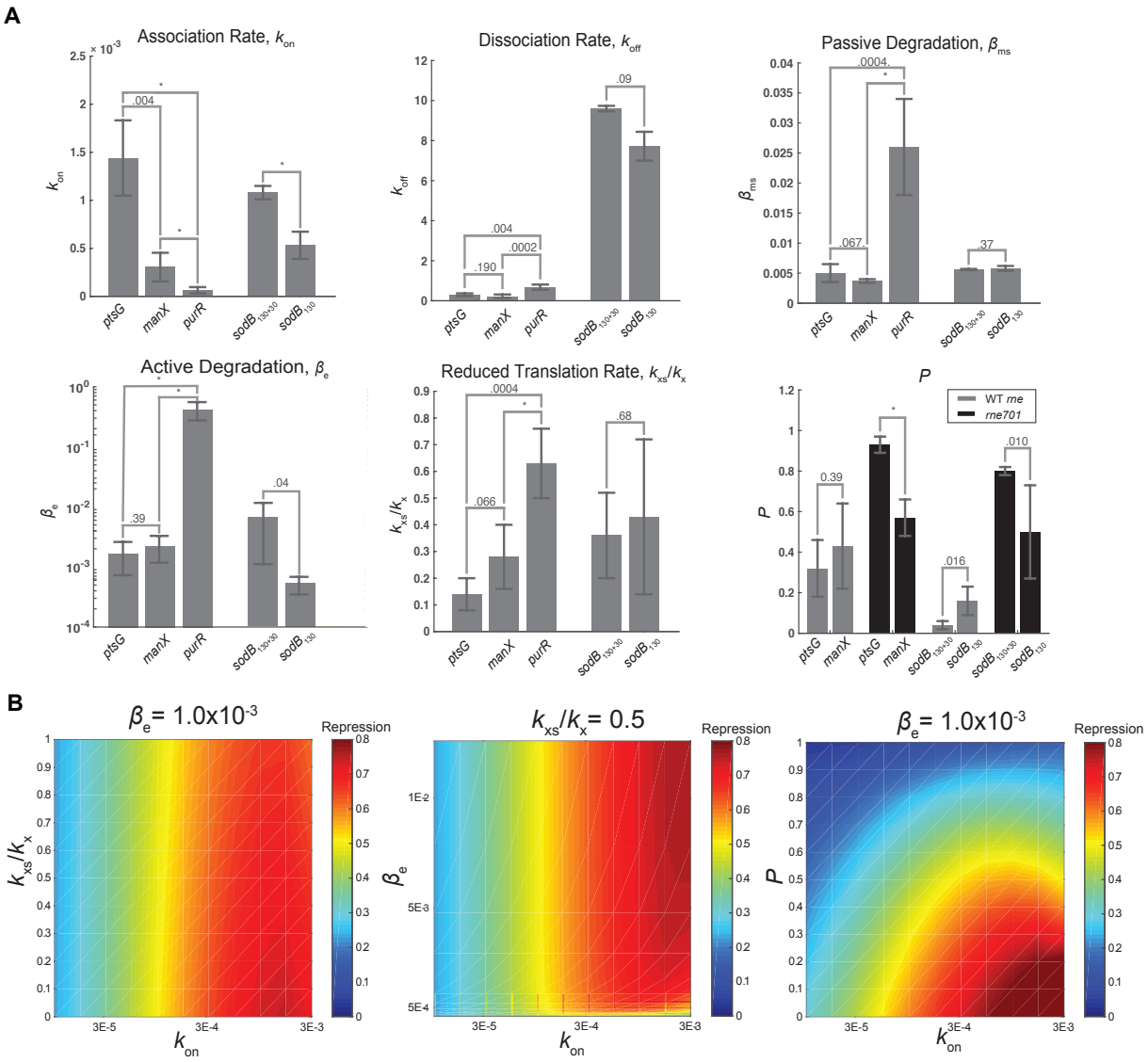


Figure 2.5 Kinetic parameters that contribute to regulation efficiency of sRNA over different targets.

(A) Parameters of sRNA regulation over different mRNA targets. Error bars represents s.d. from 2-3 biological replicates, each containing ∼500-1000 cells. p-values for two-sample t-tests are provided for pairwise comparisons. ∗indicates p<0.0001.

Figure 2.5 ***(continued)*** **(B)** Protein-level repression heatmap, calculated by screening across the listed parameters. Repression level of 1 represents complete repression of protein expression; 0 means no repression. For the left panel, β_e = 1.0x10^{-3} and P = 0.32. For middle panel, k_{xs}/k_x = 0.5 and P = 0.32. For right panel, k_{xs}/k_x = 0.5 and β_e = 1.0x10^{-3}. For all simulations, k_{init}, k_x, k_{off}, β_m, β_{ms}, β_s, and α_s are set to the measured or MAP values for *ptsG* (Tables 2.1 and 2.2).

2.3 Conclusions and Discussion

We have presented here a general approach combining imaging and modeling, which can be used to quantify the kinetic parameters underlying differential regulation of multiple mRNA targets by a single sRNA. While we focused on mRNAs that are downregulated by their corresponding sRNA in this study, we expect that this approach can easily be adapted to upregulated mRNA targets. While we initially sought to determine kinetic parameters of regulation at translation and degradation levels for an sRNA that was classically categorized as a post-transcriptional regulator, we unexpectedly found that SgrS can regulate co-transcriptionally. Similar co-transcriptional regulation was also observed for the sRNA RyhB. In our revised model, sRNAs can act on nascent transcripts as soon as their binding sites are released from the RNA polymerases (Figure 2.6A). This co-transcriptional regulation is promoted by Rho-dependent termination, but also affected by the C-terminal region of RNase E. Our model provides, so far, the most comprehensive description of sRNA-mediated gene regulation and helps dissect kinetic parameters governing hierarchical regulation. We find that binding of sRNA to the mRNA target is the rate-limiting step and the primary determinant for setting hierarchical regulation, while regulation at transcription, translation and degradation levels all contribute.

Previous models of sRNA regulation were able to reproduce mRNA repression assuming only post-transcriptional regulation[63,80–82,84,98,122–124]. The different order of sRNA and mRNA induction may explain why co-transcriptional regulation has not been noted in previous studies. Previous studies mostly either induced sRNAs in the presence of pre-existing or pre-induced mRNAs, or co-induced mRNAs and sRNAs simultan-

eously, whereas we pre-induced sRNAs to a certain level before inducing and tracking the changes of targets. Therefore, we created a time window, *i.e.*, early induction phase, when the mature mRNA level was low and the ratio between the nascent mRNA and the mature mRNA was high. Given the high abundance of pre-induced sRNA, and assuming in our model that sRNA used the same binding kinetics for both nascent and mature mRNA targets, the action of sRNA at the co-transcriptional level was enhanced compared to the cases where mature mRNAs were predominant. The effect of co-transcriptional regulation may be further enhanced by the target being plasmid-encoded. Because total target mRNA transcription was contributed by multiple plasmids in our experimental setting, the sRNA may more effectively regulate mRNA co-transcriptionally by targeting multiple transcription sites.

Although our target reporter mRNA genes are encoded by plasmids, it is very likely that sRNAs can act co-transcriptionally on the chromosomally encoded, endogenous genes. To co-transcriptionally regulate chromosomally encoded targets, sRNAs should be able to diffuse into the nucleoid region, which normally has a higher diffusion barrier. A previous report demonstrated that sRNAs have unbiased distribution between the nucleoid and cytoplasm using a few plasmid-encoded sRNAs as examples[125,126]. Here, using single-molecule localization microscopy (SMLM), we confirmed the unbiased localization for these two sRNAs under our experimental conditions (Figure 2.6B-E). In addition, the chaperone protein, Hfq, was observed to diffuse freely into the nucleoid region using single-particle tracking[127,128] and to bind to the nascent transcripts in a recent study using Chip-seq[129]. It is likely that at least part of the Hfq binding to the nascent transcripts is mediated by sRNAs.

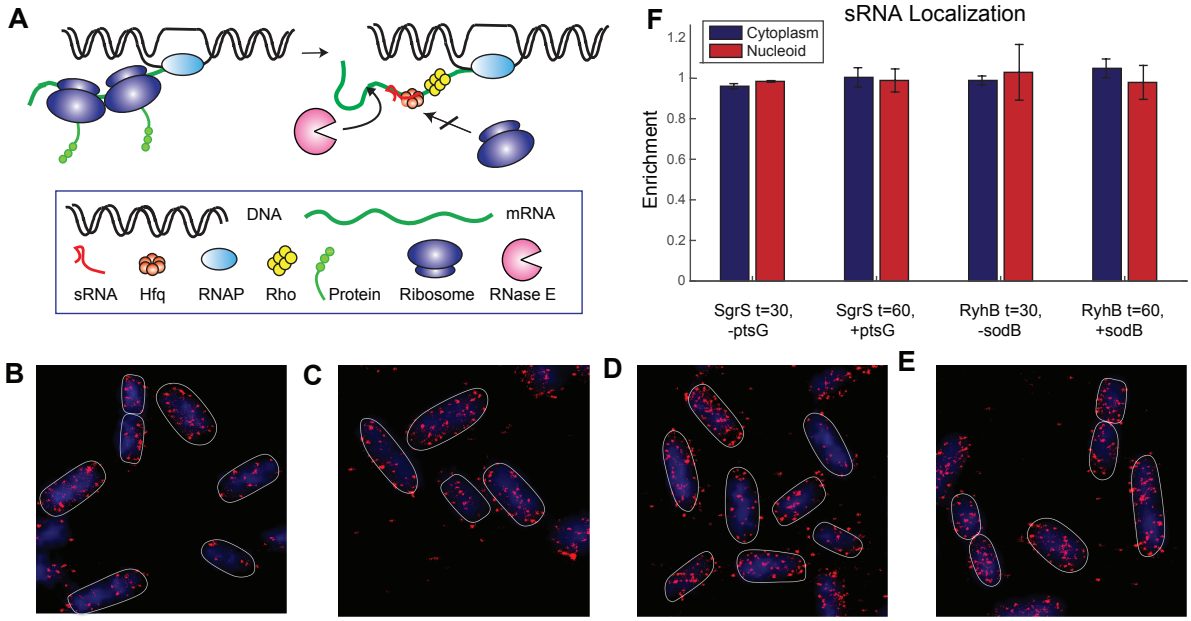


Figure 2.6 Model for co-transcriptional regulation by sRNAs.
(A) sRNAs can freely diffuse in into the nucleoid region of bacterial cells and bind to the target mRNAs as soon as the sRNA binding site is transcribed. Binding of sRNA affect transcriptional-coupled translation and increase the binding of Rho, thereby terminating transcription. Recruitment of RNase E through its C-terminal scaffold region positively contributes to the efficiency of co-transcriptional regulation. Representative SMLM images of SgrS in the absence of *ptsG-sfGFP* mRNA induction (30 minutes after sRNA induction, before mRNA induced) **(B)**, and in the presence ptsG-sfGFP mRNA induction (54 minutes after sRNA induction, 24 minutes after mRNA induction) **(C)**. Representative SMLM images of RyhB in the absence **(D)** and presence **(E)** of $sodB_{130+30}$-*sfGFP* induction. Red spots are sRNA signals detected by SMLM imaging. Blue area represents DAPI-stained nucleoid region. **(F)** The 3-dimensional localization for SgrS and RyhB in the absence and presence of the target mRNA was determined. Both SgrS and RyhB exhibited unbiased localization between the nucleoid and cytoplasm regardless of the presence of their target mRNAs. Error bars represents s.d. from 2-3 biological replicates, each containing ~100 cells.

While the majority of sRNAs are categorized as post-transcriptional regulators, cases have also been reported in which sRNAs can regulate transcription elongation, for example, by modulating the accessibility of the binding site of Rho factor, or by the conformational switch between terminator and antiterminator structures [130–133]. Interestingly, previously characterized post-transcriptional sRNA regulators, DsrA, ArcZ and RprA can also upregulate the target *rpoS* mRNA by suppressing pre-mature Rho-dependent transcription termination, a mechanism that may be widespread in bacterial genes with long

5' UTRs containing a Rho binding site[113]. In addition, ChiX sRNA was observed to co-transcriptionally regulate the distal gene in the *chiP* cistron by inducing Rho-dependent termination within the *chiP* coding region, establishing a regulation polarity of the downstream gene with the same cistron[79]. Our results demonstrate, using SgrS and RyhB as examples, that co-transcriptional regulation may be a common feature for sRNAs that can regulate at the translational level. While translational repression of the mature mRNAs by sRNAs increases their susceptibility to ribonucleases in the cytoplasm, repression of transcription-coupled translation on the nascent mRNA increases the access of Rho and therefore promotes pre-mature transcription termination. It remains to be demonstrated whether such Rho-dependent co-transcriptional regulation is pervasive throughout sRNA regulons. In addition, we observe that, for the same sRNA-mRNA pair, efficiency of co-transcriptional regulation is in general higher in the WT compared to the *rne*701 background (smaller *P* value in WT compared to *rne*701 background), indicating that the presence of the intact scaffold region of RNase E positively contributes to this process. Future investigation is needed to pinpoint the mechanistic role of RNase E in the co-transcriptional regulation.

Finally, our model suggests several kinetic steps that can determine the overall regulation efficiency. The binding kinetics between the sRNA and mRNA are the primary determinant of regulation efficiency. While k_{off} differs substantially between different sRNAs, within the regulon of a given sRNA, k_{on} changes more dramatically compared to k_{off}, and contributes to the regulation priority of different mRNAs by the same sRNA. At a constant k_{on}, the strength of translational regulation (k_{xs} /$k]_x$), sRNA-induced RNase E-mediated active degradation (β_e), and regulation efficiency at the co-transcriptional level (P) all positively contribute to the overall regulation efficiency (Figure 2.5B). However, a fast k_{on} (>10^5 $M^{-1}S^{-1}$) is generally needed to repress the target by more than 50% regardless of the rates or efficiencies at other steps (Figure 2.5B), suggesting that binding of the sRNA to the target mRNA might be the rate-limiting step. This is consistent with the observation that *purR*, which has a very low k_{on} rate, has the lowest regulation efficiency among

SgrS regulon despite a higher β_e. The binding kinetics are not correlated with the *in vitro* predicted hybridization thermodynamics (Supplementary Figure 2.7)[134], suggesting that more factors *in vivo* can affect the sRNA target search process. Interestingly, when comparing different sRNA-mRNA pairs, we found a positive correlation between k_{on} and the basal translation rate of the mRNA ($k]_x$) (Figure 3.1), as noted previously[63]. Specifically, higher k_{on} observed for $sodB_{130+30}$ compared to $sodB_{130}$is possibly due to its higher $k]_x$. This correlation implies a potential positive role of translating ribosomes in promoting sRNA binding. From a functional point-of-view, it is logical to have a higher regulation efficiency on the most translated targets under stress conditions to achieve the most effective response. One possible mechanism by which ribosome can facilitate sRNA binding is through unwinding the secondary structures at the sRNA binding site[135], while other potential mechanisms are yet to be uncovered.

2.4 Materials and Methods

2.4.1 Bacterial strains, plasmids

DB166 was made via P1 transduction by moving *lacIq*, *tetR*, *specR* cassette from JH111[67] into DJ480. Δ*ryhB*::cat was moved to DB166 from EM1453 (Jacques et al., 2006) via P1 transduction to create DB186. *rne*701-FLAG-cat was moved into strains DB166 and JH111[67] by P1 transduction from TM528[112] to create XM100 and XM101 respectively. The *ryhB*::tet allele in strain XM221 was created by using primers OXM211 and OXM212 with homology to RyhB to amplify the tetracycline resistance cassette. The PCR product was recombined into the chromosome of XM100 using λred functions provided by pSIM6[136].

Target mRNAs are all encoded by pSMART plasmid and under P_{tet} promoter. Target mRNA reporters carry the small RNA binding sequence from the endogenous mRNAs, and a sfGFP gene (Supplementary Figure 2.7). pSMART_ptsG-10aa-sfGFP ("10aa" refers

to the first 10 codons) was generated from pZEMB8[39] using site directed mutagenesis and the pSMART LCKan Blunt Cloning Kit (Lucigen, 40821-2). Briefly, the lac promoter of pZEMB8 was switched to a tet promoter to reduce leaky expression, using primers (JZ25 and JZ26) that include 5' overhangs containing the tetracycline promoter sequence. The fragment containing the entire promoter, gene of interest, and terminator was generated by PCR using primers EH1 and EH2 and ligated into the pSMART vector, following manufacturer's instructions. pSMART_*manX*-34aa-*sfGFP* was generated following the same method as pSMART_*ptsG*-10aa-*sfGFP*, with pZEMB10[69] serving as the template for the *manX*-34aa-*sfGFP* region, and primers JZ26 and EH3 containing the tetracycline promoter sequence. pSMART_*ptsG*-10aa-*sfGFP* was further used to generate pSMART_*purR*-32aa-*sfGFP* and pSMART_$sodB_{430}$-*sfGFP* using Gibson Assembly. $sodB_{430}$ contains RyhB binding site on *sodB* mRNA and additional 363 nucleotides in the coding region. pSMART_$sodB_{130}$-*sfGFP* and pSMART_$sodB_{130+30}$-*sfGFP* were generated from pSMART_$sodB_{430}$-*sfGFP* by using primers (EH390/EH391 and EH440/441) that amplify the entire plasmid, excluding the regions that were not desired in $sodB_{130}$-*sfGFP* or $sodB_{130+30}$-*sfGFP*. The PCR products were then phosphorylated (NEB M0201S) and ligated (NEB M0202S) before transformation. Each plasmid was confirmed by DNA sequencing and transformed into the various genetic backgrounds utilized in this study.

All cell strains and plasmids used in this work are listed in Table 2.3, and primers used for PCR are listed in Table 2.4.

2.4.2 Culture growth and induction for imaging experiments

For all imaging and qPCR experiments, overnight E. coli cultures were grown in LB media with 25 ug/mL Kanamycin. Overnight cultures were diluted 100-fold in MOPS-Minimal media (TEKnova, M2106) supplemented with 1% glycerol and 25 µg/mL kanamycin at 37 °C. The cells were grown to approximately OD = 0.2-0.3, at which point SgrS or RyhB was induced by adding 0.5% αMG or 500 µM DIP directly to the culture. The

stress was present for 30 minutes before induction of the reporter mRNA construct using 10 ng/mL anhydrous tetracycline (aTc, Sigma-Aldrich). The time of aTc induction marked the t=0 time point in imaging experiments. Fractions of cells were taken at different time points after mRNA induction for downstream sample treatment.

2.4.3 Fluorescence in situ hybridization (FISH)

10 FISH probes targeting the sfGFP coding region, 9 probes for SgrS and 4 probes for RyhB were designed using the Stellaris Probe Designer from Biosearch, and labeled as previously described[38]. sfGFP probes were labeled Alexa Fluor 568 NHS ester (A568, Invitrogen A20003). SgrS and RyhB probes were labeled with Alexa Fluor 647 NHS ester (A647, Invitrogen A20006). The16S rRNA probe was labeled with Alexa Fluor 405 NHS ester (A405, Invitrogen A30000). The A405 signal serves to indicate sufficient permeabilization. FISH was performed as previously described (Fei et al., 2015). 10 mL of culture of cells were taken out at the corresponding time points and fixed with 4% formaldehyde at room temperature (RT) for 30 minutes. Cells were then permeabilized with 70% ethanol for 1 hour at RT. After ethanol permeabilization, 60 µL samples were taken for each time point and cells were additionally permeabilized with 25 µg/mL lysozyme for 10 minutes (1 µg/mL lysozyme corresponds to 70 units/mL). Cells were hybridized with labeled DNA probes (Table 2.4) in the FISH Hybridization buffer (10% dextran sulfate (Sigma D8906) and 10% formamide in 2x SSC) at 30°C in the dark for overnight. The concentration of the labeled probes was 15 nM per probe for mRNAs, 50 nM per probe for sRNAs, and 10 nM for 16S rRNA. After the hybridization, samples were washed three times with 10% formamide in 2x SSC and resuspended in 4x SSC.

2.4.4 Epi-fluorescence Imaging and image analysis

Cells in 4x SSC buffer were imaged in 3D printed 2-well chambers. 1.2-1.4 µL of the sample were placed on the glass slide bottom of the chamber, with a 1% agarose gel pad

placed on top to lay the cells flat. Imaging was performed on a custom inverted microscope (Nikon Ti-E with 100x NA 1.49 CFI HP TIRF oil immersion objective)[137]. Multicolor Z-stack images were taken with 0.130 μm step size and 11 slices for each color. SgrS-A647 and RyhB-A647, mRNA-A568, sfGFP, and 16S rRNA-A405 were imaged with a 647 nm laser (Cobolt 06-01), a 561 nm laser (Coherent Obis LS), a 488 nm laser (Cobolt 06-01), and a 405 nm laser (CrystaLaser, DL405-025-O), respectively. In addition to the multicolor z-stack images, each image had a corresponding differential interference contrast (DIC) image, used for segmentation and image analysis purposes.

Cells were segmented individually based on DIC images using homemade MATLAB code[105]. The segmented cell mask was then overlaid on each color channel stack individually, and the volume-integrated fluorescence intensity was calculated by adding the area-integrated intensities of each cell for the 5 most in-focus slices (the most in-focus slice, and two slices above and below). The background intensities of the image and of the cells due to nonspecific binding of the FISH probes were subtracted from the calculated volume-integrated intensities. The signal contributed by probe nonspecific binding was measured using the same imaging conditions by calculating the volume integrated intensities of cells lacking target RNAs but in the presence of the FISH probes at the same concentration as for positive samples. ΔsgrS cells (JH111) without transformation of any mRNA-sfGFP fusion plasmids were used for background measurements in the sRNA, mRNA, and GFP channels. The 16S rRNA-A405 signal was used as an indicator of sufficiently permeabilized and labeled cells. Background A405 fluorescence intensity distribution due to probe nonspecific binding was first determined using cells labeled with the same concentration of off-target A405-labeled probes. A threshold at the 90th percentile of the background intensity distribution was then used as the 405 intensity cutoff. Cells with 16S rRNA -A405 intensities below this threshold (less than 10% of the total population) were considered not sufficiently permeabilized, and not included in further analysis.

2.4.5 SMLM Imaging and image analysis

Single molecule localization microscopy (SMLM) imaging was conducted using the same microscope as described above with super-resolution modality[137]. Fixed cells were immobilized on the 8-well chambered glass coverslip (Cellvis C8-11.5H-N) using poly-L-lysine (Sigma-Aldrich P8920) and imaged in imaging buffer (50 mM Tris-HCl, 10% glucose,1% 2-Mercapgtoethanol (Sigma-Aldrich M6250), 50 U/mL glucose oxidase (Sigma Aldrich G2133-10KU), 404 U/mL catalase (EMD Millipore 219001) in 2X SSC, pH = 8.0). Images were acquired through a custom programmed data acquisition code, which programs the laser power, camera exposure time, and spot detection threshold, using the Nik_{on} NIS JOBS function. SMLM images were reconstructed with the IDL analysis package as previously published[137].

2.4.6 RT and qPCR

Total RNA was extracted from each sample using Trizol (Thermo Fisher, 15596026) extraction. 2 mL culture of bacterial cells were collected at the desired time point and immediately spun at 12,000 g for 1 minute in cold. The cell pellet was homogenized in 200 μL of trizol incubated at RT for 5 minutes. 1/5 volume of chloroform was added to the Trizol mixture. After incubation for 2-5 minutes at RT, the mixture was centrifuged at 12,000 g for 5 minutes. The upper phase was transferred to a new tube and extracted again with chloroform. The aqueous layer was collected, from which the RNA was then precipitated by standard ethanol precipitation. The total RNA pellet is resuspended in nuclease-free water, and further desalted by a P6 microspin column (Bio-Rad, 7326221). Genomic DNA contamination in the total RNA was further removed by DNase treatment. 2 μL of Turbo DNase (Thermo Fisher, AM2238) was added to 2 μg of total RNA, and the reaction was incubated for 2 hours at 37°C. The DNase was inactivated by adding EDTA (pH = 8) at a final concentration of 15 mM and incubating at 75°C for 10 minutes. The reaction was desalted by a P6 column.

Each reverse transcription (RT) reaction was performed using 50 ng total RNA in 1 mM dNTPs (NEB N0447S), 10% DMSO (Fisher, BP231), 10 mM DTT (Sigma-Aldrich, 10197777001), 250 nM of gene specific reverse primer (IDT), and 20-fold dilution of reverse transcriptase from iScript cDNA Synthesis Kit (Bio-Rad, 1708891) and incubated following manufacturer instructions. Each qPCR reaction was prepared using 1X SsoAdvanced Universal SYBR Green Supermix (Bio-Rad 1725274), 250 nM forward and reverse primers (Table 2.4), and 1 μL of cDNA generated by the RT reaction in a final volume of 20 μL. The qPCR reactions were performed with CFX real-time PCR system (Bio-Rad), using pre-incubation of 95°C for 30 s, followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. The reported D/U ratio ($R_{D/U}$), a ratio between the downstream and upstream amplification of the mRNA target, was calculated as:

$$R_{D/U} = \frac{1}{2^{(Ct_D - Ct_U)}}$$

where Ct_D and Ct_U are the Ct values of the downstream and upstream amplicons respectively.

2.4.7 Determination of sRNA and mRNA copy numbers

To convert the mRNA and sRNA fluorescence values to molecule copy numbers, a qPCR calibration curve of RNA copy number vs. Ct value was first built. *ptsG-sfGFP* mRNA and SgrS were produced using *in vitro* transcription. PCR using forward primers harboring the T7 promoter sequence were used to produce linear dsDNA transcription templates (Table 2.4) and 1 μg template was incubated in T7 buffer (160 mM HEPES-KOH, pH 7.5, 20 mM DTT, 3 mM each rNTP, 20 mM MgCl2, 2 mM spermidine, 120 units SUPERase In RNAse inhibitor) and 10 units T7 RNA polymerase (kind gift from Yuen-Ling Chan) at 37°C for overnight. 4 units TURBO DNase was added to remove template DNA and incubated at 37°C for an additional 2 hours. RNA was extracted using standard

phenol-chloroform and confirmed on a 7% Urea-PAGE gel.

To build a calibration curve between Ct value and RNA copy number, RT reactions were performed on a series of dilutions of *in vitro* transcribed RNA, from 10 ng to 0.001 ng. Different amounts of *in vitro* transcribed RNA were spiked into collected cell samples, then subjected to the same total RNA extraction protocol as described above. Briefly, JH111 cells (Δ*sgrS* cells with the plasmid encoding the mRNA-*sfGFP*) were grown under the same conditions used for imaging and collected when cells reached OD_{600} = 0.2-0.3. Cells were spun down, then homogenized in Trizol. At this point (after adding Trizol, but before subsequently spinning down and adding chloroform) the *in vitro* transcribed RNA was added. RT was performed using iScript cDNA Synthesis Kit (Bio-Rad, 1708891) and qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad 1725274). A linear function was fit between the Ct values of the qPCR reactions and the logarithm of the input RNA copy numbers (Supplementary Figure 2.8A). The copy number of the RNA was calculated using the known molecular weight of the RNA and the amount of RNA added to the initial RT reaction.

To relate RNA copy number and arbitrary fluorescence values, cell samples with different RNA expression levels were subjected to RNA extraction, RT-qPCR, and fluorescence measurement, as described above. Based on the Ct value vs. RNA copy number calibration curve built above, *sfGFP* fusion mRNA and SgrS copy numbers were calculated for the extracted RNA of each sample, and further converted into copy number per cell based on the cell numbers measured by OD_{600} for each sample. RNA copy number per cell was then plotted against the volume-integrated cell fluorescent intensities for each corresponding sample and fit with a linear function (Supplementary Figure 2.8B). Fluorescent intensities of the cells from the imaging experiments were compared to this calibration curve of fluorescent intensity vs. RNA copy number to extract RNA copy number per cell. For RyhB, the conversion factor between SgrS fluorescence values and copy number was multiplied by 4/9, as only 4 FISH probes were used to label RyhB compared to 9 for SgrS,

which was used to create the calibration curve. We assume a linear relationship between number of probes and fluorescent intensity.

2.4.8 Simulation, fitting, and model selection

We used Markov Chain Monte Carlo (MCMC) simulation to explore the parameter spaces of our kinetic models as defined by their ordinary differential equations (ODEs). Specifically, we utilized the emcee package[138], which is a Python implementation of the Goodman-Weare Affine Invariant Ensemble Sampler[139], and integrated the ODEs with the LSODA solver[140,141]. In this approach, an ensemble of parameter sets evolves to sample a Bayesian posterior distribution, which is the product of a prior distribution and a likelihood function. Assuming Gaussian and independent errors, the logarithm of the likelihood (log-likelihood) function takes the form:

$$L = lnp(y|x,\theta) = \sum_m \left\{ -\frac{1}{2} \sum_n \left[\frac{y_{m,n} - f(x_{m,n}|\theta)^2}{\sigma^2_{(m,n)}} \right] \right\},$$

where *m* is the molecular species (mRNA, sRNA, and protein in the WT and *rne*701 strain, for a total of 6), *n* is the time point (7 in our case, t = 0, 1, 3, 6, 12, 18 and 24 min), $y_{m,n}$ is the experimental value for molecular species *m* at time t_n, (in units of copy number for sRNA and mRNA, and arbitrary fluorescent unit for protein), $f(x_{m,n}|\theta)$ is the simulated value for molecular species m at time t_n given the parameter value set θ, $\sigma^2_{(m,n)}$ is the experimental variance for molecular species m at time point t_n. The six-minute folding time of sfGFP is directly accounted for in the fitting process by introducing a six minute time delay in protein observation, meaning that the protein fluorescent signal is assumed to have been produced by mRNA transcribed six minutes earlier (e.g. $y_{\text{WT Protein,t=12}}$ corresponds to $y_{\text{WT mRNA,t=6}}$). The log-posterior distribution is the sum of the log-prior distribution and log-likelihood function.

We fit parameters by running simulations in a two-step process. First, mRNA transcription and translation rates were fit using the –sRNA experimental data, *i.e.* the data acquired from the cell samples in the absence of sRNA. The best fit parameter values and their associated errors were used as prior distributions for transcription and translation rates in the second step, where the rest of the parameters were determined by fitting to the +sRNA experimental data, acquired from cell samples in the presence of sRNA. For the co-transcriptional regulation model using the one-step transcription module, the –sRNA simulations explored a 3-dimensional parameter space: [α_m, k_x, β_m]; and the +sRNA simulations explored a 9-dimensional parameter space: [k_{on}, k_{off}, k_{xr}, β_e, β_{ms}, $k_{x,wt}$, $k_{x,rne}$, $\alpha_{ms,wt}$, $\alpha_{ms,rne}$], where k_{xr} is the ratio k_{xs}/k_x. For the post-transcriptional regulation model using the one-step transcription module, $\alpha_{ms,wt}$ and $\alpha_{ms,rne}$ were set to α_{wt} and α_{rne}, respectively. For the co-transcriptional regulation model using the two-step transcription module, the –sRNA simulations explored a 3-dimensional space: [k_{init}, k_x, β_m]. The elongation rate, k_{elon} was assumed to be a constant for each mRNA, determined by dividing a constant elongation speed (50 nucleotides per second[142]) by the length of the mRNA. The +sRNA simulations explored a 9-dimensional space [k_{on}, k_{off}, k_{xr}, β_e, β_{ms}, $k_{x,wt}$, $k_{x,rne}$, P_{wt}, P_{rne}], where P_{wt} and P_{rne} represent the probability of generating full length mRNA in WT *rne* and *rne*701 backgrounds, respectively. For the post-transcriptional regulation model using the two-step transcription module, P_{wt} and P_{rne} were set to 1. For –sRNA simulations in 3 dimensions, 50 walkers, representing 50 parameter sets, each evolved for 10000 steps, which we found to be a sufficient number of steps for the log posterior to level off. For +sRNA simulations in 9 dimensions, 100 walkers each evolved for 10000 steps. Initial positions for the walkers were chosen at random from the bounded interval of possible values defined by its prior distribution. We used the default settings for the emcee sampler, such that the each move is a "stretch" move, with stretch parameter, a = 2, giving an average acceptance fraction equal to 0.44[138,139].

