DEFINING THE CONSTELLATION OF RNA ELEMENTS THAT INTERACT WITH

BACILLUS SUBTILIS HFQ

APPROVED BY SUPERVISORY COMMITTEE

Wade C. Winkler, Ph.D.

Russell Debose-Boyd, Ph.D.

Kevin H. Gardner, Ph.D.

Benjamin Tu, Ph.D.

DEFINING THE CONSTELLATION OF RNA ELEMENTS THAT ASSOCIATE WITH *BACILLUS SUBTILIS* HFQ

By

MICHAEL DAVID DAMBACH

DISSERTATION /THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2012

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my mother Karen and my father David, who pushed me to chase my dreams where ever they may lead. Their love and support was a guiding light through many dark times. For without them none of this would have been possible. I would also like to bestow my gratitude upon my sister Megan and my Brother Mark, their encouragement and advice was critical to my completion of this thesis.

I would like to extend my sincere thanks to past and present members of the Winkler lab; in particular Dr. Arati Ramesh, Dr. Irnov, and Dr. Catherine Wakeman. I have never worked in a more fun, stimulating, and enriching environment than the one I shared when we were all together. I wish you all the very best in your future endeavors.

In addition, I would like to thank my thesis committee- Dr. Russell Debose-Boyd, Dr. Kevin Gardner, and Dr. Benjamin Tu for taking the time to participate in my studies as well as offering critical advice and guidance through this process. I must extend my sincerest thanks and gratitude to Drs. Melanie Cobb and Steve McKnight. I cannot put into words the lessons that you have

bestowed on me. I would not be the scientist or the person I am today without your guidance and leadership and for that I am eternally grateful.

And finally I must give the ut most thanks and gratitude to my mentor Dr. Wade Winkler. Wade took me in and nurtured me when I was at my lowest point during this process. His unconditional support and encouragement fostered and stoked my scientific passion and reminded me why we love this endeavor we call science. I can honestly say I'm a better person for knowing him and his family. His teachings transcend science and I will carry them with me on the journey through life.

DEFINING THE CONSTELLATION OF RNA ELEMENTS THAT ASSOCIATE WITH

BACILLUS SUBTILIS HFQ

Michael David Dambach, Ph.D

The University of Texas Southwestern Medical Center at Dallas, 2012

Wade C. Winkler, Ph.D

Bacteria utilize a wide variety of genetic regulatory strategies in order to sense and respond to various environmental fluctuations in nutrient availability, temperature, salinity, and oxygen among others. As such, bacterial species have evolved highly coordinated and tightly regulated systems as a means of

efficiently responding to potentially deleterious changes in environmental conditions. Traditionally DNA binding transcriptions factors were thought to be the primary means by which the cell executes a selective genetic response. However, the advent of microarray and next generation sequencing platforms, coupled with the wealth of sequenced genomes and powerful bioinformatics have revealed that RNA mediated post transcriptional gene regulation is wide spread in bacterial species and may in fact rival protein based regulatory systems in scope and breadth.

RNA mediated post transcriptional gene regulation is broadly divided into two categories-those in which the RNA element is transcribed with the mRNA it regulates (*cis*-acting regulatory RNAs) or those which are transcribed independently from the gene that they regulate (*trans*-acting regulatory RNAs). In general *cis*-acting RNA elements are embedded within a 5' UTR of a gene that they regulate and may or may not require a protein cofactor to execute genetic regulation. Whereas, *trans*-acting regulatory RNAs, also known as sRNAs, function via base pairing with their target mRNA and this usually requires the protein chaperone Hfq. Hfq mediated gene regulation is poorly understood in Gram-positive organism, thus I undertook studies of this protein in the model Gram-positive organism *Bacillus subtilis.* I used co-immunoprecipitation and

deep-sequencing to define the suite of RNA elements that associate with this regulatory protein. In addition I performed global transcriptomic studies on an Hfq deletion mutant in order to identify genes that are regulated via Hfq. These studies identified sRNAs that may be involved with sporulation. This led me to analyze the transcriptomic profile of *Bacillus subtilis* spores in an attempt to identify new sRNA regulators.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Adenosine
asRNA	antisense RNA
bp	Basepair
BME	β-Mercaptoethanol
BSA	Bovie Serum Albumin
CDS	Coding Sequence
DNA	Deoxyribose Nucleic Acid
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium Bromide
Fur	Ferric uptake regulator
G	Guanosine
GlcN6P	Glucosamine-6-phosphate
IPTG	Isopropyl-β-thiogalactopyranoside

Hfq	Host Factor Required for Q-beta Replication
kD	Kilo Dalton
LSm	SM like
mRNA	Messenger RNA
nM	Nanomolar
oligo	Oligonucleotide Triphosphates
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
RBS	Ribosome Binding Site
RNA	Ribose Nucleic Acid
RNAP	RNA Polymerase
RNase	Ribonuclease
RNA-seq	RNA sequencing
rRNA	Ribosomal RNA

RT	Reverse Transcriptase
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulfate
sRNA	Small RNA
T	Thymidine
TA	Toxin Antitoxin
TBAB	Tryptone Blood Agar Base
tRNA	Transfer RNA
TSS	Transcription Start Site
U	Uridine
YFP	Yellow Fluorescent Protein

CHAPTER ONE

Introduction and Literature Review

Introduction

Since the initial description of the *lac* operon over 50 years ago, genetic control in eubacteria has been presumed to result mostly from protein factors as the agents of regulatory function. However, discoveries over the past 15 years have illuminated an ever-expanding role for noncoding RNAs in the control of bacterial gene expression. These RNAs employ a variety of mechanisms to exert their regulatory functions but can be broadly segregated into two distinct classes: those that are co-transcribed (*cis*-acting) or transcribed separately (*trans*-acting) of the mRNA or protein they regulate. *Trans*-acting RNAs, also known as, small noncoding regulatory RNAs (sRNAs), consist of independently transcribed RNAs that interact with target mRNAs through base-pairing interactions (Repoila and Darfeuille 2009, Vogel and Papenfort 2006). Other sRNAs affect gene expression by associating with specific RNA-binding proteins or with RNA polymerase (Babitzke and Romeo 2007, Storz, Opdyke and

Wassarman 2006). By contrast, regulatory RNAs can also be enslaved within the target transcript that they regulate (*cis*-acting regulatory RNAs). In bacteria, examples of the latter category are largely located within the 5′ untranslated region (5′ UTR) of the transcripts that they regulate. These *cis*-acting regulatory elements individually respond to many different types of effector signals, including fluctuations in metabolites, second messengers, temperature, RNA-binding proteins and transfer RNAs (tRNAs) (Figure 1-1B)(Gottesman et al. 2006, Irnov, Kertsburg and Winkler 2006, Narberhaus, Waldminghaus and Chowdhury 2006).

Cis-encoded Regulatory RNAs

Cis-acting regulatory RNAs are signal responsive un-translated elements most often found embedded in the 5′UTR of mRNAs whose expression they control. These RNAs are capable of sensing a wide variety of substrates within the cell including proteins, tRNA, ions, and metabolites (Dambach and Winkler 2009a, Breaker 2012, Henkin 2008). Regulatory RNAs which sense ions and metabolites are particularly wide spread and are collectively known as Riboswitches (Dann et al. 2007, Dambach and Winkler 2009b, Baker et al. 2012). Riboswitches are modular RNA elements consisting of a conserved structured ligand sensing domain known as an aptamer, followed by a variable “expression

platform" which controls gene expression through formation of alternative RNA secondary structures (Dambach and Winkler 2009a, Garst, Edwards and Batey 2011, Breaker 2012). Ligand sensing by riboswitch aptamer domains are exquisitely selective and sensitive. For example, the purine riboswitches exhibit over 10,000 fold discrimination between adenine and guanine with an apparent K_D in the low nanomolar (nM) range (Mandal et al. 2003). The fact that these high affinity RNA based receptors can achieve such sophisticated discrimination between very similar chemical compounds is remarkable considering that RNA is constructed from only four different monomers. As a result, riboswitch aptamer domain sequence and structure are highly conserved over great evolutionary distances (Gelfand et al. 1999, Nahvi, Barrick and Breaker 2004, Sudarsan, Barrick and Breaker 2003).

Control of Transciption Termination

In contrast to the aptamer component of riboswitches the expression platform, which executes gene regulatory function in response to ligand binding, exhibits a great degree of variability. The expression platform functions as an allosteric binary on/ off switch leading to structural rearrangements which either activate or repress downstream gene expression in a signal responsive manner. Organisms have evolved various mechanisms to control gene expression upon

metabolite induced conformational switching. One of the most common regulatory outputs in bacteria involves control of transcription termination. In these instances ligand induced structural rearrangements result in the formation of an intrinsic transcription terminator, which is a GC rich hairpin directly followed by a poly-Uridine (U) tract. The poly-U tract destabilizes RNA polymerase due to the weak hydrogen bonding potential between adenine and uracil resulting in dissociation of RNA polymerase thus preventing transcription of the downstream open reading frame (ORF) (Figure 1-2A). Conversely, metabolite binding can also lead to the formation of an antiterminator stem loop which is an alternative helical structure that sequesters a portion of the terminator helix resulting in subsequent transcription of the downstream ORF. Consequently ligand induced formation of intrinsic terminators are "off switches" where as structural rearrangements that favor formation of antiterminator helices are genetic "on switches" (Henkin 2008, Dambach and Winkler 2009a, Garst et al. 2011, Breaker 2012).

Control of Translation Intiation

Another commonly used genetic control strategy employed by riboswitches occurs at the level of translation initiation. In this scenario, signal mediated structural rearrangements result in the exposure of an anti-Shine-

Dalgarno sequence which forms a helix with its cognate ribosome binding site (RBS), the effect of which blocks ribosome binding and prevents translation. As with transcription termination/ anti-termination mechanisms, ligand binding to the aptamer domain can result in disruption of an inhibitory conformation, in this case exposure of the RBS leading to translation of the mRNA (Figure 1-2b) (Henkin 2008, Dambach and Winkler 2009a, Garst et al. 2011, Breaker 2012). It's interesting to note that there is a phylogenetic dichotomy in regards to utilization of transcriptional versus translational control strategies in eubacteria. For reasons that are not intuitively obvious Gram-positive bacteria prefer to utilize transcription attenuation based mechanisms where as Gram-negative organisms tend to employ translation based regulatory strategies.

Control of Bacterial mRNA Stability

Though less prevalent, riboswitches can also utilize catalytic RNAs or ribozymes, in order to execute genetic regulation. The first documented case of a *cis*-acting signal responsive ribozyme was observed in the 5'UTR of the *glmS* gene of certain *Bacillus* species (Figure 1-2C) (Winkler et al. 2004). The *glmS* gene codes for glucosamine-6-phosphate (GlcN6P) synthase which catalyzes the formation of GlcN6P and glutamate from fructose-6-phosphate and glutamine (Winkler et al. 2004). Under conditions of high GlcN6P the *glmS* gene is

repressed suggesting feedback inhibition. However, the *glmS* UTR lacked sequence elements characteristic of transcriptional or translational control mechanisms. It was subsequently demonstrated that the *glmS* riboswitch underwent autocatalytic cleavage *in vitro* in response to GlcN6P and that this sugar actually serves as a cofactor for the ribozyme (Winkler et al. 2004, Klein and Ferre-D'Amare 2006, Klein, Been and Ferre-D'Amare 2007). Ribozyme activation results in a 5′ self-cleavage products that contains a 2′, 3′-cyclic phosphate and a downstream 3′ cleavage product (corresponding to the *glmS* transcript) containing a 5′ hydroxyl group. In *Bacillus subtilis*, transcripts containing a 5′ hydroxyl are selectively degraded by the 5′-to-3′ exoribonuclease RNase J1 (Condon and Bechhofer 2011). Thus, self-cleavege by the *glmS* ribozyme in response to changes in intracellular Glc6NP levels results in an unstable mRNA product which is rapidly degraded (Fig 1-2C) (Collins et al. 2007, Winkler et al. 2004).

Control of Eukaryotic Splicing

One of the major differences in genetic organization between eubacteria and eukaryotes is the presence of introns within pre-mRNAs, which must be excised in order to yield functional transcripts. Alternative splicing is the process in which introns are excised and exons are separated and rejoined in different

combinations to produce alternative mRNA transcripts (Blencowe 2006). In most instances regulation of splicing is dependent on the interaction of RNA-binding proteins and pre-mRNA sequences proximal to splice sites. However, for certain fungal, plant, and algal species, alternative splicing of genes involved in thiamine pyrophosphate (TPP) biosynthesis is directly controlled by TPP riboswitches located within intronic regions (Figure 1-2D and E). For *Neurospora crassa*, a total of three genes are regulated thusly; two are repressed in response to TPP, while expression of the third gene is activated (Cheah et al. 2007, Wachter et al. 2007, Bocobza et al. 2007, Croft et al. 2007). Binding of TPP to these riboswitches alters availability of a nucleotide tract within the riboswitch that could otherwise form base-pairing interactions with a riboswitch-proximal splice site or branch point components. In this manner, availability of TPP controls whether or not an upstream open reading frame (uORF) is formed, the presence of which decreases expression of the downstream, main ORF. TPP also controls formation of an uORF for regulation of thiamine biosynthesis genes in the algal species, *Chlamydomonas reinhardti* and *Volvox carteri* (Croft et al. 2007). Finally, TPP riboswitches have also been found to control alternative splicing of plant biosynthesis genes (Figure 1-2E) (Wachter et al. 2007, Bocobza et al. 2007). However, these TPP riboswitches are located within the 3′ UTR of plant thiamine synthesis genes. Under low thiamine conditions a short oligonucleotide stretch

within the riboswitch is capable of base pairing to a complementary sequence that overlaps the 5′ splice site, thereby preventing splicing at that site. This particular splicing pattern then results in inclusion of an RNA processing site that leads to a shortened 3′ UTR, conditions that lead to mRNA stabilization and increased translation of the target gene. Upon an increase in TPP, the riboswitch oligonucleotide tract cannot pair with the 5′ splice site and the altered splicing pattern leads to a loss of the 3′ UTR mRNA processing site and, ultimately, destabilization of the overall transcript. At the heart of these mechanisms is the ligand-mediated influence on mutually exclusive base-paired regions, conceptually common among all riboswitches. Therefore, discovery of these mechanisms would suggest there is no reason not to anticipate that other riboswitches could control eukaryotic alternative splicing and stability.

Trans-Acting Regulatory RNAs

In addition to *cis*-acting regulatory RNAs, eubacteria also depend on *trans*-acting regulatory factors, also known as small RNAs (sRNA), to regulate gene expression. sRNA regulators are short transcripts (~50-500 bases in length) that act on independently expressed targets primarily through base-pairing or via modifying protein activity (Babitzke and Romeo 2007, Wassarman 2007, Storz, Vogel and Wassarman 2011, Gottesman and Storz 2011a). The most

extensively studied sRNAs to date are known as *trans*-acting sRNAs which regulate mRNA through short imperfect base pairing, similar to micro-RNA based regulation in eukaryotes (Figure 1-3C) (Gottesman and Storz 2011a, Bartel 2009). sRNA mediated gene regulation has been best characterized in the Gram-negative enteric bacteria *E. coli* and *Salmonella* where over 100 putative sRNAs have been catalogued thus far (Gottesman and Storz 2011a). These regulatory factors are transcribed in response to a variety of stimuli such as iron limitation, outer membrane perturbation, oxidative stress and intracellular sugar-phosphate levels; resulting in the regulation of diverse cellular process such as quorum sensing, virulence, carbon metabolism, iron homeostasis, and biofilm formation (Waters and Storz 2009, Richards and Vanderpool 2011, Gottesman and Storz 2011a).

At the heart of this regulation is the ability of sRNAs to form regulatory networks. Many sRNAs are known to regulate multiple mRNAs involved in a certain physiological process or response thus forming a regulon. One example of this phenomenon is the RybB system of *E.coli* which is transcribed in response to iron limitation within the cell. Under conditions of low iron the transcriptional repressor ferric uptake regulator (Fur) disassociates from the RybB promoter allowing transcription of this sRNA. RybB then targets mRNAs of non-essential

iron containing enzymes via base pairing interactions that require the RNA binding protein Hfq, resulting in the subsequent down-regulation of these proteins (Masse, Vanderpool and Gottesman 2005, Jacques et al. 2006, Masse et al. 2007). At the same time RybB positively regulates, in a direct manner the *shiA* gene which is involved in the production of the iron-chelating siderophore, enterobactin (Prevost et al. 2007, Salvail et al. 2010).

sRNAs can also produce wide-spread changes in gene expression through indirect methods. Many sRNAs directly modulate the expression of key transcription factors within the cell that result in significant changes in global gene expression, as illustrated by the RpoS system in *E. coli.* The RpoS gene codes for an alternative sigma factor that induces a stress resistant state when activated. In fact, it's estimated that RpoS controls nearly 10% of the *E. coli* genome either directly or indirectly (Weber et al. 2005). RpoS is positively regulated at the post-transcriptional level by the action of three independent sRNAs: DsrA, RprA and ArcZ. All three of these sRNAs are under the control of different promoters and are transcribed in response to different cellular stresses resulting in enhanced translation of the RpoS protein (Battesti, Majdalani and Gottesman 2011). Thus, the RpoS leader acts as platform for signal integration by sRNAs that sense different cellular stresses.

sRNA Mediated Translational Repression

Trans-acting small RNA based regulation operates primarily at the level of translation initiation. In general most sRNAs recognized sequences encompassing the Shine-Dalgarno or AUG start codon of the message they regulate, the effect of which masks the RBS and occludes ribosome binding and subsequent translation (Figure 1-3C). However, recent studies have demonstrated that sRNA mediated translational repression can occur through targeted base pairing interactions that occur distant from the RBS (Bouvier et al. 2008, Holmqvist et al. 2010).

Recent work in *E. coli* has established that two sRNAs, OmrA and OmrB are capable of negatively regulating the expression of the transcription factor csgD, and that this regulation occurs at the translational level. The authors subsequently showed that OmrA and OmrB operated by base pairing 70 nucleotides upstream of the translational start site within a highly structured 5'UTR and that regulation did not involve modulating access to the RBS (Holmqvist et al. 2010). Translational repression by sRNAs can also be mediated by base pairing within the coding regions of open reading frames. In *Salmonella* the sRNA RybB represses translation of the outer membrane protein ompN independent of interactions with the RBS or start codon. This occurs via base

pairing with a 15 base tract downstream of the AUG, suggesting a five codon window of translational mRNA repression by sRNAs is possible (Bouvier et al. 2008).

Consequently the majority of sRNA regulators act as repressors of their target transcripts. However, this is not absolute as illustrated by the *E. coli* stress responsive sigma factor *rpoS.* The major *rpoS* transcript within the cell originates 567 base pairs up stream of the translational start site, resulting in a long, highly structured 5'UTR (Battesti et al. 2011, McCullen et al. 2010). This long leader adopts a secondary structure that results in sequestration of its RBS within an inhibitory stem-loop structure that prevents ribosome loading and translation (McCullen et al. 2010). It was subsequently demonstrated that translation is stimulated through the action of two independent sRNAs, DsrA and RprA, respectively (Majdalani et al. 1998, Majdalani, Hernandez and Gottesman 2002, Mandin and Gottesman 2010). All three of these sRNAs operate by inducing structural remodeling of the *rpoS* 5'UTR upon base-pairing, resulting in freeing of the RBS and translation of the protein (Battesti et al. 2011).

The Role of Hfq and sRNA Binding

One hallmark of small RNA regulators in Gram-negative organisms is the requirement of the protein chaperone Hfq to facilitate base pairing between

sRNAs and their targets (Figure 1-3C). Hfq is a ubiquitous RNA binding protein which has been predicted to be present in nearly 50% percent of sequenced bacterial genomes (Sun, Zhulin and Wartell 2002). It was originally discovered biochemically as a protein required for *in vitro* replication of the RNA phage Qβ (Blumenthal and Carmichael 1979). Later studies demonstrated that Hfq mutants exhibited pleiotropic effects in *E. coli* such as decreased growth rate, sensitivity to ultraviolet light, and osmosensitivity among others, and that these phenotypes could be attributed to decreased RpoS translation (Tsui, Leung and Winkler 1994, Muffler, Fischer and Hengge-Aronis 1996). It was subsequently demonstrated that Hfq associated with sRNAs, leading to increased interactions with their mRNA targets, providing a rational for the observed phenotypes (Zhang et al. 1998, Zhang et al. 2002).

It is now widely accepted that Hfq is a key player in sRNA mediated post-transcriptional gene regulation in many Gram-negative organisms. In fact, Hfq can be viewed as a proxy for sRNA based functions in many organisms do to the requirement of this protein cofactor in executing sRNA based gene silencing. This is supported by the observation that Hfq deletion, in Gram-negative bacteria in particular, results in pleiotropic phenotypes that mimic those observed by deletion of individual sRNAs. (Battesti et al. 2011, Gottesman and Storz 2011a,

Papenfort et al. 2009, Vogel 2009). In *Salmonella* it is estimated that up to 20% of the genome is regulated either directly or indirectly by Hfq, suggesting that sRNAs play a significant role in modulating gene expression (Sittka et al. 2008, Vogel 2009). In addition, Hfq is essential for virulence in many pathogenic organisms such as *Brucella*, *Vibrio*, *Pseudomonas*, *Yersinia*, *Salmonella*, and *Listeria* among others, implicating yet to be discovered sRNAs are crucial mediators of bacterial pathogenesis in these organisms (Chao and Vogel 2010).

Hfq is one of the founding members of an extensive family of RNA binding proteins that can be found in all three domains of life (Wilusz and Wilusz 2005, Vogel and Luisi 2011). Bioinformatic and structural based studies over the last decade have placed Hfq in the broad Sm family of proteins, similar to the eukaryotic Sm-like (LSm) proteins involved in alternative splicing (Brennan and Link 2007, Vogel and Luisi 2011, Zhang et al. 2002, Moller et al. 2002). Though widely divergent at the sequence level among various bacterial and archeal species, high resolution crystal structures from multiple organisms reveal that Hfq adopts a homohexameric quaternary ring structure containing a central pore and two RNA binding faces on either side of the ring (Brennan and Link 2007, Vogel and Luisi 2011). *In vitro* experiments suggest that Hfq preferentially binds single-stranded A/U rich sequences that are often adjacent to stem-loop

structures in order to facilitate base pairing between a sRNA and its target. However, the prediction of Hfq binding sites within sRNAs are often unreliable due to weak conservation and in many instances the presence of multiple A/U-rich stretches within sRNAs (Storz et al. 2011, Gottesman and Storz 2011b, Vogel and Luisi 2011).

The prediction of Hfq binding sites present on mRNA targets has proven to be even more difficult. The majority of which have been inferred through crystallographic studies and genomic SELEX experiments, however these sequence preferences correlate poorly with known physiologic substrates of Hfq *in vivo* (Link, Valentin-Hansen and Brennan 2009, Lorenz et al. 2010). Complicating matters further is the relative contribution of RNA secondary structure in addition to primary sequence in relation to Hfq recognition and binding. How this enigmatic protein is able to recognize sRNAs and their cognate mRNAs among the thousands of different transcripts within the cell is an over arching question in the field and remains to be determined.