For –sRNA fitting, the prior distributions for the free parameters were uniform distributions (Table 2.5). For the +sRNA fitting, the prior distributions of the parameters determ-

ined from the –sRNA fitting were normal distributions centered on their –sRNA maximum *a posteriori* (MAP) values, and the prior distributions for the remaining parameters were uniform distributions. After the parameter fitting, the posterior probability distributions of the fitted parameters were determined, along with their MAP values and associated errors. For experimentally determined variables, the widths of the normal distributions were determined by their experimental errors. For the remaining free parameters, the widths of the uniform distributions were set empirically, either by observing physical constraints (e.g., k_{on} is constrained by the diffusion limit) or by logical constraints (e.g., k_{xr} cannot be below 0 or above k_x).

Each experimental replicate was fit separately. $\sigma^2_{(m,n)}$ was the same across all replicates. A single set of parameter values was chosen to be the best fit for the combined samples by selecting the point estimate of the MAP parameter values for the best walker for each replicate, then averaging over the replicates. One replicate in a –sRNA simulation was one experimental dataset containing mRNA and associated protein values, with datasets for WT and *rne*701 backgrounds fit separately. One replicate in a +sRNA simulation was a combination of one experimental dataset in the WT background, and one in the *rne*701 background. The reported parameter values and their associated errors were the mean and standard deviations of the MAP values from all simulations, respectively. All simulations were performed with custom software written in Python, and parallelization was implemented using emcee. We utilize both CPU and GPU functions to maximize the efficiency of our simulations. All codes for all simulations are available publicly on GitHub: https://github.com/JingyiFeiLab/Regulation_Kinetics.

The Bayesian information criterion (BIC) was used for model selection between post-transcriptional and co-transcriptional regulation models. The BIC is defined as:

$$BIC = kln(n) - 2ln(\hat{L})$$

where $\hat{L}$ is the maximized likelihood value of the model, *k* is the number of parameters (*k* = 7 for post-transcriptional model, *k* = 9 for co-transcriptional model, accounting for the added variables P_{wt} and P_{rne}), *n* is the number of data points or observations (*n* = 42 in our case, representing 7 time points x 3 molecules x 2 *rne* backgrounds). For each target, the minimized BIC was calculated for both the post- and co-transcriptional models, and the model which produced the lowest BIC was selected.

2.5 Supplementary Information

Figure 2.7 sRNA target sequences.
of SgrS mRNA targets (*ptsG*, *manX*, and *purR*) and RyhB mRNA target (*sodB*). Vertical bars represent base pairing between the mRNA and sRNA. Purple boxes represent the Shine-Dalgarno sequence of the mRNA. Blue boxes represent the start codon of the mRNA-sfGFP fusion. The *purR* Shine-Dalgarno sequence and start codon are far upstream of the sRNA binding site. ΔG for each base pair interaction was calculated using the DINAMelt server[134]

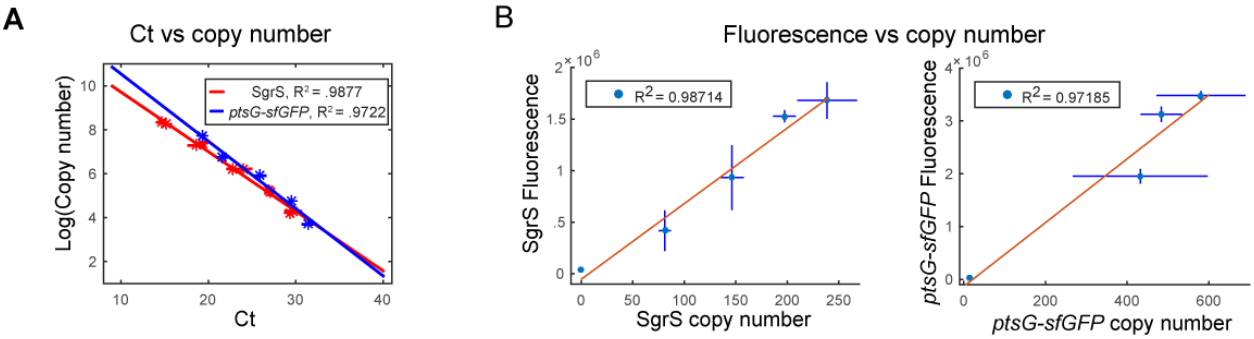


Figure 2.8 RT-qPCR calibration for RNA copy number determination.
(A) Calibration curve relating copy number of the input RNA to the Ct value. Dilution series of *in vitro* transcribed SgrS or *ptsG-sfGFP* mRNA was spiked into the cell pellet of the Δ*sgrS* strain without the plasmid encoding the target mRNA. The spike-in RNAs and the cell pellet were put through total RNA extraction, RT and qPCR reactions and Ct values were determined.. **(B)** Calibration curves relating the fluorescence value of each RNA from image to RNA copy number determined by RT-qPCR and calibration curve from (A). Error bars represent standard deviation of duplicate measurements.

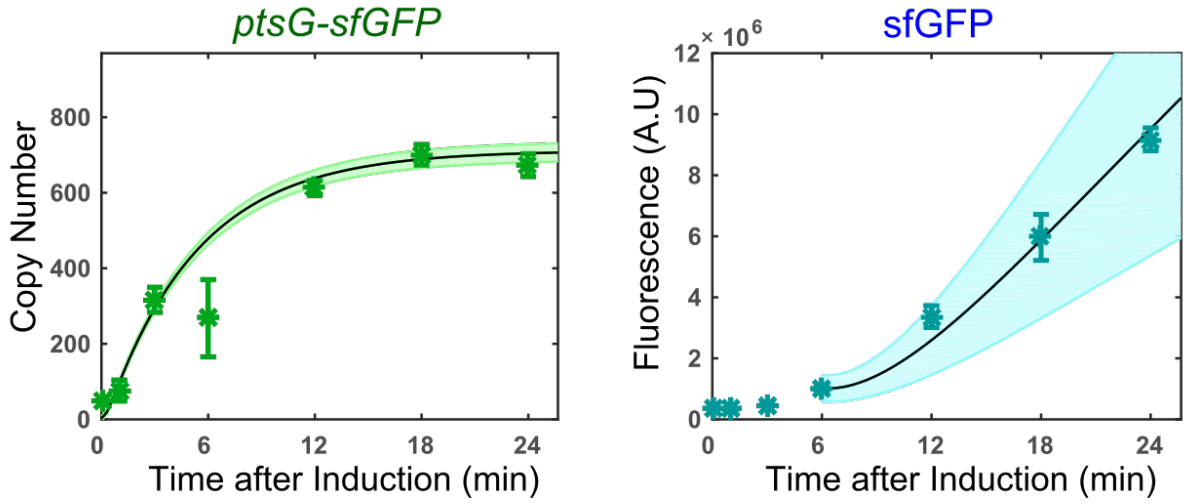


Figure 2.9 Transcription and translation of *ptsG-sfGFP* in the absence of αMG in the WT strain.
The mRNA and protein levels are significantly higher than those in the Δ*sgrS* strain in the presence of αMG experiments (Figure 2.2C). Error bars represent standard deviation of duplicate measurements.

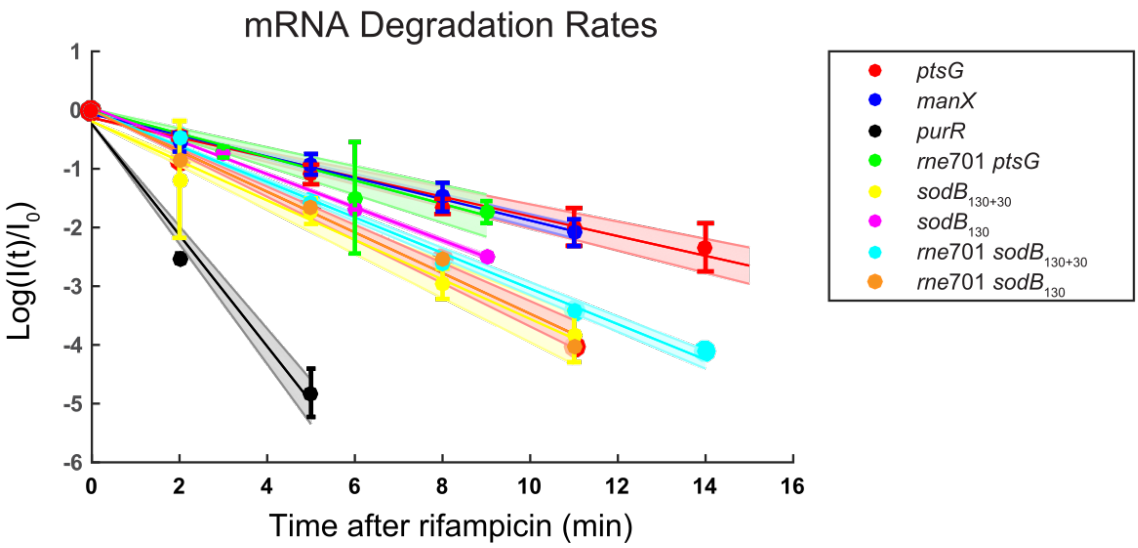


Figure 2.10 Measurements of mRNA degradation rates.
Degradation rates for all target mRNAs used in this study were determined with rifampicin pulse chase experiments. Degradation rates were measured in both WT *rne* and *rne*701 backgrounds for *ptsG* and *sodB* constructs, confirming that the degradation rates for the target mRNA fusions were the same in both backgrounds. Y-axis shows the fluorescent intensity of the target mRNA relative to its intensity when rifampicin was added. Degradation rates were determined by fitting the time-dependent relative fluorescent intensity with an exponential decay model. Error bars represent standard deviation of duplicate measurements.

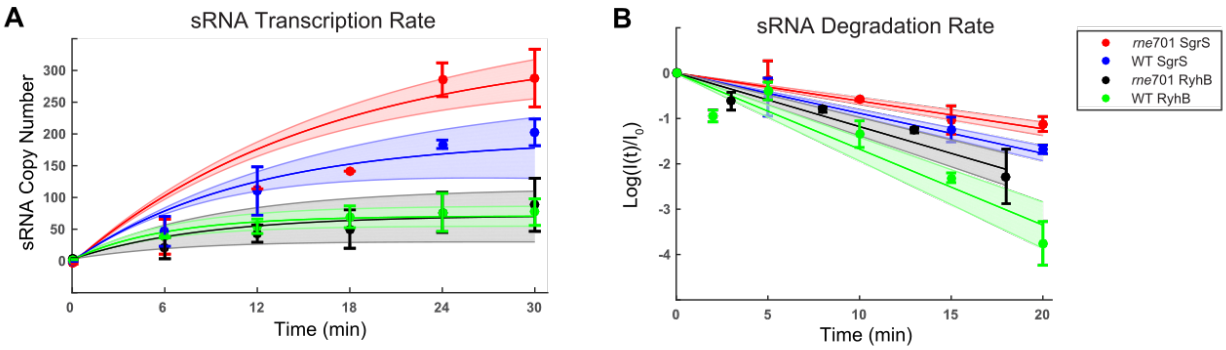


Figure 2.11 Measurements of sRNA transcription and degradation rates.
(A) SgrS and RyhB transcription rates were measured in both WT *rne* and *rne*701 backgrounds by inducing the appropriate stress in the absence of target mRNA. **(B)** SgrS and RyhB degradation rates were measured in both WT *rne* and *rne*701 backgrounds by inducing sRNA to steady state levels, then removing the inducer from the media. Degradation rates were determined by fitting the time-dependent relative fluorescent intensity with an exponential decay model. Error bars represent standard deviation of duplicate measurements.

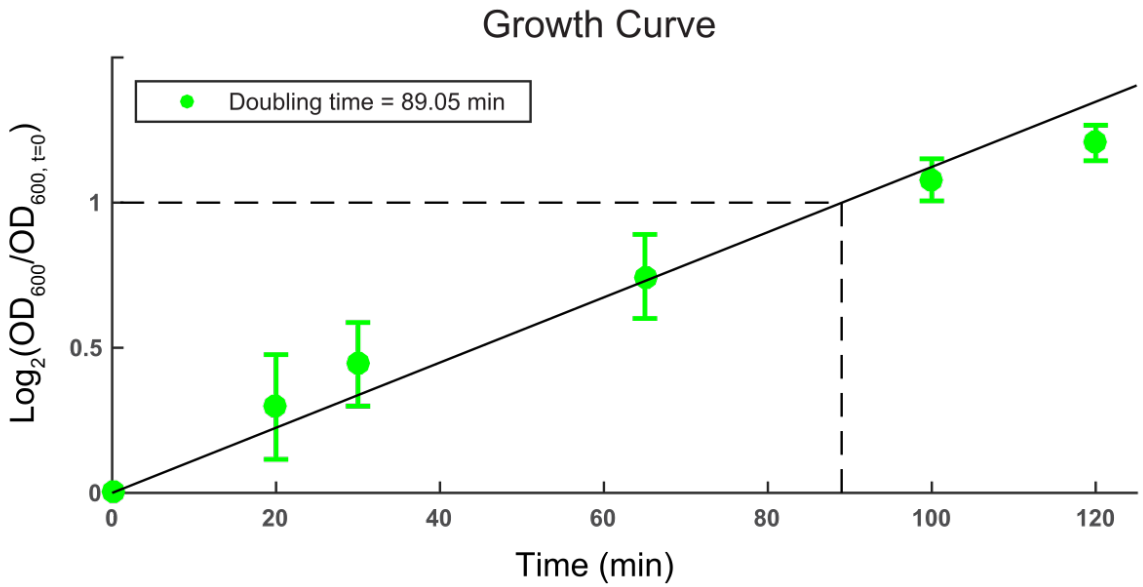


Figure 2.12 Cell doubling time measurement.
. $Log_2(OD_{600})$ relative to OD_{600} at first time point in the exponential growth phase is plotted over time to determine cell doubling time. Error bars represent standard deviation of duplicate measurements.

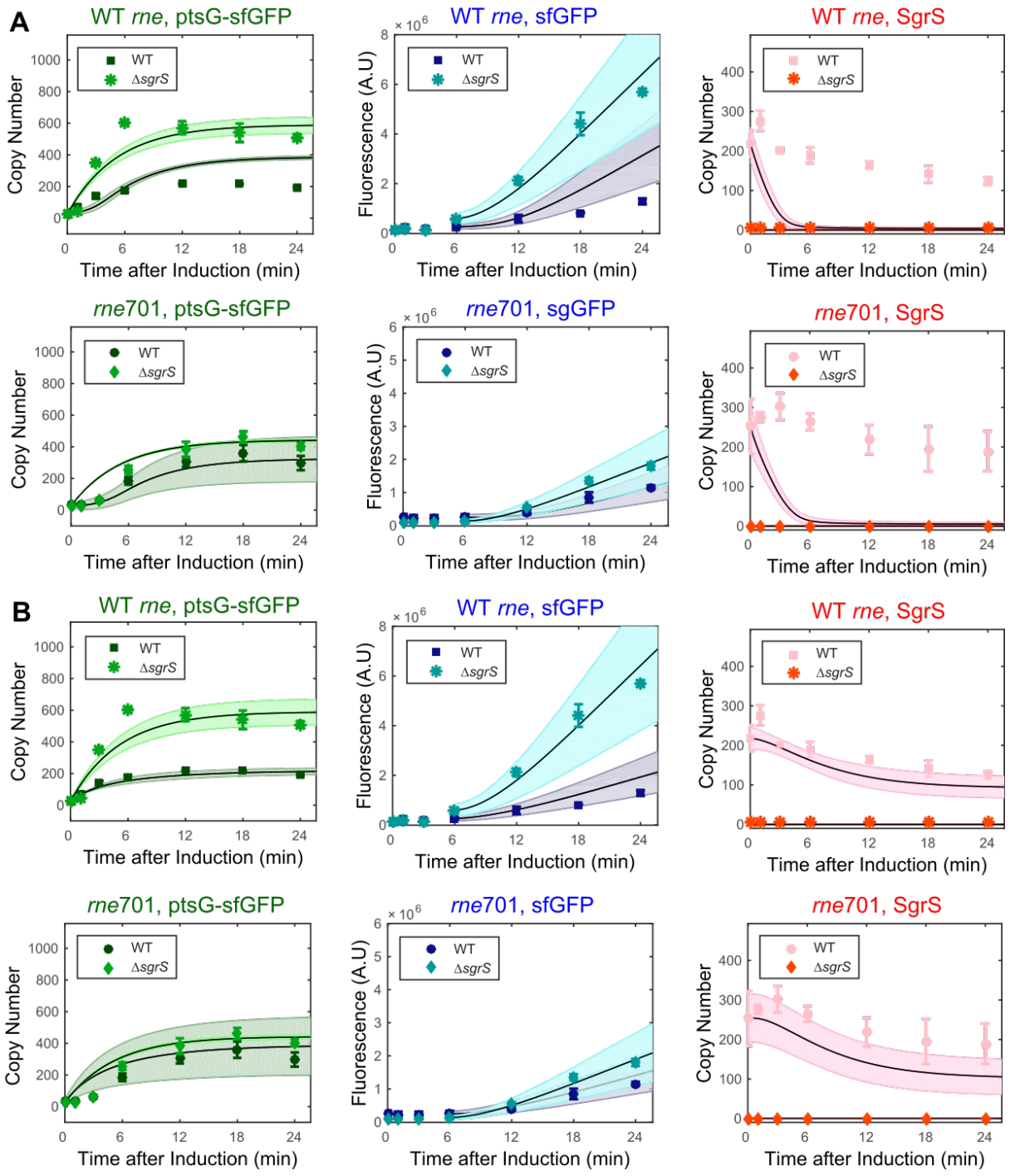


Figure 2.13 Fits of SgrS regulation on ptsG-sfGFP using one-step transcription module. **(A)** Best fit of post-transcriptional regulation model for *ptsG*. **(B)** Best fit of co-transcriptional regulation model for *ptsG* using one-step transcription module.

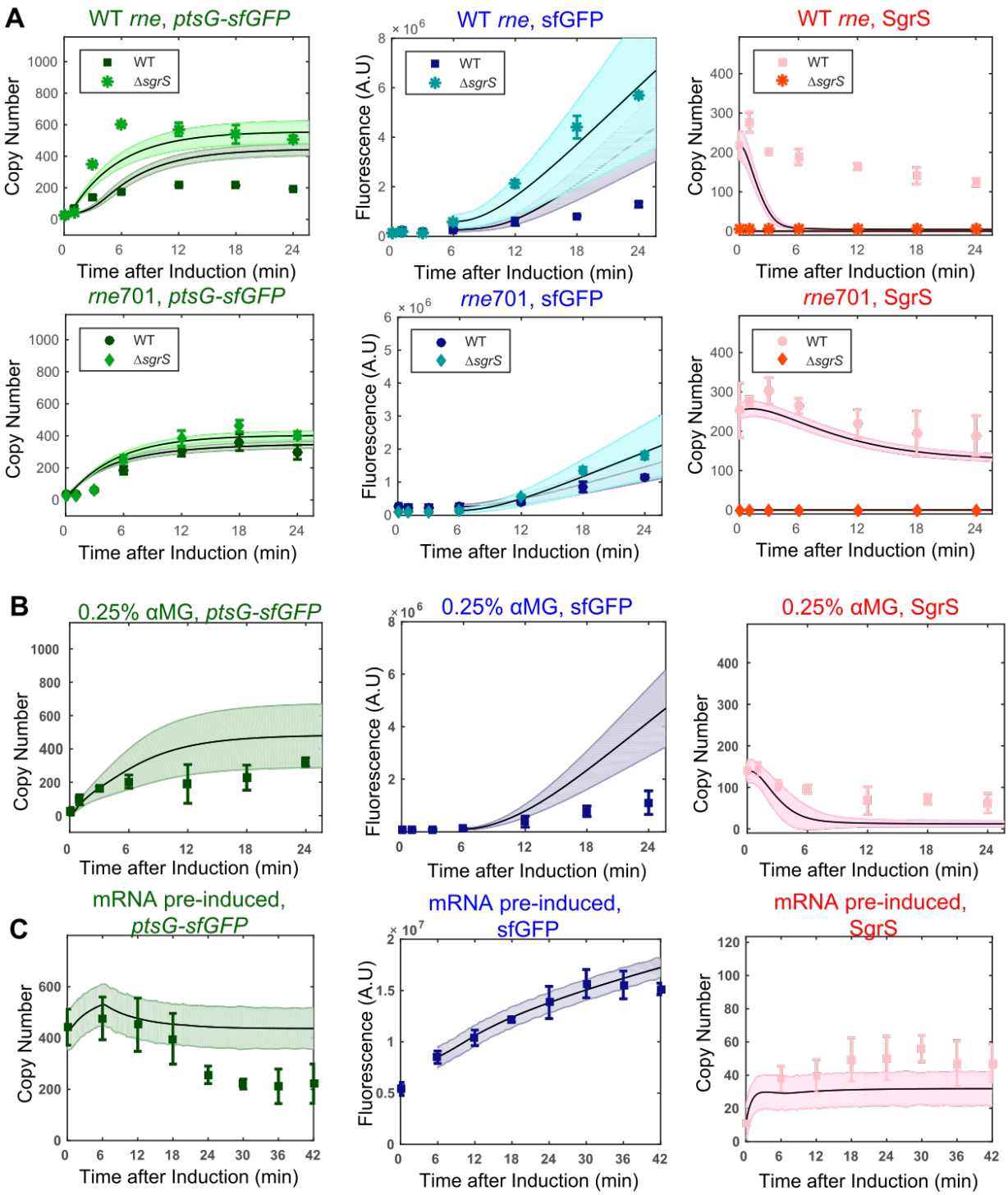


Figure 2.14 Fits and predictions of post-transcriptional regulation model for SgrS regulation on *ptsG-sfGFP* using two-step transcription module.
(A) Best fit of post-transcriptional regulation model *ptsG* in WT and *rne*701 background. **(B)** Simulation of best fit using post-transcriptional model on 0.25% αMG validation dataset. **(C)** Simulation of best fit using post-transcriptional model on pre-induced mRNA dataset.

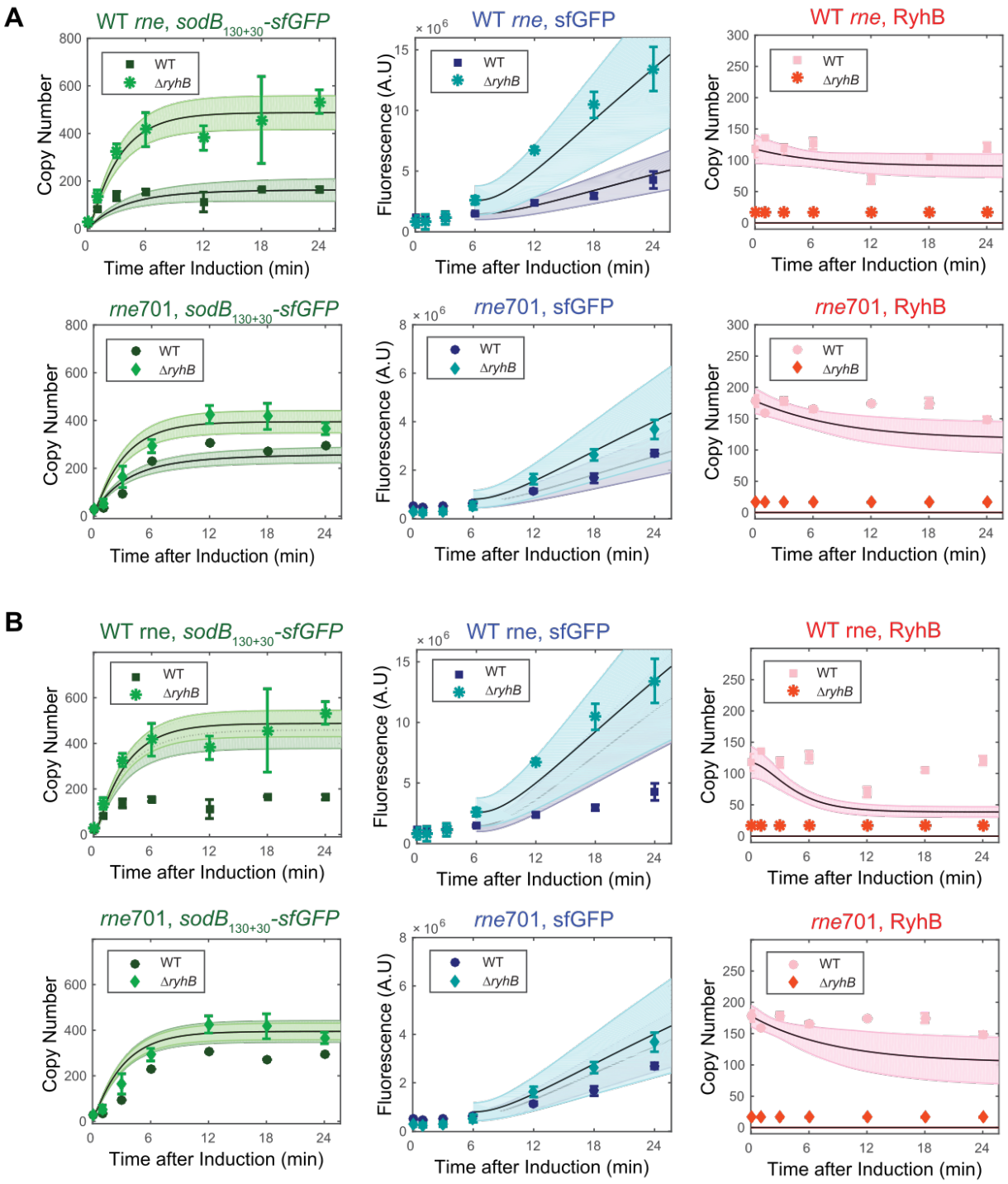


Figure 2.15 Model comparisons for *sodB*$_{130+30}$ regulation.
(A) Best fit of co-transcriptional regulation model for RyhB regulation of *sodB*$_{130+30}$-*sfGFP*. **(B)** Best fit of post-transcriptional regulation model for RyhB regulation of *sodB*$_{130+30}$-*sfGFP*.

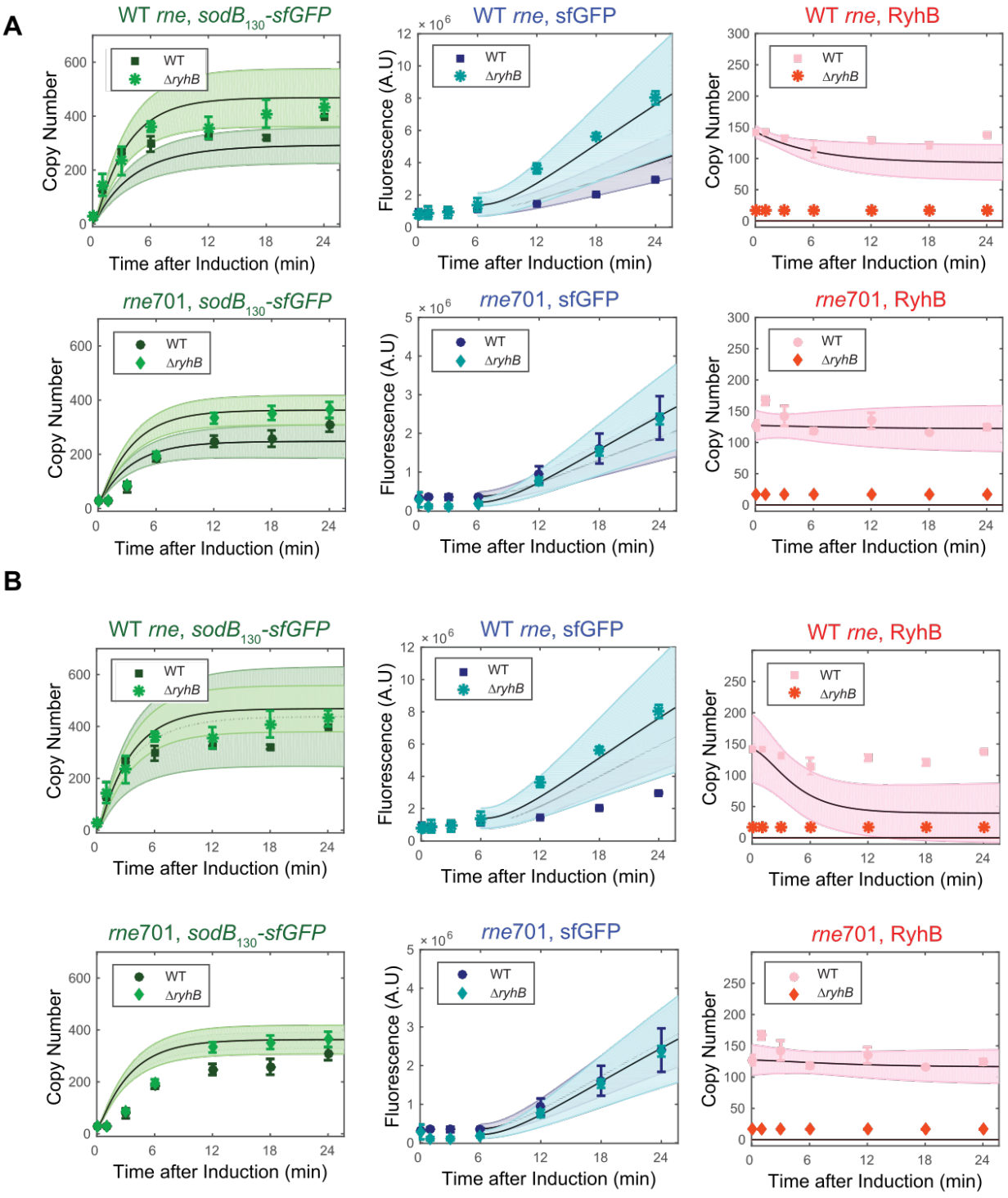


Figure 2.16 Model comparisons for *sodB*$_{130}$ regulation.
(A) Best fit of co-transcriptional regulation model for RyhB regulation of *sodB*$_{130}$-*sfGFP*. **(B)** Best fit of post-transcriptional regulation model for RyhB regulation of *sodB*$_{130}$-*sfGFP*.

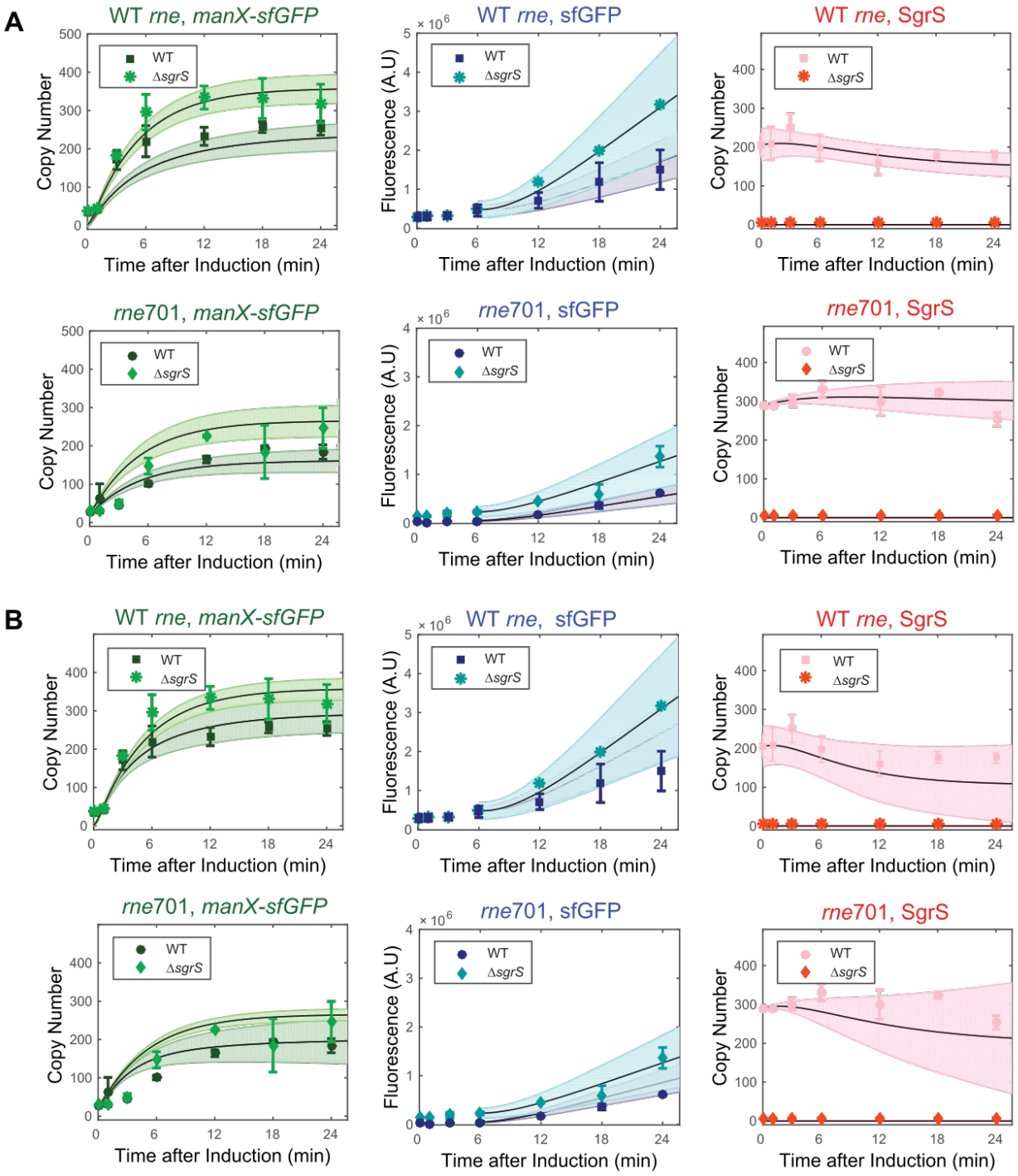


Figure 2.17 Model comparisons for *manX* regulation.
(A) Best fit of co-transcriptional regulation model for SgrS regulation of *manX-sfGFP*. **(B)** Best fit of post-transcriptional regulation model for SgrS regulation of *manX-sfGFP*.

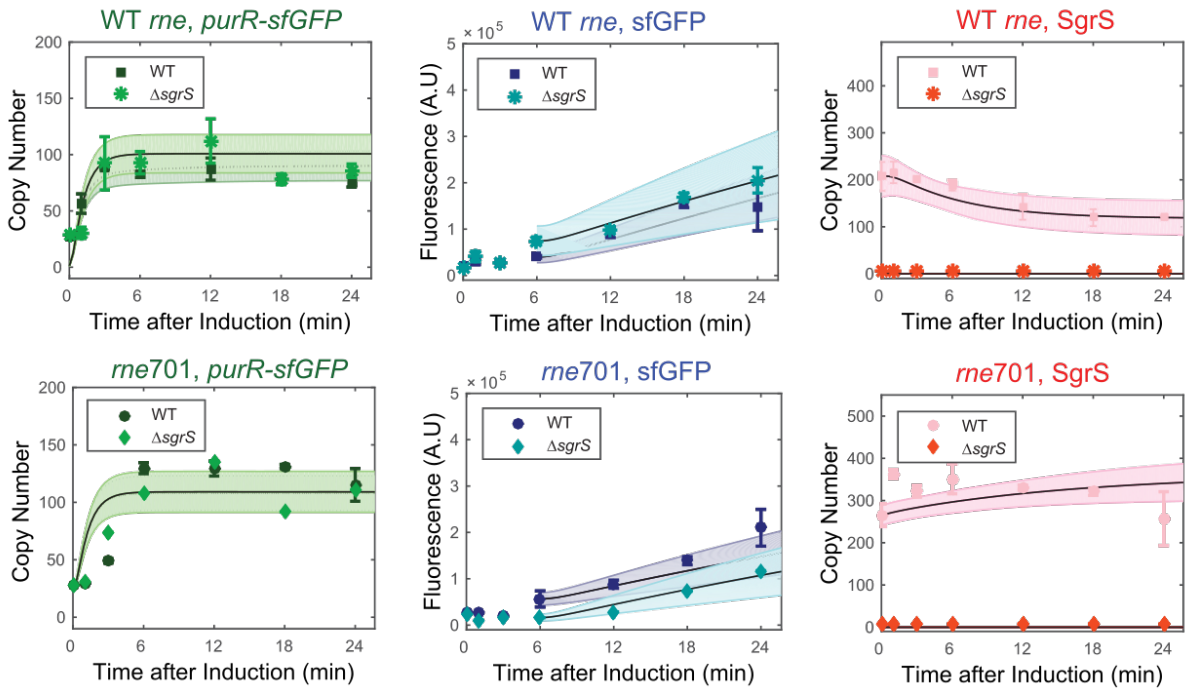


Figure 2.18 Post-transcriptional model for *purR* regulation.
Best fit of post-transcriptional regulation model for SgrS regulation of *purR-sfGFP*. As the BIC for the post-transcriptional model of *purR* regulation was lower than the BIC for co-transcriptional model, the post-transcriptional model was chosen and the co-transcriptional model is not shown here.

Table 2.1 Basal activity kinetic parameters.

SgrS			
α_s **(molecules ·s^{-1})**	0.36 ±0.03		
β_s (WT) **(s^{-1})**	(1.5 ±0.3) x10^{-3}		
β_s (*rne*701) **(s^{-1})**	(1.0 ±0.1) x10^{-3}		
Targets of SgrS	***ptsG***	***manX***	***purR***
α_m (Δ*sgrS* + *α*MG) **(molecules ·s^{-1})**	0.09 ±0.01	0.06 ±0.01	0.08 ±0.01
α_m (*rne*701ΔsgrS + *α*MG) **(molecules ·s^{-1})**	0.07 ±0.01	0.04 ±0.01	0.08 ±0.01
k_x (ΔsgrS) **(AU•s^{-1})**)	14 ±1	11 ±1	1.5 ±0.7
k_x (rne701ΔsgrS) **(AU•s^{-1})**)	6.1 ±0.4	5.46 ±0.03	0.7 ±0.3
β_m **(s^{-1})**	(3.2 ±0.1) x10^{-3}	(3.3 ±0.1) x10^{-3}	(1.7 ±0.1) x10^{-2}
RyhB			
α_s **(molecules ·s^{-1})**	0.26 ±0.07		
β_s (WT) **(s^{-1})**	(2.8 ±0.2) x10^{-3}		
β_s (*rne*701) **(s^{-1})**	(2.0 ±0.4) x10^{-3}		
Targets of RyhB	***sodB$_{130+30}$***	***sodB$_{130}$***	
α_m (Δ*ryhB* + DIP) **(molecules ·s^{-1})**	0.11 ±0.03	0.12 ±0.01	
α_m (*rne*701Δ*ryhB* + DIP) **(molecules ·s^{-1})**	0.11 ±0.01	0.09 ±0.02	
k_x (ΔryhB) **(AU•s^{-1})**	27 ±2	19.9 ±0.9	
k_x (*rne*701ΔryhB) **(AU•s^{-1})**	10 ±2	8 ±1	
β_m **(s^{-1})**	(5.2±0.1) x10^{-3}	(5.0 ±0.1) x10^{-3}	

Table 2.2 Parameter Comparison between sRNA-mRNA pairs[a].

mRNA	*pstG*	*manX*	*purR*[b]	$sodB_{130+30}$	$sodB_{130}$
sRNA	**SgrS**	**SgrS**	**SgrS**	**RyhB**	**RyhB**
Repression%	0.78 ± 0.07	0.53 ± 0.18	0.20 ± 0.05	0.68 ± 0.14	0.63 ± 0.03
Steady State Protein Repression%	0.57	0.43	0.05	0.67	0.48
Steady State mRNA Repression%	0.48	0.33	0.11	0.67	0.37
k_{on} ($M^{-1}s^{-1}$)[c]	$(1.3\pm0.6)\times10^{6}$	$(3.4\pm1.2)\times10^{5}$	$(3.3\pm1.2)\times10^{4}$	$(14\pm0.6)\times10^{5}$	$(4.6\pm0.9)\times10^{5}$
k_{off} (s^{-1})	0.30 ± 0.04	0.21 ± 0.12	0.56 ± 0.14	9.6 ± 0.3	9.1 ± 0.4
k_{xs}/k_x	0.14 ± 0.06	0.28 ± 0.12	0.63 ± 0.13	0.36 ± 0.16	0.43 ± 0.29
β_{ms} (s^{-1})	$(5.0\pm1.5)\times10^{-3}$	$(3.7\pm0.3)\times10^{-3}$	$(2.6\pm0.8)\times10^{-2}$	$(5.6\pm0.1)\times10^{-3}$	$(5.8\pm0.4\times10^{-3}$
β_e (s^{-1})	$(1.3\pm1.1)\times10^{-3}$	$(2.3\pm1.2)\times10^{-3}$	$(5.7\pm1.2)\times10^{-1}$	$(2.2\pm1.5)\times10^{-3}$	$(4.9\pm0.8)\times10^{-4}$
***P* (WT)**	0.32 ± 0.14	0.43 ± 0.21	N/A (Post)	0.04 ± 0.02	0.16 ± 0.07
***P* (*rne*701)**	0.93 ± 0.04	0.57 ± 0.09	N/A (Post)	0.80 ± 0.02	0.50 ± 0.23
BIC (co)	497.420	481.706	436.404	492.869	494.216
BIC (post)	577.487	486.493	429.181	669.051	515.328

[a] Reported parameter values and their associated errors are the mean and standard deviations of the outputted MAP values from all replicates. One replicate in a +sRNA simulation is a combination of one experimental dataset in WT background, and one in *rne*701 background

[b] b. Best fit parameters for purR were determined from experiments using 1% αMG, rather than the 0.5% used for other SgrS targets. This is due to the fact that repression was minor for purR at 0.5% αMG. The repression % for purR was determined using 0.5% αMG induction for a fair comparison with other SgrS targets.

[c] k_{on} is reported as $M^{-1}s^{-1}$ assuming 1 molecule corresponds to 1 nM in *E. coli* cells[143].

Table 2.3 List of all strains and plasmids used in this study.

Strain	Description	Reference
DB166	*WTsgrS, λattB::lacIq,tetR, specR*	This study
JH111	*ΔsgrS, λattB::lacIq,tetR, specR*	Reference[67]
XM100	*lacIq tetR specR rne701-FLAG::cat*	This study
XM101	*ΔsgrS lacIq tetR specR rne701-FLAG::cat*	This study
XM221	*lacIq, tetR, specR, rne701-FLAG::cat, ryhB::tet*	This study
DB186	*lacIq, tetR, specR, ryhB::cat*	This study
Plasmids	**Description**	**Reference**
	pSMART_*ptsG*-10aa-*sfGFP*	This study; modified from Reference[69]
	pSMART_*manX*-34aa-*sfGFP*	This study; modified from Reference[39]
	pSMART_*purR*-32aa-*sfGFP*	This study
	pSMART_*sodB*430-*sfGFP*	This study
	pSMART_*sodB*130-*sfGFP*	This study
	pSMART_*sodB*130+30-*sfGFP*	This study; modified from Reference[46]

Table 2.4 List of all oligonucleotides used in this study.

Strain/plasmid generation primers	Sequence 5'-3'	Description
JZ25	TCCCTATCAGTGATAGAGATACTG-GAGCACAGAATTCATAAATAAAGGG	Tet promoter + ptsG
JZ26	TCAATCTCTATCACTGATAGGGACTTTCTC-GAGGTGAAGACGAAA	Tet promoter + pZEMB vector
EH1	TCGTCTTCACCTCGAGAAAGTC	Amplifies tet_mRNA-sfGFP for ligation into pSMART
EH2	CGAACGCCTAGGTCTAGGGCGG	Amplifies tet_mRNA-sfGFP for ligation into pSMART
EH3	TCCCTATCAGTGATAGAGATACTGGATACT-GAGCACAGAATTC	Tet promoter + manX
EH307	TGGACCTGGGGATCCGCTGGCTCCG	sodB430 + pSMART

Table 2.4 - continued from previous page

EH308	CAGCGGATCCCCAGGTCCAGCCAGAAC-CAAAG	pSMART + sodB430
EH309	ATTGTGCGTATGAATTCTGT-GCTCCAGTATCTCTATCACTG	sodB + pSMART
EH310	GCACAGAATTCATACGCACAATAAGGCTAT-TGTACGTATG	pSMART + sodB
EH390	CTCGATGGTTTCCGCAGAAATGTG	sodB130
EH391	GGATCCGCTGGCTCCGC	sodB130
EH440	ATCAAAAACTTTGGTTCTGGCTGG	sodB130+30
EH441	CTCGATGGTTTCCGCAGAAA	sodB130+30
OXM211	GTGTTGGACAAGTGCGAAT-GAGAATGATTATTATTGTCTC CAT-TAATTCCTAATTTTTGTTGACACTCTATC	
OXM212	AAGCACTCCCGTGGATAAAT-TGAGAACGAAAGATCAAAAA GAATAA-CATCATTTGGTGACGAAATAACTA	
OXM112	ATGAGCAAAGGAGAAGAAC	pSMART
OXM113	GAATTCTGTGCTCCAGTATC	pSMART
OXM115	gttcttctcctttgctcatGAATTCGCCAGAACCAGC	purR + pSMART
OXM116	atactggagcacagaattcTACACTATTTGCGTACT-GGC	purR + pSMART
qPCR primers	**Sequence 5'-3'**	
ptsG_U_F	CAGAATTCATAAATAAAGGGCGCTTAGA	qPCR targeting ptsG-sfGFP up-stream of SgrS binding site
ptsG_U_R	TCTCACGCGTGGCAAGG	qPCR/RT targeting ptsG-sfGFP up-stream of SgrS binding site
ptsG_D_F	CCGTTCAACTAGCAGACCATTA	qPCR targeting ptsG-sfGFP down-stream of SgrS binding site
ptsG_D_R	GACAGATTGTGTCGACAGGTAA	qPCR/RT targeting ptsG-sfGFP downstream of SgrS binding site
16S rRNA_F	AGGCCTTCGGGTTGTAAAGT	qPCR targeting ribosomal RNA

Table 2.4 - continued from previous page

16S rRNA_R	ATTCCGATTAACGCTTGCAC	qPCR/RT targeting ribosomal RNA
SgrS_F	AGCGTCCCACAACGATTAAC	qPCR targeting SgrS
SgrS_R	CACCAATACTCAGTCACACATGA	qPCR targeting SgrS
In vitro transcription primers	**Sequence 5'-3'**	
SgrS +T7_F	TAATACGACTCACTATAGGGAT-GAAGCAAGGGGGTGC	
SgrS +T7_R	AAAAAAAACCAGCAGGTATAATCTGCT	
ptsG-sfGFP +T7 F	TAATACGACTCACTATAGGCACAGAATTCATAAATAAAGGGCG	
ptsG-sfGFP +T7 R	CCGCCCTAGACCTAGGCGTTCG	
FISH Probes	**Sequence 5'-3'**	
SgrS_1	GTGCTGATAAAACTGACGCA	
SgrS_2	ACTTCGCTGTCGCGGTAAAA	
SgrS_3	CTTAACCAACGCAACCAGCA	
SgrS_4	CATGGTTAATCGTTGTGGGA	
SgrS_5	ATCCCACTGCATCAGTCCTT	
SgrS_6	GTCAACTTTCAGAATTGCGG	
SgrS_7	TCAGTCACACATGATGCAGG	
SgrS_8	GCGGGTGATTTTACACCAAT	
SgrS_9	AACCAGCAGGTATAATCTGC	
sfGFP_1	ATTTGTGCCCATTAACATCA	
sfGFP_2	GAGTAGTGACAAGTGTTGGC	
sfGFP_3	TCATGTGATCCGGATAACGG	
sfGFP_4	TAGTGCGTTCCTGTACATAA	
sfGFP_5	GCCGTGATGTATACATTGTG	
sfGFP_6	GTTAGCTTTGATTCCATTCT	
sfGFP_7	GCTAGTTGAACGGAACCATC	

	Table 2.4 - continued from previous page
sfGFP_8	CGCCAATTGGAGTATTTTGT
sfGFP_9	TGTCGACAGGTAATGGTTGT
sfGFP_10	TCAAGAAGGACCATGTGGTC
RyhB_1	GCGAGGGTCTTCCTGATCGC
RyhB_2	ATGTCGTGCTTTCAGGTTCT
RyhB_3	AATACTGGAAGCAATGTGAG
RyhB_4	GCCAGCACCCGGCTGGCTAA
16S rRNA	CCC CAG TCA TGA ATC ACA AA

Table 2.5 Parameter prior distributions.

α_m **(molecules** $\cdot s^{-1}$**)**	$N(\mu_\alpha,\sigma_\alpha)$[a]
k_x **(AU•**$\cdot s^{-1}$**)**	$N(\mu_k,\sigma_k)$
k_{on} **(molecules**$^{-1}s^{-1}$**)**	$U(10^{-9},10^{-2})$[b]
k_{off} **(molecules** $\cdot s^{-1}$**)**	$U(10^{-5},10)$
k_{xs}/k_x	U(0,1)
β_{ms} **(**s^{-1}**)**	U(β_m,1.0)
β_e **(**s^{-1}**)**	$U(10^{-6},1.0)$

[a] Normal distribution with mean and standard deviation calculated from –sRNA experiments

[b] Uniform distribution

2.6 Appendix: Fitting and Simulation Code Instructions

2.6.1 Introduction and requirements

The following is an instruction manual for the pipeline I used to get from raw image data to kinetic parameter values. It can be followed exactly to reproduce the results de-

scribed in this chapter, or modified according to the needs of the user. If the user would like to run the protocol as is, the following are required:

1. A full analysis, from raw image data to kinetic parameter values requires *at least* eight imaging datasets: two replicates each of wild-type (WT) *rne*, Δ*sgrS* and *rne*701, Δ*sgrS* (in order to calculate the transcription and translation rates), and two replicates each of WT *rne*, WT SgrS and *rne*701, WT SgrS. Of course, more replicates of any sample can be acquired and accommodated, but these are the minimum set. Each fluorescent image is assumed to be a multichannel Z-stack, with a violet channel included, in .nd2 format. These .nd2 files should be converted to .tif format using the NIKON conversion program.

2. Converted fluorescent image data is organized in a tiered folder organization, following the naming format: ImageFolder = ”Date”/”Sample”/”Time”/”Replicate”. For example, the image of the second replicate from the t=24 time point of the sample ”MR156” (pSMART_*ptsG*-10aa-*sfGFP*, wt *rne* background), acquired on January 25, 2021 can be found in the folder ”January_25_2021/MR156/t24/sample2/”. Within the folder ”January_25_2021/MR156/” is a separate folder for each time point, named according to the above example.

3. Each fluorescent image has an associated DIC image, labeled according to the replicate. For example, the DIC image associated with the second image in the January 21, 2021, MR156, t = 24 dataset is called ”dic2.tif”, with full extension name ”January_25_2021/MR156/t24/dic2.tif”

4. The most in-focus slices of the Z-stack in the violet channel and one other channel are recorded. For example, for a Z-stack with the 647, 568, 488, and 405 nm channels recorded, the most in-focus slice for the violet (405 nm) channel might be slice 6, and the most in-focus slice for the 647 channel might be slice 4. These should be recorded somewhere, most likely a lab notebook. Notably, the 647, 568, and 488

channels always have the same in-focus slice (assuming there is no z-drift during the data acquisition), but that slice is often different from the most in-focus slice for the violet channel.

The ensuing instructions will be written assuming these requirements have been fulfilled and the folders and files are named according to this format.

2.6.2 Segmentation

The first step in the analysis is to segment the cells and calculate their volume-integrated fluorescent intensities using the MATLAB code ”seg2D_DIC.m”.

1. Line 4: change ”parentDir” to the folder path you would like all your analyzed fluorescent image data to be saved to. For example, all of my data (all dates, all samples) are saved within the folder called ”/Users/reyer/Data/SingleCellEpi/”. Thus, I change Line 4 to: parentDir = ”/Users/reyer/Data/SingleCellEpi/”. Data will be further organized automatically according to the strain name and date of the dataset.

2. Line 6: Input the strain numbers to be analyzed. All of my strains are named according to the format ”MRx”, where ”x” is replaced with the strain number. The descriptions corresponding to each strain number are stored in a Google Sheet.

   For example, if you are to analyze a dataset of strain ”MR156”, replace Line 6 with: strain = 156. If you would like to analyze multiple datasets at once, perhaps two ”MR156” and ”MR162”, replace Line 6 with: strain = [156,156,162], etc.

   If User X would like to label their strain names using their own initials, for example ”UX156”, they can replace the ”MR” in Line 28 with ”UX”.

3. Line 8: Input the date of each dataset. For example, to analyze a dataset acquired on January 25, 2021, change Line 8 to: Date = {’January_25_2021’}. To analyze

three datasets, two from January 25, 2021, and one from January 27, 2021, change Line 8 to: Date = {'January_25_2021','January_25_2021','January_27_2021'}.

4. Line 13: Input time points of dataset. For example, for the experiments described in Chapter 2, replace Line 13 with: time = [0,1,3,6,12,18,24].

5. Line 15: Input the number of fluorescent channels in the dataset. For example, for a dataset with images in the 647, 568, 488, and 405 nm channels, replace Line 15 with: channels = 4.

6. Line 16: Input the identity of the violet channel. Fluoresent image datasets are acquired sequentially by imaging every slice in a channel, then moving to the next channel and imaging every slice, and so on. If the violet channel is the fourth channel in this sequence, replace Line 16 with: violet_channel = 4

7. Line 17: Input the most in-focus slice for each non-violet fluorescent channel. For example, for a single dataset including only two time points, t=0 and t=1, with three replicates each, where the most in-focus slices for the three replicates of t=0 are 5, 9, and 9, respectively, and the most in-focus slices for the three replicates of t=1 are 9, 9, and 6, respectively, replace Line 17 with: ref_slice_i = {{[5,9,9],[9,9,6]}}

8. Line 18: Input the most in-focus slice for each violet fluorescent channel. For example, for a single dataset including only two timepoints, t=0 and t=1, with three replicates each, where the most in-focus slices in the violet channel for the three replicates of t=0 are 6, 10, and 10, respectively, and the most in-focus slices for the three replicates of t=1 are 10, 10, and 7, respectively, replace Line 18 with: violet_slice_i = {{[6,10,10],[10,10,7]}}

After these steps have been completed, "seg2D_DIC.m" can be run. No other steps are required. A single .mat file will be saved in the directory specified in Line 4 for each analyzed image. For example, the analyzed .mat files corresponding to the three replic-

ates of the t=0 time point of the sample ”MR156” acquired on the date January 25, 2021 will be stored automatically in the folder ”/Users/reyer/Data/SingleCellEpi/MR156/January_25_2021/t0/” as ”sample_001.mat”, ”sample_002.mat”, and ”sample_003.mat”. Each .mat file contains all the input and output variables so that they can be accessed at any time. Of particular interest are the structured array titled ”part4”, which contains the Cell ID, volume, center, and volume-integrated intensity in each channel for each cell, and the intermediate segmented images, which can be checked for quality assurance.

2.6.3 Violet channel filtration

In the experiments described in Chapter 2, the 16S rRNA-A405 signal was used as an indicator of sufficiently permeabilized cells. A threshold at the 90th percentile of the background intensity distribution was then used as the 405 intensity cutoff. Cells with 16S rRNA -A405 intensities below this threshold (less than 10% of the total population) were considered not sufficiently permeabilized, and not included in further analysis. The violet filtration can be done using ”VioletNormalize.m” following the segmentation.

1. Line 4: Input the time points, same as in ”seg2D_DIC.m”. For example, replace Line 4 with: time = [0,1,3,6,12,18,24]

2. Line 5: Input the strain name, same as in ”seg2D_DIC.m” For example, replace Line 5 with: strain = [156].

3. Line 6: Input the replicates you wish to filter. Normally, this is every replicate analyzed in ”seg2D_DIC.m” but if there is some aberration in the image that makes it not appropriate for subsequent analysis, for example xy-drift during the image acquisition, they can be skipped. For example, for a single dataset including only two time points, t=0 and t=1, with three replicates each, but the second replicate of the t=1 dataset cannot be analyzed further, replace Line 6 with: sample = {{[1,2,3],[1,3]}}.

4. Line 7: Input the dates of the datasets to be analyzed, same as in ”seg2D_DIC.m”.

5. Lines 8-11: Input the identities of the red, green, blue, and violet channels in the multichannel fluorescent image. For example, if the first channel is red, the second channel is green, the third channel is blue, and the fourth channel is violet, replace Line 8 with: red_channel = 1; replace Line 9 with: green_channel = 2;, replace Line 10 with: blue_channel = 3; and replace Line 11 with: violet_channel = 4. If the fluorescent image does not include one of the channels, that channel ID should be set to 0. For example, if in the multichannel fluorescent image the first channel is green, the second channel is blue, the third channel is violet, and there is no red channel included, replace Line 8 with: red_channel = 0; replace Line 9 with: green_channel = 1;, replace Line 10 with: blue_channel = 2; and replace Line 11 with: violet_channel = 3.

After these steps have been completed, ”VioletNormalize.m” can be run. Cells that do not meet the 405 intensity cutoff will be filtered out of the dataset, and the saved .mat file will be modified to now include the structured array titled ”Part4_V_Normalized”, which is in the same format as ”part4”.

2.6.4 Sample check

This step is optional but highly recommended, as it serves as a check of the quality of the data. If any of the replicates are anomalous in some way, for example if the background fluorescence is very high in one image, it will stand out in the figures output in this step by the code ”internalSampleCheck_VThresh.m”. The inputs for ”internalSampleCheck_-VThresh.m” are identical to those in in ”VioletNormalize.m”, but located in Lines 17-24. The code outputs a figure like Figure 2.19.

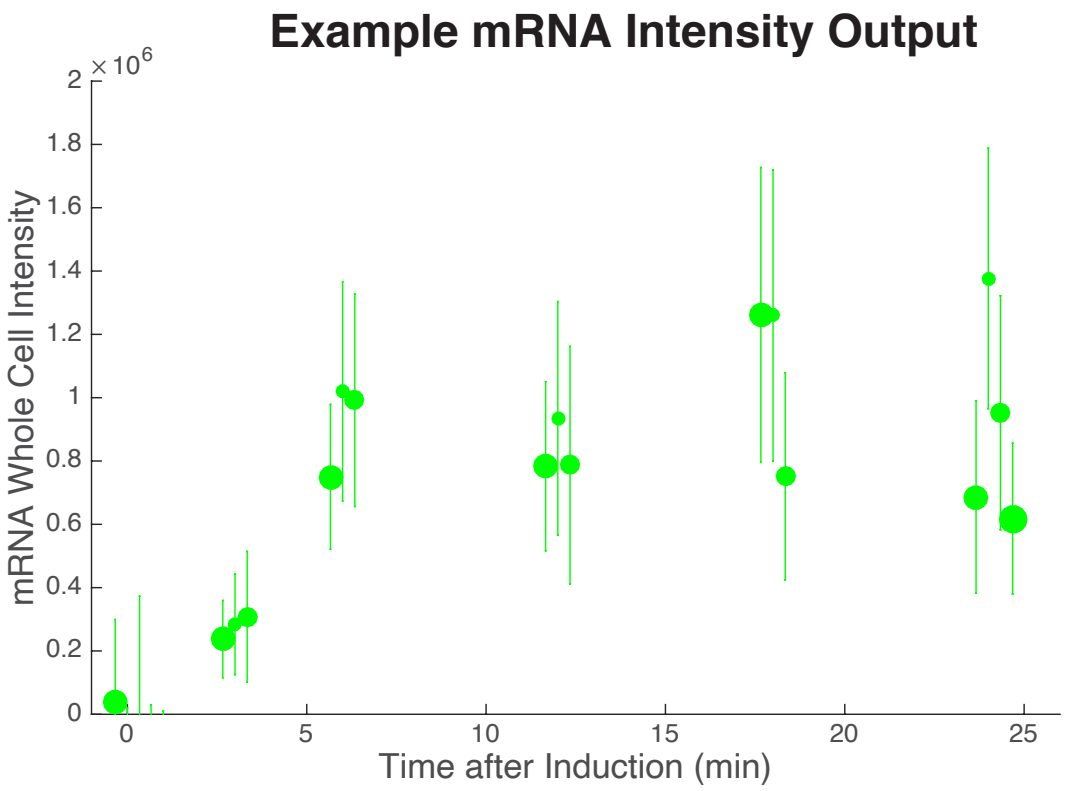


Figure 2.19 Example Output from internalSampleCheck_VThresh.
Each point with error bar represents the average fluorescent intensity and standard deviation of a single channel, in this instance the mRNA channel, of an image. In this example, there are some points at the t=18 and t=24 time point that appear to be anomalous. The user may want to carefully inspect those images.

2.6.5 Timecourse of intensities output CSV

The final step before fitting kinetic parameters is calculating and recording the mean and standard deviation of intensities for the experiment using the code ”makeSingleCell-Figs_VThresh.m”. The inputs for ”makeSingleCellFigs_VThresh.m” are identical to those in ”internalSampleCheck_VThresh.m” and ”VioletNormalize.m”, but located in Lines 18-25. The code outputs a .csv file titled ”VThresh_BackSubIntensity.csv” in the same output folder as defined earlier in the analysis. The output csv appears like Table 2.6, where each entry is in Arbitrary Fluorescent Unites (AU).