Hfq Binding and RNA Stability

Inhibition of translation by sRNA base pairing to mRNAs also results in decreased mRNA stability. This phenomenon was first recognized through mRNA decay studies of the *E. coli* transcript *sodB.* Under normal circumstances the half life for *sodB* in *E.coli* is approximately 10 minutes. However, upon iron limitation the sRNA, RyhB is transcribed and base pairs with the 5′ UTR of *sodB* in an Hfq dependant manner; resulting in a fivefold reduction in *sodB* half-life (Masse, Escorcia and Gottesman 2003). The accelerated decay of transcripts after sRNA annealing is thought to be a secondary consequence of translational inhibition, because the half-life of bacterial mRNA is strongly affected by its association with ribosomes (Deana and Belasco 2005). However, Hfq may play a role in accelerating decay of sRNA/ mRNA pairs by recruiting specific RNases. In support of this is the recent report that Hfq binds to the large C-terminus of the major endoribonuclease, RNase E; suggesting it may serve as a platform for assembly of machinery dedicated to RNA turnover (Ikeda et al. 2011).

In some instance, sRNA based regulation bypasses translational mechanisms and instead directly operates through modulating mRNA decay. This is best illustrated by the small RNA MicC in *Salmonella* which regulates one of the major outer membrane proteins *ompD.* Structural probing studies suggested that MicC annealing to *ompD* did not occur near the ribosome binding

site of the transcript, but instead occurred within the coding sequence (CDS), encompassing codons 23-26 (Pfeiffer et al. 2009). MicC regulation of *ompD* through inhibition of translation seemed unlikely due to the strong helicase activity of elongating 70S ribosomes, which are capable of disrupting a perfect 27-bp helix with a predicted melting temperature of 70˚C (Takyar, Hickerson and Noller 2005). Indeed, *in vitro* studies confirmed that internal pairing to the *ompD* CDS did not inhibit translation but instead affected RNA stability, in an Hfq and RNase E dependant manner. Interestingly the same report demonstrated that MicC regulates the other major outer membrane protein in *Salmonella, ompC,* however this regulation occurs through classical translational repression targeting the 5'UTR of this transcript (Pfeiffer et al. 2009).

This dichotomy illustrates the complexity of predicting mRNA targets of sRNAs using traditional computational approaches, as has been the gold standard to date. Furthermore, the ability of sRNAs to mediate gene expression through pairing within coding sequences greatly expands the potential regulatory capacity of these RNA elements. Also it appears that for MicC at least, the mechanism of post-transcriptional regulation is dictated by the target and not the sRNA since the same region of MicC is responsible for targeting *ompC* and *ompD.*

Antisense sRNA Regulators

In addition to *trans*-encoded sRNAs which are located in a different region of the chromosome relative to their target, many bacteria also transcribe antisense sRNAs that are located on the opposite strand of their target message (Figure 3D). These regulators are often referred to as *cis*-encoded sRNAs and fundamentally differ from *trans*-encoded sRNAs in that they display perfect complementary base pairing with their mRNA targets due to their antisense nature. Consequently, the requirement of an RNA chaperone such as Hfq is dispensable in regards to facilitating annealing between the two RNA species. The advent of sophisticated sequencing technologies, microarray gene expression profiling, and computational searches have resulted in the identification of hundreds of new bacterial antisense RNAs over the last decade (Thomason and Storz 2010). Though, the function of these RNAs are less well understood as compared to *trans*-encoded sRNAs.

Antisense sRNAs derived from plasmids, phages, and transposons were actually some of the first sRNAs discovered some 30 years ago. Subsequently these *cis*-encoded sRNAs were shown to be components of "addiction" modules which ensure maintenance of parasitic elements within the host (Thomason and Storz 2010). These addiction modules are now recognized as toxin/ antitoxin

(TA) systems, and consist of a stable toxin and an unstable anti-toxin. TA systems are categorized as either type I or type II based on their mechanism of action. Type I TA systems produce a stable mRNA (toxin) that codes for a toxic peptide, as well as an unstable antisense RNA (antitoxin); annealing of the antisense RNA to the toxin mRNA results in degradation of both RNA species. However, if the parasitic element is lost from the cell the unstable antitoxin is degraded but the stable toxin mRNA persists and can ultimately be translated resulting in cell death (Gerdes and Wagner 2007). The same principle applies to type II TA systems with the exception that they operate at the protein level.

Accumulating evidence suggests that TA systems may be part of normal cellular physiology and not just maintenance systems as they have been found to be chromosomally encoded, are tightly regulated, and are only toxic when expressed at very high-levels (Gerdes and Wagner 2007, Thomason and Storz 2010). Although the role these systems play in cellular physiology remains to be determined.

Small RNA Mediated Regulation of Proteins

Most sRNAs operate in conjunction with proteins *in vivo* to execute genetic regulation. *Trans*-acting sRNAs for instance often require the protein cofactor Hfq to facilitate annealing as mentioned above. Additionally, these

small RNAs may require cellular RNases to accomplish their regulatory function. Common to the various sRNA control mechanism discussed thus far is the requirement of base pairing between a regulator and its target. However, other *trans*-acting regulatory RNAs act directly on proteins to influence gene expression. These factors can be separated into two groups, those that modify RNA binding proteins and those that regulate enzymes (Storz et al. 2011).

Regulating RNA Binding Proteins

Regulation of RNA binding proteins by sRNAs typically involves molecular mimicry of the proteins target mRNAs. In these instances the sRNA usually contains the recognition sequence of the protein, often in multiple copies. The classic example of this is the *E. coli* carbon storage regulatory protein A (CsrA) (Figure 3A). CsrA functions by binding to recognition sequences present in the 5'UTR of transcripts that consist of repeated stem-loop structures with GGA present in the terminal loops (Babitzke and Romeo 2007). These regulatory sequences usually overlap or are adjacent to the RBS of the mRNA target. Binding of CsrA usually results in translational repression by directly masking the RBS or by inducing secondary structural rearrangements in the 5' UTR that leads to the RBS being sequestered in a helix (Babitzke and Romeo 2007, Timmermans and Van Melderen 2010). All though less prevalent CsrA can also activate gene

expression by remodeling the 5'UTR resulting in exposure of a previously sequestered RBS. CsrA is negatively regulated by the sRNA CsrB, which contains 18 repeated CsrA binding sites allowing it to interact with 9 CsrA dimers. Accumulation of the CsrB sRNA serves as a cellular protein "sink" that effectively titrates away CsrA from its mRNA targets, resulting in large scale changes in gene expression.

CsrA homologs and CsrB-like sRNAs are widely distributed among bacterial species. In fact, it is not uncommon to find multiple CsrB like sRNAs within a bacterial genome and some organisms contain multiple CsrA proteins (Babitzke and Romeo 2007, Timmermans and Van Melderen 2010). Undoubtedly, more examples of sRNA mediated regulation of proteins wait to be discovered. In fact, it was recently shown that a *Pseudomonas* sRNA, CrcZ contains five repeats of a CA rich motif which is bound by the translational repressor Crc. This suggests that sRNA control of protein activity may be more wide spread than previously thought (Moreno et al. 2009, Sonnleitner, Abdou and Haas 2009).

Regulation of Enzyme Activity

The consequences of sRNA regulation of proteins can be more complex than simple sequestration as detailed above. Some sRNAs have the potential to

inhibit, activate or modify enzyme activity, which is the case for the sRNA 6S (Figure 1-3B) (Storz et al. 2011). 6S is a highly expressed sRNA which binds the house-keeping form of RNA polymerase (σ^{70}-RNAP) in *E. coli* (Wassarman 2007, Storz et al. 2011). The secondary structure of 6S adopts a conformation similar to that of DNA in an open promoter (Figure 1-3B) (Barrick et al. 2005). This suggests that one way in which 6S could regulate gene expression is as a direct competitor for σ^{70}-RNAP *in vivo.* The effect of 6S on transcription is surprisingly complex with only a subset of σ^{70} –dependant promoters being directly down regulated, suggesting that σ^{70}-RNAP is modified rather than inhibited (Cavanagh et al. 2008, Neusser et al. 2010, Storz et al. 2011). In addition 6S serves as a template for RNA synthesis of short product RNAs (pRNAs) by σ^{70}-RNAP which leads to the dissociation of 6S from RNAP (Storz et al. 2011, Wassarman 2007). This opens the possibility that sRNA may regulate protein activity through other mechanisms such as allostery or by providing a platform for protein complexes to organize.

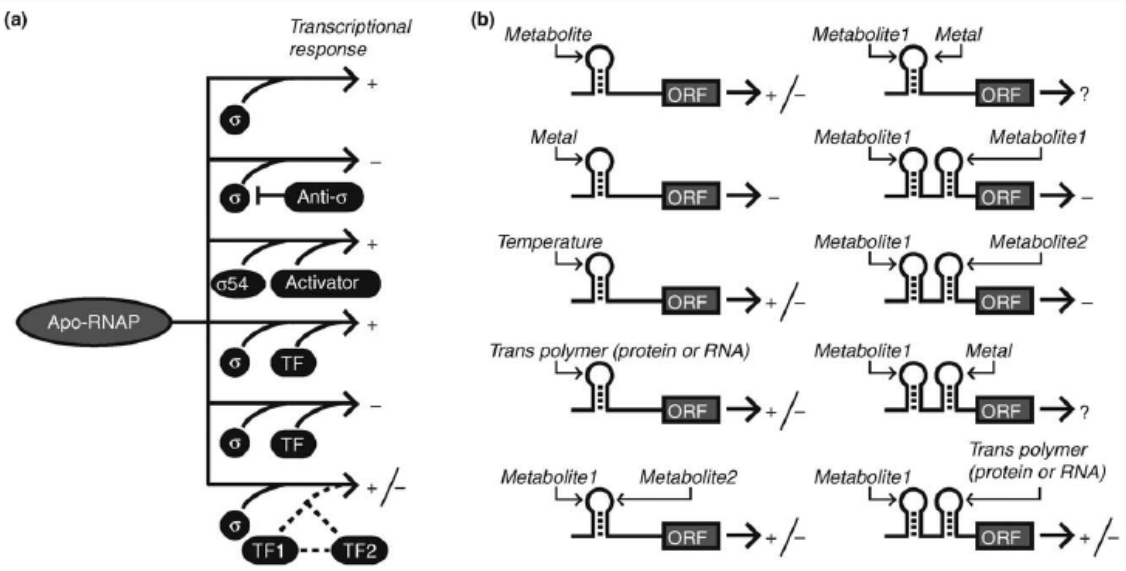


Figure 1-1. Regulation of bacterial gene expression by transcriptional and posttranscriptional mechanisms. (A) Simplified view of transcriptional regulation using alternative sigma factors and DNA-binding transcription factors. Arrows point to the genetic outcome. Plus and minus signs indicate transcriptional activation and repression, respectively. Sigma factors promote transcription initiation through sequence-specific DNA contacts. Sigma 54 is an exception in that it requires a protein factor for transcriptional activation. Anti-sigma factors prevent sigma function. Transcription factors ('TF') can individually activate or repress expression. Also, transcription factors can coordinate with one another to produce complex regulatory outcomes. (B) Simplified view of posttranscriptional regulation. Plus and minus symbols indicate increased or decreased gene expression. Hairpin structures denote individual *cis*-regulatory RNAs, which can respond to temperature, metabolites, metals, or trans-encoded polymers (RNAs or proteins). Alternatively, two different metabolites can associate with the same *cis*-regulatory RNA. It is also possible that a particular, hypothetical *cis*-regulatory RNA may respond to a combination of metals and metabolite concentrations. Metabolite-sensing riboswitches may be arranged in tandem to respond to the same metabolite or to two different metabolite ligands. It is also possible for multiple, distinct posttranscriptional regulatory mechanisms to cooperate in controlling downstream expression.

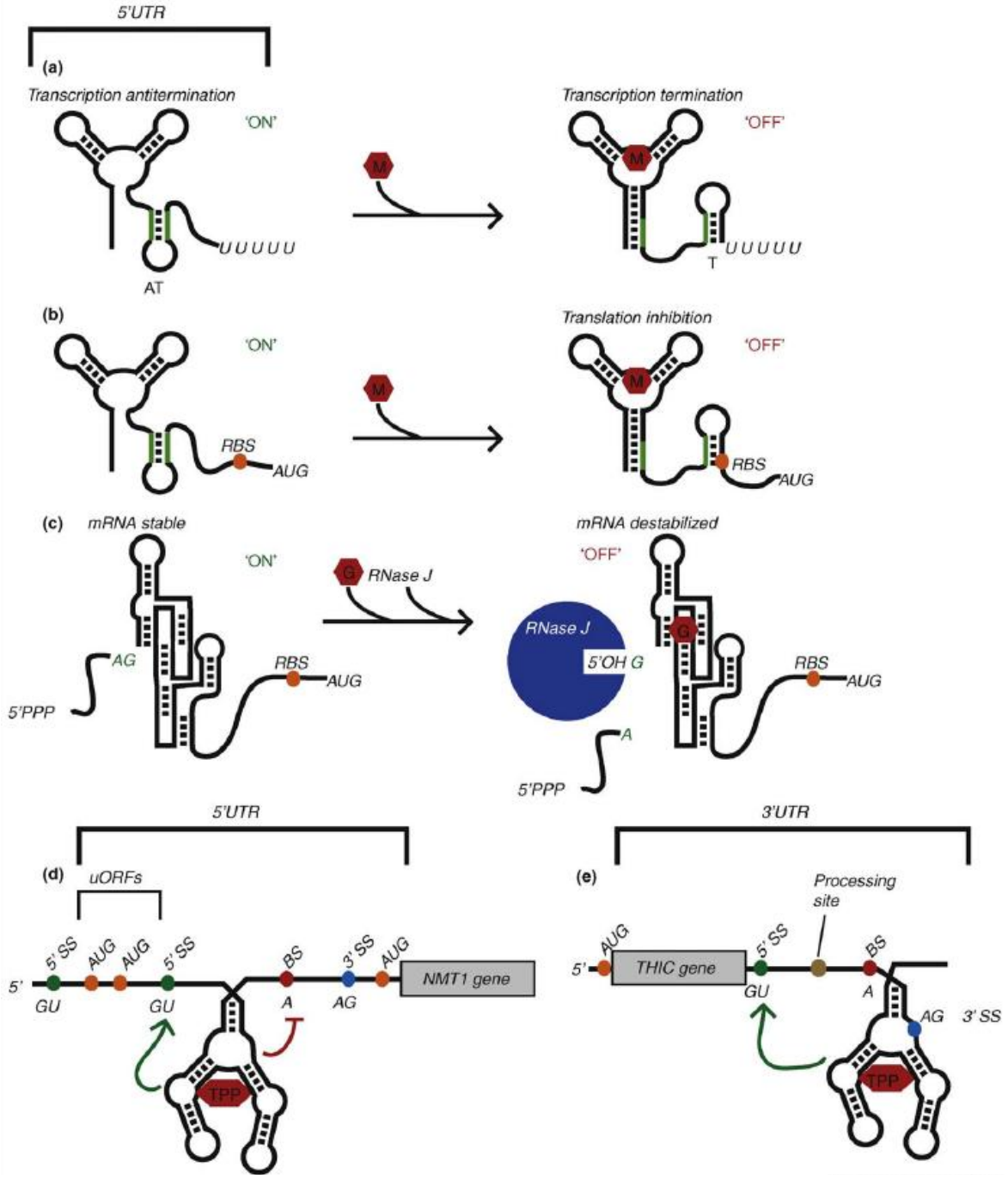

5'UTR
(a)
Transcription antitermination
'ON'
UUUUU
AT
M
Transcription termination
'OFF'
UUUUU
T
(b)
'ON'
RBS
AUG
Translation inhibition
'OFF'
(c)
mRNA stable
'ON'
G
RNase J
AG
5'PPP
RBS
AUG
mRNA destabilized
'OFF'
RNase J
5'OH G
A
(d)
uORFs
5' SS
AUG
5' SS
BS
3' SS
5'
GU
A
AG
NMT1 gene
TPP
3'UTR
(e)
Processing site
THIC gene
3' SS

Figure 1-2. Genetic control by metabolite-sensing riboswitches. (A) Control of formation of an intrinsic transcription terminator by a metabolite-binding riboswitch 'M' denotes the metabolite ligand. (B) Control of translation initiation efficiency by a metabolite-sensing riboswitch. (C) Control of mRNA stability by the GlcN6P-sensing ribozyme as described in the text. 'G' denotes the GlcN6P riboswitch ligand. (D) An example of a TPP riboswitch that controls alternative splicing of the NMT1 gene in Neurospora crassa, as discussed in the text. '50 SS' indicates the 50 splice sites. 'BS' indicates a branch site. (E) As discussed in the text, a TPP riboswitch from the 30 UTR of Arabidopsis thaliana THIC controls mRNA processing and stability. Green arrows signify regions where structural rearrangements occur upon TPP binding and that activate splicing elements. The red line identifies a region where structural rearrangements occur upon TPP binding and that repress splicing elements.

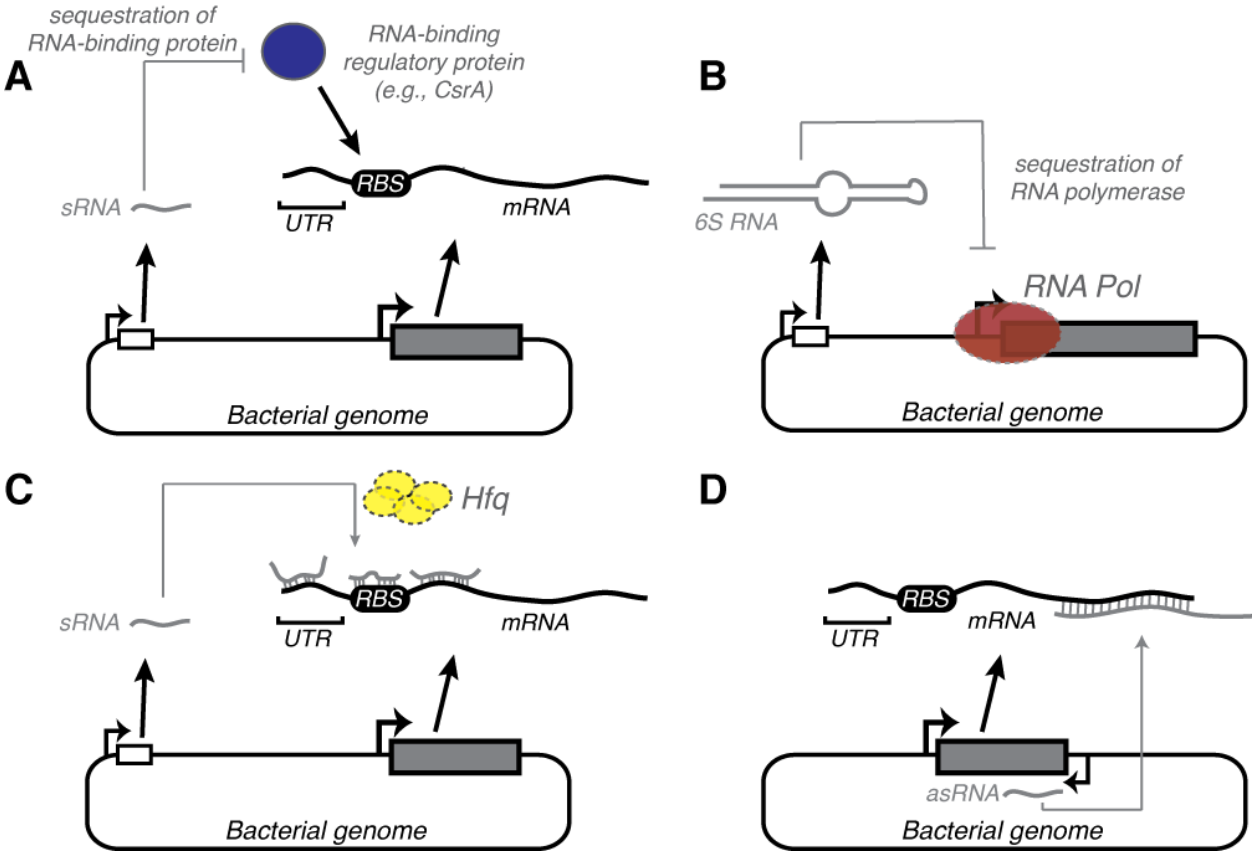


Figure 1-3. Control of gene expression by *trans*-acting regulatory RNAs. (A) sRNAs can function by acting as molecular mimics of natural RNA substrates. The regulatory protein CsrA functions by binding to structured elements embedded in the 5'UTR of RNA transcripts it regulates. Upon binding CsrA induces a conformational change that usually relieves an inhibitory helix which sequesters the RBS of thus allowing for translation initiation. As cells enter into stationary phase a molecular RNA decoy consisting of repeated CsrA binding sites is transcribed and results in sequestration of this regulatory protein thus influencing gene expression patterns within the cell. (B) Is another example of a protein sequestering sRNA known as 6S. This sRNA adopts a secondary structure that mimics an open promoter and thus titrates away a specific isoform of RNA polymerase thus influencing global gene expression patterns. (C) Depicts *trans*-acting sRNA regulation. These small RNAs are transcribed in a different location in the genome from the messages which they regulate. They display incomplete base-pairing and as a result most often require an RNA chaperone such as Hfq to

facilitate base-pairing and subsequent regulation. (D) Describes antisense sRNA based regulation. These sRNAs are transcribed from the opposite strand from the message that they regulate and as a result display perfect complimentarity to their target mRNA.

CHAPTER 2

Identification of RNA Elements that Associate with *Bacillus subtilis* Hfq *in vivo*

Introduction

Regulation by small, trans-acting regulatory RNAs (sRNAs) is critical for bacterial gene regulation (Storz *et al*., 2011). Indeed, their overall importance may eventually be found to rival that of transcription initiation-based regulatory strategies. For example, there are approximately 100 sRNA regulators in *Escherichia coli* and *Salmonella* species (Papenfort and Vogel, 2010), which contain a moderately similar number of transcription factors (~200) affecting the efficiency of transcription initiation. Moreover, both of these classes of regulatory molecules (sRNAs and transcription factors) can influence the abundance of multiple mRNA transcripts, indicating that genetic regulons can be controlled by both initiation and post-initiation regulatory strategies (Vogel 2009; Gottesman and Storz 2010). Therefore, it has become clear from these prior studies that sRNA-mediated regulation is an important “layer” of genetic control.

In Gram-negative proteobacteria, sRNA regulators are expressed in response to certain stress or physiological conditions and are stabilized against degradation through association with the homohexameric RNA-binding protein Hfq (Vogel and Luisi, 2011). Hfq is an Sm-like protein that is widespread in eubacteria (Valentin-Hansen *et al*., 2004). For instance, most α- β- and γ-proteobacterial species encode at least one copy of *hfq*. Deletion or mutation of *hfq* results in decreased fitness and reduced virulence for many Gram-negative pathogens (Chao and Vogel, 2010). For example, deletion of *Salmonella typhimurium hfq* resulted in several phenotypic changes including but not limited to reduced replication in macrophages, incomplete secretion of virulence effectors, loss of motility, and attenuated virulence in mice (Sittka *et al*., 2007; Chao and Vogel, 2010). It is generally presumed that Hfq's importance stems almost entirely from its functions in sRNA-based regulation. In general, Hfq increases the intracellular half-life of sRNAs and facilitates base-pairing interactions between sRNAs and their mRNA targets. Hfq may assist sRNA-mRNA interactions by enhancing the rate of their annealing or by instigating RNA structural remodeling (Maki *et al*., 2010; Fender *et al*., 2011; Hopkins *et al*., 2011; Hwang *et al*., 2011). Consistent with this possibility, structural analyses of Hfq hexamers suggest that the ring-shaped complex exhibits at least two RNA-binding portions for these purposes (Link *et al*., 2009).