Table 2.6 Example Image Timecourse Output.

Time	Mean_Blue	Blue_Sigma	Blue_SE	Mean_Green	Green_Sigma	Green_SE	Mean_Red	Red_Sigma	Red_SE
0	23388.0291	33948.4802	1203.27198	30174.4949	40376.8198	1431.11843	1151599.51	1879311.02	66610.413
1	17162.258	149951.909	5275.28974	14820.0972	28180.2301	991.377037	1301486.68	1930173.29	67903.2595
3	12621.4844	20156.6204	765.686695	49188.4762	57617.4171	2188.70469	888734.425	1701998.49	64653.5764
6	34618.1992	66218.5343	2920.77434	174907.248	127528.66	5625.04803	1085190.52	1917016.75	84555.9839
12	52863.2863	60751.2941	1745.75567	139805.535	96143.8918	2762.80114	231026.965	879624.013	25276.9695
18	87460.673	94507.3212	2830.27099	126775.679	87631.9686	2624.37042	301791.504	1048239.68	31392.3017
24	169019.18	156426.523	5925.07705	181014.856	119098.897	4511.19242	169274.09	623835.524	23629.4554

2.6.6 Running scripts on Midway

All kinetic parameters are calculated with Python scripts run on Midway. An easy way to transfer scripts and run them on Midway is described here. Users will need a Midway user account. These instructions are written for a Mac user.

Editing scripts on your computer:

1. Create a folder on local computer containing all scripts of interest. Write and edit scripts locally using some text editor or Python Interactive Development Environment (IDE). A good choice is the Spyder IDE.

Transferring scripts to Midway:

1. To change directories, use the command ”cd”, then enter. In order to move back one folder, type ”cd ..”. In order to move into a new folder, type ”cd new/folder/directory”. In the terminal, you will start in your Home folder. If at any time you need to check which directory you are currently located in, use the command ”pwd”. To check the contents of the current folder, type ”ls”.

2. From the home directory in the Mac terminal, change to the directory containing scripts of interest. If the scripts of interest are located in the folder ”Users/reyer/python_trans/5320/”, type ”cd python_trans/5320”. You are able to tab complete in the

Terminal, meaning you can type ”cd pyth” for example, then press ”tab” and (assuming there is only one folder that starts with ”pyth”), the rest of the folder will be auto-completed to ”cd python_trans”. This is useful for time purposes and if you can’t remember the exact name of the folder.

3. You should now be located in the directory of interest. Check with ”pwd” if necessary.

4. To transfer the entire folder to some folder in your Midway directory, use the ”scp” command. For example, if I want to copy ”Users/reyer/python_trans/5320/” to the ”project2/jingyifei/deltaSgrS/” directory on Midway2, type:
   ”scp -r . reyer@midway2.rcc.uchicago.edu://project2/jingyifei/deltaSgrS”.

5. You will be prompted to type your Midway password, which will be invisible. Press enter. You will next be asked to authenticate with 2-FA.

6. If you initiated the transfer correctly, you will see the progress of the transfer in the Terminal.

Logging on to Midway:

1. To log on to your Midway account, from any directory in the terminal use the command ”ssh”, for example ”ssh reyer@midway2.rcc.uchicago.edu”, then enter. You will again be asked to enter your password and authenticate with 2-FA

2. If successful, you will arrive in your home directory on Midway 2, for example ”/home/reyer”. Use ”pwd” to confirm

Running scripts on Midway:

1. Change to the directory you just copied to Midway. That should be somewhere in the PI-account directory on project2. NEVER RUN SCRIPTS ON THE LOGIN NODE! Only run scripts in project2.

2. To change to the directory of interest, use ”cd ../../directory/of/interest”. You usually need to move back a couple folders from you home directory first. For example, if I transferred the folder ”Users/reyer/python_trans/5320/” to ”reyer@midway2.rcc.uchicago.edu://project2/jingyifei/deltaSgrS” and I now want to change to that directory, from the Midway home directory type ”cd ../../project2/jingyifei/deltaSgrS/5320/”. Check if you are in the correct directory using ”pwd”.

3. Once in the directory of interest, either move into the subfolder containing your script, or if the script is located in the current folder, it can be run now.

4. The folder should contain your python script and a .sbatch file, which gives the directions to run the script. My .sbatch files are always called ”mcmcFitting.sbatch”, and there is one in each folder containing a python script. You can check and edit the .sbatch script by typing ”emacs mcmc.sbatch”.

5. Run the script by typing the command ”sbatch mcmcFitting.sbatch”. Let the code run to completion.

6. You can check the status of your runs using:
   ”sacct -u user –format=jobid,partition,elapsed,stat”

7. When the runs are complete, the output files ”likelihood.csv”, ”posterior.csv”, and ”walkers.csv” will be written in the same folder as the script is located. You can now transfer this whole folder back to your local computer.

Transferring Folders back to Local Computer:

1. Logout of Midway by typing ”logout”. You will now be in whatever directory you were in when you logged into Midway

2. To transfer the entire folder, containing the output files, from Midway back to your computer, first change directories to wherever you want your data deposited. For

example, if I want to copy all my data to ”Users/reyer/Data/Kinetics”, I will move back to my home directory then type ”cd Data/Kinetics”

3. Once in the directory where you would like your data deposited, copy the folder using ”scp -r reyer@midway2.rcc.uchicago.edu://project2/jingyifei/deltaSgrS/5320 .” replacing the username and file name as necessary. You will again be asked to enter your password and authenticate with 2-FA. If successful, you will see the progress of the transfer in the terminal

2.6.7 Calculating transcription and translation Rates

The scripts to calculate transcription (or initiation) and translation rates will vary slightly based on the number of replicates and the model (*e.g.* one-step vs. two-step transcription). The terms ”transcription” and ”initiation” will be used interchangeably in these instructions, as the procedure for calculating either is exactly the same. The only thing that changes is the Model ODE’s in the python script.

The following instructions are written for an example calculating the ptsG initiation and translation rates using the two-step transcription model in WT *rne* background, when two imaging replicates have been collected. Of course, the same thing will need to be done for *rne*701 background, and for each mRNA. This example uses the script titled ”ptsG_delta_MCMC.py”, which I place in its own folder with an associated .sbatch file, as described above. I write a separate script for each replicate. Assuming all this, the script can be edited according to:

1. Line 53: Input the timepoints, in seconds. If the experiment to be analyzed has timepoints, t = 0,1,3,6,12,18,24 change line 53 to: xData = [0, 60, 180, 360, 720, 1080, 1440].

2. Lines 57-65: Input protein and mRNA fluorescence intensity numbers. You can

copy these from the fluorescence intensity .csv, such as in Table 2.6, into the arrays yP1, yP2, yM1, and yM2. For replicate 2, you can switch the names yP1 and yP2, as well as yM1 and yM2. An example of what a protein data array would look like using the data from Table 2.6 is: yP1 = [23388.02909, 17162.25803, 12621.48443, 34618.19921, 52863.28629, 87460.67299, 169019.1797].

3. Line 89: Input the mRNA degradation rate, which should have been calculated with rifampicin experiment.

These are the only inputs that need to be changed for this script. Of course, if more detailed changes, for example a change to the Model or a different number of MCMC walkers, are required, the script will need to be further modified. For this example, the script can be run as is.

Once the script is edited, transfer the folder to Midway, and run it using the instructions in Section *2.6.6*. When the run is complete, transfer the folder back from Midway to your local computer. Using the posterior and walker output .csv files, we will now calculate the transcription and translation rates using ”am_kx_outputs.py” and ”am_kx_fits.py”. First, using ”am_kx_outputs.py” we will output figures that visualize the walker trajectories and posterior distributions.

1. Line 25: Input the name of the directory containing the output .csv files. For example, change Line 25 to: WalkerFolder = ”/Users/reyer/Data/Kinetics/5320/”

2. Line 27: Input the number of walkers used in the MCMC run.

3. Line 28: Input the number of steps each walker took in the MCMC run.

4. Output figures will be saved automatically in WalkerFolder.

Next, we will calculate the transcription and translation rates with ”am_kx_fits.py”.

1. Line 15: Input the name of the directory containing the output .csv files, same as in ”am_kx_outputs.py”. For example, change Line 25 to: WalkerFolder = ”/Users/reyer/Data/Kinetics/5320/”.

2. Line 27: Input the number of walkers, same as in ”am_kx_outputs.py”.

3. Line 29: Input the number of steps, same as in ”am_kx_outputs.py”.

After these steps have been completed, the script can be run without further modification. Output figures showing the MAP values for the transcription and translation rates will be saved automatically in WalkerFolder. The average MAP values for each parameter, and the associated error, should be recorded.

2.6.8 Calculating remaining kinetic parameter values

The scripts to calculate the remaining kinetic parameter values will vary slightly based on the number of replicates and the model (*e.g.* one-step vs. two-step transcription, co-transcriptional vs post-transcriptional).

The following instructions are written for an example calculating the ptsG kinetic parameter values using the two-step transcription, co-transcriptional regulation model, when two imaging replicates have been collected in both WT *rne* and *rne*701 backgrounds (4 datasets total). This example uses the script titled ”ptsG_plus_delta_MCMC.py”, which I place in its own folder with an associated .sbatch file, as described above. I write a separate script for each replicate. Assuming all this, the script can be edited according to:

1. Line 50: Input the timepoints, in seconds. If the experiment to be analyzed has timepoints, t = 0,1,3,6,12,18,24 change line 53 to: xData = [0, 60, 180, 360, 720, 1080, 1440].

2. Line 55: Input the number of replicates in the dataset. In this example, there are two

replicates, so Line 55 is: numReplicates = 2.

3. Lines 61-73: Input protein, mRNA, and sRNA fluorescence intensity numbers from both replicates. You can copy these values in the yData matrix. An example of the yData matrix is:

yData[0, 0, :] = [134917.1029, 184452.1771, 180579.3257,216245.1063, 432250.9255, 839879.6454, 1170873.494, 1903248.354]
yData[0, 1, :] = [22443.71086, 339380.1687, 774276.8132, 989754.3786, 1178364.294, 1168843.171, 1039977.762, 1152377.24]
yData[0, 2, :] = [1299487.304, 2158748.53, 1461797.642, 1479650.818, 1213052.738, 1132760.79, 930608.6544, 743058.2862]
yData[0, 3, :] = [257171.8156, 242054.4163, 248115.1836, 246027.831, 383643.0025, 802367.1203, 1131430.97, 1382649.746]
yData[0, 4, :] = [16225.77785, 7646.443305, 194166.3494, 714018.5826, 1418654.816, 1533096.423, 1617435.188, 2046367.339]
yData[0, 5, :] = [2188990.237, 2050903.278, 2390768.473, 1788277.574, 1296136.946, 1018016.032, 973140.0654, 973140.0654]

yData[1, 0, :] = [144202.1426, 333919.142, 149161.7576, 304584.7737, 738063.1426, 808832.9779, 1392651.226, 1681171.266]
yData[1, 1, :] = [33318.96697, 262752.4716, 627298.9573, 851325.6287, 1159651.072, 1151829.933, 1016580.917, 1118273.902]
yData[1, 2, :] = [1788327.436, 1787030.495, 1406293.534, 1188055.269, 1088649.355, 825880.4722, 785990.9093, 744898.0836]
yData[1, 3, :] = [64338.27187, 36900.77477, 47182.78534, 52802.52599, 298986.6166, 731084.9614, 1055270.669, 1940635.146]
yData[1, 4, :] = [28424.46802, 39375.74472, 237406.821, 1027314.934, 2003857.809, 2042997.469, 1758470.575, 2012571.706]
yData[1, 5, :] = [2547665.889, 2079076.002, 2324980.48, 2178616.805, 1944435.717, 956261.5957, 1099876.149, 1250870.797]

where the format of the data is as follows:

yData[0, 0, :] = WT Protein, Rep 1

yData[0, 1, :] = WT mRNA, Rep 1

yData[0, 2, :] = WT sRNA, Rep 1

yData[0, 3, :] = *rne*701 Protein, Rep 1

yData[0, 4, :] = *rne*701 mRNA, Rep 1

yData[0, 5, :] = *rne*701 sRNA, Rep 1

yData[1, 0, :] = WT Protein, Rep 2

yData[1, 1, :] = WT mRNA, Rep 2

yData[1, 2, :] = WT sRNA, Rep 2

yData[1, 3, :] = *rne*701 Protein, Rep 2

yData[1, 4, :] = *rne*701 mRNA, Rep 2

yData[1, 5, :] = *rne*701 sRNA, Rep 2

The MAP values of the kinetics will be calculated for the combination of datasets (WT *rne* and *rne*701) with the 0's in the first column of yData. The 0's and 1's in the first columns can be switched appropriately to calculate the MAP values for the four dataset combinations (see Section *2.4.8*)

4. Line 95-97: Input the WT initiation rate and standard deviation, calculated in Section *2.6.7*.

5. Line 99-101: Input the *rne*701 initiation rate and standard deviation, calculated in Section *2.6.7*.

6. Line 107-109: Input the mRNA degradation rates and standard deviation

7. Lines 116-121: Input the WT *rne* sRNA transcription and degradation rates, and standard deviation

8. Lines 123-128: Input the *rne*701 sRNA transcription and degradation rates, and

standard deviation

9. Lines 130-135: Input the WT *rne* and *rne*701 translation rates and standard deviations, calculated in Section *2.6.7*.

These are the only inputs that need to be changed for this script. Of course, if more detailed changes, for example a change to the Model or a different number of MCMC walkers, are required, the script will need to be further modified. For this example, the script can be run as is.

Once the script is edited, transfer the folder to Midway, and run it using the instructions in Section *2.6.6*. When the run is complete, transfer the folder back from Midway to your local computer. Using the posterior and walker output .csv files, we will now calculate the remaining kinetic parameter values using ”coTran_outputs.py” and ”coTran_fits.py”. There are different versions of ”output.py” and ”fits.py” depending on the Model being simulated. Again, here we are assuming the two-step transcription, co-transcriptional regulation model. First, using ”coTran_outputs.py” we will output figures that visualize the walker trajectories and posterior distributions.

1. Line 40: Input the name of the directory containing the output .csv files. For example, change Line 40 to: WalkerFolder = ”/Users/reyer/Data/Kinetics/5320/”

2. Line 42: Input the number of walkers used in the MCMC run.

3. Line 43: Input the number of steps each walker took in the MCMC run.

4. Output figures will be saved automatically in WalkerFolder.

Next, we will calculate the kinetic parameter values with ”coTran_fits.py”.

1. Line 45: Input the name of the directory containing the output .csv files, same as in ”coTran_outputs.py”. For example, change Line 45 to:

WalkerFolder = ”/Users/reyer/Data/Kinetics/5320/”.

2. Line 74: Input the number of walkers, same as in ”coTran_outputs.py”.

3. Line 64: Input the number of steps, same as in ”coTran_outputs.py”.

After these steps have been completed, the script can be run without further modification. Output figures showing the MAP values for the kinetic parameter values will be saved automatically in WalkerFolder. The average MAP values for each parameter, and the associated error, should be recorded.

2.6.9 BIC calculation

Calculate the BIC associated with the best fit set of parameters using ”co_BIC.py” (or ”post_BIC.py” for post-transcriptional regulation model). The only input is the WalkerFolder, same as in the previous scripts, in Line 22.

2.6.10 Checking quality of fits

Check the quality of the fits by inputting the MAP kinetic values and their associated errors into the MatLab script ”ptsG_initFit.m”. There is a separate script for each mRNA (*e.g.* ”manX_initFit.m”, ”purR_initFit.m”, etc.), and for each model being simulated. The output figures will look like those in Figure 2.4.

CHAPTER 3

INTERACTIONS BETWEEN THE SRNA ACCESSORY PROTEIN, HFQ, AND THE RIBOSOME

3.1 Introduction

In the broad view of RNA-mediated gene regulation, RNA-binding proteins are perhaps of equal importance to the regulatory RNAs themselves. Chief among the RBPs in bacteria is Hfq, which has been found in more than half of known bacterial species[52,53]. Hfq works in tandem with small RNAs as a chaperone protein, helping to regulate gene expression at the co-transcriptional, post-transcriptional, and translational levels[144–147].

Hfq aids the sRNA-mediated regulatory process mainly via three mechanisms. First, Hfq protects sRNAs from degradation by RNases[45,54,148]. Second, Hfq facilitates the sRNA target search process and sRNA-mRNA base pairing[54,149]. Third, Hfq promotes the rapid turnover of mRNA and sRNA by RNase E[45,58,127,150]. Hfq can efficiently carry out these roles because of its ability to directly bind both sRNAs and mRNAs through different surfaces[5,50,127,151], and because of its interactions with various other proteins involved in the regulation process, principally RNase E[46,58,150,152–154].

From a simplified perspective, an effective bacterial stress response can come down to whether an mRNA is translated into a gene product by the ribosome, or regulated by the sRNA-Hfq complex. It has already been established that sRNA-Hfq interacts directly with both the mRNA and RNase E, and indirect association between the ribosome and Hfq have been noted before[155–158], so it perhaps follows naturally to ask if sRNA-Hfq interacts directly with the other major player, the ribosome, and how such interactions may affect the regulatory functions of Hfq. In this chapter, inspired in part by a revelation from Chapter 2, we propose to explore the possibility of direct interactions beetween Hfq and the ribosome, and what effect those interactions may have on the sRNA-Hfq target

search process and subsequent regulation of target mRNAs.

3.2 Potential relationship between translation and association rates

We can imagine potential interactions between the ribosome and sRNA-Hfq to be either competitive or cooperative. In the former scheme, Hfq and the ribosome ”compete” for access to the mRNA. Since many target mRNAs have sRNA binding sites close to or directly overlapping the ribosome binding site (RBS)[26,55,73,159,160], one might infer that ribosome binding and Hfq binding are mutually exclusive. Therefore, rapid and regular ribosome binding and, by extension, efficient translation would decrease the effectiveness of sRNA-mediated regulation by blocking access of the sRNA-Hfq to its mRNA binding site. Such a model is analogous to the competition between the ribosome and RNase E that dictates mRNA degradation, wherein binding between the two molecules is in fact mutually exclusive and efficient translation subsequently protects mRNA from degradation[161]. If this model were real, we would expect a negative correlation between the mRNA translation rate and the binding rate of sRNA-Hfq.

In the latter scheme, ribosome binding facilitates sRNA-Hfq in its pursuit to regulate the target mRNA. Such cooperation could perhaps be the byproduct of efficient translation increasing mRNA stability[162], an increased ability of sRNA to bind stable targets[163], or some other unknown mechanism. While it is certainly true that the ribosome and Hfq cannot be bound to the same site on an mRNA at the same time, perhaps binding of the ribosome could attract Hfq to bind in quick succession. If this were the case, we would instead expect a positive correlation between translation rates and sRNA-Hfq association rates.

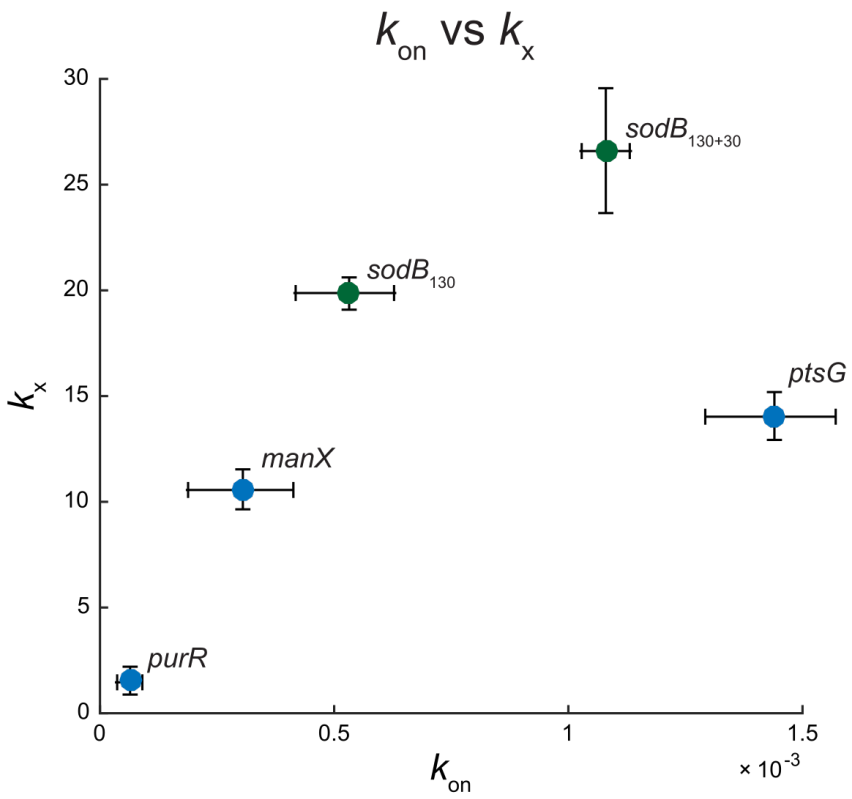


Figure 3.1 Translation rate vs on-rate of sRNA binding.
k_x vs. k_{on} for all mRNA targets. Points in blue are mRNA targets of SgrS. Points in green are targets of RyhB. Error bars represent standard deviation of calculated MAP values (see Table 2.1 and 2.2).

While the kinetic competition model of ribosome-Hfq interactions is perhaps more intuitive, mounting evidence suggests that ribosome-Hfq interactions do exist and more closely resemble the cooperation model. A quantitative study by the Levine group set out to probe an adjacent question, namely whether ribosome and *sRNA* interactions are competitive or cooperative[63]. Using a library of reporter genes with RBSs of differing predicted translation efficiencies and mathematical models representing both the competition and cooperation frameworks, they showed that translation was positively correlated with sRNA-mediated regulation. Though at the time of the study the mechanism behind the recruitment was unknown, the authors posited that it might be due to ”direct interaction between the ribosome and either the sRNA or Hfq,” motivating our search for such interactions[63].

Furthermore, a result from our own work (described in Chapter 2) also point towards the cooperative model of ribosome-sRNA interactions. In the our study on the kinetics of sRNA-mediated regulation, we were able to directly measure the translation rates of multiple target mRNAs, each with different RBSs and translation efficiencies, within single sRNA regulons. We inferred the on-rates of sRNA binding from modelling. In line with the results from (Lavi-Itzkovitz, 2014), we saw a positive correlation between translation and association rates (Figure 3.1). Furthermore, since we were deriving the association rates themselves rather than simply observing the amplitude of the sRNA regulation effect, our observation of a connection between translation and sRNA binding was an even more direct and concrete comment on the cooperation model. This result combined with the hypothesis posed by the Levine group inspired us to investigate further the mechanisms behind the correlation, and specifically the potential of a direct ribosome-Hfq interaction.

3.3 Hfq-ribosome interactions revealed by imaging and staining

We sought to utilize both traditional molecular biology techniques and imaging to explore potential interactions between Hfq and the ribosome.

Using our platform designed to detect the binding and interaction states of Hfq using diffusivity as a proxy[127], we discovered a connection between translation activity and Hfq spatial dynamics. Under normal conditions (exponential growth, no treatment, NT), Hfq is equally distributed and diffuses freely throughout the cell (Figure 3.2A). Imaging results show that Hfq can be trapped in low-diffusivity, high density clusters under two conditions under which ribosomes become locally condensed. In the first condition, *E. coli* cells are exposed to cloramphenicol (CM), an antibiotic that blocks translation by inhibiting elongation and causes ribosome stalling[164]. The same phenomenon was not observed when the antibiotic kasugamycin (Ksg), which inhibits translation by preventing ribosome loading, was added[165] (Figure 3.2B).

In the second condition, we overexpressed a translation-efficient (*i.e.* has a strong RBS) fusion mRNA, *sodB-lacZ*, from a plasmid. Again, in the presence of high translation activity, we observed low-diffusivity Hfq clusters (Figure 3.2B, C). Interestingly, these clusters remained present even when WT Hfq was replaced with a mutant Hfq (Hfq_{65}) with its C-terminal, unstructured region truncated, suggesting that the clustering is not driven by intrinsically disordered regions (IDRs), a common requisite for *in vivo* clustering. Furthermore, a second Hfq mutant (Y25D), which is a distal face mutant deficient in mRNA binding, still formed clusters and co-localized with the ribosome, though its diffusivity decreased. In total, these result suggest that translation activity, but specifically the clustering of ribosomes, attracts and condenses Hfq, and that this trapping effect is enhanced by mRNA binding, but neither dependent on mRNA binding nor IDRs.

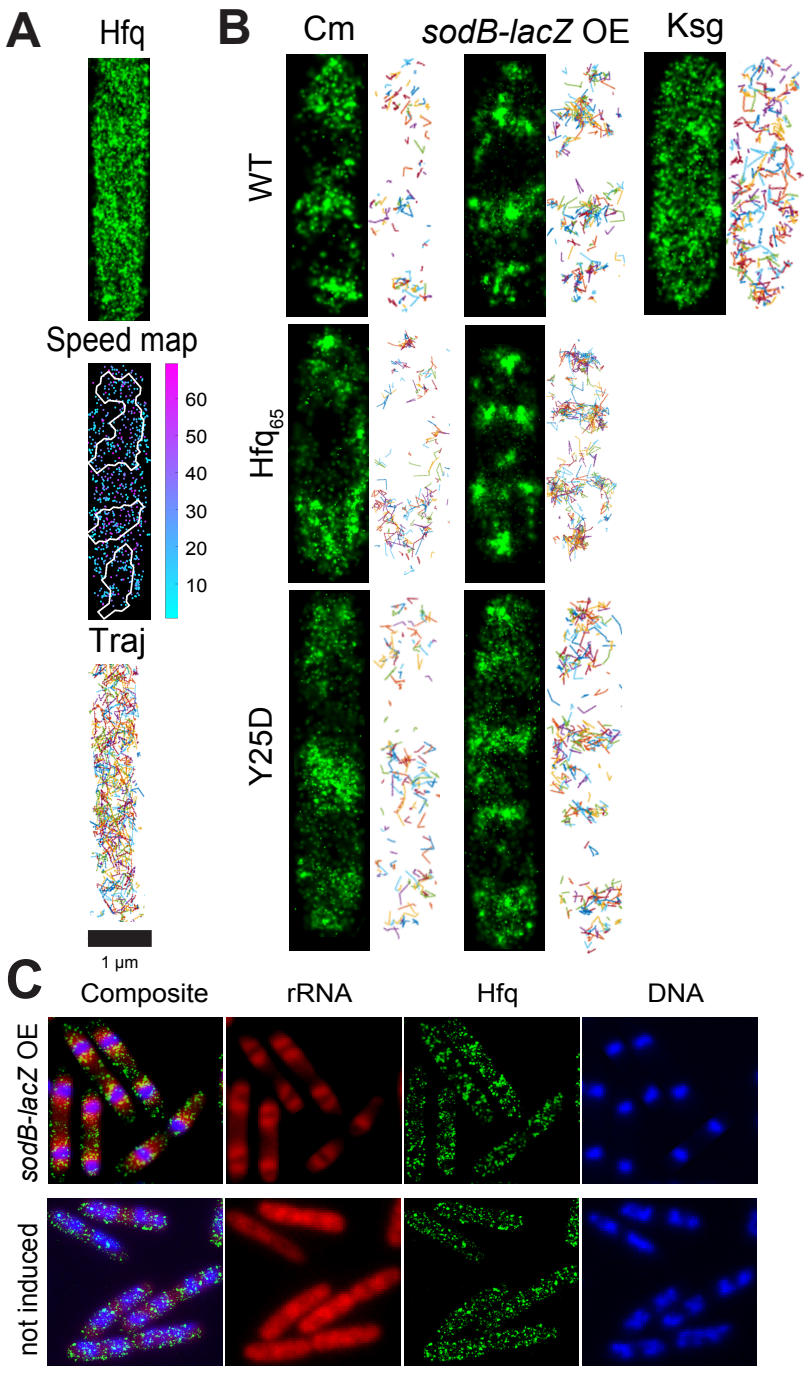


Figure 3.2 Hfq is Trapped by Condensed Ribosomes.
(A) A representative example of WT Hfq-mMaple3 in WT *rne* background in a single cell during exponential growth under no treatment (NT) condition. 2D reconstructed image of Hfq-mMaple3 is shown in the black background. One-step displacement (osd) speed map (unit: μm/s) is shown as a scatter plot where different colors represent different speeds at each position, and the white curves represent the nucleoid regions detected by Hoechst staining. Different diffusion trajectories from tracking algorithm are shown in different colors ('Traj').

Figure 3.2 ***(continued)*** **(B)** WT Hfq is trapped in low-diffusivity clusters under conditions where the ribosome becomes condensed, including in the presence of CM and by overexpressing *sodB-lacZ*. Trapping is not observed in the presence of Ksg. Notably, the Hfq_{65} and Y25D mutations do not affect the trapping behavior. **(C)** Representative cell images of fixed Hfq-mMaple3 co-stained with rRNA FISH and DAPI to illustrate co-compartmentalization of Hfq and the ribosome.

Next, we utilized Western blotting to detect if there existed an interaction between Hfq and the ribosome. The Western blot was conducted with purified ribosome, Hemagglutinin (HA)-tagged Hfq, and HA antibody to detect the tagged Hfq. Ribosome-containing fractions were collected from a sucrose gradient and analyzed for the presence of Hfq. The size of Hfq is 52 kDa; therefore, the presence of Hfq should be communicated by a band around the 50 kDa marker (Figure 3.3). When left untreated, we see faint bands at the 50 kDa marker. However, in the presence of CM, we see much darker bands at the 50 kDa marker, indicating an interaction between Hfq and the ribosome is enhanced under conditions where the ribosome becomes condensed, consistent with our imaging results.

Seeing as the apparent interaction between Hfq and the ribosome is dependent neither on the Hfq IDR or mRNA binding, we looked elsewhere for potential mechanisms. It has been observed in a study by Schavemaker *et al.* that positively charged proteins (specifically, a charged GFP) diffuse more slowly than negatively charged proteins due to electrostatic interactions with the ribosome[166]. *E. coli* Hfq, notably, is highly positively charged. We therefore hypothesized that the trapping effect ribosomes have on Hfq may be due to electrostatic interactions. However, when the Western blot was performed with the reduced-charge mutant of Hfq, R17D, there is no subsequent decrease in the Hfq-ribosome interaction, which potentially complicates this hypothesis. Therefore, the search for a mechanism continues.

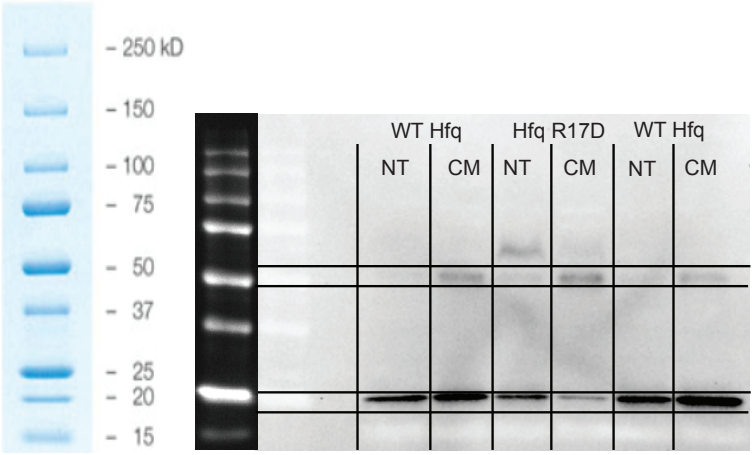


Figure 3.3 Hfq association with Ribosome Revealed by Western Blot.
Western blot was conducted with purified ribosome and HA antibody against HA-tagged Hfq. Bands at 25 kDa likely represent Hfq dimers.

Before we can conclusively determine the mechanism, though, we first ask: if there does exist an interaction between the ribosome and Hfq, what effect might it have on the sRNA target search process? We attempt to address this question using simulations.

3.4 Simulations on sRNA target search as a function of Hfq recruitment

Disregarding for a moment the mechanism behind the Hfq-ribosome interaction, we now ask what are the functional implications of it and specifically, what effect it may have on sRNA target search process. To test the feasibility of Hfq-ribosome interaction-facilitated target search, we first turned to a spatial stochastic simulation. Using the Lattice Microbes platform developed by the Luthey-Schulten lab[167], we attempted to computationally recapitulate the observed correlation between translation and sRNA association rates, assuming that the ribosome has a trapping effect on Hfq.

Lattice Microbes uses chemical and reaction-diffusion master equations to simulate the cell containing various compartments (e.g., cytoplasm, membrane, nucleoid) com-

posed by many 3D sub-volumes (or lattices), which particles (representing biomolecules) can diffuse into and react if co-occupying with other particles. The likelihoods of diffusion and reaction steps are determined by a transition matrix, which is defined by diffusion coefficients and kinetic rates as inputs. We are able to set some of these inputs from literature[38,168–174] and results from both our Hfq tracking experiments and sRNA regulation kinetics study (Chapter 2).

The model we enacted simulates a regulatory sRNA searching for a target mRNA *in vivo*, assuming that the target mRNA is greatly outnumbered by non-target mRNAs, that Hfq is limited relative to all cellular sRNAs and mRNAs[156,175], and that ribosomes can trap Hfq in a translation-dependent manner. We do not consider the downstream regulatory effects of the sRNA-mediated regulation, only the target search and binding processes. The model includes specific kinetic modules for:

(A) Translation of target and non-target mRNAs;

(B) Transient interactions between the sRNA-Hfq complex (referred to as "SHC") and the ribosome/polysomes on both target and non-target mRNAs in a translation rate-dependent manner (*i.e.* the higher translation rate, the larger polysome density on the mRNA, and the stronger interactions in kinetic module (B));

(C) Association of SHC with target and non-target mRNAs when they are in the same lattice, with the former having a higher reaction rate than the latter.

A detailed description of the model parameters can be found in Section 3.7.3.

This model simulates an intrinsic linkage between basal translation activity and the effective target search rate, in order to see if and how Hfq-ribosome interactions may play a role. Module (B) connects the basal translation rate of the mRNA to the dwell time of SHCs in close proximity. Module (C) links the longer dwell time to an increased chance of SHC binding to the mRNAs. These scenarios are meant to demonstrate that if in fact

translating ribosomes trap Hfq, the effective search time of the SHC can be reduced. Here we define the effective search time (t) as the time required for 50% of the target mRNAs bound by a sRNA. Our preliminary results from the simulation demonstrate that a higher basal translation rate (k_X) shortened the search time (t) by orders of magnitude (Figure 3.4B). Our simulation demonstrates a potential function of Hfq-ribosome interactions in facilitating the sRNA target search.

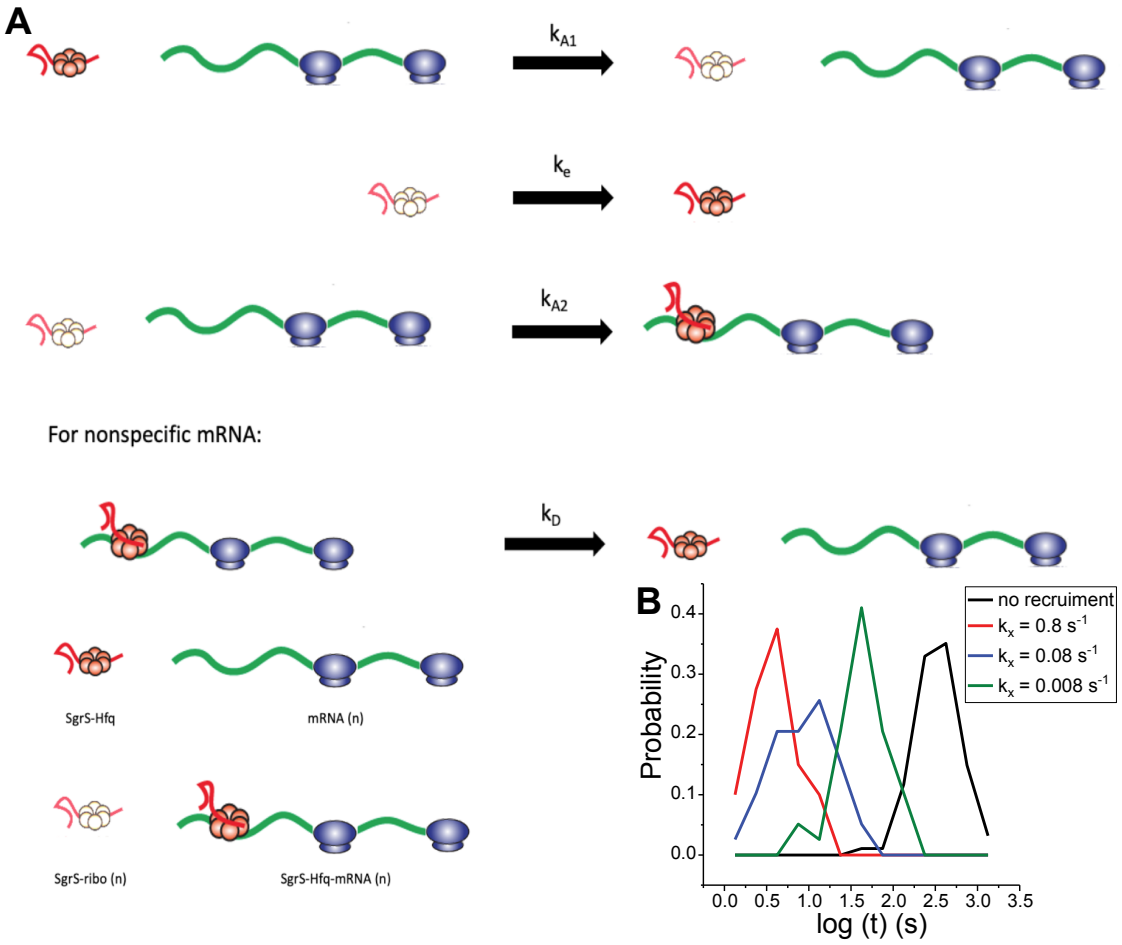


Figure 3.4 Ribosome Recruitment of Hfq Simulation.
(A) A detailed description of model parameters can be found in Section 3.7.3. Briefly, we model the ribosome trapping of SgrS-Hfq, and the subsequent binding of SgrS-Hfq to mRNA as a two-step process. The first step locally traps the SgrS-Hfq, and the second constitutes the actual binding step. The faint SgrS-Hfq cartoon represents the locally trapped SgrS-Hfq complex which we consider a separate species called SgrS-ribo. We consider dissociation only of non-specific mRNAs. **(B)** t is the mean target search time, which decreases as a function of translation initiation rate in the presence of ribosome-driven recruitment.

3.5 Proposed single molecule experiments to test effect of translation on sRNA-Hfq binding to mRNA

In order to experimentally test the effect of translation on the sRNA target search process, we propose a single-molecule microscopy experiment. Using Cy3 and Cy5 fluorescent labels to tag RyhB and the *sodB-GFP*, we will perform an *in vitro* translation experiment. *sodB-GFP* will be anchored to a glass slide, and we can test the speed of the target search process as a function of translation rate using FRET experiments. The exact protocol for this experiment is currently under development.

We intend to introduce two perturbations to the experiment. First, we will change the salt or polyamine conditions in the imaging solution to disrupt the Hfq-ribosome interactions. We are currently performing electrophoretic mobility shift assay (EMSA) and *in vitro* translation assays in order to make sure that the perturbed imaging conditions do not affect the Hfq-RNA interaction or the basal translation rate, respectively. This assay can help determine if the Hfq-ribosome interaction is in fact driven by electrostatics.

Second, we will modify the anchored mRNA to contain a stop codon located only a short distance downstream from the start codon. The ribosome should unload from the mRNA soon after initiating translation, thus preventing polysome buildup. In this way, we can reduce the ribosome density without affecting the translation initiation rate. If Hfq-ribosome interactions do in fact facilitate the sRNA target search process, as the simulations in Section 3.4 predict, this assay can help distinguish if that is due to the ribosome density itself, or if it is merely a byproduct of the ribosome removing the secondary structure of the mRNA, for instance.

3.6 Discussion

In this chapter, we discussed our findings regarding the trapping effect that condensed ribosomes have on the accessory protein Hfq. We believe that this effect is driven specifically by translation and polysome clustering. Though the exact mechanism of the trapping effect is still unknown to us, we show via simulation that such a trapping effect can aid the sRNA-Hfq target search process. In the immediate future, we intend to experimentally test if translation does in fact reduce the sRNA target search time.

3.7 Appendix: Running Lattice Microbes simulation

The following is an instruction manual for running the Lattice Microbes simulation of ribosome recruitment described in Section 3.4. The probability distributions of ribosome density (probability of *n* number of ribosomes per mRNA) were calculated analytically for three initiation rates (Table 3.7.4). Those distributions take the form of discrete cumulative distribution functions (CDF), representing the probability of having *n* ribosomes or fewer bound to an mRNA at a given initiation rate. Those CDFs are written as arrays in the simulation scripts.

All simulations are written as Python scripts and are run on Midway. Instructions for how to edit scripts locally and transfer and run them on Midway are identical to those written about in Section 2.6.

3.7.1 Running the ribosome recruitment simulation

We simulated the search times, defined as the time until half of the target mRNAs are bound by SgrS-Hfq complex, in 20 cells. To run this simulation using ”riboSim.py” and its associated .sbatch file, follow:

1. Copy the folder containing ”riboSim.py” and subsequent processing codes to Midway

2. Line 13: Set the initiation rate (either 0.833, 0.0833, or 0.00833 molecules^{-1}s^{-1})

3. Line 14: Set the recruitment option. If simulating a recruitment model, we set α= 0.95. For the no recruitment model, set α= 0.0

4. Line 494: Name the output file to be saved when the simulation is done, for example ”riboOutput.lm”.

5. Initiate the simulation run with ”sbatch lmBatch.sbatch”.

In order to run the simulations as described, these are the only inputs that need to be changed. The output file, which records the trajectories and reactions of every particle within the lattice at each time step, is saved in the same folder as a .lm file, which can be interpreted with built-in Lattice Microbes functions.

3.7.2 Processing ribosome recruitment simulations

After running ”riboSim.py”, process the data using ”Processing.py”. Unlike the analysis in Chapter 2, the analysis here should be performed on Midway before transferring files back to local computer. The reason for this is that the Lattice Microbes environment has been (or should have been) previously built on Midway. Therefore, some of the built-in functions associated with the environment are not accessible in a standard Python environment.

1. To initiate a Python environment while on Midway, simply type ”python” and press enter.

2. Open ”Processing.py” by typing ”emacs Processing.py”.

3. Line 5: Input the output file from ”riboSim.py”. For example, change Line 5 to: filename = riboOutput.lm

4. Run ”Processing.py” by typing ”python Processing.py” and pressing enter

The processing code will output a .png figure displaying the traces of sRNAs, target mRNA, and non-target mRNAs. The sRNA .png’s will display the ”SHp” species, which is the complex formed when sRNA-Hfq binds to target mRNA. These traces can be used to visualize the search time. In addition to a figure for each trace, a figure called ”distribution.png”, which displays the distribution, mean, and median of search times across all cells, is saved automatically. These figures can be used for further analysis.

3.7.3 Description of parameters used in ribosome recruitment simulations

Table 3.1 Basal Kinetic Parameters used in Ribosome Recruitment Simulations.

Parameter	Value
Initial Target mRNA Transcript Molecular Count	10
Initial Nonspecific mRNA Transcript Molecular Count	2000
Initial Free Ribosome Molecular Count	10000
Initial SgrS-Hfq Molecular Count	100
β	5×10^{9} s^{-1}
k_{A1}	1 $molecules^{-1}s^{-1}$
k_{A2}	1 $molecules^{-1}s^{-1}$
k_{D}	100 s^{-1}
D_{mRNA}	1×10^{-12} m^2/s
$D_{SgrS\text{-}Hfq}$	0.124×10^{-12} m^2/s
$D_{ribosome}$	0 m^2/s

In this model, ribosome recruitment of the SgrS-Hfq complex is modeled as a two-step process. The first step is an electrostatic recruitment at the rate k_{A1}, which in the language

of the simulation turns the molecular species ”SgrS-Hfq” into the molecular species ”SgrS-Ribo”. The second is the actual binding between the electrostatically recruited SgrS-Ribo species and the target mRNA at rate k_{A2}, which results in the new molecular species ”SgrS-Hfq-mRNA”. In our simulations, both of these steps in the recruitment and binding process are diffusion-limited. The rate of this two-step process is separate from the search process required for SgrS-Hfq to diffuse into the same volume as the target mRNA so that the recruitment step can occur.

Binding of SgrS-Hfq to non-target mRNAs happens via the same two-step process, but the subsequent ”SgrS-Hfq-non-specific-mRNA” molecular species can also dissociate with the rate k_D. As we were only interested in the time required for target search and binding to target mRNAs, we did not include a dissociation step between SgrS-Hfq and target mRNAs.

Instead, the electrostatically trapped SgrS-Ribo species can escape, thus returning to the freely diffusing SgrS-Hfq species, via the rate k_e. k_e is based on the Arrhenius Equation, and decreases exponentially with the number of ribosomes bound on the mRNA associated with the SgrS-Ribo:

$$k_e = \beta e^{-\alpha n}$$

where β is a constant used for unit conversion, α quantifies the strength of recruitment (α= 0 for no recruitment, α> 0 for recruitment), and n is the number of ribosomes bound to the mRNA. Thus, the rate at which SgrS-Hfq escapes the electrostatic trap is proportional to the number of bound ribosomes.

The diffusion rates for unbound mRNA (D_{mRNA}) and the SgrS-Hfq complex ($D_{SgrS\text{-}Hfq}$) come from literature [176] and our own tracking data. The diffusion rate of any species bound to the ribosome ($D_{ribosome}$) is set to 0.

The mean search times as a function of initiation rate, in both ribosome recruitment simulations and non-recruitment simuations, are shown in Table 3.2. Each recruitment simulation was replicated over 20 cells. Each non-recruitment simulation was replicated over at least 12 cells, rather than 20, because the simulations took much longer in the absence of recruitment.

Table 3.2 Search Times as a function of Initiation Rate.

Conditions	**Mean Search times (Seconds)**
k_{init} = 0.833 molecules^{-1}s^{-1} , α = 0.95 (Recruitment)	3.62
k_{init} = 0.0833 molecules^{-1}s^{-1}, α = 0.95 (Recruitment)	14.51
k_{init} = 0.00833 molecules^{-1}s^{-1}, α = 0.95 (Recruitment)	45.78
k_{init} = 0.833 molecules^{-1}s^{-1}, α = 0 (No Recruitment)	366.8
k_{init} = 0.0833 molecules^{-1}s^{-1}, α = 0 (No Recruitment)	333.9
k_{init} = 0.00833 molecules^{-1}s^{-1}, α = 0 (No Recruitment)	426.9

In the presence of ribosome recruitment, the SgrS-Hfq search time decreases as the initiation rate increases. In the absence of recruitment, there is no relationship between search time and initiation rate. These simulations serve to show how ribosome-driven trapping of the SgrS-Hfq complex could aid the sRNA target search process.

CHAPTER 4

IMAGE ANALYSIS TOOLS TO STUDY RNA-MEDIATED REGULATION

4.1 Introduction

The forte of the Fei lab is our advanced fluorescence and super-resolution imaging approaches to studying RNA-mediated regulation. The contexts, environments, and systems in which we study RNA-mediated regulation are varied, but at the core, all of the studies share these common goals: to detect RNAs, find where they are, see how they move, and determine how those things contribute to the tasks they carry out. The detection we leave to the microscopes (and those who operate them). As for the where and how they move, we rely on computer programs to extract that information from the microscope-acquired images and movies. But with novel detection techniques comes a void in such programs to analyze them. My efforts to fill this void, for the good of the lab and the community at large, are the subject of this chapter.

4.2 An automated image analysis method for segmenting fluorescent bacteria in three dimensions

4.2.1 Introduction

Single-cell fluorescence microscopy has become a powerful method for studying the stochasticity of cellular activities and heterogeneity within a population[177,178]. To achieve single-cell resolution, an efficient and accurate segmentation method is a critical tool in the data analysis. Bacteria serve as model systems for investigating the fundamental mechanisms of many biological processes. In many experiments, bacteria are immobilized in two-dimensional (2D) surfaces or trapped and aligned in microfluidic devices;[178] therefore, many automatic segmentation methods are available to identify and track bacterial

cells lying in two dimensions[179–181]. However, many interesting behaviors of microbes may not be observed in a 2D setting. Bacterial pathogens invading host cells[182] or bacteria undergoing a 2D to 3D transition during biofilm formation[183], for instance, are often clustered with different 3D orientations. In such examples, changes in the phenotypes and cellular activities at the single-cell level may be not only time-dependent but also 3D space-dependent. Therefore, a 3D segmentation method designed for these specific conditions is necessary. Currently, a fully automatic 3D segmentation method for bacterial cells with high cell identification and segmentation accuracy is not available. We considered several well-cited programs used for 3D segmentation, including ImageJ[184], Imaris (Bitplane, St. Paul, MN), Icy Spot Detector[185], and Cell Profiler[186], and found two general shortcomings. (1) Most of the commercial software can perform automatic 2D segmentation but not automatic 3D reconstruction to generate 3D segmentation, and (2) the performance on segmenting clustered cells is not satisfying (Figure 4.7).

Here we report a new analysis method, Seg-3D, for the segmentation of bacterial cells in three dimensions that can also be flexibly applied to 2D images. Seg-3D is based on local thresholding, 2D shape and concavity analysis, concavity-based cluster splitting in two dimensions, and morphology-based 3D reconstruction (Figure 4.1). First, we use a two-step local thresholding technique to minimize background noise and the additive background characteristic of bacterial clustering. Second, the likelihood of identified 2D objects being single cells is calculated using 2D-shape and concavity analysis. Third, we split multicell clusters at concave points along the object borders, which denote cell boundary intersections[187–189]. The single-cell identification step and concavity-based splitting iterate until all objects can be either identified as single cells or discarded. Finally, 2D objects are combined with their most likely partners in adjacent slices to form 3D cells. The parameters used to reconstruct 3D cells can be determined using low-density, single-cell images, which improves the flexibility and accuracy of automatic 3D segmentation.

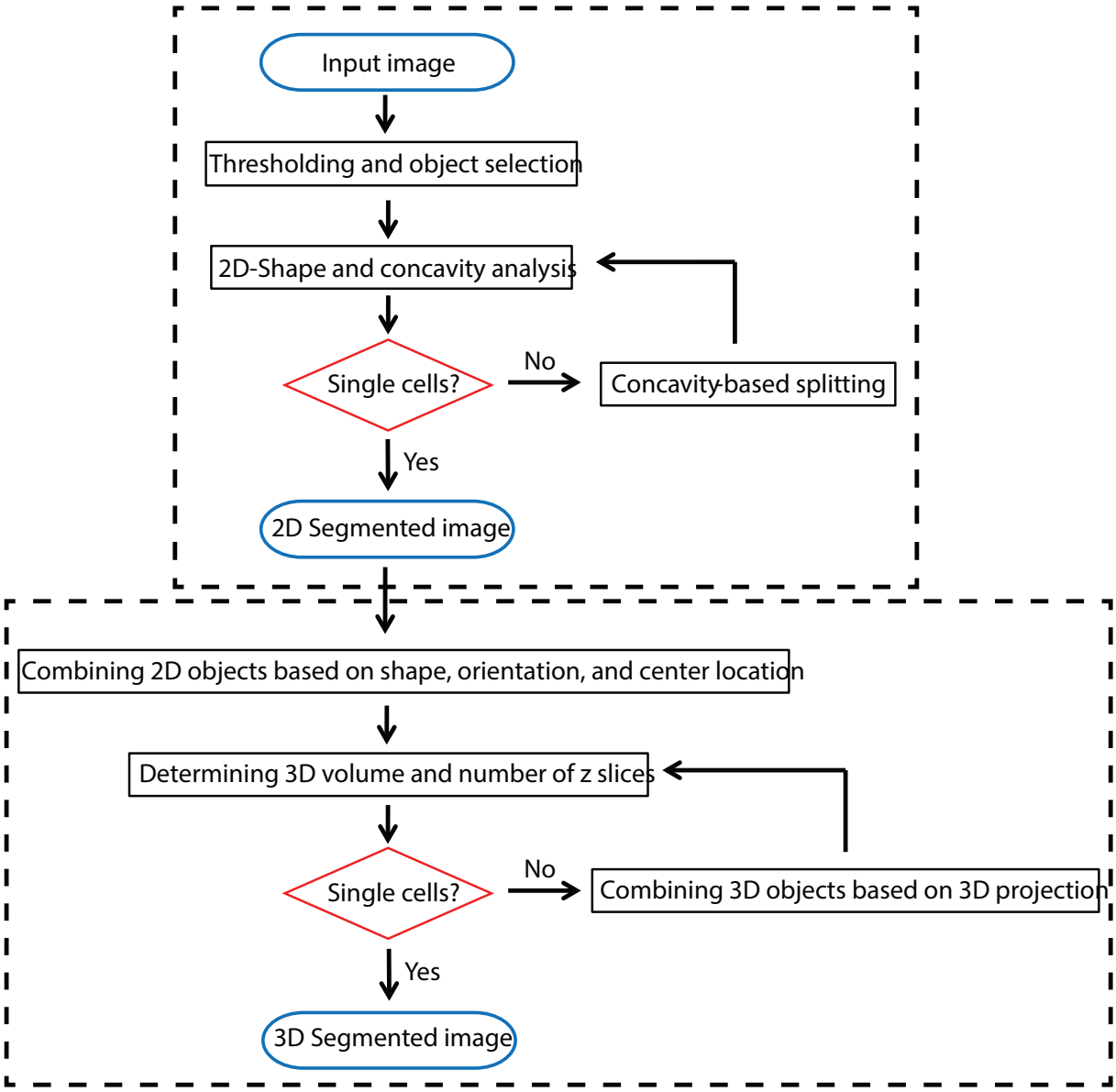


Figure 4.1 Workflow of the image analysis program.
Details explained in text.

4.2.2 Image smoothing and thresholding

3D image stacks are loaded as 3D matrices and converted to grayscale. The gray-scale image stacks are minimally smoothed using an anisotropic diffusion algorithm[190], with previously published MATLAB code[191]. We compared the effect of the number of

smoothing iterations using experimental images containing densely packed cells (Figure 4.8). Without smoothing, the object outlines remain jagged to the point of creating many identifiable concave points. After three smoothing iterations, the object outlines become smooth and, mostly, significant concave points, representing actual intersections between cells, remain. After 10 smoothing iterations, objects become indistinguishable and meaningful concave points are obscured. Depending on the actual properties of the input images, especially the signal-to-noise ratio, the users may need to optimize the number of smoothing iterations empirically. However, we generally recommend one to three iterations, because, on one hand, without any smoothing, signal variations within the same cell can cause more roughness on the edges that leads to oversplitting of cells into multiple small fragments (Figure 4.9); on the other hand, oversmoothing can lead to the merging of multiple cells into one.

After image smoothing, the first stage in cell identification is implemented through a two-step local thresholding technique. First, the original image is sharpened by subtracting its corresponding low-frequency image, in which only the low-frequency signals of the Fourier-transformed image are kept, and then rescaling[192]. This helps correct for the slow-sweeping changes in background illumination characteristic of bacteria clustering around distinct puncta in images. Second, objects pass through a Bradley adaptive threshold[193,194], wherein the intensity of pixels of candidate objects must be a defined percentage higher than the mean intensity of the pixels in an $n \times n$ neighborhood (we set n equal to 10, 11, or 12, depending on the size of the image). We chose a local, adaptive threshold with a fixed percentage, rather than a single, global threshold, to avoid false positives and false negatives due to uneven illumination, autofluorescence from the host cells, and/or additive fluorescence from neighboring cells. We tested the threshold parameter sensitivity of the 2D object selection on experimental input images with or without smoothing. The initial thresholding results are robust across a range from 25 to 400% of the default intensity thresholding value (0.00001), leading to no discernible differences in the initial characterization of the objects (Figure 4.10).

4.2.3 Single-cell identification by shape and concavity analysis

For each identified object in a 2D slice, we perform shape and concavity analysis to distinguish single cells from clustered cells. For shape analysis, we fit each 2D object to an ellipse using a least-squares criterion[195,196]. We then compare the actual outline of the thresholded object to the ellipse and calculate the deviation of the outline from the best-fit ellipse, using an analogue of the Hausdorff distance[196] (Figure 4.2A). The assumption is that bacterial cells are roughly elliptical, and if a 2D slice of an object deviates too far from its best-fit ellipse, it is likely not a single cell (Figure 4.2C). The deviation of the object from its best-fit ellipse is calculated by the sum of the distance between each object edge pixel and the nearest edge pixel in the best-fit ellipse. The sum is then normalized by the total number of pixels on the periphery of the object to give an error value ($Err_{2D\text{-}shape}$) (Figure 4.2A).

To calculate the concavity of each edge pixel (Figure 4.2B), the edge coordinates are first arranged into a two-column array and then smoothed with a Savitzky–Golay filter[197]. After the smoothing, the tangent line of a given edge pixel is approximated by fitting a straight line to the edge pixel and two pixels before and after the edge pixel. The angle of the approximate normal line, perpendicular to the tangent line, is then calculated. Finally, the difference between the angles of the normal lines for pixeli and pixeli–1 is recorded, which we define as a parameter for concavity (Conc) (Figure 4.2B). With our definition, a positive Conc value marks a concave point (Figure 4.2B). In the case in which multiple cells are clustered into the same binary object, the point at which the cells intersect is a concave point in the binary map (Figure 4.2C).

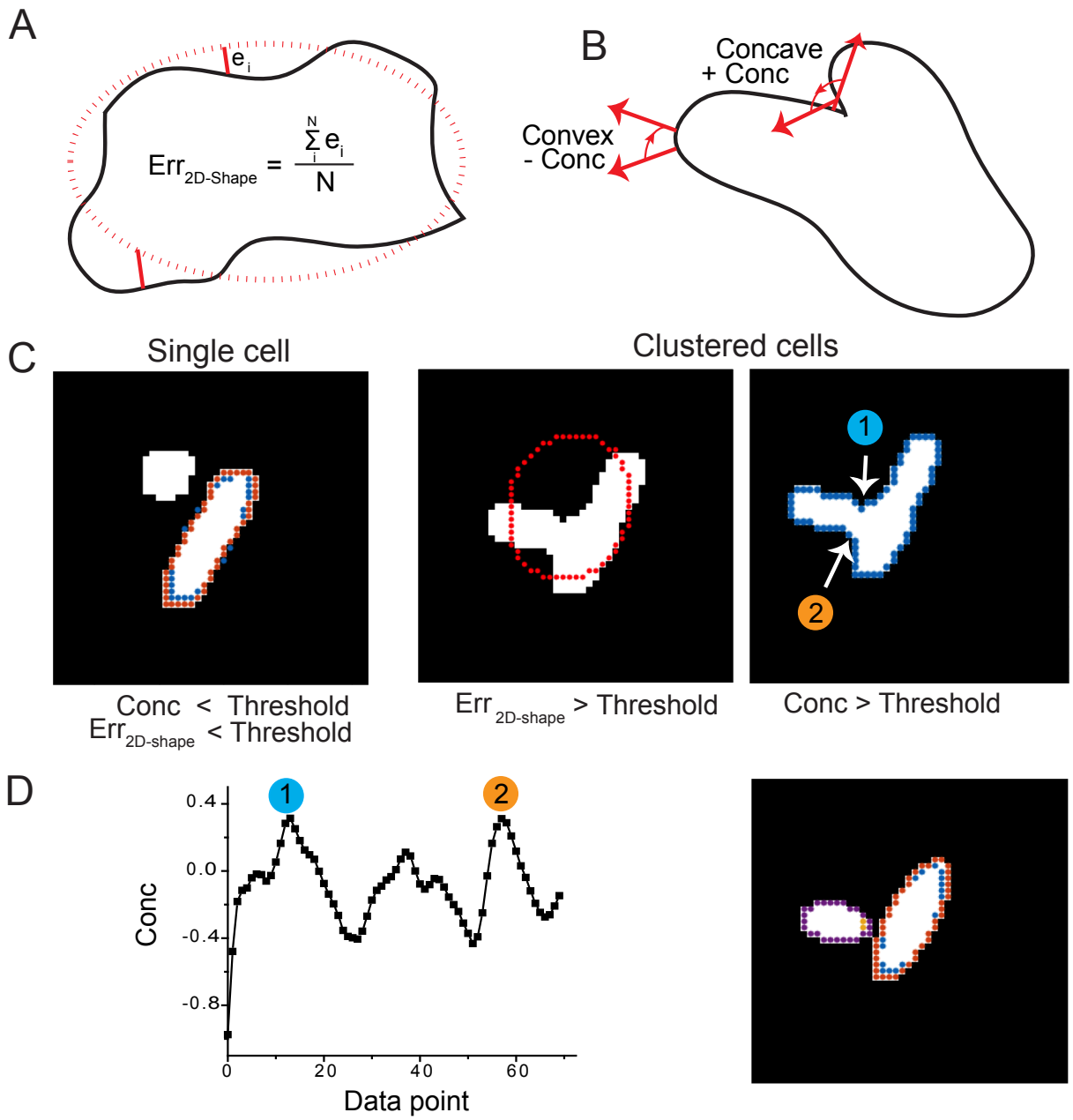


Figure 4.2 Shape and concavity analysis for single-cell identification in two dimensions. **(A)** Illustration of the calculation of $Err_{2D\text{-}shape}$. **(B)** Illustration of the calculation of Conc. **(C)** Example of single cell vs clustered cells, showing a difference in $Err_{2D\text{-}shape}$ and Conc. The outline of the isolated, single cell closely resembles its best-fit ellipse, resulting in a subthreshold $Err_{2D\text{-}shape}$ value. The outline of the clustered cells, in contrast, deviates significantly from their best-fit ellipse, resulting in a high $Err_{2D\text{-}shape}$ value and leading to this object being labeled as a nonsingle cell. Similarly, the isolated single cell is a mostly convex object, whereas the clustered cells can be easily identified by the presence of one or more concave points. **(D)** Concavity values for every border pixel in the clustered cell object are shown in the plot (left). The two most concave points are marked and are made candidates for splitting location. These splitting coordinates result in a successful segmentation.