Thus far, most sRNAs have been found to affect gene expression by binding to sites within 5' leader regions and altering translational efficiency of the downstream gene. However, some sRNAs associate within mRNA coding regions, while still other sRNAs control expression by altering mRNA stability (Pfeiffer *et al*., 2009; Waters and Storz, 2010). Regardless, Hfq is required for virtually all sRNA-mediated gene regulation in *E. coli* and *Salmonella*. Correspondingly, multiple experimental methods have been employed to identify the full catalog of Hfq-associated sRNAs for several Gram-negative bacteria (Ansong *et al*., 2009; Sharma and Vogel, 2009). For example, one global study analyzed RNAs that co-immunoprecipitated with *E. coli* Hfq, using Hfq-specific antisera; the protein-associated RNAs were subsequently detected using high-density oligonucleotide microarrays (Zhang *et al*., 2003). Other studies have identified Hfq-associated RNAs through subcloning and sequencing of cDNAs (Christiansen *et al*., 2006; Huttenhofer, 2006; Sonnleitner *et al*., 2008; Nielsen *et al*., 2010). More recently, the development of high-throughput sequencing methodologies has expanded the sensitivity and scale of these analyses (Sharma and Vogel, 2009). To create a comprehensive catalog of sRNAs and mRNAs that co-immunoprecipitated with epitope-tagged Hfq in *Salmonella*, high-throughput pyrosequencing was used to identify the protein-associated RNAs (Sittka *et al*., 2008). This method successfully recovered known *Salmonella* sRNA genes and led to the discovery of

new sRNAs, raising the total number of sRNAs in this organism to greater than 100. When Hfq-associated mRNAs were examined against the transcriptomic profile of an *hfq* mutant strain, these data also allowed for the preliminary prediction of global mRNA targets. In general, the high-throughput sequencing-based approach is rapid, global, and exhibits a wide dynamic range for quantification of protein-associated RNA species.

In contrast, less is known regarding the relative importance of sRNAs in Gram-positive bacteria. Similarly, the role(s) of Hfq in Gram-positive bacteria have also been poorly characterized, relative to the level of attention that has been oriented towards Gram-negative Hfq (Romby and Charpentier, 2010). However, recent data suggests that there are likely to be similarities and differences between Gram-positive and Gram-negative bacteria. In *Staphylcoccus aureus*, at least a dozen sRNAs have been identified, whose intracellular abundance is independent of Hfq, in contrast to the general requirement for Hfq by proteobacteria (Geissman *et al*., 2009). Most of the *S. aureus* sRNAs included a C-rich motif that is likely to be important for association with their target mRNAs. Interestingly, one of these sRNAs, coined RsaE, was found to be conserved in other Bacillaceae, including *B. subtilis*, and is predicted to regulate expression of metabolic genes through a C-rich motif (Geissman *et*

al., 2009; Rasmussen *et al*., 2009). The most thoroughly studied *S. aureus* sRNA, RNAIII, utilizes an Hfq-independent mechanism to regulate expression of multiple mRNAs involved in virulence. Therefore, it is still unclear what role(s) Hfq might play in this organism. *Streptococcal* species, which do not encode a known Hfq homologue, have also been found to produce multiple putative sRNA regulators (Romby and Charpentier, 2010). Similarly, over 50 putative sRNAs have been identified in *Listeria monocytogenes* (Christiansen *et al*., 2006; Toledo-Arana *et al*., 2009); of these, a single sRNA, LhrA, associates with its target mRNA in an Hfq-dependent manner (Nielsen *et al*., 2009). Therefore, it is still unclear what, if any, role Hfq plays with sRNA-mediated regulation in these bacteria. The fact that Hfq does not appear to be required or associate with sRNAs from these bacteria has suggested that it may not play a general role in sRNA regulation, as is the case for Gram-negative proteobacteria. Perhaps, the sRNA-mRNA interactions within these bacteria are thermodynamically sufficient without Hfq; however, this hypothesis remains to be fully explored. Alternatively, other RNA chaperones may be redundant with Hfq in these organisms, effectively masking a functional requirement for protein-mediated assistance of sRNA-mRNA interactions. To begin to explore these issues, we specifically investigated in this study whether intracellular RNA molecules associated with Hfq in the model microorganism, *B. subtilis*.

Approximately 100 putative sRNAs have been identified in *B. subtilis* (Rasmussen *et al*., 2009; Irnov *et al*., 2010). However, very few of these putative regulatory RNAs have been characterized, either in regards to their target mRNA(s) or mechanism of action. One such sRNA, called RatA, is an antisense sRNA that associates with a toxin-encoding mRNA, *txpA*, and is believed to down-regulate expression of a toxic peptide (Silvaggi *et al*. 2006). At least three other so-called type I toxin-antitoxin systems have been identified in this bacterium (Fozo *et al*., 2010; Irnov *et al*., 2010). Another cis-encoded antisense RNA has been found to base pair with an uncharacterized gene, *yabE* (Eiamphungporn and Helmann, 2010). Three sRNAs have been characterized to varying degrees with respect to their regulatory functions: SR1, FsrA, and RsaE. SR1 has been demonstrated to control expression of *ahrC*, which encodes a transcriptional activator of arginine catabolism genes (Heidrich *et al*., 2006). FsrA is expressed upon derepression of the iron-regulatory protein, Fur, and affects expression of iron-containing proteins (Gaballa *et al*., 2009). Finally, RsaE, an unusually widespread sRNA in diverse Gram-positive species, has been predicted to target central metabolism genes and a 'carbon starvation' gene, *cstA* (Geissmann *et al*., 2009; Rasmussen *et al*., 2009). Notably, the potential role for RNA chaperones, such as Hfq, has not been fully explored for these sRNAs. Indeed, it was not known whether Hfq (or other RNA-binding proteins) affect the

intracellular abundance of any *B. subtilis* sRNAs, or even if it was expressed under typical bacterial growth conditions. In this study, we find that Hfq is expressed in minimal and rich media, but that expression is increased during stationary phase conditions. To identify potential RNA ligands for association with Hfq we incorporated an epitope-tagged copy of *hfq* into the genome and co-immunoprecipitated Hfq from stationary phase cells. Protein-associated RNAs were converted to cDNA and then identified using high-throughput sequencing. We find that *B. subtilis* Hfq broadly associates with different classes of RNA molecules in vivo, including portions of mRNAs, numerous 5' leader regions, and approximately 25% of the putative sRNAs that were identified previously Collectively, these data demonstrate that *B. subtilis* Hfq broadly associates with intracellular RNA molecules and correspondingly may influence multiple regulatory RNA networks.

RESULTS

Expression of Hfq in Bacillus subtilis

The role(s) of Hfq in Gram-positive bacteria have not been as well studied as for selected Gram-negative bacteria. It is currently unclear whether Hfq

functions in Gram-positive bacteria as a chaperone that assists mRNA-sRNA interactions and stabilizes sRNAs, as has been observed previously for Gram-negative bacteria, or is involved in other unidentified functions. For example, Hfq has yet to be examined in *Bacillus subtilis*, a model system for the low-GC Gram-positive bacteria. In *Staphylococcus aureus*, the putative *hfq* gene is only weakly expressed (Geisinger *et al*., 2006) and is not required for stabilization or regulation of any of the identified *S. aureus* sRNAs (Romby and Charpentier, 2009). Interestingly, the Hfq sequences for many Gram-positive bacteria appear to lack a short, positively-charged stretch of amino acids at the C-terminus. For example, the *B. subtilis hfq* gene (originally annotated as *ymaH*) encodes a 73 amino acid Hfq homolog while, in contrast, *E. coli* Hfq contains an additional 29 amino acids at its C-terminus (Figure 2-1A). Indeed, the absence of this region has been proposed to be a key functional limitation for Hfq in many Gram-positive species (Vecerek *et al*., 2008). However, recent data demonstrated that an *E. coli* Hfq containing a truncation of this C-terminal extension was still proficient in sRNA-mediated regulation (Olsen *et al*., 2010). Some bacteria (*e.g., Moraxella catarrhalis* and *Acinetobacter baylyi*) contain even larger C-terminal regions (Sun *et al*., 2002; Attia *et al*., 2008; Schilling and Gerischer, 2009). Therefore, the general importance and functional roles of the C-terminal portion of Hfq are currently unclear. It is therefore possible that *B. subtilis* Hfq may share

functional similarities with its Gram-negative Hfq counterparts, despite the lack of the C-terminal extension.

In many bacteria, *hfq* is co-transcribed with an adjacent tRNA modification enzyme, *miaA*. Indeed, *miaA* is located immediately upstream of *hfq* in the *B. subtilis* genome, suggesting a similar arrangement in this bacterium (Figure 2-1B). However, we previously used a high-throughput sequencing method to identify transcription start sites (TSS) for *B. subtilis* during stationary phase growth (Irnov *et al*., 2010) and noted a single start site located upstream of *hfq* and downstream of *miaA* (Figure 2-1B); therefore, *B. subtilis hfq* is likely to be expressed as a monocistronic transcript. Next we cultured cells in rich and minimal media and extracted total RNA at varying points during growth (Fig 2-2A and B). The relative abundance of the *hfq* transcript was determined by quantitative real-time RT-PCR (qPCR) (Fig 2-2E and F). Very little (<2-fold) change in mRNA abundance was observed as bacteria transitioned from exponential to stationary phase growth, suggesting that *hfq* is not likely to be transcriptionally regulated. To investigate whether protein levels were altered under these conditions, we replaced the endogenous *hfq* with an epitope (FLAG)-tagged copy and monitored protein abundance by Western blot analysis (Fig 2-2C and D). We engineered the FLAG-tagged copy into the C-terminus of Hfq, as crystal

structures have shown this region to be on the exterior of the hexameric quaternary structure, where as the N-terminus has been demonstrated to mediate multimer formation (Brennan and Link 2007, Link et al. 2009, Moller et al. 2002, Someya et al. 2012). Interestingly, Hfq expression was significantly increased upon transition in to stationary phase growth. Therefore, we postulate that the *hfq* gene is likely to be subjected to either translational or post-translational regulation, although the exact mechanism remains to be identified. However, we cannot rule out the possibility that addition of the epitope tag at the C-terminus of Hfq renders the protein recalcitrant to degradation resulting in its accumulation over time. Overall, these data demonstrated that Hfq is expressed in *B. subtilis*, and that its function is most likely to be relevant during stationary phase conditions.

General approach for identifying Hfq-associated RNAs

In *Listeria monocytogenes*, Hfq has been reported to play a role in stress tolerance and pathogenesis in mice, and has been shown to facilitate association of an antisense sRNA to its corresponding mRNA target (Christiansen *et al*., 2004; Nielsen *et al*., 2010). However, a broad role for Hfq has still yet to be determined in other Gram-positive bacteria, even though it is present in approximately 50% of sequenced genomes and contains conserved residues that are important for

RNA binding activity (Sun et al. 2002). Therefore, in an effort to better understand the function of Hfq in *Bacillus subtilis*, we co-immunoprecipitated Hfq and employed massively parallel sequencing for identification of protein-associated RNA molecules (Fig. 2-3). Specifically, a FLAG-epitope tag was translationally coupled to the carboxy terminus of Hfq for the purposes of co-immunoprecipitation (coIP), and a wild-type strain lacking the epitope tag was used as a negative control. The cells were cultured to stationary phase in defined medium prior to preparation of cellular extracts; these conditions were chosen based upon the expression pattern of Hfq. In addition, the culture conditions and media composition used for our coIP experiments mirrored those used in a previous study in our lab which mapped global transcription start sites within the *Bacillus subtilis* genome; allowing us to use these two data sets in conjunction to identify novel RNA based regulatory elements (Irnov et al. 2010).

Illumina-based sequencing of cellular RNAs that were recovered from coIP of Hfq resulted in approximately 25 million and 10 million total cDNA reads for the Hfq-FLAG and control samples respectively. The average length of the cDNA sequences in this experiment was approximately 35 nucleotides. Approximately 10 million and 1.5 million cDNA reads were unambiguously matched to the *B. subtilis* reference genome for the Hfq-FLAG and control samples, respectively.

Approximately 98% of the unmapped cDNA reads corresponded to the adapter oligonucleotides that are employed for Illumina sequencing. At the time that these deep-sequencing studies were conducted, sample preparation required the use of gel purification in order to remove unligated adaptor sequences prior to performing the sequencing reaction. Gel purification is a non-issue when dealing with copious amounts of genetic material such as total RNA. However, phenol/ chloroform extraction, resuspension in large volumes, ethanol and buffer washes to eliminate salts and non-specific proteins and RNAs that bind to Hfq all lead to significant reductions in material that was severely compounded by gel purification, resulting in substantial loss of RNA. Taking this into consideration we elected to omit the gel purification step entirely and proceed with sequencing, resulting in the large number of unligated adaptor sequence observed in our data set. In fact, the amount of material recovered from the coIP was so minimal in relation to ribosomal RNA (rRNA) and transfer RNA (tRNA) we could not observe a difference in OD_{260} between the Hfq pull down and the mock control, illustrating the power and sensitivity of next-generation sequencing technologies.

To identify clusters of cDNA reads corresponding to cellular RNAs we searched for peaks that exhibited at least five cDNA reads for a minimum of 50

consecutive genomic positions. RPKM values (cDNA reads per kilobase of genomic sequence per total mapped reads; Mortazavi *et al*., 2008) were determined for all peaks identified from the Hfq-FLAG and control samples, and used to estimate the degree of enrichment by Hfq, which ranged from < 1 to > 1,000 (Fig. 2-5B; Tables 2-1-5). Almost all of the peaks exhibiting an RPKM ratio of less than two corresponded to tRNAs or rRNAs (Fig. 2-5B), suggesting that those RNA molecules were not likely to be enriched by co-immunoprecipitation with Hfq under our experimental conditions. Based in part on this observation, we applied an RPKM ratio of two as an arbitrary cut-off for more detailed analysis of the remaining coIP peaks.

One potential disadvantage of the method we used for Illumina sample preparation was that information regarding the genomic strand from which an RNA species originated was lost upon conversion to cDNA. However, our lab previously used a different high-throughput sequencing method for identification of approximately 600 *B. subtilis* transcription start sites (Figure 2-4) (Irnov *et al* 2010). Importantly, the culture conditions for the prior TSS analysis were identical to the culture condition that was used for the Hfq coIP experiment described herein. Therefore, the combination of these data sets allowed us to determine the directionality of the majority of the Hfq-enriched

peaks. As a result, we were able to categorize most Hfq-enriched peaks into the following groups: (1) those associated with open reading frames (ORF), (2) those associated with 5' leader regions, (3) those corresponding to sRNAs, and (4) those corresponding to tRNA or rRNAs (Fig. 2-5C). Also, a small but notable class of Hfq-enriched peaks could not be easily placed into these categories (Table 2-2).

mRNAs and mRNA coding region fragments that associate with Hfq

The most abundant class of RNAs enriched by Hfq corresponded to the ORF portion of mRNAs (Table S2). Specifically, over 60% of the Hfq-enriched RNA molecules mapped to either a portion of an ORF or encompassed an entire ORF. As such, we subdivided these peaks into four distinct categories based on the position of the mapped reads relative to the message: full ORF coverage, overlap with the start codon, overlap with the stop codon, and peaks that were internally located in ORFs (Figures 2-5C; 2-6; Table 2-4). Hfq-enriched peaks fully encompassed 17 ORFs, which mainly corresponded to genes of unknown function (Fig. 2-6A). All of these ORFs are short and appear to encode peptides or small proteins. The longest mRNA to be enriched by Hfq was 276 nucleotides in length and the shortest was 110 bases; the average length was 203 nucleotides (Table 2-4).

17 Hfq-enriched RNA fragments corresponded to a portion of the ORF that included the start codon (Figure 2-6B; Table 2-4). These fragments ranged from 54 to 178 nucleotides in length (with a mean of 104). Many sRNAs affect expression of target mRNAs through base-pairing interactions with or near the ribosome-binding site (Vogel and Luisi 2011, Storz et al. 2011)Gottesman and Storz, 2010). It has also been demonstrated that sRNAs need not necessarily target ribosome binding site sequences directly in order to affect translation efficiency; instead, base-pairing within a five codon window downstream from the AUG is sufficient to reduce translation (Bouvier *et al*., 2008). Therefore, this class of Hfq-bound RNAs is intriguing as it may conceivably correspond to sites of Hfq-mediated sRNA action in *B. subtilis*.

Another group of Hfq-associated RNA molecules corresponded to peaks that overlapped with the stop codon (Fig. 2-6C; Table 2-4). In total 27 peaks fit this criterion, which ranged in size from 70 to 271 nucleotides in length. It is currently unclear what relationship, if any, may exist between Hfq and this region of mRNA ORFs. However, a potential interaction between Hfq and both start and stop codons has been postulated previously (Lorenz *et al*., 2010), which is corroborated by our aggregate data. Interestingly, several Hfq-enriched peaks appeared to correspond to portions of mRNAs that included the 3' UTR.

However, since we could not unambiguously determine the directionality of these particular peaks, we placed them into a category with other intriguing but unknown RNA molecules (Table 2-2). Finally, a large group consisted of enriched reads located within internal regions of ORFs. Approximately 45% of the ORF-associated peaks were placed into this category, ranging from 52 to 425 nucleotides in length (Fig. 2-6D). These Hfq-enriched RNA molecules may correspond to stable degradation intermediates, alternative ORFs, new sRNAs, or pairing sites for Hfq-assisted sRNAs. In fact, most small RNAs exert regulatory function through modulating translation initiation; however sRNAs can also affect target mRNA expression through pairing within the coding region of an ORF, resulting in decreased RNA stability (Pfeiffer et al. 2009). Therefore it is possible that reads which mapped to the internal portions of coding sequences may actually represent targets of sRNA action, however further experimentation is required to determine the physiological relevance of this class of Hfq associated RNA species.

Long mRNA leader regions associate with Hfq

Another large class of Hfq-enriched signals that we found in our data corresponded to *cis*-acting, signal-responsive regulatory RNA elements, located

within 5' leader regions. Approximately 80 such signal-responsive RNA sequences have been previously identified within the *B. subtilis* genome (Irnov *et al*., 2006; Winkler, 2007). Virtually all of these leader regions are greater than 100 nucleotides in length and are predicted to fold into complex secondary and tertiary structure arrangements. Tertiary folding of the sensory-responsive portion of these sequences (aptamer domain) typically correlates with association of the appropriate ligand. Binding of the ligand molecule(s) subsequently leads to repression or activation of downstream gene expression through modulation of transcription attenuation or the efficiency of translation initiation (Dambach and Winkler, 2009). In *B. subtilis*, there are classes of *cis*-acting, signal-responsive regulatory RNAs that respond to different types of ligand molecules, including metabolites, metals, RNA-binding proteins, and selected tRNAs (Irnov *et al*., 2006; Winkler, 2007). Overall, 19% of the Hfq-enriched peaks corresponded to previously characterized leader regions (Figures 2-5C; 2-7A-D). Interestingly, there did not appear to be a strong bias with respect to the classes of signal-responsive leader regions that were enriched by Hfq. In other words, Hfq-enriched peaks were observed for leader regions that responded to all types of ligand molecules including magnesium, metabolites, proteins, and tRNAs; however, not all members of a particular class were

enriched by Hfq. For example, only four of the 19 known tRNA-sensing regulatory RNAs were enriched by Hfq (Figure 2-7D).

One aspect of massively parallel sequencing technology that contrasts to most conventional RNA profiling methods is the ability to resolve data at individual nucleotide resolution. This level of resolution allowed for the examination of exactly which portions of signal-responsive leader regions co-immunoprecipitated with Hfq. Signal-responsive regulatory RNAs can be thought of as consisting of two portions: a highly structured ligand-sensing aptamer domain, followed by an expression platform which consists of downstream sequence elements that control gene expression in response to ligand-induced conformational changes. Interestingly, all of the Hfq-enriched peaks for signal-responsive leader regions corresponded almost exclusively to the aptamer domain (see Figure 2-7 for representative examples). It is possible that these structured regions are recalcitrant to digestion by cellular RNases, and therefore accrue within cells to the degree that they associate nonspecifically with Hfq hexamers. Alternatively, Hfq may associate with these leader regions because it has a yet-to-be determined role in their regulatory functions. Nonetheless, it still remains to be determined whether the putative interaction between Hfq and mRNA leader regions is functionally meaningful.

Moreover, in addition to associating with previously identified cis-acting regulatory sequences, Hfq also co-immunoprecipitated with 5' leader regions that are likely to correspond to new regulatory elements. Specifically, several uncharacterized 5' leader regions exhibited similar coIP patterns as with known riboswitches (Table 2). For example, the 5' leader region for the *guaA* gene exhibited a coIP pattern that was essentially identical to several proximally located guanine-sensing riboswitches, suggesting that it might also correspond to a signal-responsive regulatory RNA. Indeed, manual inspection of the *guaA* leader region revealed the presence of a putative intrinsic transcription terminator site, consistent with a transcription attenuation-based regulatory element located within the 5' leader region (Figure 2-8). Therefore, we speculate that the *guaA* leader region is likely to contain a newly identified but uncharacterized transcription attenuation system. *B. subtilis* already utilizes several transcription attenuation-based regulatory mechanisms for regulation of purine and pyrimidine levels (Turnbough and Switzer, 2008; Winkler, 2007). For example, UMP levels are sensed by an RNA-binding protein, PyrR, to control expression of pyrimidine biosynthesis genes via transcription attenuation. Also, the *pyrG* gene is regulated by a unique transcription attenuation mechanism, whereby conditions of low CTP stimulate reiterative addition of G residues, which then act to stabilize an antiterminator helix and promote downstream

expression. Additionally, other *B. subtilis* purine biosynthesis genes are regulated by a total of five guanine-sensing and adenine-sensing riboswitches (Winkler, 2007). We therefore speculate that the *guaA* leader constitutes yet another post-initiation regulatory mechanism dedicated to nucleotide homeostasis. Unlike guanine- and adenine-sensing riboswitches, we did not observe any purine-induced changes in *guaA* secondary structure by preliminary structural probing assays (data not shown); therefore, the *guaA* leader region is likely to use a mechanism other than direct sensing or purine levels as is the case for riboswitches In general, these data suggest that the putative relationship between Hfq and leader regions is close enough that Hfq co-immunoprecipitation profiles may assist discovery of new signal-responsive regulatory RNAs within the genome.

Signal-responsive *cis*-acting regulatory RNAs typically act through modulation of transcription elongation (attenuation) or by affecting the efficiency of translation initiation. Curiously, Gram-positive bacteria preferentially utilize transcription attenuation-based mechanisms while Gram-negative bacteria prefer to modulate translation initiation (Irnov *et al*., 2006; Winkler, 2007). The molecular basis for this evolutionary preference is not yet known. However, one potential consequence could be that Gram-positive bacteria may accumulate

many more prematurely terminated 5' leader regions relative to Gram-negative bacteria, due to their greater reliance on transcription attenuation-based regulatory strategies. These 5' leader regions usually do not encode for ORFs and are typically 150-400 nucleotides in length with an intrinsic terminator helix at their 3' terminus. In fact, recent studies in *E. coli* concerned with defining the functional sRNA binding module recognized by Hfq determined that this protein preferentially recognized double and single stranded hairpin structures followed by poly-U tracts analogous to intrinsic transcription terminators (Ishikawa et al. 2012, Otaka et al. 2011). The overall arrangement of these RNAs resembles closely that of sRNA regulatory molecules. Therefore, it is interesting to speculate that Hfq hexamers from Gram-positive bacteria must associate specifically with sRNAs within an intracellular environment containing potentially high concentration of sRNA-like 5' leader regions.

sRNAs that associate with Hfq

Over 100 putative sRNAs have been discovered in *B. subtilis* (Rasmussen *et al*., 2009; Irnov *et al*., 2010), although only a few have been experimentally validated. Moreover, none of the *B. subtilis* sRNAs with proven mRNA targets have been found to be influenced by Hfq (Silvaggi *et al*., 2006; Heidrich *et al*., 2007; Gaballa *et al*., 2008; Romby and Charpentier, 2010; Schmalisch *et al*.,

2010). Of the Hfq-enriched peaks identified in this study, approximately 11% corresponded to putative sRNAs (Fig. 2-5C; Table 1). One interesting class of sRNAs involved type I toxin/ antitoxins modules. Prior studies have identified at least four type I toxin/ antitoxin systems (type I TA) have been discovered in *B. subtilis* (Figure 2-9A) (Silvaggi *et al*., 2005; Irnov *et al*., 2010). These enigmatic systems are usually encoded on plasmids or are integrated into the host genome in prophage regions, where they are hypothesized to serve as "addiction" modules ensuring that the parasitic genetic element is retained in the genome (Figure 2-9B). However, it has also been speculated that they may also participate in other, unknown physiological functions (Fozo *et al*., 2008). They consist of a stable toxin gene, which encodes for a small hydrophobic peptide capable of inserting into the plasma membrane, and an unstable antitoxin which pairs through antisense interactions with the toxin mRNA (Figure 2-9C). Interestingly, sRNAs corresponding to the four TA systems were all enriched by Hfq in this study (Figure 2-10A-D). More specifically, the Hfq-enriched portion mostly corresponded with the sequences involved in antisense pairing between the toxin and antitoxin transcripts. Significantly more cDNA reads corresponded overall to the antisense transcript than with the sense toxin transcript. Together, these observations suggest that the antisense sRNAs may be specifically co-immunoprecipitated by Hfq. Recently a bioinformatics approach was utilized to

identify new type I TA systems in bacteria (Fozo *et al*., 2010). One new toxin candidate that was identified in *B. subtilis* was *yonT*, which is a gene of unknown function encoded within the SPβ prophage region of the *B. subtilis* 168 genome. The authors also detected an anti-sense transcript corresponding to a putative antitoxin to *yonT*, which is also tentatively supported by our prior TSS mapping data (Figure. 2-11). Interestingly, an Hfq-enriched peak corresponded specifically to the 3' region of *yonT*, in the region that would be expected to correspond to an antitoxin transcript, similar to the other four type I TA systems. We speculate from these aggregate data that Hfq may play a role in generally associating with antitoxin transcripts, although further experimentation will be required to test this hypothesis.

Among the putative sRNAs that have been discovered in *B. subtilis* (Rasmussen *et al*., 2009; Irnov *et al*., 2010), only a few validated mRNA targets have been identified. Among these are: (1) FsrA, which was identified as a Fur-regulated sRNA involved in the iron sparing response, (2) SR2, which is activated by the global regulator CodY in response to branched chain amino acids and GTP levels, and (3) RsaE, which is a regulator of central metabolism genes that is widespread among diverse Gram-positive bacteria (Heidrich *et al*., 2006; Gabella *et al*., 2008; Geissmann *et al*., 2009; Bohn *et al*., 2010). In this study, FsrA was

enriched (11-fold) upon coIP of Hfq, with cDNA reads fully encompassing the sRNA (Table 2-1; Fig. 2-13A). FsrA could also be easily detected by northern blot analysis of RNA molecules that coIP with Hfq (Figure 2-13B), confirming the enrichment by Hfq. In *B. subtilis*, FsrA functions to suppress enzymes that require iron as a cofactor by pairing with the ribosome binding sites of these transcripts resulting in a reduction in translation initiation efficiency (Gabella *et al*., 2008). One of these mRNA targets is the iron-containing enzyme, succinate dehydrogenase (*sdh*), which participates in the tricarboxylic acid cycle (TCA) and is a common post-transcriptional regulatory target in other microorganisms. In this study, the *sdh* leader region was also identified as a putative ligand for association with Hfq, therefore, both an sRNA and its putative mRNA target locus were shown to associate with Hfq (Figures 2-12B; 2-13C and D). It is therefore tempting to speculate that these Hfq-enriched sRNA and mRNA peaks specifically corresponded to processing of the RNAs after sRNA:mRNA base-pairing interactions. If so, it is possible that other ORF-associated Hfq-enriched peaks in this data set may also correspond to sites of sRNA action.

In total, only a small subset of sRNAs that have been identified previously co-immunoprecipitated with Hfq in this study (Table 1). Moreover, 32% of these Hfq-bound RNAs corresponded to antisense transcripts or type I antitoxins

rather than “classical” sRNAs. It is also important to note that these Hfq-enriched RNAs did not simply correspond to those sRNAs that exhibited the highest expression levels (Irnov *et al*., 2010 and data not shown). In contrast, at least half of *Salmonella* sRNA candidates were co-immunoprecipitated by Hfq in a similar study (Sittka *et al*., 2008). Therefore, unlike the γ-proteobacterial species, *B. subtilis* Hfq appeared to associate broadly, yet selectively, with intracellular sRNA molecules. The molecular basis for recognition of this sRNA subset by Hfq still remains to be identified, as does its functional consequences. In addition to these Hfq-associated RNA molecules, several Hfq-enriched peaks did not correspond to previously identified sRNAs but still exhibited sRNA-like features (Table 1). We therefore designated these signals as putative sRNAs. One of these putative sRNAs is particularly noteworthy because the Hfq-enriched peak overlapped extensively with a portion of the genome that included an orphan riboswitch located upstream of an unknown gene, *ylbH* (Figure 2-14A and B) (Barrick et al., 2004). At first glance, it was tempting to categorize this Hfq-enriched peak as corresponding to the 5' leader region of the *ylbH* gene; however, this RNA sequence could not be detected by northern blot analysis (Fig. 2-15B). Instead, a strong signal was detected using an oligonucleotide probe that hybridized to the reverse complement sequence. This result suggested that the earlier identification of the orphan riboswitch was incorrect, and that,

instead, the conserved RNA element was potentially a sRNA expressed from the opposite genomic strand. Indeed, closer inspection of the sequences corresponding to this RNA element (Rfam entry RF00516; Gardner *et al*., 2008) revealed several features suggesting that it was likely to be positioned in the opposite genomic orientation. Specifically, a strong intrinsic terminator element was located at the 3' of the conserved region for the reverse complement sequence, which is a common feature of sRNAs (Figure 2-14A). Also, a re-evaluation of the comparative sequence alignment revealed the presence of a helical element at the putative 5' portion (P1) of the conserved region that was missing from the earlier sequence alignments (Fig. 2-15C and D). Also, when presented in the reverse orientation it became apparent that the region exhibiting the highest degree of sequence conservation corresponded to CU-rich oligonucleotide stretches located within terminal loops or interhelical regions. This is a feature that has been observed with many sRNAs, including FsrA and RsaE in *B. subtilis*, and is a critical component for recognition of target mRNAs (Figure 2-15D). During this analysis, another study was published describing discovery of this putative sRNA, which they coined CsfG (Marchais et al., 2011).

In general, the discovery of several new putative regulatory RNAs by coIP of Hfq, including new putative sRNAs and transcription attenuation systems,

demonstrates a clear utility for examination of Hfq ligands in Gram-positive bacteria. Even though the full range of Hfq cellular function(s) remain to be identified in these bacteria, it can potentially be employed as a tool for discovery of new candidate regulatory RNA molecules.

Concluding Remarks and Discussion

The recent development of tools enabling global transcriptome analyses, coupled with the staggeringly large number of fully sequenced eubacterial genomes has uncovered a vast role for post-transcriptional gene regulation in bacteria. These technologies have fueled the discovery of many RNA-based regulatory mechanisms, most notably riboswitches and sRNAs, that in aggregate appear to rival protein-based transcriptional regulatory mechanisms in their breadth and scope. However, identifying putative RNA regulators in bacteria has become much easier than identification of their targets and mechanisms of action. For instance, in *Bacillus subtilis*, over 100 putative sRNAs have now been reported (Rasmussen *et al*., 2009; Irnov *et al*., 2010), yet defined targets and functional roles have only been defined for a few. Indeed, much more is known about the roles and mechanisms of sRNAs for post-transcriptional gene

regulation in Gram-negative organisms. This is partly due to the close, interdependent relationship between sRNAs and Hfq that is exhibited by Gram-negative bacteria, where Hfq has been extensively investigated. Given that Hfq has been less studied overall in Gram-positive bacteria, and is not required thus far for sRNA-mediated regulation in certain bacteria, such as *Staphylococcus aureus*, we investigated in this study the possibility that *B. subtilis* Hfq associates with intracellular RNAs.

At the onset of our experiments we expected to find one of three possibilities: (1) Hfq does not generally interact with cellular RNAs, and therefore might be involved in other, unknown cellular duties, (2) Hfq specifically associates with some but not all sRNAs, and (3) Hfq associates with the majority of sRNAs that have been identified. Our data support the second hypothesis – that Hfq associates with only a subset of the putative sRNAs that have been discovered. Moreover, the interaction between Hfq and sRNAs was strong enough under our conditions for discovery of new, putative sRNAs, including the widely conserved CsfG sRNA. Together, this important information will aid future characterization of post-transcriptional regulatory networks in *B. subtilis*. These data also provide direct evidence that *B. subtilis* Hfq indeed exhibits RNA-binding

properties that are sufficient in theory for facilitating sRNA-mRNA interactions in vivo.

Intriguingly, Hfq also generally associated with the ligand-binding portions of the majority of cis-acting, signal-responsive regulatory RNAs in *B. subtilis*. This finding was unexpected, although *E. coli* Hfq has previously been found to interact with a few such RNA elements. In fact, the association of Hfq with the 5' leader regions of signal-responsive RNAs was so consistent in this study that we predict our data is likely to include new signal-responsive RNA elements, such as the 5' leader region of the *guaA* gene that we speculate to be a *cis*-acting regulatory RNA. However, the potential purpose, and consequences, of the interactions between Hfq and 5' leader regions remains to be investigated. *B. subtilis* expresses over 80 known signal-responsive, *cis*-acting regulatory RNAs, and the majority of these RNA elements control gene expression through transcription attenuation (Winkler, 2007).

Therefore, each of these RNA elements is likely to be highly expressed as an independent transcript under conditions that favor transcription termination rather than antitermination. In our study, we found that proportionally more signal-responsive 5' leader regions appeared to coIP with Hfq than sRNAs. We speculate from this observation that Hfq may face a greater challenge in

associating with sRNA regulators in *B. subtilis*, where sRNAs may broadly compete for Hfq access with 5' leader regions, than in *E. coli*, where only a few signal-responsive, transcription attenuation systems are utilized. Together, our findings suggest a much broader role than previously appreciated for the in vivo RNA-binding activity of *B. subtilis* Hfq*,* The catalog of Hfq-associated intracellular RNAs, as presented here, can now be explored for the functional significance of these individual interactions.

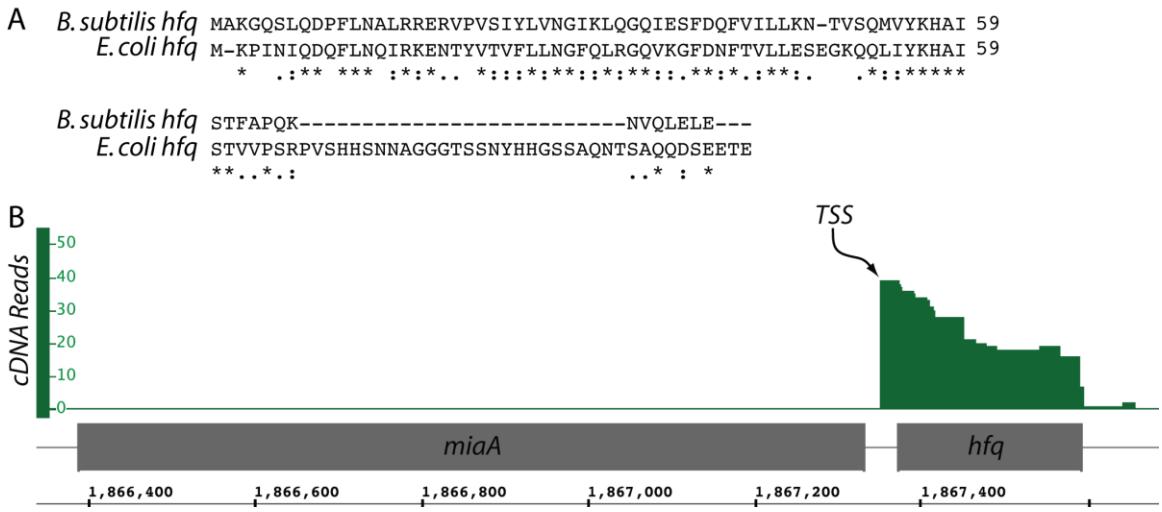


Figure 2-1. Expression of the *hfq* locus in *Bacillus Subtilis*. (A) The Hfq sequences from *B. subtilis* and *E. coli* are shown as an alignment to highlight the C-terminal truncation in *B. subtilis* (and many other Gram-positive bacteria). (B) Approximately 600 *B. subtilis* transcription start sites were determined in a previous study (Irnov *et al*., 2010). We examined these data and found that they supported a single transcription start site upstream of *hfq*, suggesting that it is a monocistronic transcript.

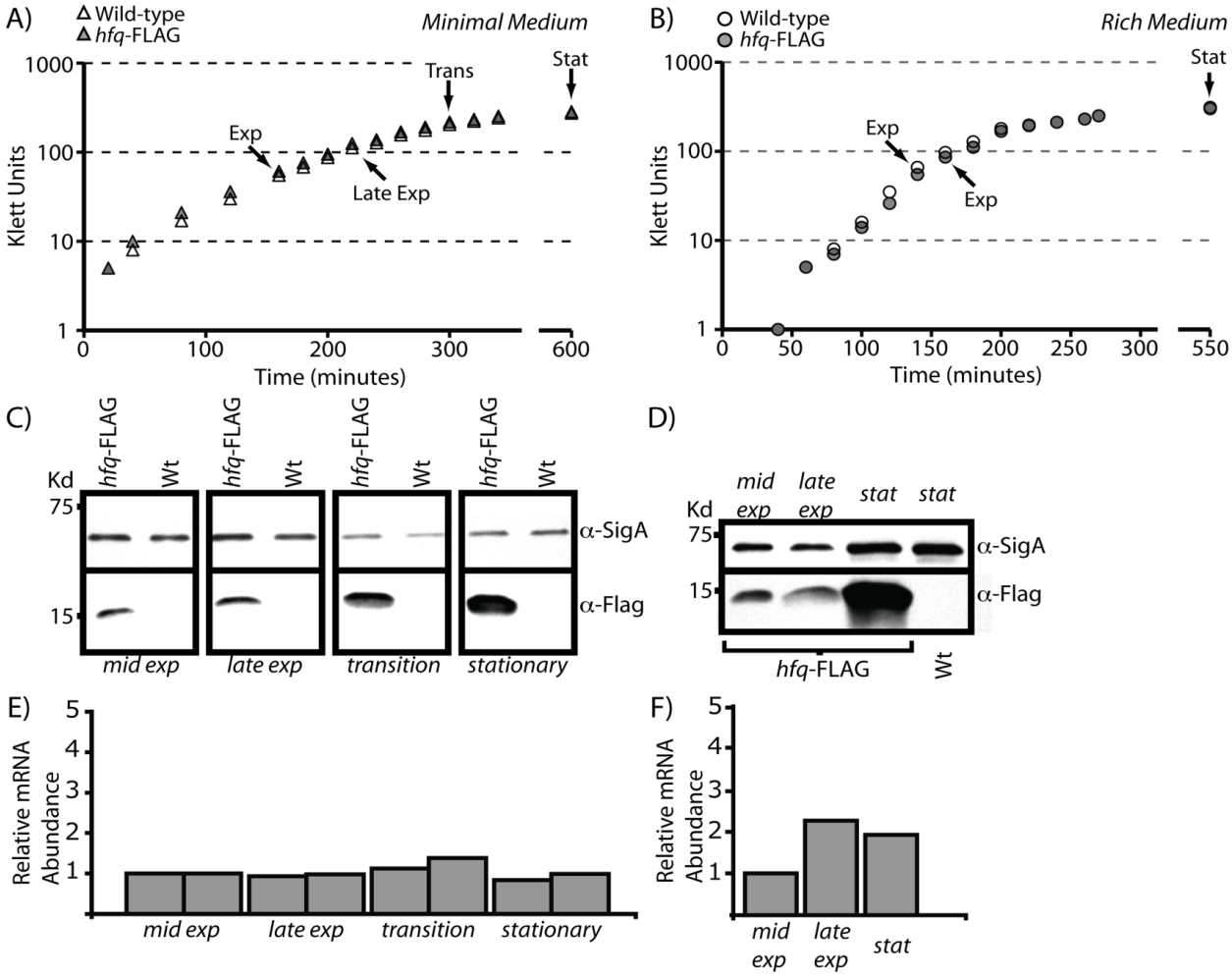


Figure 2-2. Expression Dynamics of Hfq during *B. subtilis* growth in liquid culture. (A, B) Growth curves for wild-type *B. subtilis* 168 and *hfq*^{FLAG} in glucose minimal media (GMM) and rich media respectively. Cells were diluted to an OD_{600} of 0.01 from overnight starter cultures in rich media and Klett readings were collected every 20 minutes. Each arrow corresponds to the position in the growth phase where samples were collected for subsequent analysis by western blotting and QPCR. (C, D) Samples collected from the positions denoted above were ran as whole cell lysates on a 4-20% SDS-PAGE mini-gel and subsequently transferred to nitrocellulose membranes. Each blot was probed with α-FLAG M2 monoclonal antibodies and developed using enhanced chemiluminescence (ECL). After development the blots were stripped and re-probed with α-Sigma A rabbit serum as a loading control. (E, F) QPCR expression profiles for *hfq* in wild type

168 cells grown in GMM and rich media respectively. Each data point was first normalized to the value obtained at mid exponential phase followed by Sigma A transcript levels corresponding to the respective phase on the growth curve. Values are reported as the change in number of cycles.

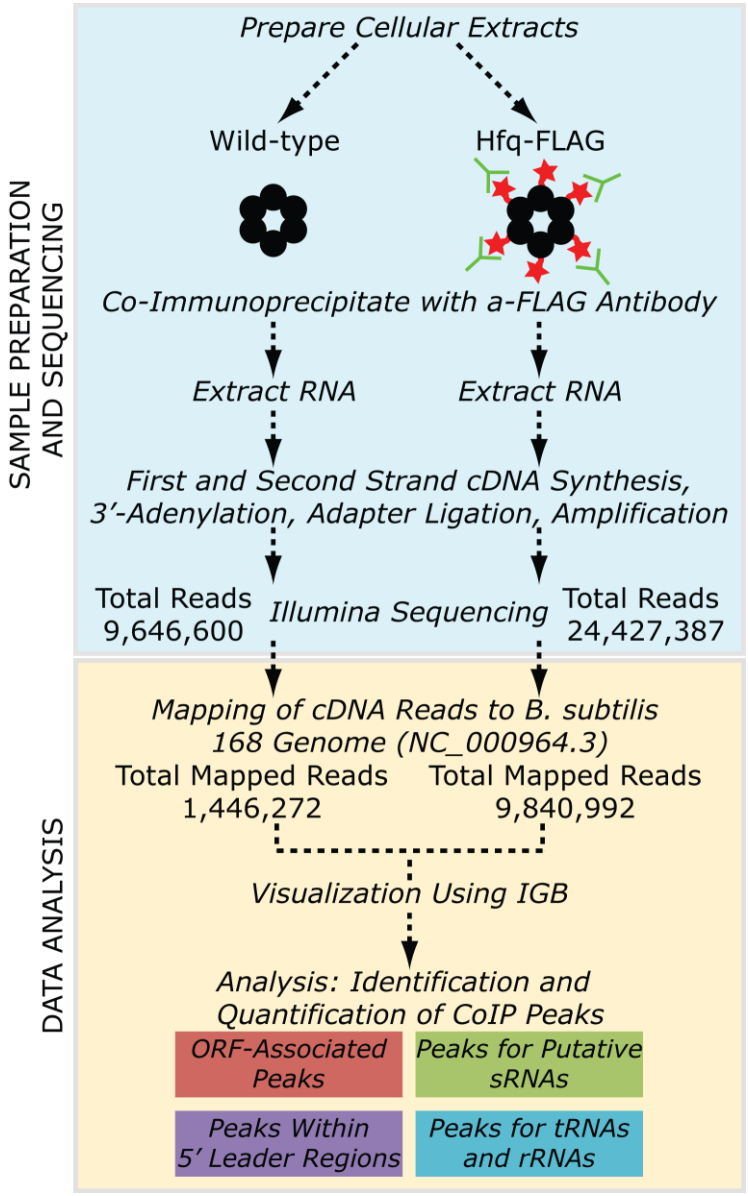


Figure 2-3. Co-immunoprecipitation of cellular RNAs that associate with *Bacillus subtilis* Hfq. Hfq^{FLAG} was co-immunoprecipitated from cellular extracts of stationary phase cells using a α-FLAG monoclonal antibody. A cDNA library was then created from the Hfq-associated RNA molecules, which was subjected to high-throughput sequencing using an Illumina Genome Analyzer.

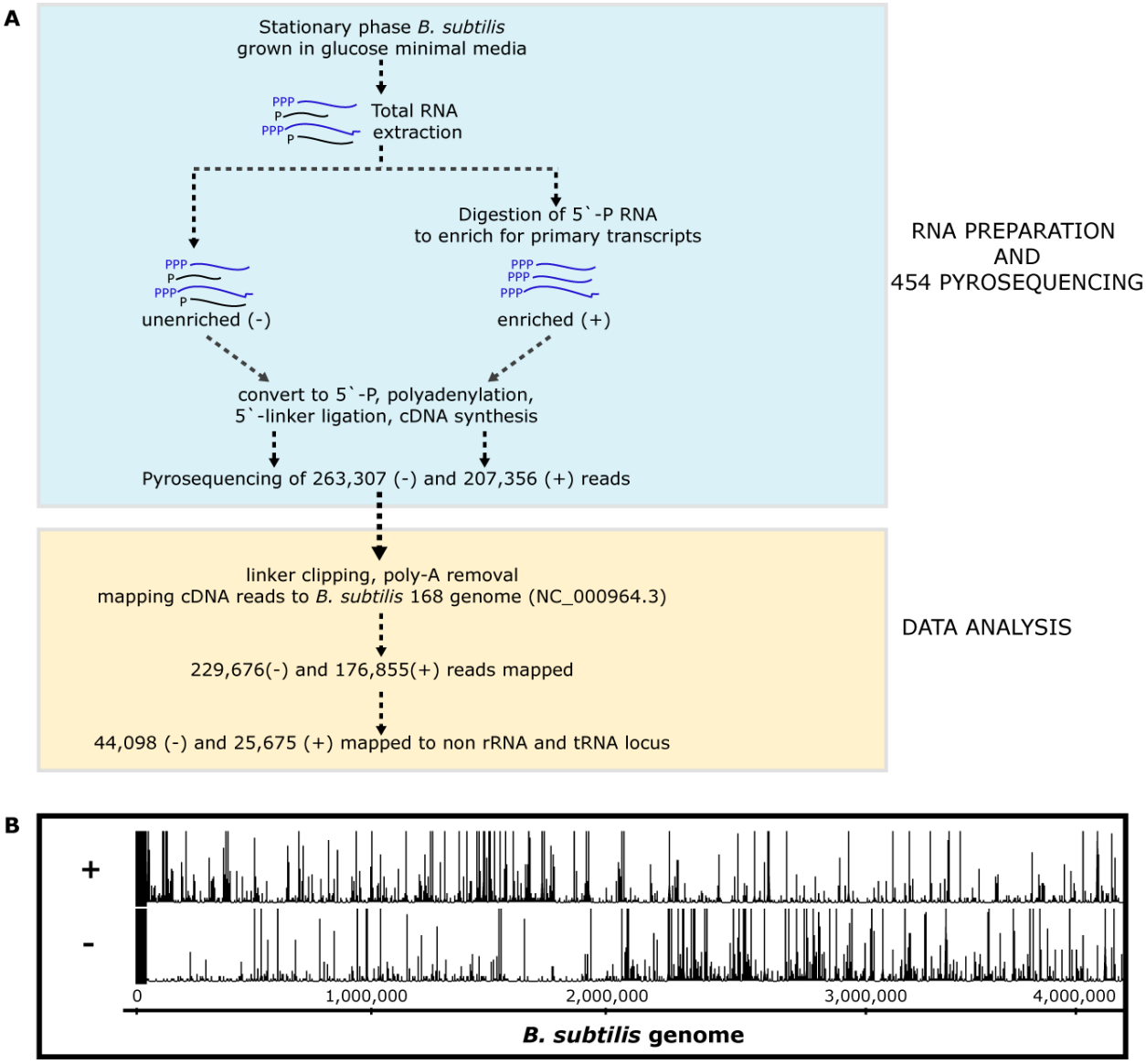


Figure 2-4 Overview of Transcription Start Site Mapping. (A) Schematic for differential RNA-sequencing (dRNA-Seq) analysis of *Bacillus subtilis* transcriptome (Adapted from Irnov et al. 2010). Total RNA was extracted from stationary phase culture of *B. subtilis* grown in minimal media. Total RNA was then converted into cDNA libraries with (+) or without (-) Terminator exonuclease treatment. The resulting libraries were sequenced on a Roche (454) FLX sequencer. cDNA reads were mapped onto *B. subtilis* 168 genome following the removal of linker sequence and poly-A tail. (B) Distribution of cDNA reads

across the *B. subtilis* genome shown using IGB. cDNA reads from TSS-enriched samples were mapped onto both positive (+) and negative (-) genomic DNA strands and shown as the number of cDNA hits per nucletotide. The distribution of cDNA reads mimics the genes distribution from each genomic strand (ie. Genes in the positive strand are enriched in the first half of the genome while genes in the negative strand are enriched in the second half).