Therefore, we could distinguish single cells from clustered cells using their $Err_{2D\text{-}shape}$ and Conc values. If $Err_{2D\text{-}shape}$ and Conc are below the user-defined thresholds, the object is classified as a single cell; otherwise, the object is deemed a nonsingle cell and is subjected to concavity-based splitting (Figure 4.2C).

Objects that do not pass the single-cell identification described above are passed through a concavity-based splitting algorithm, adapted from previously published methods[187–189]. After calculating and storing the concavity of every edge pixel for nonsingle cells, we arrange them from highest to lowest Conc value (x_1 to x_n, respectively), disregarding convex points (Conc < 0), and select the largest Conc values along the object borders as candidate locations for splitting (Figure 4.2D). Starting from the pair of x_1 and x_2, we draw a straight line between the points, splitting the cell. After the cell is split, we characterize the fragments using the single-cell identification procedure described above. If one or both of the fragments now pass the single-cell identification, we move on to the next nonsingle cell object. If neither of the fragments passes the threshold, we draw a new line between points x_1 and x_3, x_1 and x_4, ..., and x_1 and x_n, and then between points x_2 and x_3, etc. In the interest of time, if the nonsingle cell is not split into objects that pass the single-cell identification in k attempts (k can be adjusted by the user), the algorithm passes over this object onto the next one, and the object remains classified as a nonsingle cell and discarded in the end.

4.2.4 Morphology-based 3D reconstruction

After each slice of the 3D image has been fully segmented and characterized, the 2D slices of the cells are combined into 3D volumes. We introduce three parameters for 3D reconstruction: D_{Center}, the distance between the geometric centers of the candidate objects from two consecutive *z* slices (Figure 4.3A); $Err_{3D\text{-}shape}$, the shape deviation between two candidate objects, a parameter equivalent to $Err_{2D\text{-}shape}$, substituting the best-fit ellipse with a potential partner 2D object (Figure 4.3B); and θ_z, the angle between

the *z* axis and the line connecting the centers of the candidate objects from two consecutive *z* slices (Figure 4.3C). The use of these parameters is based on the assumption that each projection on the x–y plane from a single cell should have similar localization and shape, and the orientation of the cell relative to the *z* axis should be a constant. Therefore, we can define thresholds for these parameters to determine whether 2D objects belong to the same cell. If multiple objects in slice_{i+1} meet the criteria for 3D recombination with a 2D cell in slice_i, the object in slice_{i+1} that minimizes these parameters is chosen as the partner.

The reconstructed 3D objects are checked for two additional criteria: *V*, the total voxel number occupied by the cell, and N_Z , the number of *z* slices occupied by the cell. Considering all possible orientations, a single cell should still occupy a minimum number of *z* slices and 3D volume. Therefore, by applying these two criteria, we can eliminate incompletely reconstructed cell fragments due to mis-segmentation in two dimensions. The identified 3D fragments based on *V* and N_Z can then be combined into a complete single-cell candidate if they have matching θ_Z and small D_{Center} values. Fragments that remain after attempted 3D recombination will be disregarded if they do not exceed the *V* and N_Z thresholds.

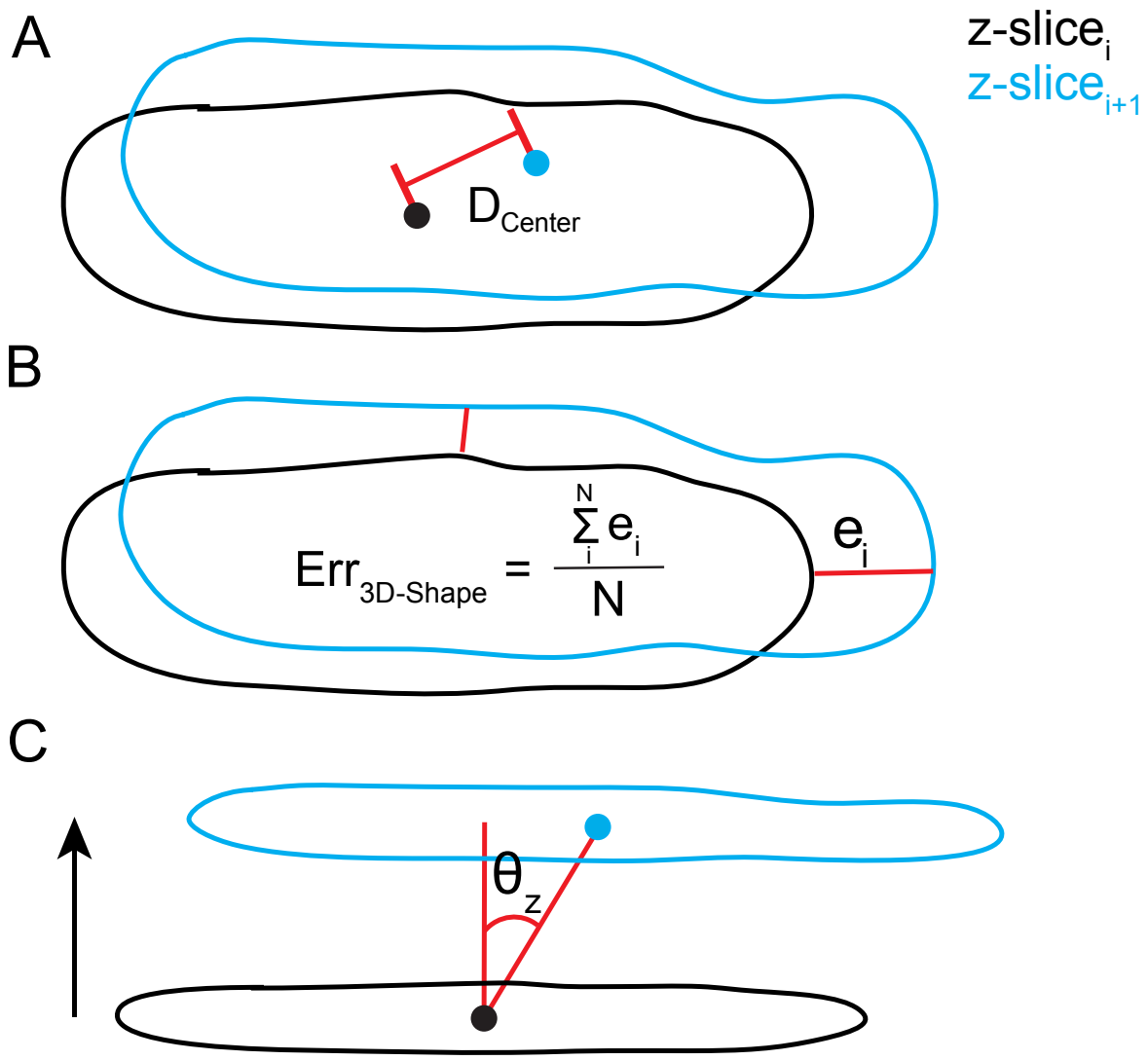


Figure 4.3 Parameters for 3D reconstruction.
(A) Illustration of the calculation of D_{Center} between potential partner objects in adjacent z slices. **(B)** Illustration of the calculation of $Err_{3D\text{-}shape}$ for potential partners in adjacent z slices. **(C)** Illustration of the calculation of θ_z for a 3D cell.

4.2.5 Method validation with synthetic data

To validate Seg-3D, we first tested it on synthetic data (Figure 4.4). To generate the synthetic data, we modeled bacterial cells as 3D rod-shaped objects comprised of individual voxels with an *xy* pixel size and *z* step interval of 130 nm × 130 nm × 130

nm, the parameters of our microscope and camera setup[198]. The 3D objects were then convolved with the point-spread function to better represent the 3D image of the bacteria. Several synthetic bacteria were randomly placed in 3D space with certain interbacterial space to make synthetic images with certain crowdedness such that cells can touch but cannot intersect (Figure 4.4B).

Seg-3D requires several user-input parameters for single-cell identification and 3D reconstruction as described above, including $Err_{2D\text{-}shape}$, Conc, D_{Center}, $Err_{3D\text{-}shape}$, θ_z, V, and N_z. To correctly decide the thresholds for these parameters, we generated synthetic single cells with a random orientation and extracted all the parameters from these single cells (Figure 4.4A). A histogram of each parameter represents the expected range from single cells. Thresholds were then set on the basis of the histograms to include at least 90% of the single-cell population. Specifically, we picked values marked by the red lines in Figure 4.4A as parameters for analyzing all of our images.

Seg-3D was then applied to the synthetic data (Figure 4.4B). We compared Seg-3D with previously published methods used to study bacterial biofilm based on the watershed algorithm without splitting or user correction[183,199]. Seg-3D showed improved accuracy in segmentation of the clustering bacteria. With 50 randomly generated synthetic images each containing five clustered bacteria, Seg-3D correctly segmented all five cells in □76% of the synthetic images. Of 250 cells in the synthetic data, Seg-3D correctly segmented □92% of them, while the watershed algorithm correctly segmented only 36% of the cells (Figure 4.4C). Several cases contributed to the incorrect segmentation (□8% of the total), including failure in splitting two significantly merging cells in two dimensions (missing splitting), failure to identify objects in two dimensions (missing objects), and failure to combine two 3D fragments belonging to one cell (oversplitting). Very rarely (two of 250 cells from 50 synthetic images), we observed misidentification of an object from the background, which was characterized as a cell. We found tuning parameters such as $Err_{2D\text{-}shape}$ could eliminate this error.

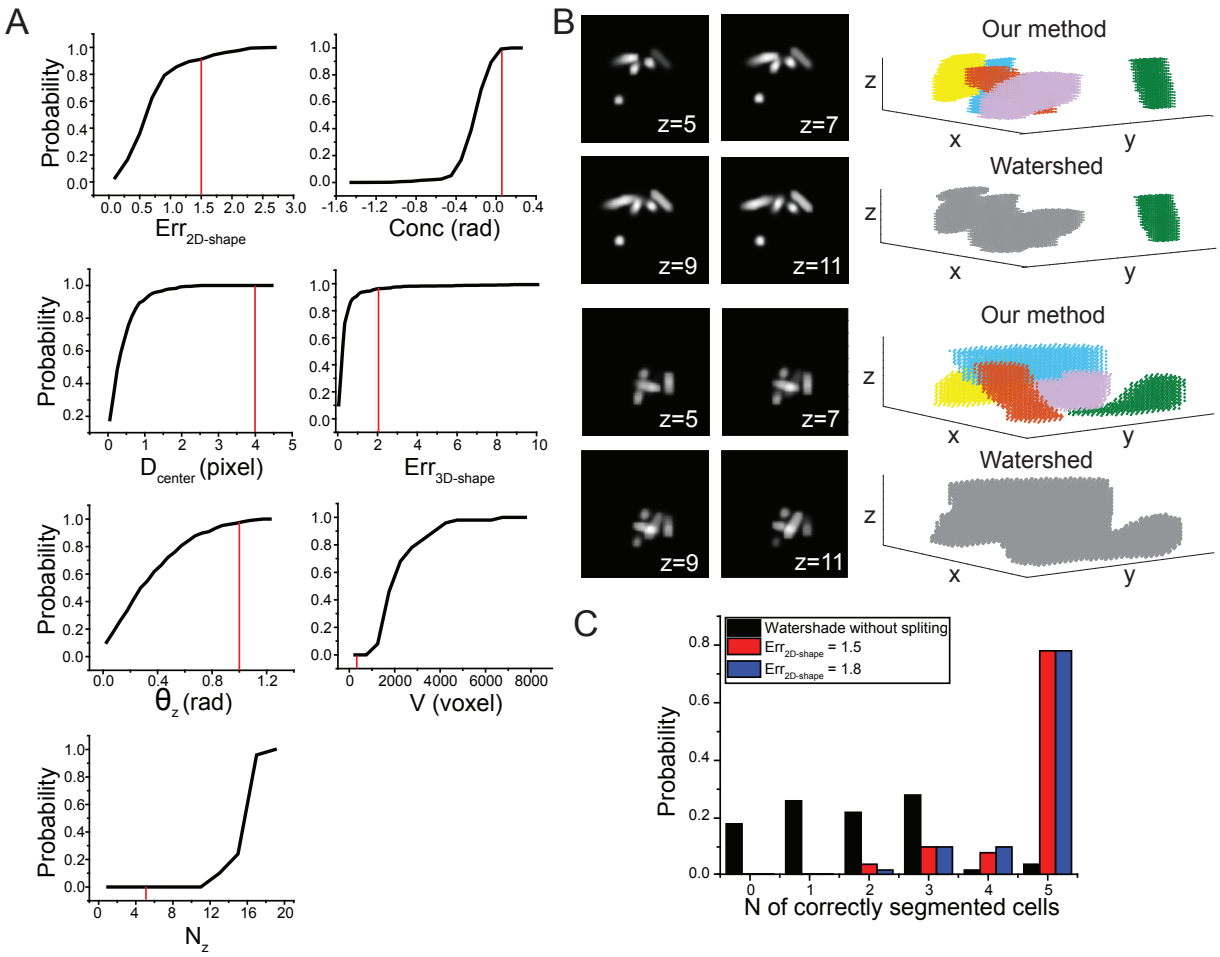


Figure 4.4 Examination of Seg-3D on synthetic images.
(A) Distribution of all parameters from synthetic images containing single cells. Red lines mark the thresholds used in analyzing synthetic images. **(B)** Two examples of 3D segmentation on synthetic images with Seg-3D and the watershed-based method without object splitting or manual fixation. Each group of successfully segmented cells is color-coded, while cells that failed to segment correctly are colored gray. **(C)** Success probability of Seg-3D (black) and the watershed-based method (red). Each synthetic image contains five randomly oriented but closely neighboring cells.

4.2.6 Test of parameter sensitivity

To evaluate the parameter sensitivity of the segmentation accuracy, we picked four other values smaller than (minimum of 25% of) and larger than (maximum of 400% of) the standard values for each parameter and tested them on the same synthetic data sets. Our results show that the Seg-3D values are robust to the changes in these user-input parameters as long as they exclude the expected ranges for single cells. Specifically, the

results are not sensitive to the change in Conc (0.1–0.4), as single cells should generate only negative Conc values, or sensitive to the change in θ_Z (0.6–1.4). Mis-segmentation happens frequently when the choice of the parameters gets close the single-cell parameter range. For example, when $Err_{2D\text{-}shape}$ is set to 0.5 (too stringent), 2D objects more often fail to be considered as single cells after a few trials of concavity-based splitting and are therefore discarded. When $Err_{3D\text{-}shape}$ or D_{Center} is set to be too small (too stringent), 2D objects more often fail to be considered to come from the same 3D cell, causing frequent oversplitting in three dimensions. The choice of a large value for V or N_Z causes rejection of small cells and therefore undercounts the cell number (Figure 4.5).

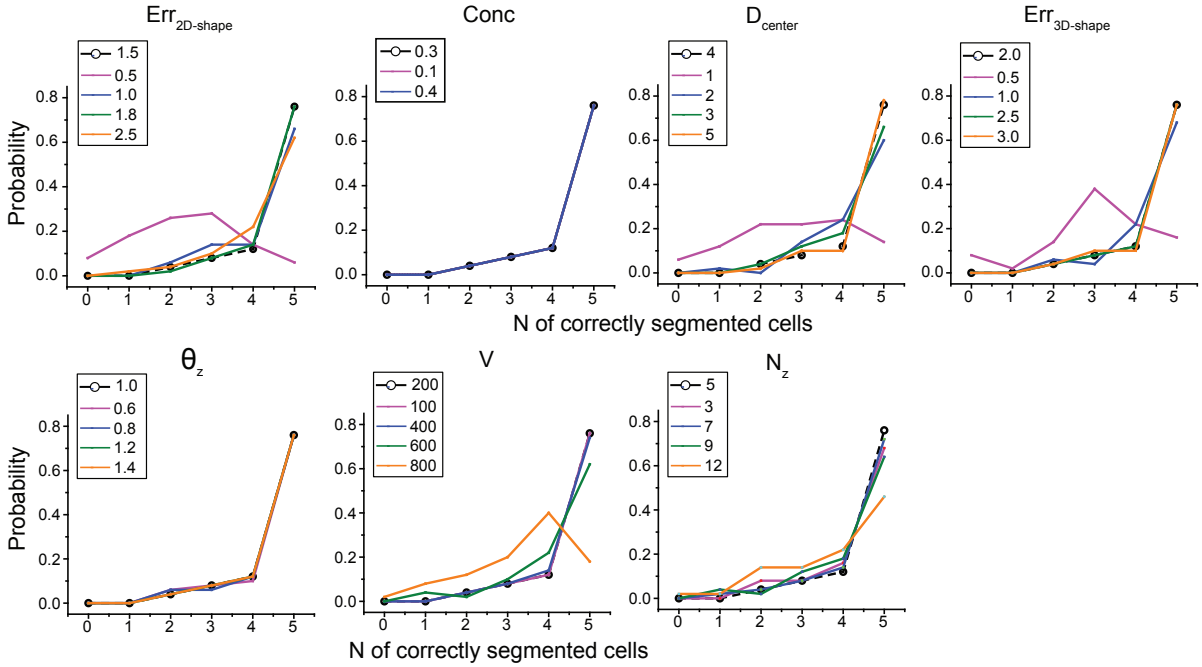


Figure 4.5 Test of parameter sensitivity on synthetic images.
Parameters are varied relative to the default values marked in Figure 4.4. The success probability using default parameters is shown as dashed lines.

4.2.7 Application of method to intrahost pathogenic bacteria

We then applied our 3D segmentation method to sample images of bacteria invading macrophages (Figure 4.4). Salmonella cells expressing GFP from a constitutive

promoter[200] were used to infect murine macrophages (RAW 264.7), and macrophages were then fixed and imaged under the fluorescence microscope. Similar to the synthetic data analysis, we first analyzed the distributions of all critical parameters corresponding to single cells from low-cell density images. Relative to the histograms derived from synthetic data, we found that ranges expected from single cells were very similar for all seven parameters, suggesting the parameters would be robust for applications to bacterial species with similar 3D shape and size. Therefore, we directly applied the same thresholds to analyze the real data. Figure 4.4 shows two examples of clustered intramacrophage *Salmonella* cells, and Seg-3D effectively segmented individual *Salmonella* in three dimensions.

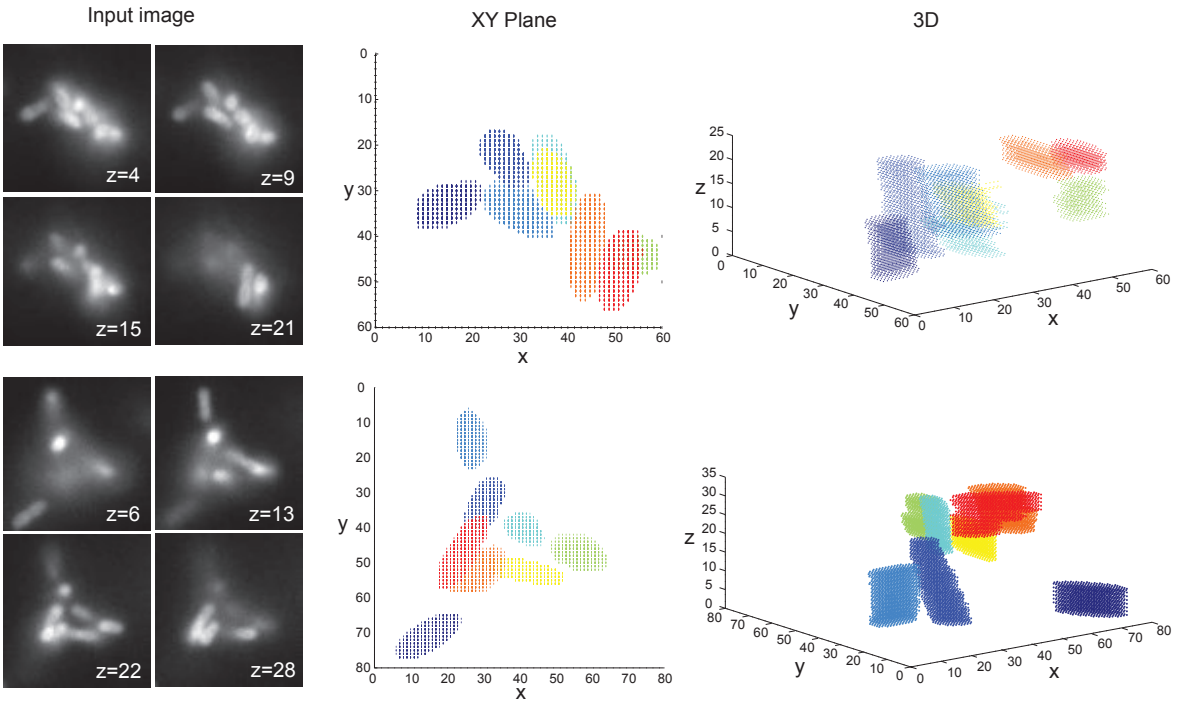


Figure 4.6 Examination of Seg-3D on experimental images.
Segmented cells are color-coded.

4.2.8 Additional features in the user interface

While Seg-3D has greatly improved the efficiency and accuracy of 3D segmentation, mis-segmentation still occurs. For example, when two cells happen to be completely touching each other with the exact same orientation, it is very likely that they can pass the single-cell criteria in 2D segmentation, or the concavity-based automatic splitting fails to split them. Therefore, we add in the manual proofread and correction feature after automatic segmentation. Possible corrections include switching positive selections to negative, switching negative selections to positive, manually drawing split lines for remaining clustered objects, manually drawing borders for missed objects, and deleting objects.

Moreover, the algorithm allows analysis for multichannel images. 3D segmentation will be performed on a user-defined channel with a uniformly stained fluorescent signal that can represent the full cell volume well. The fluorescence signal from other channels with staining on biomolecules of interest will be allocated into each segmented cell, allowing further single-cell quantification of biomolecules. Finally, the 3D segmentation code can be very flexibly adapted to analyze surface-attached cells in two dimensions. Seg-3D will benefit single-cell imaging and analysis under complex conditions such as bacterial pathogen infection and biofilm formation. The full package of Seg-3D, coded in MATLAB, with a user manual can be found in the Supporting Information or downloaded as open source code (https://github.com/JingyiFeiLab/Cell_seg).

4.2.9 *Supplementary information*

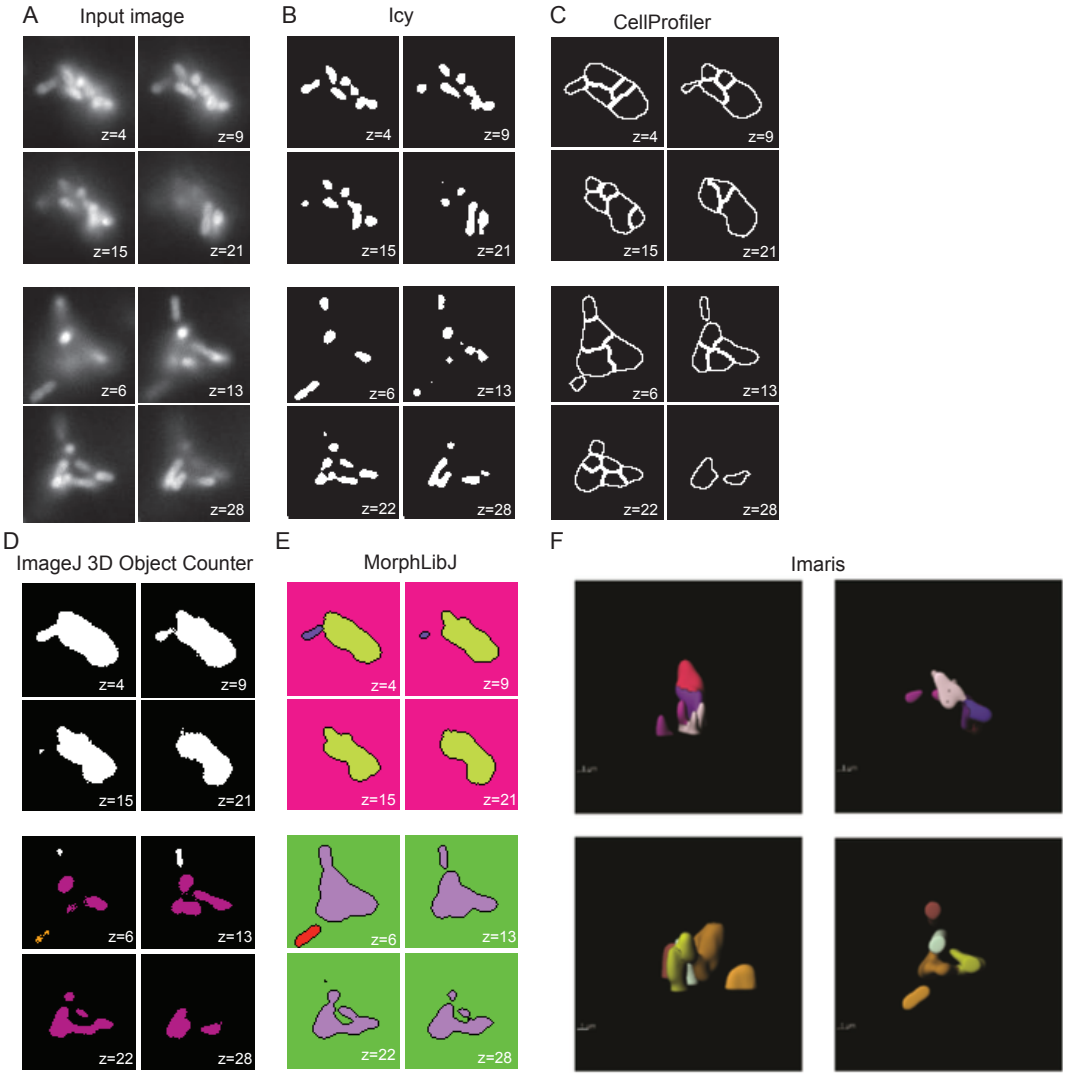


Figure 4.7 Segmentation results from available commercial and free software.
(A) The same two experimental images as in Figure 4.4 are used to test the performance of popular available software packages. Since most available packages do not have 3D reconstruction or visualization capabilities, we show their performance on 2D slices. **(B)** Results from Icy Spot Detector. All disconnected objects in a slice are given unique IDs, and any objects connected by at least a single pixel are identified as a single object. **(C)** Results from CellProfiler. Each enclosed space is given a unique ID. **(D)** Results from the ImageJ Plug-In, 3D Object Counter. Each color represents a unique 3D object. **(E)** Results from ImageJ Plug-In, MorphoLibJ. Each color represents a unique 3D object. **(F)** 3D reconstruction results from Imaris. The two views for each image are the same as those in Figure 4.4. Each color represents a unique 3D object.

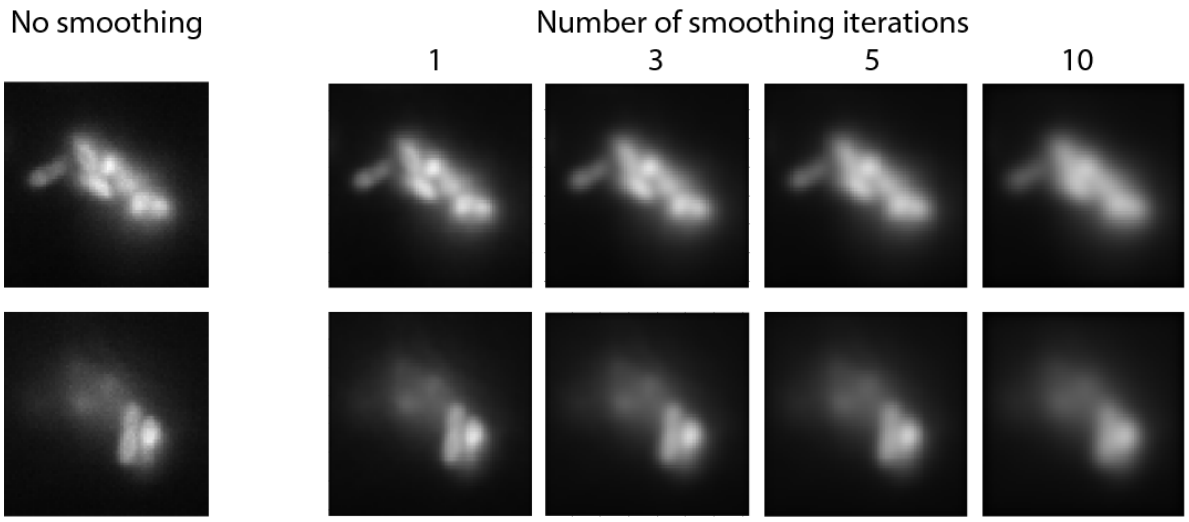


Figure 4.8 Effect of different number of smoothing iterations.
(Left) Two 2D slices before any image processing, and (Right), the image after processing by the Anisotropic Diffusion smoothing algorithm for different numbers of iterations. The output image from smoothing will be passed directly to the initial, local-intensity thresholding. It is clear that the difficulties arise when the image is over-smoothed. Aside from the number of smoothing iterations, other parameters for Anisotropic Diffusion smoothing algorithm are the default parameters provided by reference [191].

Figure 4.9 Example of an over-split cell due to no smoothing.
(Left) An example of two clustered cells, without any image processing. (Middle) An attempt at segmentation without smoothing the image at all. Each color represents a unique object. The cell on the left is incorrectly split into two, likely due to the jagged edges of its initial outline before concavity-based splitting. (Right) The segmentation of the same two cells after three smoothing iterations. The cells are correctly segmented.