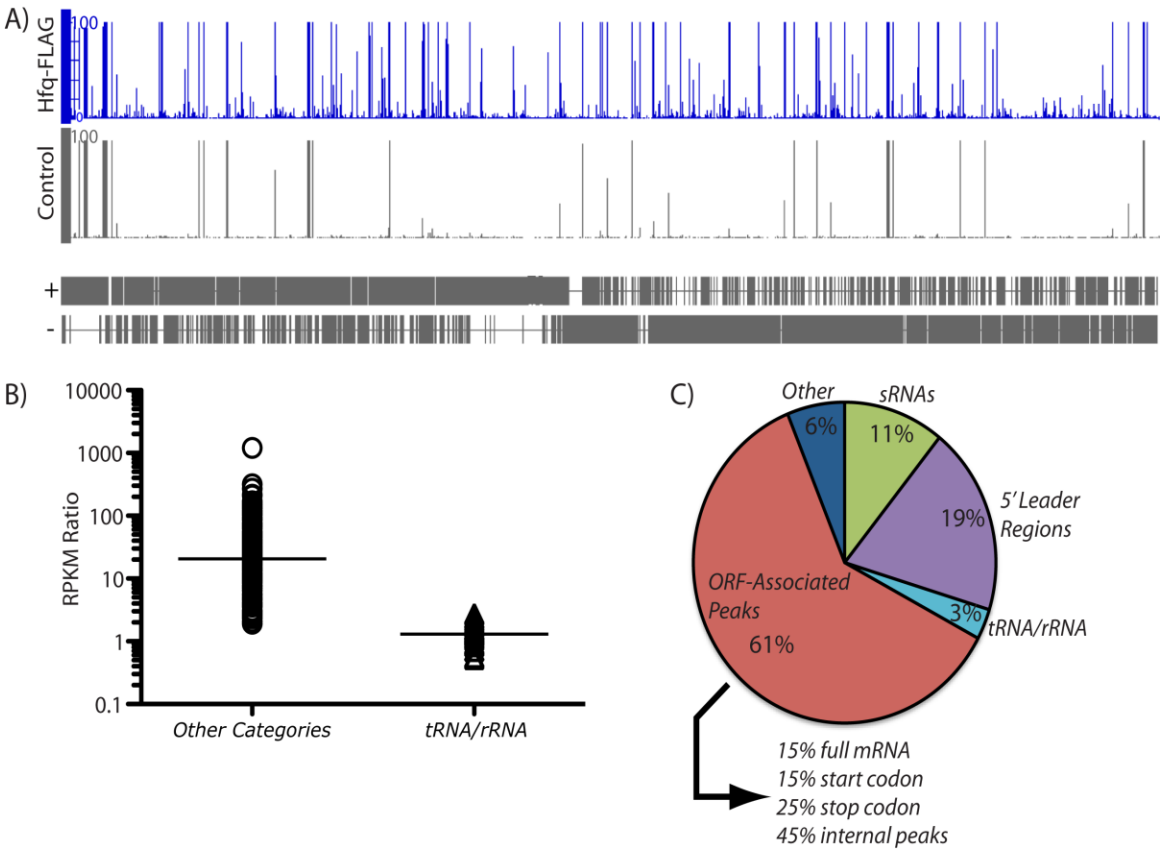


FIGURE 2-5. Co-immunoprecipitation of cellular RNAs that associate with *Bacillus subtilis* Hfq. (A) Hfq^{FLAG} was co-immunoprecipitated from cellular extracts of stationary phase cells using an α-FLAG monoclonal antibody. A cDNA library was then created from the Hfq-associated RNA molecules, which was subjected to high-throughput sequencing using an Illumina Genome Analyzer. The cDNA peaks that resulted from this analysis were evenly distributed across the genome for both the Hfq^{FLAG} and mock control samples. Many more peaks were obtained from the Hfq^{FLAG}-associated material. (B) The expression of each peak was quantified in reads per kilobase per million mapped reads, or 'RPKM' (Mortazavi *et al.*, 2008). The ratio of these values for the Hfq^{FLAG} and mock control samples was calculated in order to identify the peak subset that was enriched by association with Hfq. Almost all peaks that exhibited an RPKM ratio of < 2 (median = 1.2) corresponded to "housekeeping" RNAs, such as tRNAs, rRNAs, and the RNA subunit of RNase P (see Table S1 for details). The other peaks combined exhibited a range of RPKM ratios (median = 20). (C) The majority of the Hfq-enriched peaks (exhibiting an RPKM > 2) corresponded to portions of open reading frames (ORFs). Smaller subsets of peaks corresponded to 5' leader regions and putative small RNAs (sRNAs).

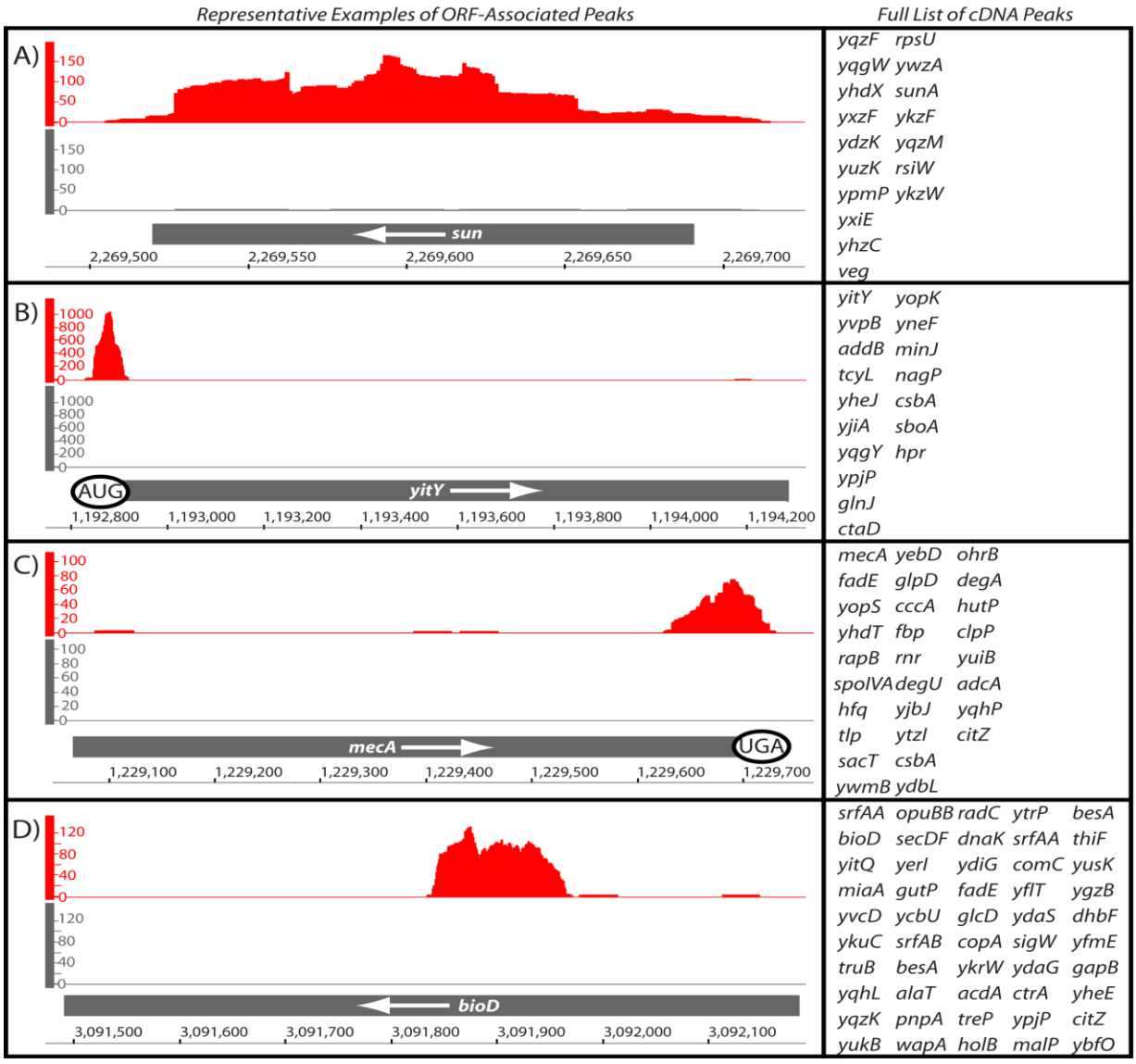


Figure 2-6. Representative Hfq-enriched peaks associated with open reading frames (ORF). (A) Representative data for a full mRNA sequence that was enriched by coIP of Hfq. A full list of genes sharing this pattern is shown to the right. (B) Representative data for an Hfq-enriched peak that overlaps with the translational start codon. Other genes sharing this arrangement are listed to the right. (C) Representative data for an Hfq-enriched peak that overlaps with the translation stop codon. Other genes sharing this arrangement are listed to the right. (D) Representative data for an Hfq-enriched peak located within an ORF.

Other genes sharing this arrangement are listed to the right. More details on these peaks are included in Table S2. The y-axis shows the volume of cDNA reads as a function of genomic position (x-axis) as determined by Illumina-based sequencing of Hfq-associated RNA molecules. Data shown in gray correspond to the negative control reaction (co-immunoprecipitation of Hfq that lacked an epitope tag).

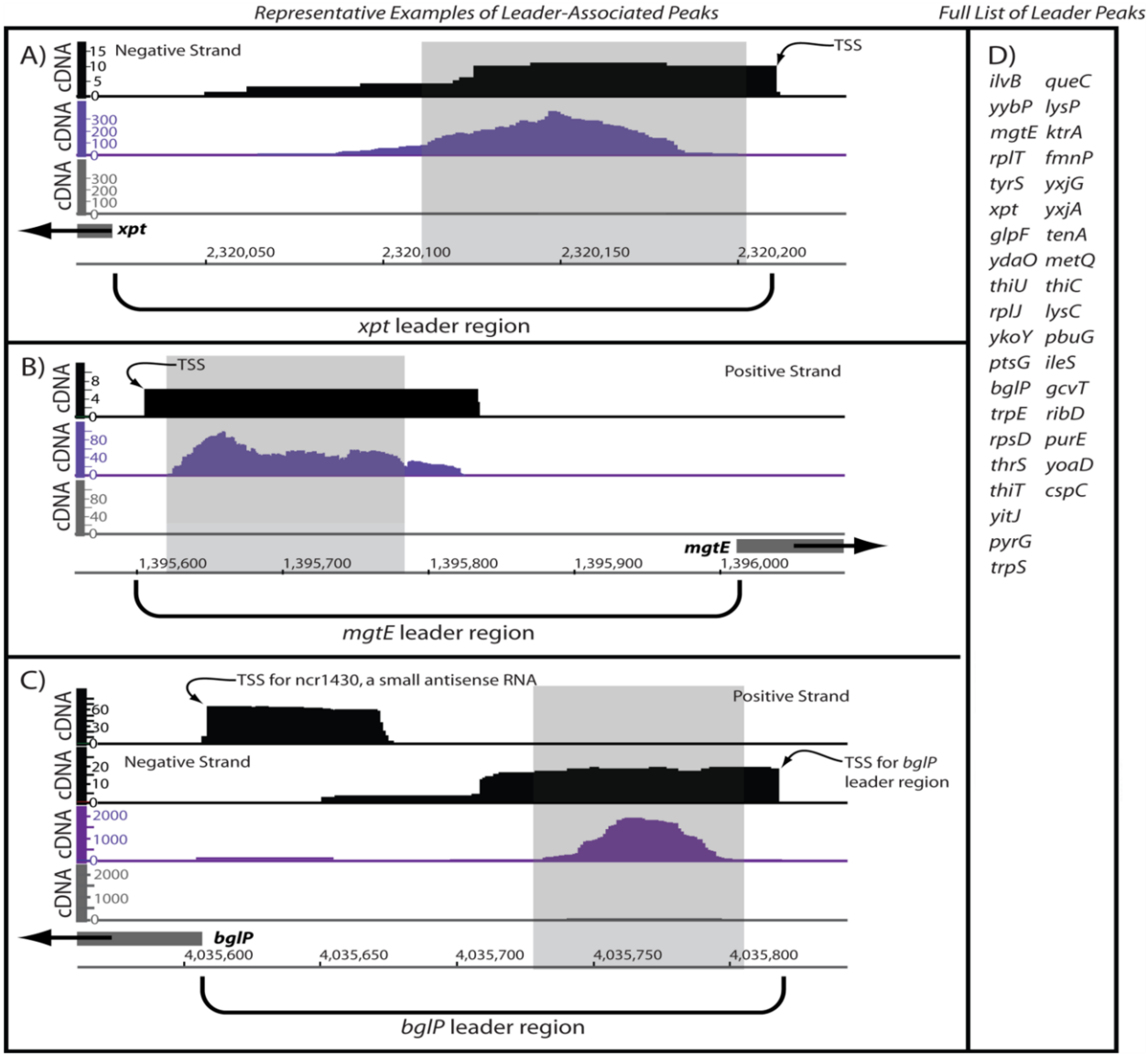


Figure 2-7. Hfq enrichment of signal-responsive, *cis*-acting regulatory RNAs. (A) Representative data are shown for Hfq-mediated enrichment of the leader region of a purine-sensing riboswitch (Mandal *et al*., 2004). The ligand-binding aptamer region is denoted by a shaded box to indicate the overlap between this sensory domain and the RNA sequences found to associate with Hfq. Previously measured transcription start site (TSS) mapping data (Irnov *et al*., 2010; shown in black) is also shown in the figure as a means of denoting the 5' end of the leader

region. (B) Representative data are shown for Hfq-mediated enrichment of the leader region of a magnesium-sensing riboswitch (Ramesh and Winkler, 2010). The magnesium-sensing domain (Ramesh *et al*., 2011) is denoted by a shaded box. (C) Representative data are shown for Hfq-mediated enrichment of the leader region of a protein-responsive *cis*-acting regulatory RNA. Specifically, the *bglP* operon has been previously postulated to be subjected to multiple layers of regulatory control, including by transcription initiation factors, a protein-responsive leader region, and a small, putative antisense RNA that appears to base pair with the ribosome binding site. Although the latter antisense RNA was modestly enriched by coIP with Hfq, the protein-binding aptamer region was highly enriched by Hfq in this study (shown in purple). (D) Other leader regions that associated with Hfq are listed herein. More details on these RNAs are included in Table 5. Throughout the figure, the y-axis shows the volume of cDNA reads as a function of genomic position (x-axis) as determined by Illumina-based sequencing of Hfq-associated RNA molecules. Data shown in gray correspond to the negative control reaction (co-immunoprecipitation of Hfq that lacked an epitope tag). Data shown in black correspond to a prior transcription start site (TSS) mapping study (Irnov *et al*., 2010).

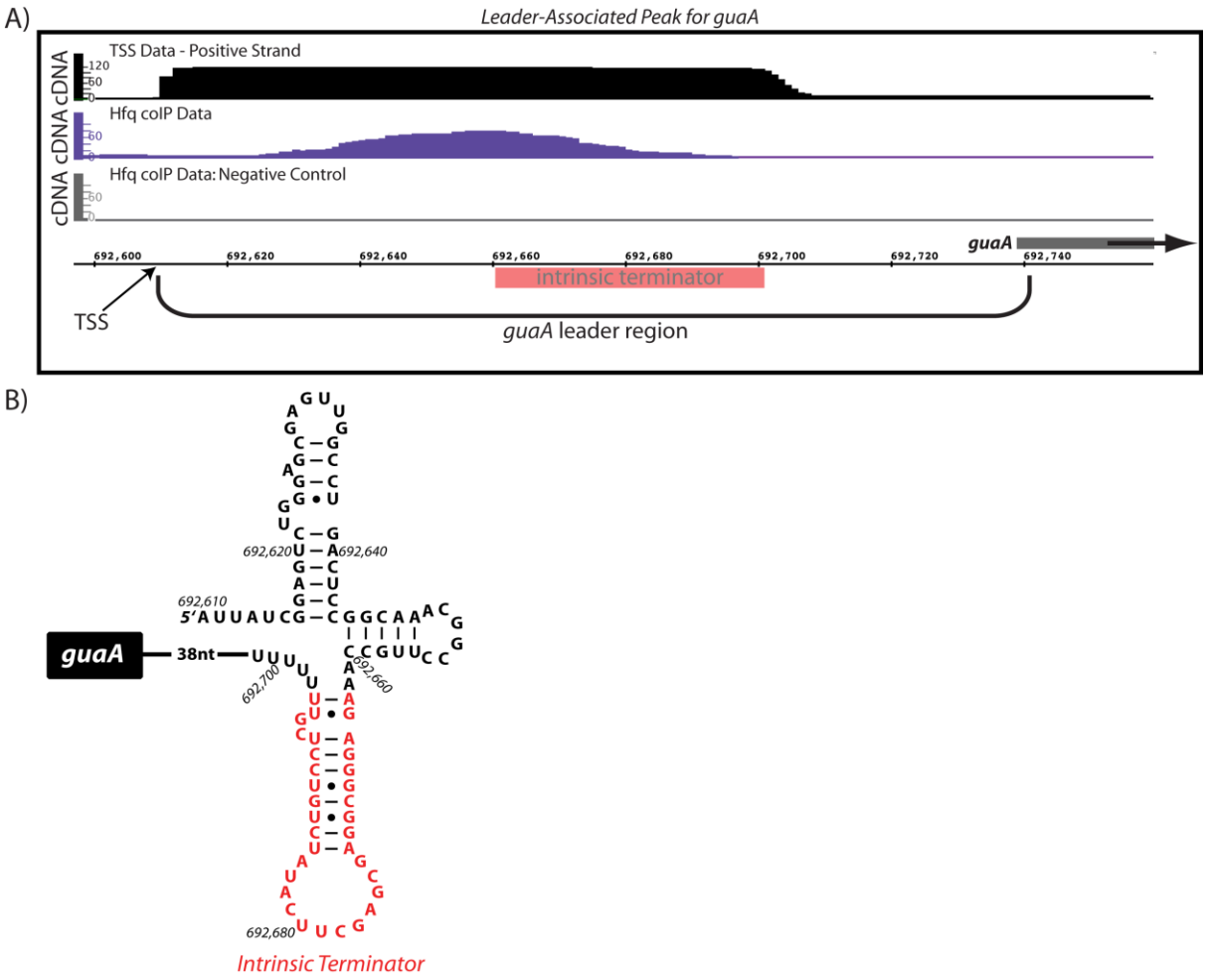


Figure 2-8. Discovery of a putative *cis*-acting transcription attenuation system upstream of *guaA*. (A) Hfq co-immunoprecipitation revealed a coIP peak upstream of the *guaA* gene. Our prior transcription start site mapping data (Irnov *et al*., 2010) revealed that a long leader region is situated upstream of *guaA*, consistent with earlier experimental evidence (Mäntsälä and Zalkin, 1992). This leader region essentially encompasses the Hfq coIP peak. This pattern is consistent with the many signal-responsive leader regions that were found to co-IP with Hfq in this study (see Figure 7 and Table 5 for more details). (B) Inspection of the *guaA* leader region revealed the presence of several putative secondary structural elements including a putative intrinsic transcription termination site. Most *cis*-acting regulatory RNAs in *B. subtilis* control gene expression by modulating transcription attenuation within a 5' leader region.

Therefore, we speculate that the presence of a premature termination site upstream of the *guaA* coding region is consistent with a transcription attenuation system, although experimentation will be required to test this hypothesis.

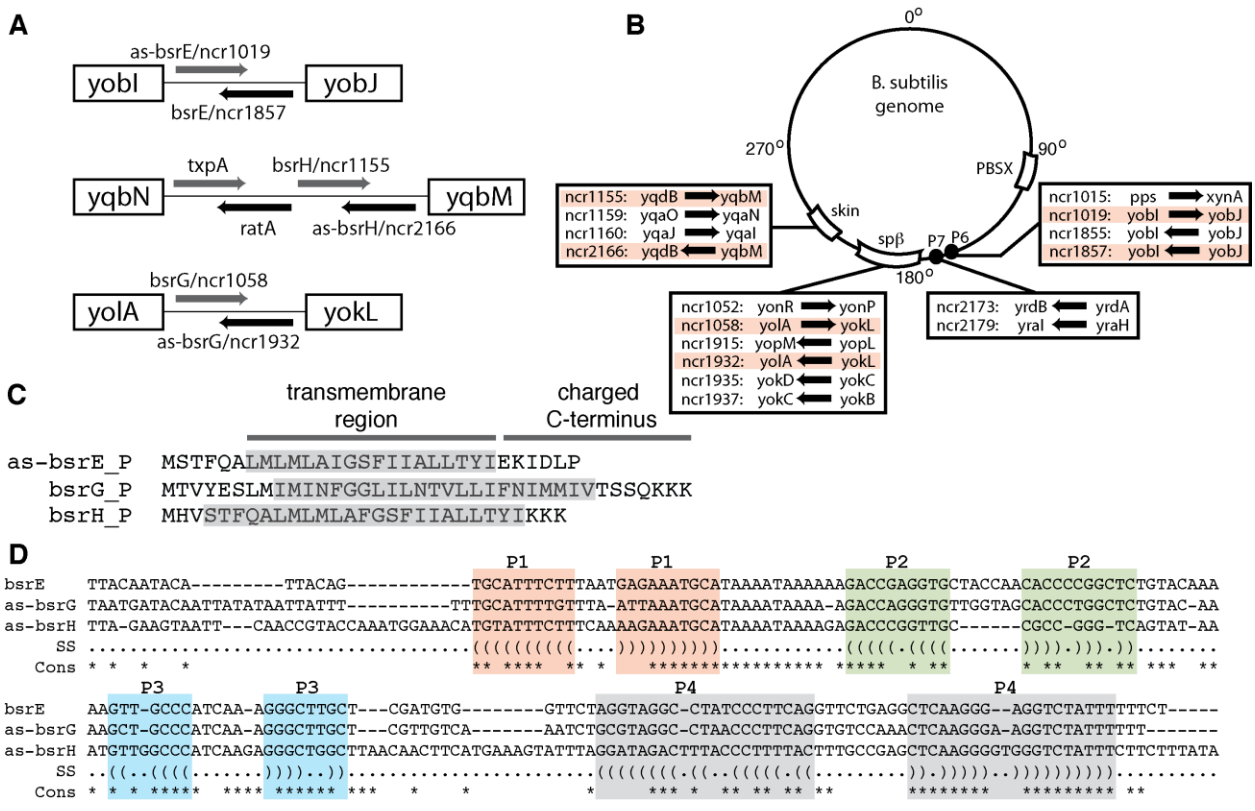


Figure 2-9. The intergenic regions, and genetic configuration of the four previously identified Type-I toxin/ antitoxin systems in *B. subtilis* (Adapted from Irnov *et al*., 2010). (A) Depicts the arrangement of the four type-I toxin/ antitoxin systems, in all instances the toxic protein (gray arrow) and the RNA antitoxin (black arrow) are arranged in a tail-to-tail configuration with the toxin residing on the sense strand and the antitoxin on the corresponding antisense strand (Silvaggi *et al*., 2005; Irnov *et al*., 2010). (B) Putative sRNAs encoded within prophage regions. 16 sRNA candidates (denoted by arrow) originating from prophage or prophage-like regions (*SPβ*, *skin*, *P6*, *P7*) are shown relative to their genomic location. Genes immediately upstream and downstream of the sRNA are also listed. (C) Putative sequences for *as-bsrE*, *bsrG* and *bsrH* toxins. Predicted membrane spanning regions are highlighted in gray. (D) Sequence alignment of the *bsrE*, *as-bsrG* and *as-bsrH* RNA antitoxins. Regions with base-pairing potentials are shown with different colors and labeled as P1–P4.

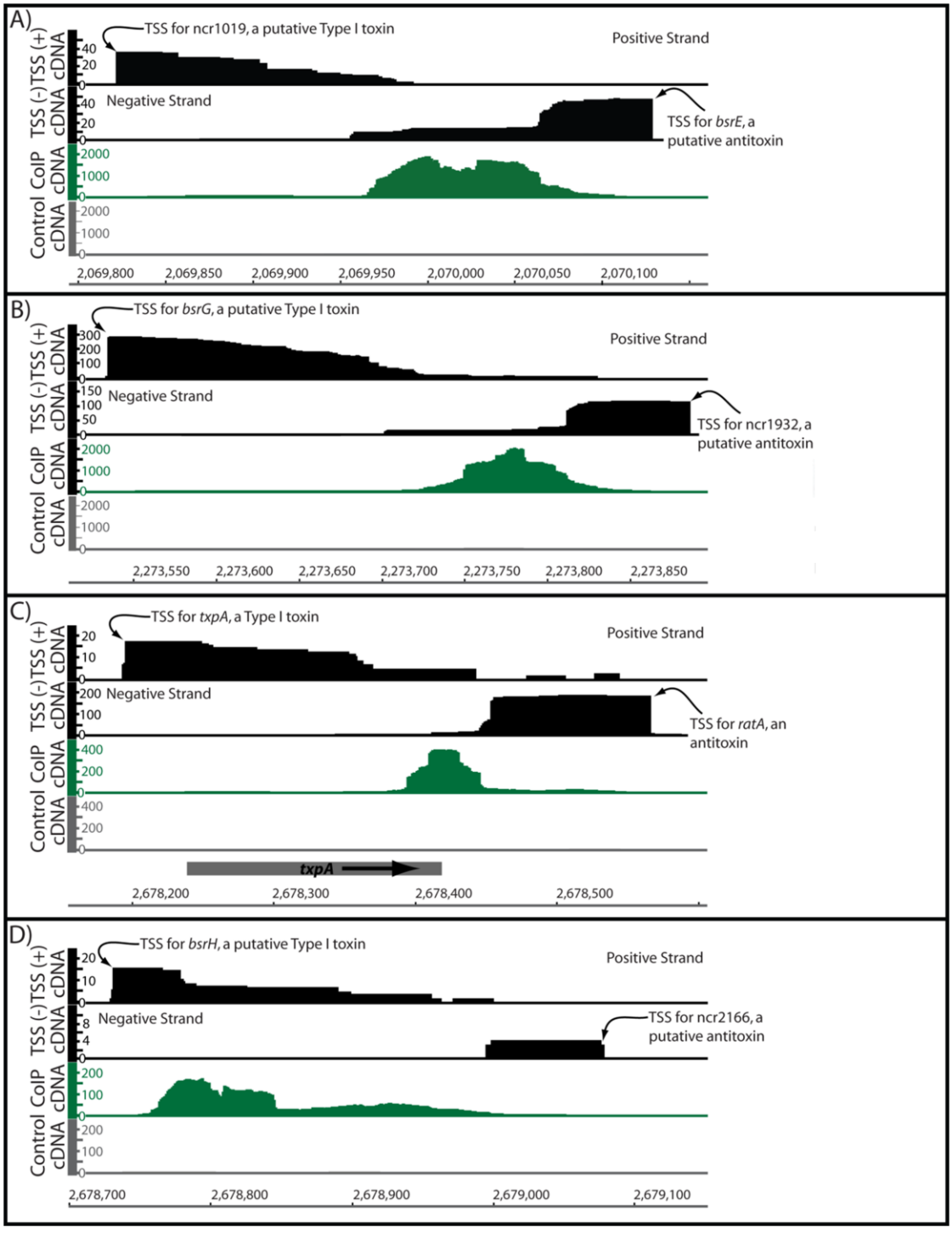

A)
TSS for ncr1019, a putative Type I toxin
Positive Strand
TSS (+) cDNA
40
20
0
TSS (-) cDNA
Negative Strand
40
20
0
TSS for bsrE, a putative antitoxin
Control cDNA
2000
1000
0
CoIP cDNA
2000
1000
0
2,069,800
2,069,850
2,069,900
2,069,950
2,070,000
2,070,050
2,070,100
B)
TSS for bsrG, a putative Type I toxin
Positive Strand
TSS (+) cDNA
300
200
100
0
TSS (-) cDNA
Negative Strand
150
100
50
0
TSS for ncr1932, a putative antitoxin
Control cDNA
2000
1000
0
CoIP cDNA
2000
1000
0
2,273,550
2,273,600
2,273,650
2,273,700
2,273,750
2,273,800
2,273,850
C)
TSS for txpA, a Type I toxin
Positive Strand
TSS (+) cDNA
20
10
0
TSS (-) cDNA
Negative Strand
200
100
0
TSS for ratA, an antitoxin
Control cDNA
400
200
0
CoIP cDNA
400
200
0
txpA
2,678,200
2,678,300
2,678,400
2,678,500
D)
TSS for bsrH, a putative Type I toxin
Positive Strand
TSS (+) cDNA
20
10
0
TSS (-) cDNA
Negative Strand
8
4
0
TSS for ncr2166, a putative antitoxin
Control cDNA
200
100
0
CoIP cDNA
200
100
0
2,678,700
2,678,800
2,678,900
2,679,000
2,679,100

Figure 2-10. Hfq associates with type I antitoxin transcripts. (A-D) A previous study of *B. subtilis* transcription start sites (TSS) (Irnov *et al*., 2010) revealed the presence of at least four putative type I toxin/ antitoxin systems. The TSS data from this study are shown in black for the purpose of indicating the orientation of these RNA transcripts, relative to regions of Hfq enrichment. In this study, Hfq appeared to preferentially associate with antisense transcripts. In particular, the sequences that appeared to coIP with Hfq corresponded most often to the portions of the antitoxin transcripts that are predicted to base pair with the toxin-encoding mRNAs. Throughout the figure, the y-axis shows the volume of cDNA reads as a function of genomic position (x-axis) as determined by Illumina-based sequencing of Hfq-associated RNA molecules. Data shown in gray correspond to the negative control reaction (co-immunoprecipitation of Hfq that lacked an epitope tag). Data shown in black correspond to a prior transcription start site (TSS) mapping study (Irnov *et al*., 2010).

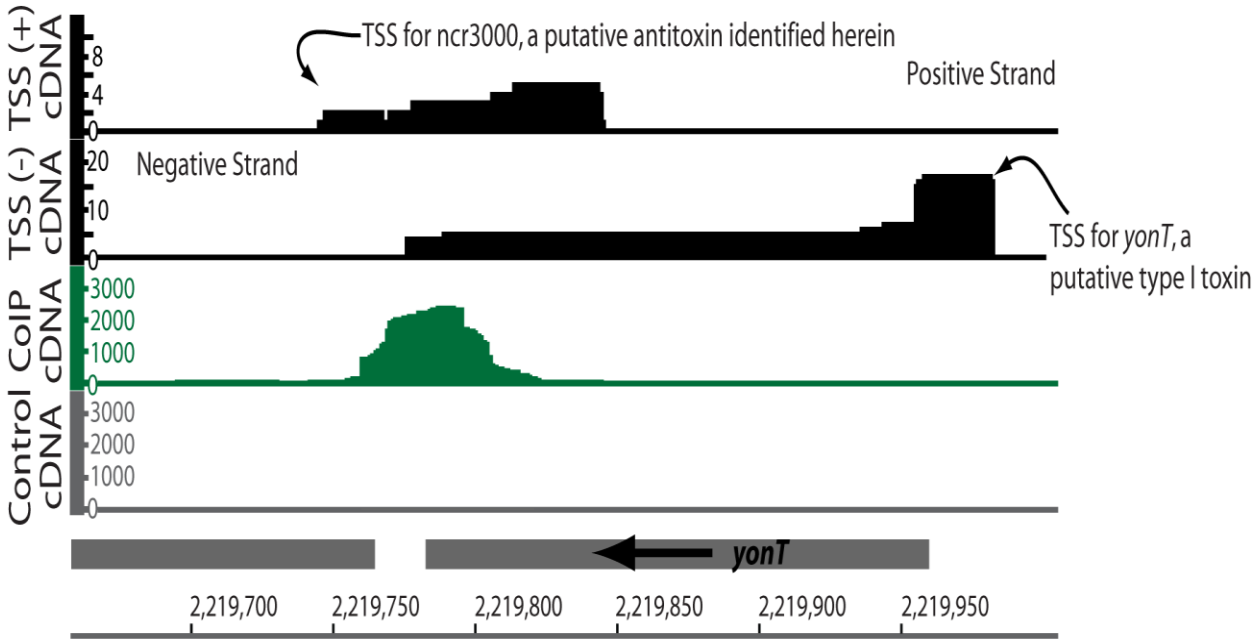


Figure 2-11. Expression of a possible antitoxin for the *yonT* type I toxin. Recently, several putative type I toxins were identified in the *B. subtilis* genome (Fozo *et al*., 2010), including the *yonT* gene. Our analysis of the Hfq coIP data revealed that four examples of previously identified type I antitoxins appeared to exhibit preferential enrichment by Hfq. Inspection of the putative *yonT* toxin gene revealed an Hfq-associated peak located in the region where an antitoxin transcript would be most likely to occur. Therefore, we speculate that the Hfq-associated peak that overlaps *yonT* might correspond to an antitoxin transcript.

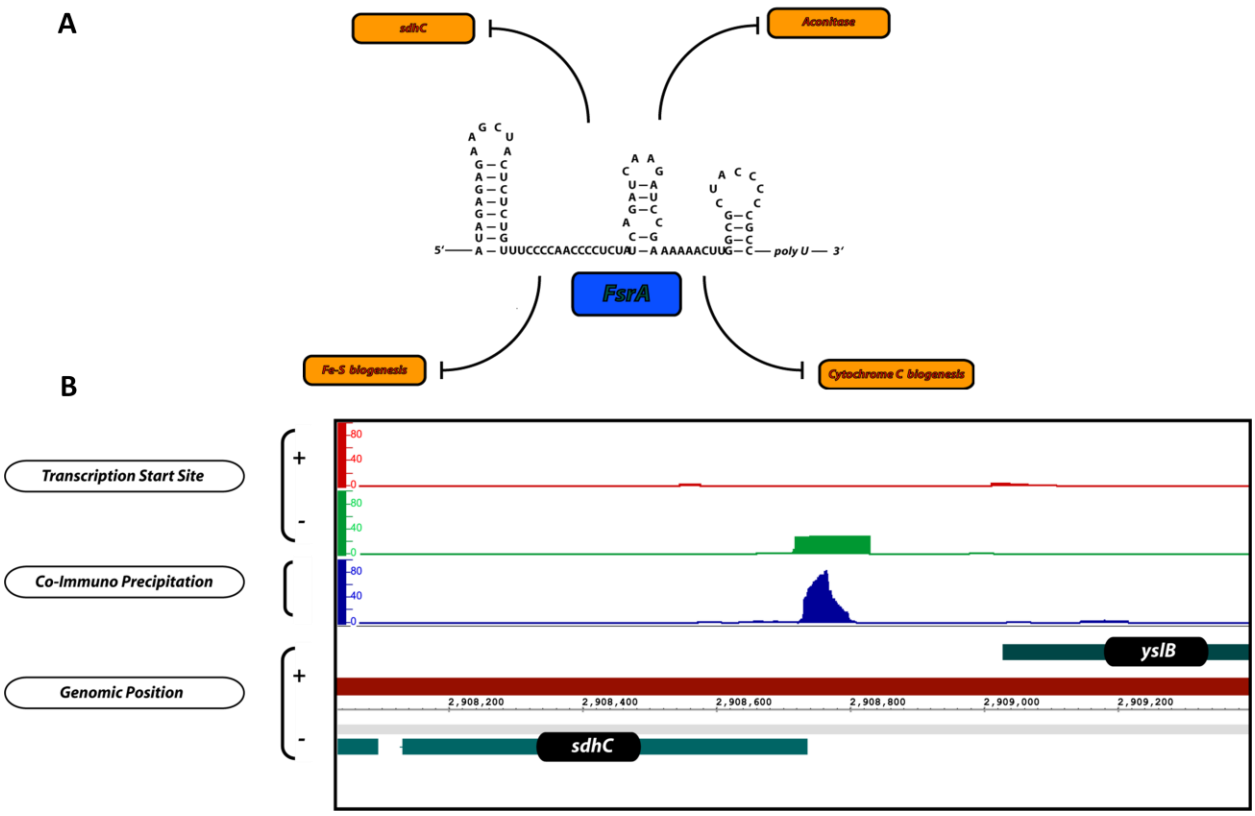


Figure 2-12. The *sdhC* 5′ UTR associates with Hfq and is a target of the sRNA regulator FsrA. (A) Previously, a sRNA regulator (FsrA) was identified as an important regulator of iron homeostasis genes in *Bacillus subtilis* (Gaballa *et al*., 2008). FsrA is derepressed in response to iron-limitation and down regulates Iron transcripts encoding proteins which utilize iron, such as *sdhC.* (B) The well-characterized mRNA target of the FsrA sRNA is the ribosome binding site region of *sdhC,* which is enriched in the Hfq co-IP (Gaballa *et al*., 2008).

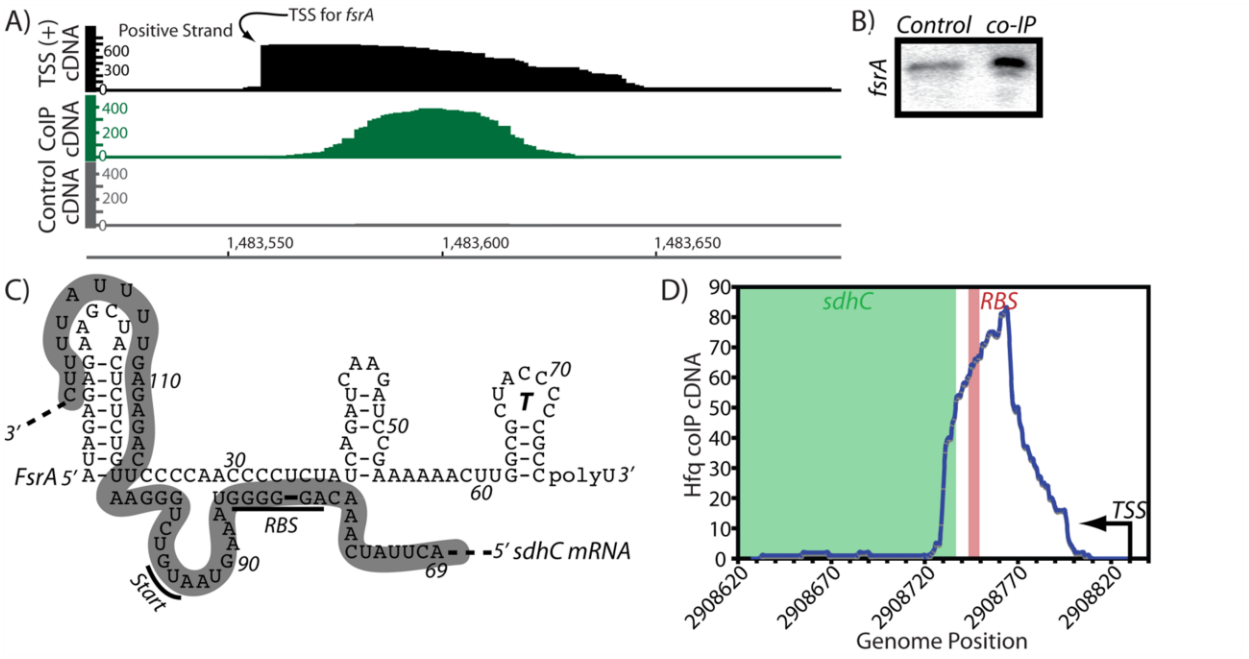


Figure 2-13. Hfq associates with the iron homeostasis sRNA regulator, FsrA. The transcription start site of *fsrA* was identified previously (Gaballa *et al*., 2008; Irnov *et al*., 2010) and is shown in panel (A). In this study, Hfq is shown to coIP with the FsrA sRNA, although the functional relevance of this interaction remains to be examined. In this panel, the y-axis shows the volume of cDNA reads as a function of genomic position (x-axis) as determined by Illumina-based sequencing of Hfq-associated RNA molecules. Data shown in gray correspond to the negative control reaction (coimmunoprecipitation of Hfq that lacked an epitope tag). Data shown in black correspond to a prior transcription start site (TSS) mapping study (Irnov *et al*., 2010). (B) Deletion of *hfq* resulted in moderately lowered abundance of FsrA, as ascertained by Northern Blotting analysis. (C) The predicted sRNA:mRNA pairing region is shown herein. The *sdhC* leader region is indicated in this figure by gray shading. (D) The region of the *sdhC* transcript that is predicted to pair with FsrA also appeared to coIP with Hfq in this study. The *sdhC* transcription start site is indicated with an arrow. The ribosome binding site is denoted by a red box and the *sdhC* coding region is shown in green. One speculative explanation for these data is that Hfq might assist FsrA:*sdhC* intermolecular interactions, although other explanations are still possible.

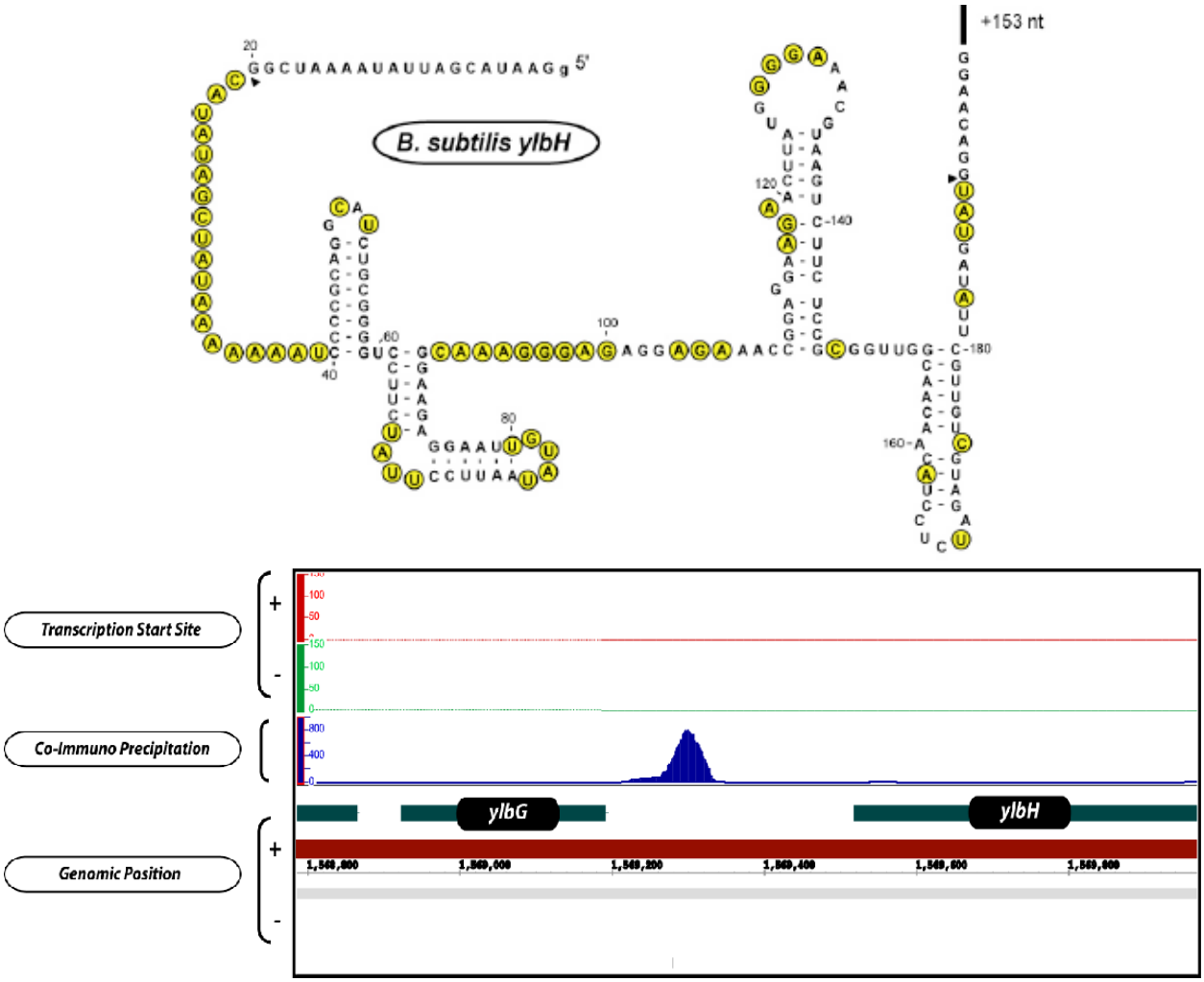


Figure 2-14. The orphan riboswitch ylbH co-immunoprecipitates with Hfq. (A) The proposed secondary structure of a previously reported putative orphan riboswitch identified upstream of the *ylbH* gene in certain *Bacillaceae* (Barrick *et al*., 2004; Gardner *et al*., 2009). Yellow residues correspond to areas of flexibility within the RNA by structural probing (adapted from Barrick *et al*., 2004). (B) Depicts the mapped reads associated with Hfq corresponding to this RNA element.