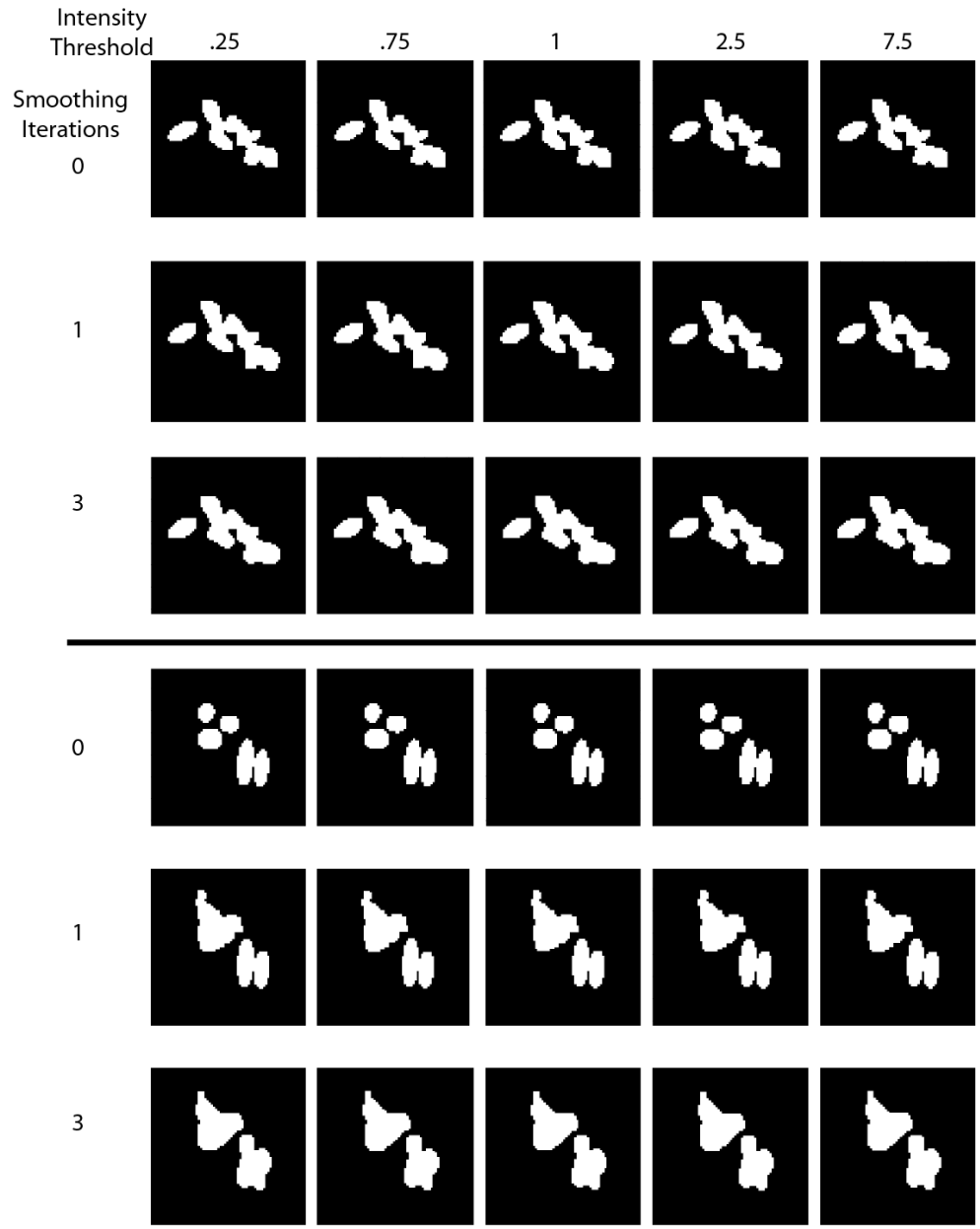


Figure 4.10 Sensitivity of 2D segmentation to intensity threshold after different numbers of smoothing iterations.
The same two z slices shown in Figure 4.8 are used as examples. It is clear that the results of the local-intensity-based thresholding depend on the number of smoothing iterations the original image is subjected to, but they are robust to a wide range of intensity threshold parameter values. We varied this value from 25 to 400% of the default parameter value, and the differences in the results are indiscernible.

4.3 An improved method for bacterial immunofluorescence staining to eliminate antibody exclusion from the fixed nucleoid

In this study, we introduce an improved immunofluorescence (IF) staining method that avoids one of the common disadvantages of the general technique. IF involves many steps, such as fixation and permeabilization, that may affect cell morphology and introduce potential imaging artifacts. These artifacts manifest in the images as changes in the geometry of the cells, or in altered fluorescent signals relative to the ground truth. These are both things that could be measured with image analysis, which is presented here.

In response to the first concern (distorted cell geometry), I adapted the code from Section 4.2 such that we were able to both automatically segment cells in DIC images and quantify their geometry. Again, an adaptive, local threshold was used to distinguish objects from the background of the images. In order to further distinguish non-cell selected objects from true cells, each threshold-selected object was fit, using a least-squares criterion, this time to a superellipse of best fit[201], rather than simply an ellipse. This modification was carried forth to all future image analysis code involving *E. coli*, as we found the 2D profile of *E. coli* matched the superellipse more closely than the ellipse. The formula for the best-fit superellipse for the identified cells was then used to calculate the area, length, and width of the cells. Using this analysis, we were able to determine the effect of various fixation and permeabilization steps on the morphology of the cell and, importantly, find the protocol that affected the natural morphology the least (Figure 4.11).

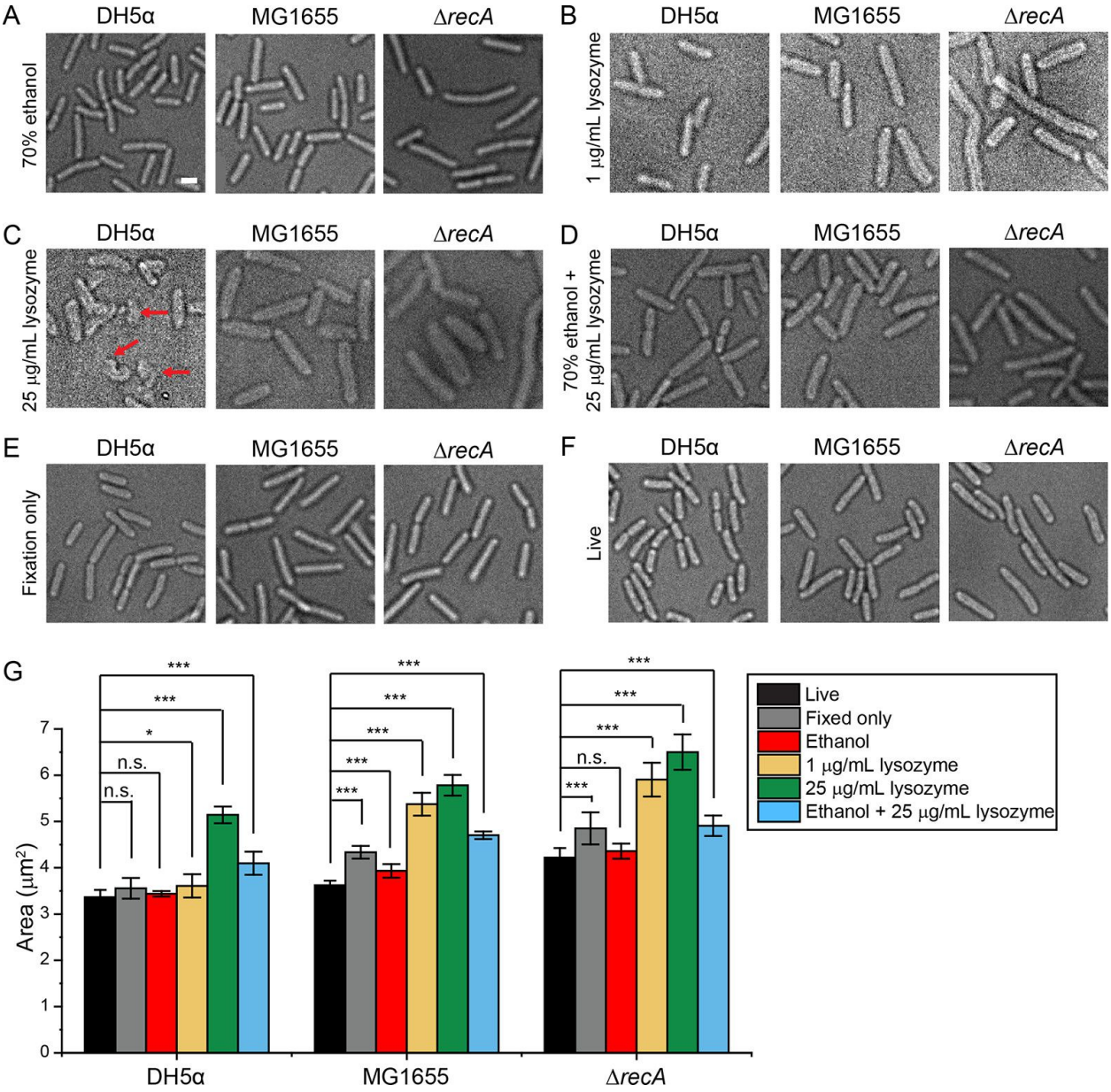


Figure 4.11 DIC images of E. coli cells with different permeabilization conditions. **(A)** 70% ethanol, **(B)** 1 μg/mL lysozyme, **(C)** 25 μg/mL lysozyme (red arrows indicate severely disrupted cells), **(D)** 70% ethanol with 25 μg/mL lysozyme, **()** fixed only, and **(F)** live cells. The scale bar represents 2 μm. **(G)** Areas of cells under different permeabilization conditions, quantified by cell segmentation based on the DIC images[105]. Error bars represent means and standard deviations of 5−10 images, with each image containing 70−300 cells. p values from the t test are added to indicate significance. n.s. stands for "not significant".

To address the second concern, we visualized the same protein in two different ways; comparing the two signals revealed which permeabilization methods led to staining arti-

facts. Using RecA-GFP, we were able to compare the fluorescent protein signal directly from the expressed protein with stained-antibody, tagged with Alexa Fluor 647 NHS Ester dye (A647) fluorescent signal. The GFP signal served as the ground truth for the protein distribution. If the antibody staining worked well, the A647 fluorescent signal should correlate well with the GFP signal; the level of divergence from that correlation served as a reference for how disruptive a particular permeabilization method was. In order to quantify that divergence, and therefore the quality of the method, I created a program to automatically calculate the correlation between fluorescent signals in bacteria. Using the technique described above to identify cells and fit their profiles to superellipses. Then, using the identified tilt and major axis of the superellipse of best fit, a line scan was conducted in the fluorescent channels of interest along the middle length of the cell in order to create intensity profiles. Since *E. coli* cells are close to uniform width under a given growth condition and narrow relative to the length, the fluorescent signal of a localized protein generally takes up the entire width of the cell; therefore a single length-wise line is sufficient to accurately characterize the intensities. Correlations between two intensity profiles were calculated by assuming each intensity profile was an independent random variable with *N* observations, where *N* is the number of pixels in the line scan. The Pearson correlation coefficient is calculated according to:

$$\rho(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}(\frac{A_i-\mu_A}{\sigma_A})(\frac{B_i-\mu_B}{\sigma_B})$$

To further test if different permeabilization methods affected the IF signals, we used that same code to calculate the correlations between the A647 fluorescent signal and DAPI staining of DNA. These analyses revealed that a permeabilization method with a lysozyme only treatment lead to an accurate depiction of protein distribution, but greatly affected the cell morphology. In contrast, a method including 70% ethanol followed by lysozyme treatment preserved the cell morphology, but affected the protein distribution, specifically

excluding the cell from the nucleoid region, as seen by the negative correlation between the A647 and DAPI profiles. Further, we found that the antibody exclusion effect could be overcome by a subsequent DNase I treatment, all the while preserving cell morphology. Finally, we were able to land on an optimal protocol involving 70% ethanol, lysozyme, and DNase I treatment, which preserves both protein distribution and cell morphology. These revelations were made possible in part because of the image analysis programs.

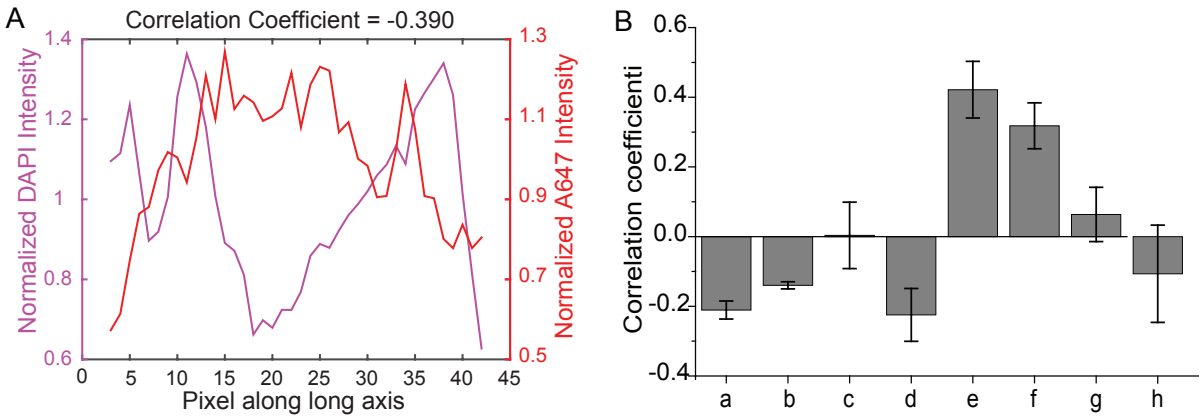


Figure 4.12 Correlation analysis between RecA-A647 and DAPI signals.
Cells are first segmented based on the DIC image, and the two fluorescent channels are manually aligned. An intensity profile for each fluorescent channel is determined by scanning across that long axis of each cell. The correlation coefficient between the two intensity profiles is calculated for each cell. **(A)** A647 and DAPI intensity profiles from a representative cell from the 70% ethanol + 25 µg/mL lysozyme treatment condition. **(B)** Average correlation coefficient between A647 and DAPI signals under various permeabilization conditions. (a) 70% ethanol + 25 µg/mL lysozyme; (b) 1 µg/mL lysozyme; (c) 25 µg/mL lysozyme; (d) 70% ethanol + 25 µg/mL lysozyme, with Atto 655 labeled secondary antibody (e) 70% ethanol + 25 µg/mL lysozyme + 100 U/mL DNase I; (f) 70% ethanol + 25 µg/mL lysozyme + 0.8 U/mL DNase I; (g) 70% ethanol + 25 µg/mL lysozyme + 0.16 U/mL DNase I; (h) 70% ethanol + 25 µg/mL lysozyme + 0.032 U/mL DNase I. Except for (d), all other conditions were prepared with A647 labeled secondary antibody. Error bars represent mean and standard deviations of 3-8 images, with each image containing 70-300 cells.

4.4 Analysis of Localization and Diffusion of RNA and Proteins in Bacterial Cells

In this study[127], we establish a platform to determine the RNA binding states of the RNA chaperone protein, Hfq, using diffusivity as a pro*xy*. Using this platform, we were

able to shed light on the complex interplay between Hfq, mRNAs, sRNAs, and RNase E, and reveal how Hfq is able to efficiently prioritizes different RNAs for regulation, when Hfq is greatly outnumbered. The diffusion speeds and coordinates of Hfq molecules were acquired with single-molecule localization microscopy (SMLM), specifically PALM imaging of Hfq tagged with a photo-switchable fluorescent protein, mMaple3[202,203] in 2D, though the corresponding analysis code is capable of analyzing both 2D and 3D image datasets. The insights we gained into the mechanism of Hfq-sRNA-mediated regulation were dependent upon localizing and tracking fluorescently-tagged Hfq in live cells. New image analysis programs were required to transform the tracking and localization data into a comprehensible form. This section describes my contribution in the form of creating those programs.

4.4.1 Diffusion and localization of Hfq

The code to localize individual Hfq molecules and their respective one-step diffusion coordinates requires only two inputs. The first is a DIC image in order to segment individual cells and assign image volumes to them. The second is a .txt file listing the one-step diffusion speeds and locations. Optionally, a single image of DNA staining can be included if the nucleoid is also a region of interest.

Again, segmented cells or fit to a superellipse in order to calculate their center, length, width, and tilt (relative to vertical in the image). Next, the coordinates of the cell boundaries (representing the outer cell membrane) are stored as 2D vectors (x- and y-coordinates of boundary pixels represented by columns, each pixel represented by a separate row). Using the tilt angle, Θ, the cell boundary coordinates are transformed to vertical using the 2D rotation matrix:

$$R_{\Theta} = \begin{bmatrix} \cos(\Theta) & -\sin(\Theta) \\ \sin(\Theta) & \cos(\Theta) \end{bmatrix} \tag{4.1}$$

Every Hfq localization is assigned to a segmented cell, based on whether or not the spot falls within the cell boundaries. All Hfq localizations within a given cell are also stored in another 2D vector (x- and y-coordinates of each localization represented by columns, each localization represented by a separate row). The corresponding one-step diffusion rates for each localization are stored in a separate vector, with the rows equal to the particle ID in both the localization and rate vectors. Next, all Hfq localizations are rotated using the same rotation matrix as the one used for the cell boundaries. The cells and the corresponding Hfq localizations are transformed to a vertical alignment to simplify subsequent distance and compartmentalization calculations. For the sake of downstream analysis, information regarding cells and spots are stored in separate structured arrays, containing information regarding The location, size, alignment, etc. of the cells, and localization, cell assignment, distance to membrane, etc. for the spots.

Next, the cells are segmented further into separate compartments of interest. Using information on the major and minor axes of the cells and their vertically aligned boundaries, the cells can be broken down into a number of sub-cellular compartments, including the membrane, poles, cytoplasm, and nucleoid (if a DNA staining image was included). The number of compartments included in the analysis is up to the user. In this study, we were interested only in the nucleoid, cytoplasm, and membrane of the cell. In this analysis, we found that under normal growth conditions (exponential growth, no treatment applied) Hfq diffuses and localizes relatively uniformly throughout the cell, with a slight over-enrichment in the cytoplasm and a slighlty slower diffusion in the membrane (Figure 4.13).

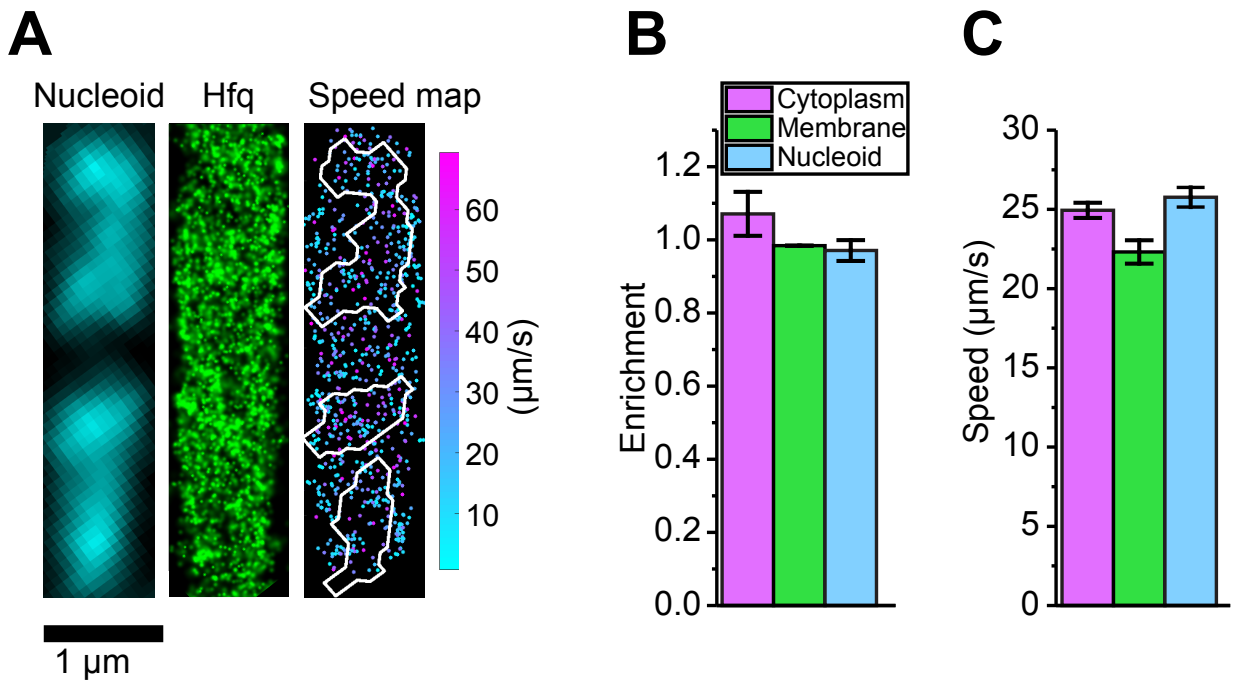


Figure 4.13 Diffusion and localization of Hfq during exponential growth.
(A) A representative example of WT Hfq-mMaple3 in WT rne background in a single cell during exponential growth under no treatment (NT) condition. Nucleoid is stained with Hoechst in live cells. 2D reconstructed image of Hfq-mMaple3 is shown in the black background. One-step displacement (osd) speed map (unit: μm/s) is shown as a scatter plot where different colors represent different speeds at each position, and the white curves represent the nucleoid regions detected by Hoechst staining. **(B)** Enrichment of Hfq localization is calculated for cytoplasm, membrane, and nucleoid regions under NT condition. **(C)** Average osd speed of Hfq within the cytoplasm, membrane, and nucleoid regions under NT condition. Error bars in all plots represent the standard deviation (s.d.) from two experimental replicates, with each data set containing ~20,000 trajectories from 80 cells.

4.4.2 3D bacterial cell region projection and enrichment

In a separate study (unpublished) we were interested in the 3D organization and compartmentalization of various sRNAs under different stress conditions. The code described above can be easily adapted for this purpose. Rather than one-step diffusion speeds acquired from PALM imaging, our input data instead was 3D coordinates of FISH-labeled sRNAs acquired from STORM imaging. Cells are segmented and geometrically transformed as described above; sRNA spots are similarly transformed with the corresponding rotation matrix. The z-dimension represents an added difficulty for the geometric trans-

formation, but here we see the advantage of the vertical cell realignment by the rotation matrix. Using a step-wise transformation we first transform only the x- and y-coordinates of the sRNA spots. Then, since the segmented cells are already projected onto 2 dimensions (we assume bacteria cells are roughly cylindrical and thus have circular width-wise cross sections), the x/y transformed coordinates can easily be collapsed onto the 2D cell plane by considering only their z-coordinate as one side of a right triangle, and the distance off the center axis (represented by the transformed x-coordinate) as the other side. Then, simple trigonometry can be used for the second transformation. The results of one such analysis are shown here (Figure 4.14).

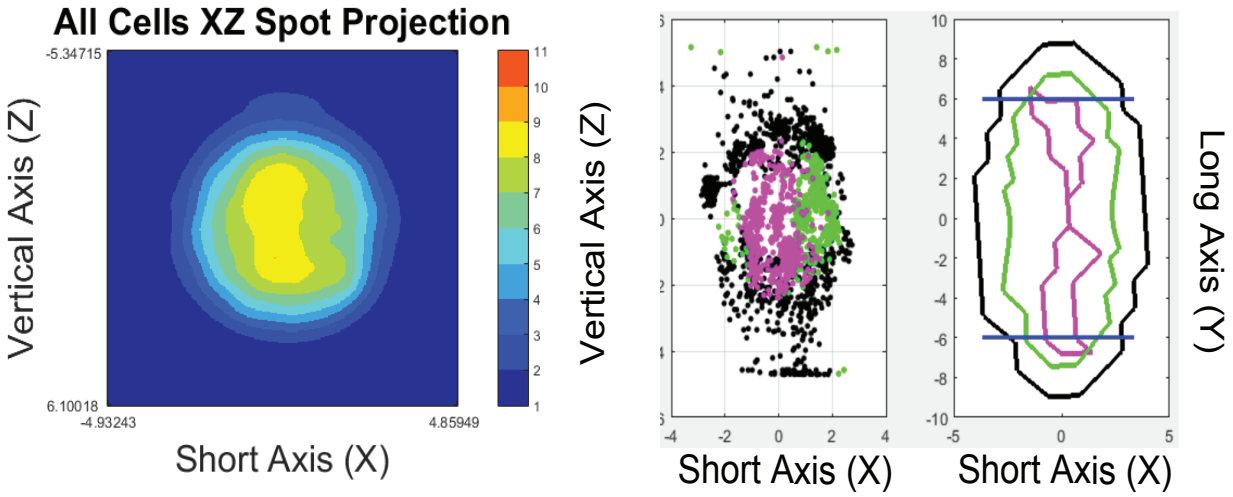


Figure 4.14 3D Bacterial Cell Region Projection and Enrichment.
(Left) Average sRNA spot density projected along the short, circular cross-section of *E. coli* cells, and (Right) Assignment of sRNA spots and boundary definitions of subcellular regions for a single cell. Purple represents the nucleoid, green the cytoplasm, blue the poles, and black the membrane.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

The work I have described in this dissertation describes my contributions to to the study of RNA-mediated regulation. Specifically, I have attempted to provide mathematical and computation methods and models to describe such regulation. I developed a general, quantitative model of sRNA-mediated regulation, and in the process, discovered that certain sRNAs that are canonically described as post-transcriptional regulators are able to regulate their mRNA targets co-transcriptionally; I helped define which kinetics contribute to the hierarchy of an sRNA regulon; I contributed numerous image analysis methods that have helped and will continue to help future studies of the localization, dynamics, and kinetics of RNA-mediated regulation; and I have described our early ventures into the study of interactions between Hfq and the ribosome.

I am excited by the new questions my research has inspired. In particular, I hope to apply the kinetic description of sRNA-mediated regulation to an expanded set of sRNA regulons, and attempt to describe other forms of regulation, such as translation activation. I believe that the modelling and analysis work described especially in Chapter 2 can be broadly applied to other kinetic studies, and I hope that Bayesian MCMC methods can provide deeper physical insights into gene regulatory networks.

In the immediate future, we will dive directly into and expand the studies proposed in Chapter 3. We hope to settle the mechanisms behind the trapping effect that condensed ribosomes have on Hfq, and to determine whether or not this effect facilitates the sRNA target search process, which could provide further evidence in support of the kinetic co-operation model of sRNA-mediated regulation.

And finally, I hope to apply the computational and mathematical methods I have developed more broadly to biological phenomena outside of sRNA-mediated regulation. I

believe that the general framework we described here - moving from imaging, to image analysis, to mathematical analysis - can be a powerful and broad tool for a more physical understanding of biology.

REFERENCES

1. Balleza, E.; López-Bojorquez, L. N.; Martínez-Antonio, A.; Resendis-Antonio, O.; Lozada-Chávez, I.; Balderas-Martínez, Y. I.; Encarnación, S.; Collado-Vides, J. Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiology Reviews* **2009**, *33*, 133–151.

2. Lee, D. J.; Minchin, S. D.; Busby, S. J. W. Activating transcription in bacteria. *Annual Review of Microbiology* **2012**, *66*, 125–152.

3. Seshasayee, A. S. N.; Sivaraman, K.; Luscombe, N. M. An overview of prokaryotic transcription factors : a summary of function and occurrence in bacterial genomes. *Sub-cellular biochemistry* **2011**, *52*, 7–23.

4. Holmqvist, E.; Vogel, J. RNA-binding proteins in bacteria. *Nature Reviews. Microbiology* **2018**, *16*, 601–615.

5. Updegrove, T. B.; Zhang, A.; Storz, G. Hfq: the flexible RNA matchmaker. *Current Opinion in Microbiology* **2016**, *30*, 133–138.

6. Helder, S.; Blythe, A. J.; Bond, C. S.; Mackay, J. P. Determinants of affinity and specificity in RNA-binding proteins. *Current Opinion in Structural Biology* **2016**, *38*, 83–91.

7. Editors, P. M. Antimicrobial resistance: is the world unprepared? *PLoS Medicine* **2016**, *13*, e1002130.

8. Sommer, F.; Anderson, J. M.; Bharti, R.; Raes, J.; Rosenstiel, P. The resilience of the intestinal microbiota influences health and disease. *Nature Reviews. Microbiology* **2017**, *15*, 630–638.

9. Chevrette, M. G.; Carlson, C. M.; Ortega, H. E.; Thomas, C.; Ananiev, G. E.; Barns, K. J.; Book, A. J.; Cagnazzo, J.; Carlos, C.; Flanigan, W.; Grubbs, K. J.; Horn, H. A.; Hoffmann, F. M.; Klassen, J. L.; Knack, J. J.; Lewin, G. R.; McDonald, B. R.; Muller, L.; Melo, W. G. P.; Pinto-Tomás, A. A.; Schmitz, A.; Wendt-Pienkowski, E.; Wildman, S.; Zhao, M.; Zhang, F.; Bugni, T. S.; Andes, D. R.; Pupo, M. T.; Currie, C. R. The antimicrobial potential of Streptomyces from insect microbiomes. *Nature Communications* **2019**, *10*, 516.

10. Mizuno, T.; Chou, M. Y.; Inouye, M. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81*, 1966–1970.

11. Andersen, J.; Delihas, N.; Ikenaka, K.; Green, P. J.; Pines, O.; Ilercil, O.; Inouye, M. The isolation and characterization of RNA coded by the micF gene in Escherichia coli. *Nucleic Acids Research* **1987**, *15*, 2089–2101.

12. Andersen, J.; Forst, S. A.; Zhao, K.; Inouye, M.; Delihas, N. The function of micF RNA. micF RNA is a major factor in the thermal regulation of OmpF protein in Escherichia coli. *The Journal of Biological Chemistry* **1989**, *264*, 17961–17970.

13. Inouye, M.; Delihas, N. Small RNAs in the prokaryotes: a growing list of diverse roles. *Cell* **1988**, *53*, 5–7.

14. Delihas, N. Discovery and characterization of the first non-coding RNA that regulates gene expression, micF RNA: A historical perspective. *World journal of biological chemistry* **2015**, *6*, 272–280.

15. Storz, G.; Vogel, J.; Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Molecular Cell* **2011**, *43*, 880–891.

16. Scaria, V.; Hariharan, M.; Maiti, S.; Pillai, B.; Brahmachari, S. K. Host-virus interaction: a new role for microRNAs. *Retrovirology* **2006**, *3*, 68.

17. Fu, X.-D. Non-coding RNA: a new frontier in regulatory biology. *National science review* **2014**, *1*, 190–204.

18. Fatica, A.; Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews. Genetics* **2014**, *15*, 7–21.

19. Storz, G.; Altuvia, S.; Wassarman, K. M. An abundance of RNA regulators. *Annual Review of Biochemistry* **2005**, *74*, 199–217.

20. Statello, L.; Guo, C.-J.; Chen, L.-L.; Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews. Molecular Cell Biology* **2021**, *22*, 96–118.

21. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **2004**, *116*, 281–297.

22. Kiss, T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **2002**, *109*, 145–148.

23. Argaman, L.; Hershberg, R.; Vogel, J.; Bejerano, G.; Wagner, E. G.; Margalit, H.; Altuvia, S. Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. *Current Biology* **2001**, *11*, 941–950.

24. Rivas, E.; Klein, R. J.; Jones, T. A.; Eddy, S. R. Computational identification of noncoding RNAs in E. coli by comparative genomics. *Current Biology* **2001**, *11*, 1369–1373.

25. Wassarman, K. M.; Repoila, F.; Rosenow, C.; Storz, G.; Gottesman, S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes & Development* **2001**, *15*, 1637–1651.

26. Chao, Y.; Papenfort, K.; Reinhardt, R.; Sharma, C. M.; Vogel, J. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *The EMBO Journal* **2012**, *31*, 4005–4019.

27. Sittka, A.; Lucchini, S.; Papenfort, K.; Sharma, C. M.; Rolle, K.; Binnewies, T. T.; Hinton, J. C. D.; Vogel, J. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genetics* **2008**, *4*, e1000163.

28. Sharma, C. M.; Vogel, J. Differential RNA-seq: the approach behind and the biological insight gained. *Current Opinion in Microbiology* **2014**, *19*, 97–105.

29. Zhang, A.; Wassarman, K. M.; Rosenow, C.; Tjaden, B. C.; Storz, G.; Gottesman, S. Global analysis of small RNA and mRNA targets of Hfq. *Molecular Microbiology* **2003**, *50*, 1111–1124.

30. Sharma, C. M.; Vogel, J. Experimental approaches for the discovery and characterization of regulatory small RNA. *Current Opinion in Microbiology* **2009**, *12*, 536–546.