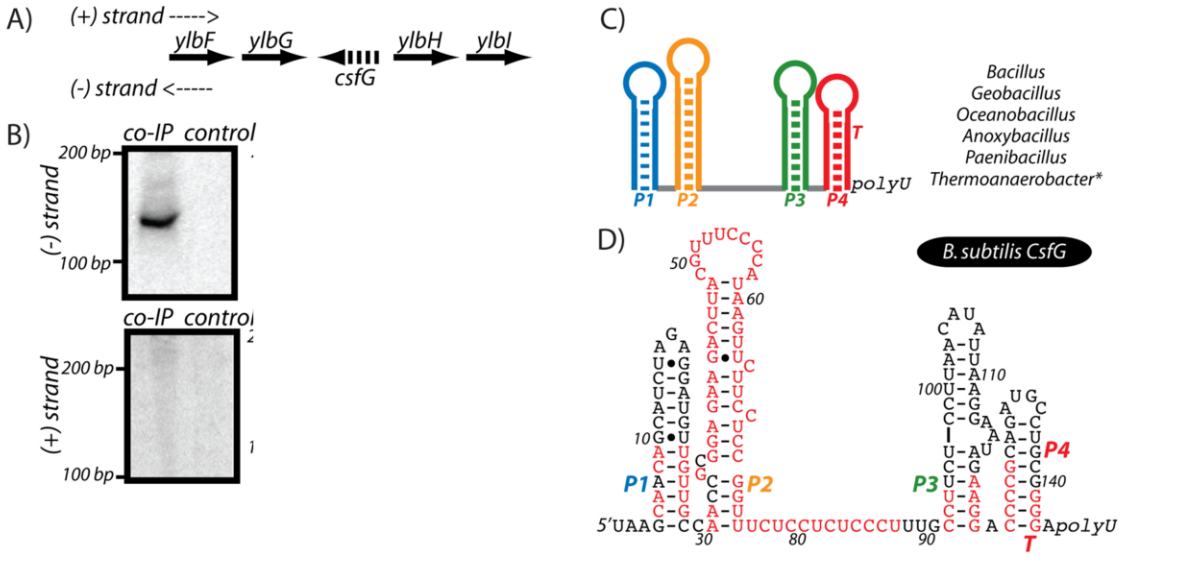


Figure 2-15. Co-immunoprecipitation of CsfG, an sRNA conserved in *Bacillaceae*, with Hfq. (A-B) Previously, an orphan riboswitch was identified upstream of the *ylbH* gene in certain *Bacillaceae* (Barrick *et al*., 2004; Gardner *et al*., 2009). In this study, we find that this RNA element is not likely to be conserved as a 5' leader region for the *ylbH* gene, but is instead transcribed from the reverse genomic strand as a putative sRNA, which has been recently coined *csfG* (Marchais *et al*., 2011). This region was detected by northern blot analyses using oligonucleotide probes to hybridize to the region upstream of the *ylbH* gene (+ strand) and to the reverse complement of that sequence (- strand). A signal was only observed for an RNA sequence transcribed from the reverse complement strand, consistent with synthesis of CsfG. (C-D) Consensus secondary structure diagram of the recently identified sRNA, CsfG (Marchais *et*

al., 2011; data not shown). Residues in red represent 95% conservation. This primary sequence and secondary structure consensus pattern was generated from covariance model searches (data not shown).

Table 2-1 Putative sRNAs that coIP with Hfq.

Annotation	Peak Length	Peak Coord	Left-Right Genes	RPKM Ratio[a]	Ref[b]
bsrC	146	474265-474410	*ydaG-ydaH*	53.6	(1)
bsrE/ncr1857[e]	275	2069840-2070114	*yoyA-yobJ*	50.9	(2)
shd60[c]	87	2190671-2190757	*Anti-yoqZ*	48.2	(3)
ncr982	81	1917500-1917580	*fosB-lexA*	47.7	(2)
ncr952	106	1780422-1780527	*mutL-ymzD*	45.0	(2)
ncr1430	50	4035605-4035654	*bglP-yxxE*	36.6	(2)
ncr3000[c,e]	103	2219742-2219844	*Anti-yonT*	35.7	New
RatA[e]	199	2678350-2678548	*Anti-txpA*	33.5	(2)
ncr3001	125	972865-972989	*yhbF-prkA*	24.4	New
csfG	136	1569212-1569347	*ylbG-ylbH*	22.9	(7)

ncr2184/ncr60	234	2779115-2779348	*yrzI-yrhG*	22.2	(2, 3)
ncr2166/asbsrH [e]	301	2678750-2679050	*yqdB-yqbM*	20.7	(2)
ncr1932/asbsrG [e]	311	2273552-2273862	*yolA-yokL*	16.4	(2)
ncr560/ncr18	217	1056393-1056609	*yhaZ-yhaX*	13.9	(2, 3)
FsrA	73	1483560-1483632	*ykuI-ykuJ*	11.1	(4)
ncr3002 [d]	80	1447035-1447114	*ykzR-ykvR*	11.1	New
SR2/BsrF	93	2079102-2079194	*yobO-csaA*	7.75	(2, 5)
RsaE	83	1233446-1233528	*yizD-yjbH*	3.97	(2, 3, 6)
ncr471	151	820667-820817	*yfmI-yfmG*	3.56	(2)
ncr3003	101	1609113-1609213	*ylmC-ylmD*	3.09	New
ncr1670	256	1077037-1077292	*hinT-ecsA*	2.77	(2)
ncr1015	75	2054012-2054086	*pps-xynA*	2.35	(2)

[a] The expression of each peak (described in the main text) was quantified in reads per kilobase per million mapped reads, or 'RPKM' (Mortazavi et al., 2008). The ratio of these values for the Hfq^{FLAG} and mock control samples was taken as an indicator of Hfq-mediated enrichment.

[b] References:

(1) Saito S, Kakeshita H, Nakamura K (2009) Novel small RNA-encoding genes in the intergenic regions of *Bacillus subtilis*. *Gene* 428:2-8;

(2) Irnov I, Sharma CM, Vogel J, Winkler WC (2010) Identification of regulatory RNAs in *Bacillus subtilis. Nucleic Acids Res* 38:6637-6651;

(3) Rasmussen S, Nielsen HB, Jarmer H (2009) The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol Microbiol* 73:1043-1057;

(4) Gaballa A, Antelmann H, Aguilar C, Khakh SK, Song KB, Smaldone GT, Helmann JD (2008) The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci USA* 105:11927-11932;

(5) Preis H, Eckart RA, Gudipati RK, Heidrich N, Brantl S (2009) CodY activates transcription of a small RNA in *Bacillus subtilis*. *J Bacteriol* 191:5446-57;

(6) Geissmann T, Chevalier C, Cros MJ, Boisset S, Fechter P, Noirot C, Schrenzel J, Francois P, Vandenesch F, Gaspin C, Romby P (2009) A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic Acids Res* 37:7339-57;

(7) Marchais A, Duperrier S, Durand S, Gautheret D, Stragier P (2011) CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers. *RNA Biol* 8:358-364;

[c] These sRNAs are arranged such that they appear to consist of putative antisense transcripts.

[d] This peak is adjacent but not overlapping with the previously identified sRNA candidate, ncr721/ncr34 (Irnov *et al*., 2010; Rasmussen *et al*., 2009).

[e] these peaks correspond to previously identified type I toxin:antitoxin systems (Irnov *et al*., 2010; Fozo *et al*., 2010).

Table 2-2 Other peaks.

Annotation	Peak Length	Peak Coord	Genes	RPKM Ratio[a]
New sRNA or *yrhE* leader region[b]	139	2781029-2781167	*yrhF-yrhE*	170
New sRNA or between *xre-xkdA*[c]	87	1321203-1321289	*xkdA-xre*	150
Uncharacterized *srfAA* leader region	150	376704-376853	*hxlR-srfAA*	134
Uncharacterized *sdhC* leader region	87	2908722-2908808	*sdhC-yslB*	106
New sRNA or *yybS* leader region[b]	92	4166607-4166698	*yybS-yyzH*	86.1
Uncharacterized *ypzK* leader region	54	2427780-2427833	*ypzK-ribH*	75.3
Uncharacterized *guaA*	125	692572-	*yebA-*	73.2

leader region		692696	*guaA*	
Uncharacterized *yqhQ* leader region	171	2540947-2541117	*yqhR-yqhQ*	23.2
Unknown; Possibly a 3' UTR for *folD*	69	2528316-2528384	*yqiB-folD*	22.1
Conserved RNA in the *dagK* 3' UTR[d]	112	737436-737547	*dagK-yefA*	18.1
Unknown or *speD* leader region[b]	79	2966841-2966919	*speD-gapB*	14.1
Uncharacterized *yxjB* leader region	174	4005159-4005332	*yxjB-yxjA*	7.05

[a] The expression of each peak (described in the main text) was quantified in reads per kilobase per million mapped reads, or 'RPKM' (Mortazavi *et al*., 2008). The ratio of these values for the Hfq^{FLAG} and mock control samples was taken as an indicator of Hfq-mediated enrichment.

[b] A previous report suggested that these genes have a leader region which encompasses the Hfq CoIP peaks reported herein (Rasmussen *et al*., 2009)

[c] This peak appears to correspond to either a previously unidentified sRNA candidate, or a portion of the intercistronic region between *xre* and *xkdA*.

[d] A conserved RNA element was discovered within the 3' UTR of several transcripts in *B. subtilis*, including *dagK*. The function of this RNA structural element remains unknown (Rasmussen *et al*., 2009).

Table 2-3 tRNA peaks and Hfq coIP peaks with an RPKM ratio of <2.

Hfq coIP peaks with an RPKM ratio of <2				
Gene	**Peak Length**	**Coordinates**	**Coordinates**	**RPKM Ratio[a]**
trnQ-Arg	97	3545879	3545975	2.00
trnSL-Ser1	164	22289	22452	1.89
trnS-Leu2	123	529311	529433	1.78
trnD-Asp	254	952036	952289	1.77
trnD-Phe	254	952036	952289	1.77
trnS-Asn	570	528691	529260	1.74
trnS-Glu	570	528691	529260	1.74
trnS-Gln	570	528691	529260	1.74
trnS-Lys	570	528691	529260	1.74
trnB-Thr	98	3173596	3173693	1.68
trnSL-Met1	187	70168	70354	1.56
trnD-Gly	289	952699	952987	1.56
trnD-Leu1	289	952699	952987	1.56
trnD-Glu	251	951782	952032	1.54
trnD-Val	251	951782	952032	1.54
trnJ-Gly	455	95772	96226	1.51

trnJ-Arg	455	95772	96226	1.51
trnJ-Pro	455	95772	96226	1.51
trnJ-Ala	455	95772	96226	1.51
trnI-Arg	273	166061	166333	1.49
trnI-Pro	273	166061	166333	1.49
trnI-Ala	273	166061	166333	1.49
trnB-Met3	793	3172791	3173583	1.48
trnB-Ala	793	3172791	3173583	1.48
trnB-Arg	793	3172791	3173583	1.48
trnB-Leu2	793	3172791	3173583	1.48
trnB-Gly1	793	3172791	3173583	1.48
trnB-Leu1	793	3172791	3173583	1.48
trnSL-Ala1	91	3194446	3194536	1.48
yxzE	85	3983070	3983154	1.47
trnSL-Glu2	466	194160	194625	1.43
trnSL-Tyr1	466	194160	194625	1.43
trnA-Ile	188	31915	32102	1.39
trnA-Ala	188	31915	32102	1.39
trnD-Tyr	178	952299	952476	1.37
trnB-Phe	366	3172424	3172789	1.37
trnB-Asp	366	3172424	3172789	1.37

trnB-Met2	366	3172424	3172789	1.37
trnB-Ser1	366	3172424	3172789	1.37
trnD-Leu2	109	953192	953300	1.34
trnO-Ile	179	11455	11633	1.34
rrnI-16S	1559	160888	162446	1.33
rrnJ-16S	1592	90500	92091	1.32
rrnH-16S	1593	166461	168053	1.32
rrnA-16S	1559	30275	31833	1.32
rrnW-16S	1560	96388	97947	1.31
rrnD-16S	1563	946690	948252	1.31
component	327	26391	26717	1.31
rrnO-16S	1564	9803	11366	1.30
trnB-Ile2	273	3172150	3172422	1.30
trnB-Gly2	273	3172150	3172422	1.30
trnB-His	273	3172150	3172422	1.30
rrnE-16S	1569	635421	636989	1.30
rrnG-16S	1593	171459	173051	1.27
rrnB-16S	1569	3177084	3178652	1.27
trnD-His	174	952484	952657	1.25
trnD-Gln	174	952484	952657	1.25
trnSL-Arg2	108	2899808	2899915	1.24

trnSL-Gly1	93	967051	967143	1.23
trnI-Asn	152	165754	165905	1.18
trnB-Asn	274	3171873	3172146	1.18
trnB-Glu	274	3171873	3172146	1.18
trnE-Arg	170	635109	635278	1.13
trnE-Gly	170	635109	635278	1.13
trnSL-Val2	115	1262757	1262871	1.10
rrnJ-5S	507	95232	95738	1.03
trnJ-Val	507	95232	95738	1.03
SAM	200	1424502	1424701	0.96
trnD-Asn	309	951451	951759	0.92
rrnD-5S	309	951451	951759	0.92
rrnD-5S	309	951451	951759	0.92
trnB-Val	231	3173710	3173940	0.88
rrnB-5S	231	3173710	3173940	0.88
trnB-Val	231	3173710	3173940	0.88
rrnB-5S	231	3173710	3173940	0.88
trnSL-Arg1	78	2003272	2003349	0.86
rrnE-5S	220	640132	640351	0.83
trnE-Met	220	640132	640351	0.83
rrnE-5S	220	640132	640351	0.83

trnE-Met	220	640132	640351	0.83
rrnO-5S	140	14673	14812	0.71
rrnO-5S	140	14673	14812	0.71
rrnH-5S	136	171180	171315	0.71
rrnH-5S	136	171180	171315	0.71
rrnA-5S	133	35224	35356	0.71
rrnA-5S	133	35224	35356	0.71
rrnI-5S	127	165585	165711	0.71
rrnI-5S	127	165585	165711	0.71
rrnW-5S	136	101077	101212	0.71
rrnW-5S	136	101077	101212	0.71
rrnG-5S	127	176190	176316	0.70
rrnG-5S	127	176190	176316	0.70
rrnI-23S	2925	162610	165534	0.56
rrnE-23S	2945	637138	640082	0.56
rrnE-23S	2945	637138	640082	0.56
rrnB-23S	2914	3173995	3176908	0.56
rrnB-23S	2914	3173995	3176908	0.56
rrnJ-23S	2935	92249	95183	0.56
rrnJ-23S	2935	92249	95183	0.56
rrnW-23S	2915	98120	101034	0.56

rrnW-23S	2915	98120	101034	0.56
rrnO-23S	2944	11691	14634	0.56
rrnO-23S	2944	11691	14634	0.56
rrnD-23S	2923	948418	951340	0.56
rrnD-23S	2923	948418	951340	0.56
rrnG-23S	2916	173225	176140	0.55
rrnG-23S	2916	173225	176140	0.55
rrnA-23S	2921	32190	35110	0.55
rrnH-23S	2913	168227	171139	0.52
rrnH-23S	2913	168227	171139	0.52
RNase P RNA	382	2331325	2331706	0.12
tRNA peaks with an RPKM ratio of >2				
Gene	**Peak Length**	**Coordinates**	**Coordinates**	**RPKM Ratio[a]**
trnY-Glu	230	4155048	4155277	2.76
trnY-Lys	230	4155048	4155277	2.76
trnSL-Gln1	74	2563888	2563961	2.61
trnE-Asp	131	640376	640506	2.35
trnY-Phe	217	4154761	4154977	2.25
trnI-Gly	90	165953	166042	2.12

[a] The expression of each peak (described in the main text) was quantified in reads per kilobase per million mapped reads, or 'RPKM' (Mortazavi *et al*.,

2008). The ratio of these values for the Hfq^{FLAG} and mock control samples was taken as an indicator of Hfq-mediated enrichment.

Table 2-4 Hfq coIP peaks associated with mRNA coding regions.

Peaks That Fully Encompass mRNA Coding Regions			
Gene	**Peak Length**	**Peak Coord**	**RPKM Ratio[a]**
yqzF	256	2507015-2507270	36.5
yqgW	214	2565687-2565900	31.4
yhdX	176	1038635-1038810	29.0
yxzF	113	3964110-3964222	24.3
ydzK	191	490817-491007	20.7
yuzK	183	3360983-3361165	18.3
ypmP	226	2292419-2292644	17.1
yxiE	460	4031798-4032257	16.3
yhzC	244	1116582-1116825	11.8
veg	276	52761-53036	10.1
rpsU	158	2620383-2620540	9.00
ywzA	229	3918808-3919036	7.50
sunA	210	2269505-2269714	5.64
ykzF	265	1485069-1485333	4.70
yqzM	160	2637373-2637532	4.70
rsiW	675	195360-196034	3.85

ykzW	110	1534108-1534217	2.94
Peaks Associated With Start Codons			
Gene	**Peak Length**	**Peak Coord**	**RPKM Ratio**
yitY	88	1192831-1192918	1200
yvpB	64	3589594-3589657	160
addB	146	1136313-1136458	120
tcyL	152	3006440-3006591	51.1
yheJ	93	1045004-1045096	45.9
yjiA	66	1290892-1290957	26.2
yqgY	82	2564619-2564700	26.0
ypjP	83	2299076-2299158	25.7
glnJ	56	265494-265549	25.4
ctaD	121	1561556-1561676	23.5
yopK	102	2210061-2210162	21.9
yneF	66	1922546-1922611	16.9
minJ	72	3621478-3621549	16.9
nagP	178	840647-840824	13.7
csbA	133	3615078-3615210	11.5
sboA	166	3836036-3836201	3.67
hpr	102	1073612-1073713	2.64

Peaks Associated With Stop Codons			
Gene	**Peak length**	**Peak Coord**	**RPKM Ratio**
mecA	106	1229626-1229731	102
fadE	271	3367001-3367271	76.9
yopS	192	2203847-2204038	72.5
yhdT	70	1036862-1036931	71.1
rapB	117	3770985-3771101	32.1
spoIVA	63	2386180-2386242	27.4
hfq	76	1867638-1867713	24.1
tlp	136	1930562-1930697	23.9
sacT	72	3906125-3906196	21.2
ywmB	91	3779267-3779357	20.1
yebD	134	697224-697357	18.9
glpD	103	1006530-1006632	14.8
cccA	136	2599474-2599609	14.2
fbp	79	4130048-4130126	14.1
rnr	109	3451848-3451956	13.9
degU	73	3644596-3644668	13.9
yjbJ	95	1235156-1235250	13.9
ytzI	102	3136927-3137028	11.9

csbA	133	3615078-3615210	11.5
ydbL	221	504809-505029	11.0
ohrB	189	1382249-1382437	10.7
degA	154	1164013-1164166	9.90
hutP	235	4041723-4041957	9.51
clpP	133	3546708-3546840	5.58
yuiB	126	3299316-3299441	5.00
adcA	155	309152-309306	3.67
yqhP	152	2542248-2542475	2.34
citZ	77	2981133-2981209	1.91
Peaks Within mRNA Coding Regions			
Gene	**Peak Length**	**Peak Coord**	**RPKM Ratio[a]**
srfAA	162	377729-377890	270
bioD	137	3091835-3091971	217
yitQ	98	1185464-1185561	142
miaA	64	1866987-1867050	108
yvcD	77	3576599-3576675	104
ykuC	228	1476174-1476401	103
truB	97	1737649-1737745	94.3
yqhL	61	2545230-1737745	66.6

yqzK	108	2449501-2449608	63.0
yukB	109	3273707-3273815	55.0
opuBB	89	3461854-3461942	49.1
secDF	60	2828778-2828837	42.3
yerI	54	725248-725301	41.4
gutP	52	668773-668824	41.0
ycbU	80	288459-288538	40.7
srfAB	258	389722-389979	31.3
besA	107	3292960-3293066	31.3
alaT	134	3225984-3226117	30.3
pnpA	102	1739618-1739719	29.9
wapA	54	4025841-4025894	28.2
radC	54	2862144-2862197	26.3
dnaK	89	2627350-2627438	25.1
ydiG	126	646894-647019	25.0
fadE	65	3367929-3367993	23.4
glcD	97	2934073-2934169	23.0
copA	64	3443019-3443082	20.6
ykrW	125	1427086-1427210	18.7
acdA	96	3814111-3814206	18.0

treP	114	851440-851553	17.8
holB	97	40752-40848	17.8
ytrP	71	3034960-3035030	17.2
srfAA	425	386241-386665	17.0
comC	60	2864721-2864780	16.9
yflT	164	827484-827647	16.1
ydaS	94	492709-492802	15.1
sigW	379	194889-195267	14.2
ydaG	225	473804-474028	13.5
ctrA	160	1023440-1023599	13.3
ypjP	159	2298907-2299065	12.1
malP	101	892290-892390	12.1
besA	138	3292745-3292882	11.8
thiF	188	1246288-1246475	11.3
yusK	126	3369240-3369365	10.5
ygzB	155	945000-945154	8.52
dhbF	61	3287221-3287281	7.79
yfmE	66	823759-823824	6.47
gapB	121	2967347-2967467	2.79
yheE	103	1050457-1050559	2.50

citZ	71	2981810-2981880	2.06
ybfO	56	251096-251151	2.06

[a] The expression of each peak was quantified in reads per kilobase per million mapped reads, or 'RPKM' (Mortazavi *et al*., 2008). The ratio of these values for the Hfq^{FLAG} and mock control samples was taken as an indicator of Hfq-mediated enrichment.

Table 2-5 mRNA leader regions that coIP with Hfq.

Name	Ligand	Peak Length	Peak Coordinates	RPKM Ratio[a]
ilvB	tRNA	76	2897048-2897123	317
yybP	Unknown	101	4169826-4169926	211
mgtE	Magnesium	203	1395622-1395824	130
rplT	Protein	177	2953417-2953593	115
tyrS	tRNA	219	3037940-3038158	88.6
xpt	Purine	135	2320065-2320199	68.3
glpF	Protein	96	1002367-1002462	56.1
ydaO	Unknown	195	486090-486284	41.5
thiU	TPP	172	1391680-1391851	38.7
rplJ	Protein	100	119850-119949	37.6
ykoY	Unknown	130	1410625-1410754	36.9
ptsG	Protein	96	1456978-1457073	31.8

bglP	Protein	121	4035698-4035818	31.6
trpE	Protein	108	2377504-2377611	30.3
rpsD	Protein	118	3035565-3035682	30.1
thrS	tRNA	233	2961244-2961476	28.3
thiT	TPP	123	3179111-3179233	28.1
yitJ	SAM	203	1180645-1180847	27.2
pyrG[b]	Protein	76	3812400-3812475	26.7
trpS	tRNA	223	1219144-1219366	24.1
queC	Purine	82	1439273-1439354	18.6
lysP	Lysine	229	3421133-3421361	18.4
ktrA	Unknown	172	3188194-3188365	18.3
fmnP	FMN	199	2410680-2410878	12.3
yxjG	SAM	123	3999167-3999289	12.2
yxjA	Purine	149	4005557-4005705	11.6
tenA	TPP	140	1242260-1242399	11.2
metQ	SAM	166	3364355-3364520	11.1
thiC	TPP	165	955648-955812	9.92
lysC	Lysine	171	2910880-2911050	5.44
pbuG	Purine	131	694450-694580	4.56
ileS	tRNA	197	1613087-1613283	4.14

gcvT	Glycine	172	2549415-2549586	3.82
ribD	FMN	132	2431468-2431599	3.08
purE	Purine	140	698396-698535	2.93
yoaD	SAM	125	2025141-2025265	2.79
cspC[b]	Protein	45	559567-559611	1.91

[a] The expression of each peak was quantified in reads per kilobase per million mapped reads, or 'RPKM' (Mortazavi et al., 2008). The ratio of these values for the Hfq^{FLAG} and mock control samples was taken as an indicator of Hfq-mediated enrichment.