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

32. Georg, J.; Hess, W. R. cis-antisense RNA, another level of gene regulation in bacteria. *Microbiology and Molecular Biology Reviews* **2011**, *75*, 286–300.

33. Lloréns-Rico, V.; Cano, J.; Kamminga, T.; Gil, R.; Latorre, A.; Chen, W.-H.; Bork, P.; Glass, J. I.; Serrano, L.; Lluch-Senar, M. Bacterial antisense RNAs are mainly the product of transcriptional noise. *Science Advances* **2016**, *2*, e1501363.

34. Lasa, I.; Toledo-Arana, A.; Gingeras, T. R. An effort to make sense of antisense transcription in bacteria. *RNA Biology* **2012**, *9*, 1039–1044.

35. Vogel, J.; Wagner, E. G. H. Target identification of small noncoding RNAs in bacteria. *Current Opinion in Microbiology* **2007**, *10*, 262–270.

36. Brosse, A.; Guillier, M. Bacterial small rnas in mixed regulatory networks. *Microbiology spectrum* **2018**, *6*.

37. Guillier, M.; Gottesman, S. The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Research* **2008**, *36*, 6781–6794.

38. Fei, J.; Singh, D.; Zhang, Q.; Park, S.; Balasubramanian, D.; Golding, I.; Vanderpool, C. K.; Ha, T. RNA biochemistry. Determination of in vivo target search kinetics of regulatory noncoding RNA. *Science* **2015**, *347*, 1371–1374.

39. Bobrovskyy, M.; Azam, M. S.; Frandsen, J. K.; Zhang, J.; Poddar, A.; Ma, X.; Henkin, T. M.; Ha, T.; Vanderpool, C. K. Determinants of target prioritization and regulatory hierarchy for the bacterial small RNA SgrS. *Molecular Microbiology* **2019**, *112*, 1199–1218.

40. Wright, P. R.; Georg, J.; Mann, M.; Sorescu, D. A.; Richter, A. S.; Lott, S.; Kleinkauf, R.; Hess, W. R.; Backofen, R. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Research* **2014**, *42*, W119–23.

41. Busch, A.; Richter, A. S.; Backofen, R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **2008**, *24*, 2849–2856.

42. Richter, A. S.; Schleberger, C.; Backofen, R.; Steglich, C. Seed-based INTARNA prediction combined with GFP-reporter system identifies mRNA targets of the small RNA Yfr1. *Bioinformatics* **2010**, *26*, 1–5.

43. Papenfort, K.; Bouvier, M.; Mika, F.; Sharma, C. M.; Vogel, J. Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 20435–20440.

44. Beisel, C. L.; Storz, G. Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiology Reviews* **2010**, *34*, 866–882.

45. Massé, E.; Escorcia, F. E.; Gottesman, S. Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes & Development* **2003**, *17*, 2374–2383.

46. Prévost, K.; Desnoyers, G.; Jacques, J.-F.; Lavoie, F.; Massé, E. Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes & Development* **2011**, *25*, 385–396.

47. Guillier, M.; Gottesman, S. Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. *Molecular Microbiology* **2006**, *59*, 231–247.

48. Papenfort, K.; Pfeiffer, V.; Mika, F.; Lucchini, S.; Hinton, J. C. D.; Vogel, J. SigmaE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. *Molecular Microbiology* **2006**, *62*, 1674–1688.

49. Hammann, P.; Parmentier, D.; Cerciat, M.; Reimeg, J.; Helfer, A.-C.; Boisset, S.; Guillier, M.; Vandenesch, F.; Wagner, E. G. H.; Romby, P.; Fechter, P. A method to map changes in bacterial surface composition induced by regulatory RNAs in Escherichia coli and Staphylococcus aureus. *Biochimie* **2014**, *106*, 175–179.

50. Melamed, S.; Peer, A.; Faigenbaum-Romm, R.; Gatt, Y. E.; Reiss, N.; Bar, A.; Altuvia, Y.; Argaman, L.; Margalit, H. Global Mapping of Small RNA-Target Interactions in Bacteria. *Molecular Cell* **2016**, *63*, 884–897.

51. Bandyra, K. J.; Said, N.; Pfeiffer, V.; Górna, M. W.; Vogel, J.; Luisi, B. F. The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Molecular Cell* **2012**, *47*, 943–953.

52. Sobrero, P.; Valverde, C. The bacterial protein Hfq: much more than a mere RNA-binding factor. *Critical reviews in microbiology* **2012**, *38*, 276–299.

53. Valentin-Hansen, P.; Eriksen, M.; Udesen, C. The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Molecular Microbiology* **2004**, *51*, 1525–1533.

54. M, T.; Franch, T.; H, P.; Keene, D. R.; Bächinger, H. P.; Brennan, R. G.; Valentin-Hansen, P. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Molecular Cell* **2002**, *9*, 23–30.

55. Wagner, E. G. H.; Romby, P. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Advances in genetics* **2015**, *90*, 133–208.

56. Mohanty, B. K.; Maples, V. F.; Kushner, S. R. The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. *Molecular Microbiology* **2004**, *54*, 905–920.

57. Vytvytska, O.; Moll, I.; Kaberdin, V. R.; von Gabain, A.; Bläsi, U. Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. *Genes & Development* **2000**, *14*, 1109–1118.

58. Morita, T.; Maki, K.; Aiba, H. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes & Development* **2005**, *19*, 2176–2186.

59. Miczak, A.; Kaberdin, V. R.; Wei, C. L.; Lin-Chao, S. Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 3865–3869.

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

61. Py, B.; Higgins, C. F.; Krisch, H. M.; Carpousis, A. J. A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature* **1996**, *381*, 169–172.

62. Azam, M. S.; Vanderpool, C. K. Translational regulation by bacterial small RNAs via an unusual Hfq-dependent mechanism. *Nucleic Acids Research* **2018**, *46*, 2585–2599.

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

64. Massé, E.; Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 4620–4625.

65. Vanderpool, C. K.; Gottesman, S. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Molecular Microbiology* **2004**, *54*, 1076–1089.

66. Hantke, K. Iron and metal regulation in bacteria. *Current Opinion in Microbiology* **2001**, *4*, 172–177.

67. Rice, J. B.; Vanderpool, C. K. The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes. *Nucleic Acids Research* **2011**, *39*, 3806–3819.

68. Bobrovskyy, M.; Vanderpool, C. K. The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria. *Frontiers in cellular and infection microbiology* **2014**, *4*, 61.

69. Bobrovskyy, M.; Vanderpool, C. K. Diverse mechanisms of post-transcriptional repression by the small RNA regulator of glucose-phosphate stress. *Molecular Microbiology* **2016**, *99*, 254–273.

70. Papenfort, K.; Sun, Y.; Miyakoshi, M.; Vanderpool, C. K.; Vogel, J. Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* **2013**, *153*, 426–437.

71. Holmqvist, E.; Unoson, C.; Reimeg, J.; Wagner, E. G. H. A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. *Molecular Microbiology* **2012**, *84*, 414–427.

72. Corcoran, C. P.; Podkaminski, D.; Papenfort, K.; Urban, J. H.; Hinton, J. C. D.; Vogel, J. Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Molecular Microbiology* **2012**, *84*, 428–445.

73. Kawamoto, H.; Koide, Y.; Morita, T.; Aiba, H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Molecular Microbiology* **2006**, *61*, 1013–1022.

74. Majdalani, N.; Chen, S.; Murrow, J.; St John, K.; Gottesman, S. Regulation of RpoS by a novel small RNA: the characterization of RprA. *Molecular Microbiology* **2001**, *39*, 1382–1394.

75. Majdalani, N.; Hernandez, D.; Gottesman, S. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Molecular Microbiology* **2002**, *46*, 813–826.

76. Majdalani, N.; Cunning, C.; Sledjeski, D.; Elliott, T.; Gottesman, S. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 12462–12467.

77. Morita, T.; Mochizuki, Y.; Aiba, H. Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 4858–4863.

78. Fröhlich, K. S.; Papenfort, K.; Fekete, A.; Vogel, J. A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *The EMBO Journal* **2013**, *32*, 2963–2979.

79. Bossi, L.; Schwartz, A.; Guillemardet, B.; Boudvillain, M.; Figueroa-Bossi, N. A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes & Development* **2012**, *26*, 1864–1873.

80. Arbel-Goren, R.; Tal, A.; Friedlander, T.; Meshner, S.; Costantino, N.; Court, D. L.; Stavans, J. Effects of post-transcriptional regulation on phenotypic noise in Escherichia coli. *Nucleic Acids Research* **2013**, *41*, 4825–4834.

81. Levine, E.; Zhang, Z.; Kuhlman, T.; Hwa, T. Quantitative characteristics of gene regulation by small RNA. *PLoS Biology* **2007**, *5*, e229.

82. Levine, E.; Hwa, T. Small RNAs establish gene expression thresholds. *Current Opinion in Microbiology* **2008**, *11*, 574–579.

83. Shimoni, Y.; Friedlander, G.; Hetzroni, G.; Niv, G.; Altuvia, S.; Biham, O.; Margalit, H. Regulation of gene expression by small non-coding RNAs: a quantitative view. *Molecular Systems Biology* **2007**, *3*, 138.

84. Mehta, P.; Goyal, S.; Wingreen, N. S. A quantitative comparison of sRNA-based and protein-based gene regulation. *Molecular Systems Biology* **2008**, *4*, 221.

85. Hussein, R.; Lim, H. N. Direct comparison of small RNA and transcription factor signaling. *Nucleic Acids Research* **2012**, *40*, 7269–7279.

86. Caldelari, I.; Chao, Y.; Romby, P.; Vogel, J. RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harbor perspectives in medicine* **2013**, *3*, a010298.

87. Felden, B.; Cattoir, V. Bacterial Adaptation to Antibiotics through Regulatory RNAs. *Antimicrobial Agents and Chemotherapy* **2018**, *62*.

88. Gerdes, K.; Wagner, E. G. H. RNA antitoxins. *Current Opinion in Microbiology* **2007**, *10*, 117–124.

89. Holmqvist, E.; Wagner, E. G. H. Impact of bacterial sRNAs in stress responses. *Biochemical Society Transactions* **2017**, *45*, 1203–1212.

90. Mika, F.; Hengge, R. Small Regulatory RNAs in the Control of Motility and Biofilm Formation in E. coli and Salmonella. *International Journal of Molecular Sciences* **2013**, *14*, 4560–4579.

91. Gottesman, S.; Storz, G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspectives in Biology* **2011**, *3*.

92. Carrier, M.-C.; Lalaouna, D.; Massé, E. Broadening the definition of bacterial small rnas: characteristics and mechanisms of action. *Annual Review of Microbiology* **2018**, *72*, 141–161.

93. Arnold, T. E.; Yu, J.; Belasco, J. G. mRNA stabilization by the ompA 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA (New York)* **1998**, *4*, 319–330.

94. Braun, F.; Le Derout, J.; Régnier, P. Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of Escherichia coli. *The EMBO Journal* **1998**, *17*, 4790–4797.

95. Yarchuk, O.; Jacques, N.; Guillerez, J.; Dreyfus, M. Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *Journal of Molecular Biology* **1992**, *226*, 581–596.

96. Nitzan, M.; Rehani, R.; Margalit, H. Integration of bacterial small rnas in regulatory networks. *Annual review of biophysics* **2017**, *46*, 131–148.

97. Papenfort, K.; Vogel, J. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Research in Microbiology* **2009**, *160*, 278–287.

98. Jost, D.; Nowojewski, A.; Levine, E. Regulating the many to benefit the few: role of weak small RNA targets. *Biophysical Journal* **2013**, *104*, 1773–1782.

99. Fang, F. C.; Frawley, E. R.; Tapscott, T.; Vázquez-Torres, A. Bacterial Stress Responses during Host Infection. *Cell Host & Microbe* **2016**, *20*, 133–143.

100. Faigenbaum-Romm, R.; Reich, A.; Gatt, Y. E.; Barsheshet, M.; Argaman, L.; Margalit, H. Hierarchy in hfq chaperon occupancy of small RNA targets plays a major role in their regulation. *Cell reports* **2020**, *30*, 3127–3138.e6.

101. Richards, G. R.; Patel, M. V.; Lloyd, C. R.; Vanderpool, C. K. Depletion of glycolytic intermediates plays a key role in glucose-phosphate stress in Escherichia coli. *Journal of Bacteriology* **2013**, *195*, 4816–4825.

102. Vanderpool, C. K.; Gottesman, S. The novel transcription factor SgrR coordinates the response to glucose-phosphate stress. *Journal of Bacteriology* **2007**, *189*, 2238–2248.

103. Rice, J. B.; Balasubramanian, D.; Vanderpool, C. K. Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, E2691–8.

104. Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* **2006**, *24*, 79–88.

105. Reyer, M. A.; McLean, E. L.; Chennakesavalu, S.; Fei, J. An automated image analysis method for segmenting fluorescent bacteria in three dimensions. *Biochemistry* **2018**, *57*, 209–215.

106. Fender, A.; Elf, J.; Hampel, K.; Zimmermann, B.; Wagner, E. G. H. RNAs actively cycle on the Sm-like protein Hfq. *Genes & Development* **2010**, *24*, 2621–2626.

107. Maurizi, M. R. Proteases and protein degradation inEscherichia coli. *Experientia* **1992**, *48*, 178–201.

108. Urban, J. H.; Vogel, J. Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Research* **2007**, *35*, 1018–1037.

109. Bandyra, K. J.; Wandzik, J. M.; Luisi, B. F. Substrate recognition and autoinhibition in the central ribonuclease rnase E. *Molecular Cell* **2018**, *72*, 275–285.e4.

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

111. Mohanty, B. K.; Kushner, S. R. Regulation of mRNA Decay in Bacteria. *Annual Review of Microbiology* **2016**, *70*, 25–44.

112. Morita, T.; Kawamoto, H.; Mizota, T.; Inada, T.; Aiba, H. Enolase in the RNA degradosome plays a crucial role in the rapid decay of glucose transporter mRNA in the response to phosphosugar stress in Escherichia coli. *Molecular Microbiology* **2004**, *54*, 1063–1075.

113. Sedlyarova, N.; Shamovsky, I.; Bharati, B. K.; Epshtein, V.; Chen, J.; Gottesman, S.; Schroeder, R.; Nudler, E. sRNA-Mediated Control of Transcription Termination in E. coli. *Cell* **2016**, *167*, 111–121.e13.

114. Baek, Y. M.; Jang, K.-J.; Lee, H.; Yoon, S.; Baek, A.; Lee, K.; Kim, D.-E. The bacterial endoribonuclease RNase E can cleave RNA in the absence of the RNA chaperone Hfq. *The Journal of Biological Chemistry* **2019**, *294*, 16465–16478.

115. Jiang, X.; Belasco, J. G. Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9211–9216.

116. Mackie, G. A. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **1998**, *395*, 720–723.

117. Cho, H.; Park, H.-g.; Zhang, X.; Riba, I.; Gaskell, S. J.; Widger, W. R.; Kohn, H. Design, Syntheses, and Evaluations of Bicyclomycin-Based Rho Inactivators. *The Journal of Organic Chemistry* **1997**, *62*, 5432–5440.

118. Kohn, H.; Widger, W. The molecular basis for the mode of action of bicyclomycin. *Current drug targets. Infectious disorders* **2005**, *5*, 273–295.

119. Kass, R. E.; Wasserman, L. A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association* **1995**, *90*, 928–934.

120. Geissmann, T. A.; Touati, D. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *The EMBO Journal* **2004**, *23*, 396–405.

121. Vecerek, B.; Moll, I.; Afonyushkin, T.; Kaberdin, V.; Bläsi, U. Interaction of the RNA chaperone Hfq with mRNAs: direct and indirect roles of Hfq in iron metabolism of Escherichia coli. *Molecular Microbiology* **2003**, *50*, 897–909.

122. Arbel-Goren, R.; Tal, A.; Parasar, B.; Dym, A.; Costantino, N.; Muñoz-García, J.; Court, D. L.; Stavans, J. Transcript degradation and noise of small RNA-controlled genes in a switch activated network in Escherichia coli. *Nucleic Acids Research* **2016**, *44*, 6707–6720.

123. Mitarai, N.; Benjamin, J.-A. M.; Krishna, S.; Semsey, S.; Csiszovszki, Z.; Massé, E.; Sneppen, K. Dynamic features of gene expression control by small regulatory RNAs. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 10655–10659.

124. Schmiedel, J. M.; Axmann, I. M.; Legewie, S. Multi-target regulation by small RNAs synchronizes gene expression thresholds and may enhance ultrasensitive behavior. *Plos One* **2012**, *7*, e42296.

125. Fei, J.; Sharma, C. M. RNA localization in bacteria. *Microbiology spectrum* **2018**, *6*.

126. Sheng, H.; Stauffer, W. T.; Hussein, R.; Lin, C.; Lim, H. N. Nucleoid and cytoplasmic localization of small RNAs in Escherichia coli. *Nucleic Acids Research* **2017**, *45*, 2919–2934.

127. Park, S.; Prevost, K.; Heideman, E. M.; Carrier, M.-C.; Reyer, M. A.; Liu, W.; Masse, E.; Fei, J. Dynamic interactions between the RNA chaperone Hfq, small regulatory RNAs and mRNAs in live bacterial cells. *BioRxiv* **2020**,

128. Persson, F.; Lindén, M.; Unoson, C.; Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature Methods* **2013**, *10*, 265–269.

129. Kambara, T. K.; Ramsey, K. M.; Dove, S. L. Pervasive targeting of nascent transcripts by hfq. *Cell reports* **2018**, *23*, 1543–1552.

130. Brantl, S.; Birch-Hirschfeld, E.; Behnke, D. RepR protein expression on plasmid pIP501 is controlled by an antisense RNA-mediated transcription attenuation mechanism. *Journal of Bacteriology* **1993**, *175*, 4052–4061.

131. Giangrossi, M.; Prosseda, G.; Tran, C. N.; Brandi, A.; Colonna, B.; Falconi, M. A novel antisense RNA regulates at transcriptional level the virulence gene icsA of Shigella flexneri. *Nucleic Acids Research* **2010**, *38*, 3362–3375.

132. Novick, R. P.; Iordanescu, S.; Projan, S. J.; Kornblum, J.; Edelman, I. pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. *Cell* **1989**, *59*, 395–404.

133. Tran, C. N.; Giangrossi, M.; Prosseda, G.; Brandi, A.; Di Martino, M. L.; Colonna, B.; Falconi, M. A multifactor regulatory circuit involving H-NS, VirF and an antisense RNA modulates transcription of the virulence gene icsA of Shigella flexneri. *Nucleic Acids Research* **2011**, *39*, 8122–8134.

134. Markham, N. R.; Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research* **2005**, *33*, W577–81.

135. Hoekzema, M.; Romilly, C.; Holmqvist, E.; Wagner, E. G. H. Hfq-dependent mRNA unfolding promotes sRNA-based inhibition of translation. *The EMBO Journal* **2019**, *38*.

136. Datta, S.; Costantino, N.; Court, D. L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **2006**, *379*, 109–115.

137. Park, S.; Bujnowska, M.; McLean, E. L.; Fei, J. Quantitative Super-Resolution Imaging of Small RNAs in Bacterial Cells. *Methods in Molecular Biology* **2018**, *1737*, 199–212.

138. Foreman-Mackey, D.; Hogg, D. W.; Lang, D.; Goodman, J. emcee : the MCMC hammer. *Publications of the Astronomical Society of the Pacific* **2013**, *125*, 306–312.

139. Goodman, J.; Weare, J. Ensemble samplers with affine invariance. *Communications in applied mathematics and computational science* **2010**, *5*, 65–80.

140. Petzold, L. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM Journal on Scientific and Statistical Computing* **1983**, *4*, 136–148.

141. Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Contributors, S. . SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **2020**, *17*, 261–272.

142. Young, R.; Bremer, H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *The Biochemical Journal* **1976**, *160*, 185–194.

143. Alon, U. *An introduction to systems biology: design principles of biological circuits*; Chapman and Hall/CRC: Second edition. |Boca Raton, Fla. : CRC Press, [2019], 2019.

144. Chen, J.; Gottesman, S. Hfq links translation repression to stress-induced mutagenesis in E. coli. *Genes & Development* **2017**, *31*, 1382–1395.

145. Desnoyers, G.; Massé, E. Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes & Development* **2012**, *26*, 726–739.

146. Ellis, M. J.; Trussler, R. S.; Haniford, D. B. Hfq binds directly to the ribosome-binding site of IS10 transposase mRNA to inhibit translation. *Molecular Microbiology* **2015**, *96*, 633–650.

147. Sonnleitner, E.; Bläsi, U. Regulation of Hfq by the RNA CrcZ in Pseudomonas aeruginosa carbon catabolite repression. *PLoS Genetics* **2014**, *10*, e1004440.

148. Sledjeski, D. D.; Whitman, C.; Zhang, A. Hfq is necessary for regulation by the untranslated RNA DsrA. *Journal of Bacteriology* **2001**, *183*, 1997–2005.

149. Zhang, A.; Wassarman, K. M.; Ortega, J.; Steven, A. C.; Storz, G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Molecular Cell* **2002**, *9*, 11–22.

150. Pfeiffer, V.; Papenfort, K.; Lucchini, S.; Hinton, J. C. D.; Vogel, J. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nature Structural & Molecular Biology* **2009**, *16*, 840–846.

151. Vogel, J.; Luisi, B. F. Hfq and its constellation of RNA. *Nature Reviews. Microbiology* **2011**, *9*, 578–589.

152. Afonyushkin, T.; Vecerek, B.; Moll, I.; Bläsi, U.; Kaberdin, V. R. Both RNase E and RNase III control the stability of sodB mRNA upon translational inhibition by the small regulatory RNA RyhB. *Nucleic Acids Research* **2005**, *33*, 1678–1689.

153. Ikeda, Y.; Yagi, M.; Morita, T.; Aiba, H. Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. *Molecular Microbiology* **2011**, *79*, 419–432.

154. Bruce, H. A.; Du, D.; Matak-Vinkovic, D.; Bandyra, K. J.; Broadhurst, R. W.; Martin, E.; Sobott, F.; Shkumatov, A. V.; Luisi, B. F. Analysis of the natively unstructured RNA/protein-recognition core in the Escherichia coli RNA degradosome and its interactions with regulatory RNA/Hfq complexes. *Nucleic Acids Research* **2018**, *46*, 387–402.

155. Sukhodolets, M. V.; Garges, S. Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry* **2003**, *42*, 8022–8034.

156. Kajitani, M.; Kato, A.; Wada, A.; Inokuchi, Y.; Ishihama, A. Regulation of the Escherichia coli hfq gene encoding the host factor for phage Q beta. *Journal of Bacteriology* **1994**, *176*, 531–534.

157. Butland, G.; Peregrín-Alvarez, J. M.; Li, J.; Yang, W.; Yang, X.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; Davey, M.; Parkinson, J.; Greenblatt, J.; Emili, A. Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* **2005**, *433*, 531–537.

158. Andrade, J. M.; Dos Santos, R. F.; Chelysheva, I.; Ignatova, Z.; Arraiano, C. M. The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity. *The EMBO Journal* **2018**, *37*.

159. Hüttenhofer, A.; Noller, H. F. Footprinting mRNA-ribosome complexes with chemical probes. *The EMBO Journal* **1994**, *13*, 3892–3901.

160. Udekwu, K. I.; Wagner, E. G. H. Sigma E controls biogenesis of the antisense RNA MicA. *Nucleic Acids Research* **2007**, *35*, 1279–1288.

161. McAdams, H. H.; Arkin, A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 814–819.

162. Deana, A.; Belasco, J. G. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes & Development* **2005**, *19*, 2526–2533.

163. Larsson, E.; Sander, C.; Marks, D. mRNA turnover rate limits siRNA and microRNA efficacy. *Molecular Systems Biology* **2010**, *6*, 433.

164. Pestka, S. Inhibitors of ribosome functions. *Annual Review of Biochemistry* **1971**, *40*, 697–710.

165. Weisblum, B.; Davies, J. Antibiotic inhibitors of the bacterial ribosome. *Bacteriological Reviews* **1968**, *32*, 493–528.

166. Schavemaker, P. E.; Śmigiel, W. M.; Poolman, B. Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome. *eLife* **2017**, *6*.

167. Roberts, E.; Stone, J. E.; Luthey-Schulten, Z. Lattice Microbes: high-performance stochastic simulation method for the reaction-diffusion master equation. *Journal of Computational Chemistry* **2013**, *34*, 245–255.

168. Bakshi, S.; Siryaporn, A.; Goulian, M.; Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology* **2012**, *85*, 21–38.

169. Volkov, I. L.; Lindén, M.; Aguirre Rivera, J.; Ieong, K.-W.; Metelev, M.; Elf, J.; Johansson, M. tRNA tracking for direct measurements of protein synthesis kinetics in live cells. *Nature Chemical Biology* **2018**, *14*, 618–626.

170. Golding, I.; Cox, E. C. RNA dynamics in live Escherichia coli cells. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 11310–11315.

171. Sanamrad, A.; Persson, F.; Lundius, E. G.; Fange, D.; Gynnå, A. H.; Elf, J. Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 11413–11418.

172. Taniguchi, Y.; Choi, P. J.; Li, G.-W.; Chen, H.; Babu, M.; Hearn, J.; Emili, A.; Xie, X. S. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **2010**, *329*, 533–538.

173. Bremer, H.; Dennis, P. P. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus* **2008**, *3*.

174. Bartholomäus, A.; Fedyunin, I.; Feist, P.; Sin, C.; Zhang, G.; Valleriani, A.; Ignatova, Z. Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* **2016**, *374*.

175. Ali Azam, T.; Iwata, A.; Nishimura, A.; Ueda, S.; Ishihama, A. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *Journal of Bacteriology* **1999**, *181*, 6361–6370.

176. Oeffinger, M.; Zenklusen, D. To the pore and through the pore: a story of mRNA export kinetics. *Biochimica et Biophysica Acta* **2012**, *1819*, 494–506.

177. Vera, M.; Biswas, J.; Senecal, A.; Singer, R. H.; Park, H. Y. Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. *Annual Review of Genetics* **2016**, *50*, 267–291.

178. Skylaki, S.; Hilsenbeck, O.; Schroeder, T. Challenges in long-term imaging and quantification of single-cell dynamics. *Nature Biotechnology* **2016**, *34*, 1137–1144.

179. Elowitz, M. B.; Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **2000**, *403*, 335–338.

180. So, L.-H.; Ghosh, A.; Zong, C.; Sepúlveda, L. A.; Segev, R.; Golding, I. General properties of transcriptional time series in Escherichia coli. *Nature Genetics* **2011**, *43*, 554–560.

181. Sadanandan, S. K.; Baltekin, O.; Magnusson, K. E. G.; Boucharin, A.; Ranefall, P.; Jalden, J.; Elf, J.; Wahlby, C. Segmentation and Track-Analysis in Time-Lapse Imaging of Bacteria. *IEEE journal of selected topics in signal processing* **2016**, *10*, 174–184.

182. Weiner, A.; Mellouk, N.; Lopez-Montero, N.; Chang, Y.-Y.; Souque, C.; Schmitt, C.; Enninga, J. Macropinosomes are key players in early shigella invasion and vacuolar escape in epithelial cells. *PLoS Pathogens* **2016**, *12*, e1005602.

183. Yan, J.; Sharo, A. G.; Stone, H. A.; Wingreen, N. S.; Bassler, B. L. Vibrio cholerae biofilm growth program and architecture revealed by single-cell live imaging. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, E5337–43.

184. Abràmoff, D. M. D. Image Processing with ImageJ. *Biophotonics International* **2004**, *11*, 36–42.

185. de Chaumont, F.; Dallongeville, S.; Chenouard, N.; Hervé, N.; Pop, S.; Provoost, T.; Meas-Yedid, V.; Pankajakshan, P.; Lecomte, T.; Le Montagner, Y.; Lagache, T.; Dufour, A.; Olivo-Marin, J.-C. Icy: an open bioimage informatics platform for extended reproducible research. *Nature Methods* **2012**, *9*, 690–696.

186. Carpenter, A. E.; Jones, T. R.; Lamprecht, M. R.; Clarke, C.; Kang, I. H.; Friman, O.; Guertin, D. A.; Chang, J. H.; Lindquist, R. A.; Moffat, J.; Golland, P.; Sabatini, D. M. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology* **2006**, *7*, R100.

187. Indhumathi, C.; Cai, Y. Y.; Guan, Y. Q.; Opas, M. An automatic segmentation algorithm for 3D cell cluster splitting using volumetric confocal images. *Journal of Microscopy* **2011**, *243*, 60–76.

188. Bai, X.; Sun, C.; Zhou, F. Touching cells splitting by using concave points and ellipse fitting. 2008 Digital Image Computing: Techniques and Applications. 2008; pp 271–278.

189. Song, H.; Zhao, Q.; Liu, Y. Splitting touching cells based on concave-point and improved watershed algorithms. *Frontiers of Computer Science* **2014**, *8*, 156–162.

190. Perona, P.; Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE transactions on pattern analysis and machine intelligence* **1990**, *12*, 629–639.

191. Lopes, D. S. Anisotropic Diffusion (Perona & Malik). 2007; https://www.mathworks.com/matlabcentral/fileexchange/14995-anisotropic-diffusion–perona—malik-.

192. Arce, S. H.; Wu, P.-H.; Tseng, Y. Fast and accurate automated cell boundary determination for fluorescence microscopy. *Scientific Reports* **2013**, *3*, 2266.

193. Bradley, D.; Roth, G. Adaptive Thresholding using the Integral Image. *Journal of Graphics Tools* **2007**, *12*, 13–21.

194. Motl, J. Bradley Local Image Thresholding. 2015; https://www.mathworks.com/matlabcentral/fileexchange/40854-bradley-local-image-thresholding.

195. Gal, O. fit_ellipse. 2003; https://www.mathworks.com/matlabcentral/fileexchange/3215-fit-ellipse.

196. Huttenlocher, D.; Klanderman, G.; Rucklidge, W. Comparing images using the Hausdorff distance. *IEEE transactions on pattern analysis and machine intelligence* **1993**, *15*, 850–863.

197. Savitzky, A.; Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* **1964**, *36*, 1627–1639.

198. Portal, R.; Dias, J.; de Sousa, L. Contact detection between convex superquadric surfaces. *Archive of Mechanical Engineering* **2010**, *LVII*.

199. Meijering, E. Cell segmentation: 50 years down the road [life sciences]. *IEEE signal processing magazine* **2012**, *29*, 140–145.

200. Westermann, A. J.; Förstner, K. U.; Amman, F.; Barquist, L.; Chao, Y.; Schulte, L. N.; Müller, L.; Reinhardt, R.; Stadler, P. F.; Vogel, J. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* **2016**, *529*, 496–501.

201. Weisstein, E. W. Superellipse – from Wolfram MathWorld. https://mathworld.wolfram.com/Superellipse.html.

202. Manley, S.; Gillette, J. M.; Patterson, G. H.; Shroff, H.; Hess, H. F.; Betzig, E.; Lippincott-Schwartz, J. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature Methods* **2008**, *5*, 155–157.

203. Wang, S.; Moffitt, J. R.; Dempsey, G. T.; Xie, X. S.; Zhuang, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 8452–8457.