[b] Almost all of the Hfq CoIP peaks shown in this table directly correspond to the aptamer portions of the cis-acting regulatory RNAs. There are two exceptions – *pyrG* and *cspC*. *pyrG* does not have a discrete aptamer domain according to current literature. Instead, it regulates downstream gene expression through reiterative incorporation of nontemplated G residues under conditions of pyrimidine limitation, which in turn affects formation of a downstream intrinsic terminator. The Hfq CoIP peak for *pyrG* exactly corresponds to the entire regulatory RNA, from the start of transcription to the transcription terminator element. Therefore, despite the absence of a recognizable aptamer domain, the ligand-sensing portion of the *pyrG* regulatory RNA still corresponds to the Hfq CoIP peak. Expression of *cspB* and *cspC* genes is auto-regulated through a transcription attenuation mechanism (Graumann and Marahiel, 1999). The aptamer portion of these leader regions has not been fully identified; however, the *cspC* Hfq CoIP peak covers the region between the transcription start site and the downstream transcription terminator, and is therefore likely to encompass the putative CspC binding site.

CHAPTER 3

Analysis of the Consequence of Loss of Hfq on the Global Transcriptional Landscape of *Bacillus subtilis*

Introduction

In Gram-negative organisms deletion of *Hfq* results in a number of phenotypic consequences such as decreased virulence, biofilm formation, loss of motility, and impaired resistance to a variety of environmental stresses. In fact, it is estimated that nearly 20 percent of the *Salmonella typhinerium* genome is regulated either directly or indirectly by Hfq (Chao and Vogel 2010, Vogel 2009). The primary function of Hfq within the bacterial cell is to facilitate base-pairing interactions between sRNA regulators and their cognate mRNA targets, thus eliciting changes in gene expression. Consequently, Hfq can be viewed as a proxy for sRNA mediated genetic regulation in eubacterial species. Curiously, loss of Hfq in Gram-positive organisms does not result in the profound phenotypic changes reminiscent of their Gram-negative counterparts. The results presented in chapter 2 of this dissertation catalogues the multitude of RNA elements that associate with Hfq in *Bacillus subtilis.* However, the functional consequence of loss of Hfq on the various RNA elements we observe

interacting with this protein remains unclear. Additionally, the physiologic consequence of deletion of Hfq on the global transcriptional landscape of *B. subtilis* is undefined. Therefore, in an attempt to answer some of these unresolved questions in regards to Hfq function we assessed the relative consequence of loss of Hfq on the global transcriptomic profile of *B. subtilis* using next-generation sequencing.

Results and Discussion

Phenotype of Hfq deletion mutant

After generating an extensive list of putative sRNAs which associated with Hfq we next wished to validate that the protein / RNA interactions we observed were physiologically relevant or possibly an artifact due to over expression of FLAG-tagged Hfq. In fact, the addition of 1mM IPTG to cells containing an ectopic Hfq-FLAG expression construct did show significant over production as compared to the endogenous chromosomally encoded Hfq (Figure 3-1A). Furthermore, addition of the FLAG epitope tag did not obstruct the ectopic protein from forming a complex with endogenous Hfq as we were able to detect untagged Hfq bound to the tagged version by mass spectrometry after co-immunoprecipitation and subsequent SDS-PAGE (Figure 3-1B). Thus the actual

amount of Hfq protein present in the over expression strain is more accurately reflected by adding the signal observed for the total cell lysate containing the FLAG-tagged Hfq at the endogenous locus to the signal observed for the over expression strain.

Previous studies in Gram-negative bacteria have detailed a relationship between RNA stability and Hfq binding. In general, it is thought that Hfq binding to sRNAs increases their stability by preventing access to binding sites within the sRNA by intracellular RNases. In support of this view is the fact that many sRNAs have a reduced half life in Hfq deletion strains, and the decreased half-lives of these sRNAs can be rescued by deletion of portions of the RNA degradeosome (Condon and Bechhofer 2011, Gottesman and Storz 2011a, Storz et al. 2011). However, upon base pairing with their targets, sRNA regulators are now permissive for degredation by RNAses due to their double-stranded nature.

Based on the well documented affect of Hfq deletion on sRNA stability we set out to examine the consequence of Hfq deletion on the intracellular levels of several of the sRNAs we identified through our pull down studies. We rationalized that if we observed changes in sRNA abundance in an Hfq null back ground this may be indicative of a physiologically relevant interaction *in vivo*, thus helping to validate the candidate sRNAs we detected associated with Hfq.

As mentioned previously Hfq is often found in a dicistronic operon with the tRNA modifying enzyme *miaA* (Figure 2-1A). In order to avoid any potential issues with polar effects due to the presence of an antibiotic resistance cassette, I generated a markerless deletion of Hfq using the pMAD system (Arnaud, Chastanet and Débarbouillé 2004). In addition, instead of deleting a portion of the protein which can lead to partial translation and possibly dominant-negative or gain of function mutant effects, I placed a stop codon after the f-Met codon of the *hfq* ORF. In addition, a *not*I restriction site was engineered into the deletion construct following the stop codon. This allowed me to screen for cells that had replaced the endogenous copy of *hfq* with the deletion construct through homologous recombination via diagnostic PCR of the *hfq* locus and subsequent digestion with *not*I. It is interesting to note that when placed in a WT *Bacillus subtilis* strain capable of forming biofilms an intermediate phenotype is produced (Figure 3-2). Cells that are incapable of forming a biofilm appear flat and mucoid, devoid of the complex arial projections which serve as preferential sites of sporulation, which are characteristic of wild type biofilms (see EAR deletion strain, Figure 4-9). Loss of the complex colony architecture and subsequently the arial projections leads to deficits in sporulation as compared to wild type biofilms (Vlamakis et al. 2008). The *hfq* mutant appears larger and flatter with a decrease in raised projections, however these cells did not display

any deficits in sporulation resulting from these phenotypic changes (data not shown). In fact, we did not observe any deficits in sporulation attributed to loss of *hfq* in cells grown in liquid sporulation media either, which is consistent with previously reported results (Silvaggi, Perkins and Losick 2006).

To assess the relationship between Hfq and the catalog of RNAs we detected bound to this protein we initially looked at several of the sRNA species we identified as interacting with this protein. In particular we examined four previously discovered type-I toxin/ antitoxin systems, one previously discovered sRNA (FsrA) which is reported to be Hfq independent, and CsfG which was a newly discovered sRNA at the time we initiated these studies but was subsequently reported while our manuscript detailing the discovery of this sRNA was in preparation (Irnov et al. 2010, Gaballa et al. 2008, Marchais et al. 2011). As a preliminary test to examine the relationship between loss of *hfq* and sRNA stability we performed Northern Blots on total RNA extracted under the same glucose minimal media conditions used in our Hfq coIP and a separate study in our lab that examined transcription start sites in *B. subtilis* (Figure 2-4) (Irnov et al. 2010). Using antisense probes for both FsrA and CsfG we were able to detect a single robust band for each of these respective sRNAs corresponding to the predicted size of each of these species (Figure 3-3B and 3-4B). However upon

deletion of *hfq* a clear reduction in signal for each of these sRNAs was observed (Figure 3-3B and 3-4B). In fact quantification of signal intensity corrected for loading variation between wild type and the Hfq knockout strain yielded a 1.8 fold decrease in FsrA signal and 6.8 fold decrease in CsfG relative to wild type. In addition, we were able to detect both FsrA and CsfG via northern bloting using RNA extracted from co-immunoprecipitated Hfq after recapitulating the conditions used for the deep-sequencing experiments discussed in the previous chapter (Figure 3-3C and 3-4B).

As mentioned previously, one consequence of loss of Hfq from the cell is an overall decrease in sRNA half-life which is presumably the reason we observe decreased signals in the total RNA Northen Blots for CsfG and FsrA corresponding to the *Δhfq* deletion strain. However, another plausible explanation for the decreased signal intensity for these sRNAs may be a byproduct of reduced transcription at these loci and not attributed to an overall decrease in sRNA stability. In an attempt to resolve these questions I performed RNA half-life experiments for CsfG and FsrA in WT and the *Δhfq* mutant background. These assays are essentially pulse-chase experiments wherein cells are grown to stationary phase and then the antibiotic rifampicin, which inhibits RNA polymerase, is added to the cell culture and samples are harvested and

flash frozen in liquid nitrogen (Coller 2008). Total RNA is then extracted from each cell pellet and northern blots are performed to assess the impact on the RNA species of interest. However, since deletion of *Hfq* resulted in reductions in FsrA and CsfG *de novo,* I constructed plasmids containing these small RNAs under the control of a xylose inducible promoter, and ectopically integrated them into the genome. This allowed me to induce expression of these sRNAs to similar levels in the WT and Hfq mutant and then assess the consequence of loss of Hfq on sRNA half-life. Figure 3-5 is a representative mRNA half-life Northern Blot for CsfG. At first glance it appears that deletion of Hfq does indeed accelerate RNA decay in comparison to WT for this sRNA. However, quantification of the blot and subsequent normalization to 5S rRNA revealed no differences in the rate of sRNA decay and these results held true for FsrA as well (data not shown). In fact, the reason for the qualitative differences between the WT and mutant for both of these sRNA regulators is due to the increased intracellular abundance of CsfG and FsrA originating from the endogenous locus within the cell. Although, we believe that these results do not necessarily rule out decreased sRNA stability attributed to the loss of Hfq; but instead may be reflective of technical limitations of this assay format, as mRNA half-life experiments are notoriously finicky and not very reproducible.

Inspired by the results obtained with FsrA and CsfG we next turned our attention to the four type-I TA systems in *B. subtilis* that were previously identified by our lab and others (Irnov et al. 2010, Silvaggi et al. 2006, Saito, Kakeshita and Nakamura 2009). Type I toxin/ antitoxin systems were initially discovered nearly 30 years ago as "addiction modules" required for maintenance of parasitic genetic elements such as plasmids and integrated prophages (Fozo, Hemm and Storz 2008, Fozo et al. 2010, Gerdes and Wagner 2007). They consist of a stable toxin which usually codes for a hydrophobic peptide that inserts into the cell membrane, and an unstable antitoxin which is an antisense RNA to the toxin mRNA in type-I systems. When the parasitic element is present in the genome the labile antitoxin is abundantly expressed and pairing with the toxin results in degradation of both sRNA and mRNA. However, if this foreign element is lost from the genome the antitoxin can no longer be transcribed and is rapidly turned over, but the toxin which is recalcitrant to RNA degradation persists and is translated resulting in cell death. This scenario clearly holds true for some of the first systems characterized, however the role of these enigmatic proteins in normal bacterial physiology are just now beginning to be appreciated.

Conventional wisdom is that Hfq is not required for toxin/ antitoxin base-pairing or function due to the complete antisense nature of these *cis*-encoded

sRNAs, and to date this view has held true. However, we recovered both toxin and antitoxin RNAs bound to Hfq in our pulldown experiments, with an apparent preference for the antitoxin versus the toxin with the exception of *bsrH*/ncr2166. Interestingly we recover the greatest number of reads corresponding to sites overlapping regions of predicted base pairing between the toxin and antitoxin, suggesting that Hfq may facilitate base pairing between these two RNAs, or conversely base pairing is required for Hfq to bind to these RNAs *in vivo* (Figure 3-6).

As an initial test of Hfq affinity for antitoxins versus type I toxins I chose to further investigate the toxin BsrG and its putative cognate antitoxin ncr1932. We chose this TA pair because it was highly enriched in our coIP data set, with over 2000 cDNA reads mapping to the region of predicted base pairing between the toxin mRNA and the antitoxin sRNA. Additionally manual inspection of the antitoxin revealed a large AU rich stretch at the 5′ end of this sRNA, and Hfq is known to bind AU rich regions within sRNAs. We subsequently *in vitro* transcribed the full length toxin, antitoxin, and a truncated antitoxin missing the AU rich region and performed electrophoretic gel mobility shift assays (EMSA) to examine Hfq binding (Figure 3-7).

Our preliminary results confirm that Hfq seems to exhibit a higher affinity for the antitoxin versus the toxin; and that deletion of the AU rich stretch from the antitoxin ameliorates some of the binding affinity. However, these gel shift experiments were not always reproducible, probably attributed to contaminating *E. coli* RNA that was nonspecifically bound by Hfq during recombinant protein purification. Indeed, purifying recombinant Hfq requires harsh heating conditions as well as, high salt and urea concentrations in order to remove exogenous DNA and RNA that is nonspecifically bound by this protein (Zhang et al. 2002). Undoubtedly, purifying this protein in the presence of high concentrations of chaotropic salts at 80˚C will lead to the loss of activity of some portion of the purified protein. It is difficult to assess even a rough estimate of recombinant Hfq loss of function due to the fact that there were no known *B. subtilis* RNA elements that bound Hfq prior to this study so we lack a positive control to determine the extent to which the purification protocol we used influenced activity.

These initial results were indeed exciting but we were cautious not to over interpret them because these RNAs are highly expressed within the cell, and we were artificially driving production of Hfq, raising concerns that we were witnessing an artifact. Therefore we performed northern blots for both the toxin

and antitoxin RNAs of these four systems in WT and *Δhfq* mutant backgrounds. We were able to detect signals for three out of the four TA systems with the exception being *bsrH*/ncr2166 (figure 3-8). Loss of Hfq had a clear effect on the antitoxins (top panel of each northern) as these RNAs clearly decreased.

Quantification of these blots revealed that deletion of Hfq resulted in a modest ~1.5 fold decrease in signal for the *Δhfq* in comparison to WT. The effect of loss of *hfq* in regards to the toxins were more subtle *ncr1019* decreasing greater than two fold in the *hfq* deletion strain and the toxins bsrG and txpA essentially unchanged. As mentioned previously, the cellular function of these small hydrophobic peptides remains unclear. For instance, true plasmid maintenance systems which use type I TA's are indeed lethal upon loss of the antitoxin, however phenotypic effects caused by deletion of antitoxins from the genome are more nuanced. In most cases lethality is only achieved through over expression of the toxic peptide (Fozo et al. 2010, Thomason and Storz 2010). In *Bacillus subtilis* deletion of the RNA antitoxin RatA leads to a cell lysis phenotype but only after several days growth on rich media agar plates (Silvaggi, Perkins and Losick 2005). Interestingly the authors deleted *hfq* from the genome and did not observe recapitulation of the lysis phenotype suggesting that Hfq was not required for antitoxin mediated neutralization of the toxin TxpA. Studies in our own lab confirm these findings as we do not observe this lysis phenotype in our

hfq deletion strain. Whether or not Hfq is playing an active role in type-I TA regulation remains to be determined. It is worth noting that the decreases in antitoxin expression levels we detected in the Hfq deletion strain were indeed modest (figure 3-8). It is possible that the observed association of these type-I TAs with Hfq is not physiologically relevant, and that the decrease in antitoxin abundance can be attributed to secondary effects of deleting Hfq that result in diminished transcription of these sRNAs.

Another explanation for a lack of a lysis phenotype could be compensation by a functionally redundant protein(s) within the cell, as there are many uncharacterized small basic proteins encoded in the *B. subtilis* genome. Clearly these modules serve some function in normal bacterial physiology. A recent bioinformatic study examining the distribution of known type-I TA systems found multiple copies across 774 new host genomes (Fozo et al. 2010). Moreover, many genomes contained multiple type-I TA families which are themselves comprised of many members. In one extreme case a strain of pathogenic *E. coli* contained 26 of these loci (Fozo et al. 2010). Interestingly the phylogenic tree of type-I TAs correlates with the host taxonomy suggesting that these elements were not freely disseminated by horizontal gene transfer but rather evolved from a common ancestor (Fozo et al. 2010).

Global transcriptomic profile of an Hfq deletion mutant

In an attempt to systematically and efficiently examine the effect of *hfq* deletion on the various RNA species that we detected in our coIP experiments we next extracted total RNA from wild type and *Δhfq* cells grown to stationary phase in glucose minimal media and performed total RNA-seq (Figure 3-9). Astonishingly, nearly 60% of the over 4,000 annotated *Bacillus subtilis* ORFs exhibited a twofold or greater change in expression between WT and the Hfq mutant (Figure 3-10A). In fact, over 2000 genes representing 46% of the genome were up regulated at least twofold in the *Δhfq* strain in comparison to the WT, while 542 genes corresponding to over 12% of the genome were decreased respectively. Closer examination of the genes up regulated in the Hfq deletion strain revealed that many of these genes were significantly elevated as compared to WT. In fact, 39% of these ORFs were increased fivefold or greater which represented over 18% of the annotated genome (Figure 3-10A). This trend was consistent in regards to genes that were down regulated in *Δhfq* as well where nearly 35% of ORFs that decreased in expression were fivefold or greater which represented approximately 5% of protein coding genes (Figure 3-10A).

The magnitude and extent of differential gene expression between the wild-type and Hfq deletion mutant were very surprising for two reasons. First, I and others have failed to detect a phenotype in *B. subtilis* Hfq mutants (Silvaggi et al. 2005, Gaballa et al. 2008). The *Δhfq* mutant has no defect in growth, sporulation, germination, motility and only a slight defect in biofilm formation. One would anticipate that such large changes in gene expression would reveal some phenotypic variation as compared to WT *B. subtilis*. Secondly, microarray studies looking at global transcriptional changes in a *Salmonella* Hfq deletion strain surmised that nearly one-fifth of the genome was affected by the loss of Hfq. However, deletion of *hfq* in this organism results in gross pleiotropic phenotypes including loss of motility, virulence, protein secretion, and response to various stresses (Chao and Vogel 2010, Vogel 2009).

The discrepancies between these two studies are likely due to the over estimation of changes in gene expression in our study and the under estimation in the *Salmonella* report. For instance when comparing two ORFs with a low number of reads, one must take caution when interpreting changes in expression levels between two samples because small changes in gene expression will manifest as large differences in the fold change between the two samples. When analyzing our data we did not set an arbitrary RPKM threshold

for what constituted a minimal number of reads that would produce confidence in the differences we observed in gene expression. Instead, the data represented in Figure 3-10 illustrates the total changes in gene expression between the two samples irrespective of a minimal threshold for number of reads at a particular gene.

An additional source of the inflated changes in gene expression we observed can be attributed to the way that the samples were prepared for sequencing. Part of the procedure for producing samples for Illumina sequencing involves generation of cDNAs using reverse transcriptase followed by PCR. This results in the loss of information pertaining to the DNA strand from which the RNA species originated. For example, a gene with very low expression that is located on the sense strand will display an artificially inflated expression level, if for instance there is a highly expressed ORF adjacent to it on the antisense strand. Finally, deep sequencing is orders of magnitude more sensitive than microarray platforms. Thus, the overall percentage of changes in the *Salmonella ⊿hfq* mutant is probably under estimated.

sRNA abundance and deletion of Hfq

Of the 119 putative small RNAs thus far cataloged in the *B. subtilis* genome, deletion of Hfq affected the expression of 51% of these sRNAs. In fact

25% percent exhibited a decrease of at least twofold in the Hfq deletion strain relative to wild type, whereas 26% percent were up regulated in the *Δhfq* strain in comparison to wild type (Figure 3-10C). However, due to the limitations described above identifying global trends in sRNA expression in an Hfq null mutant is not possible and instead individual sRNAs must be manually inspected to evaluate the potential impact of Hfq on the RNA species in question.

This is readily apparent when examining the CsfG sRNA. CsfG was enriched nearly 23 fold (Table 2-1) in our Hfq coIP experiment relative to the mock control and deletion of *hfq* results in a nearly seven fold reduction in signal via northern blot analysis (Figure 3-4). Yet our global transcriptome data yielded a perplexing six fold increase for CsfG in the Hfq knockout strain compared to WT. Visual inspection of this locus within the *Bacillus subtilis* genome offers several possible explanations for the discrepancies between the northern blot data and the results of the total RNA-seq from the Hfq deletion mutant. The first issue upon closer examination of the two total RNA-seq data sets is the low number of reads corresponding to the CsfG genomic position for both samples. We observe an RPKM value for CsfG of only 1.4 in the WT and nearly 9 in the Hfq deletion mutant. In fact there is an 18 nucleotide stretch at the 3′ end of CsfG that partially overlaps the intrinsic terminator in which we do not detect any

reads in the wild-type data set. As mentioned previously relatively small changes in the RPKM values for genes with low coverage can result in large fold differences between two samples; leading to the perception of large changes in gene expression. Unfortunately we cannot make quantitative statements about the expression of CsfG purely from the deep-sequencing data. At first glance one might assume that the low number of reads from CsfG was due to a lack of coverage at this genomic position and this is a fair assumption considering that the genes that flank this sRNA display similar RPKM values. It is also possible that this sRNA is expressed at low levels within the cell since our coIP experiment serves to enrich for sRNAs that associate with Hfq. However, *a priori* one would think that if this sRNA was truly very lowly expressed, detection by northern blotting would be difficult. Regardless, quantification of the expression level of CsfG or any other genomic locus would need to be validated using alternative measures such as quantitative PCR.

Further complicating matters is the genomic arrangement of CsfG relative to the adjacent ORF's (Figure 3-11). CsfG was originally annotated as an orphan riboswitch for the *ylbH* gene based on the sequence conservation among members of the family Bacillaceae, as well as the presence of an intrinsic terminator located within this putative UTR (Barrick et al. 2004). However, it

was subsequently demonstrated that this RNA element is actually a small RNA located on the opposite strand relative to the sequence corresponding to the predicted orphan riboswitch (Marchais et al. 2011). In fact, computational studies focused on predicting operons in *B. subtilis* using the presence of intrinsic transcription terminators as one of the defining criteria annotates *ylbG* as a monocistronic transcript (De Hoon et al. 2004). However, this intrinsic terminator is actually on the opposite strand and functions to terminate transcription of CsfG. Therefore, *ylbG* appears to be part of a three gene operon that includes *ylbH* and *ylbI*. This operon is clearly up regulated in the *Δhfq* strain and since we are not able to assign which strand the cDNA reads originated from in our sequencing experiment it is likely that observed fold enrichment in the Hfq KO relative to WT is probably an artifact of increased transcription of the *ylbG-ylbH-ylbI* operon located on the opposite strand from CsfG (Figure 3-11).

However, the global transcriptome data did corroborate the trends we observed for FsrA and the TA systems by northern blotting. In fact, quantification of the FsrA blot yielded a 1.8 fold decrease in the *Δhfq* strain and we observed a 1.5 fold decrease by deep sequencing. Manual inspection of this genomic locus gives us confidence that the changes that we observe are real and not an artifact for reasons mentioned above. In particular, the gene *ykuI* which

is adjacent to the 5′ of FsrA contains a strong intrinsic transcription terminator and we do not observe read through transcription of this terminator (Figure 12).

In fact, there is a span of over 100 nucleotides between the end of the *ykuI* gene and the transcription start site of FsrA that is completely devoid of any mapped reads. In addition, this sRNA is in the middle of a 12,000 base segment of the genome that lacks any annotated ORFs on the opposite strand so any potential differences arising from transcription originating on the opposing strand seems implausible (Figure 3-12B). Further examination of the FsrA locus provides intriguing insight into transcriptional noise, as well as the perils of interpreting gene expression changes. Whereas the terminator at the end of the *ykuI* gene results in complete termination of transcription, the terminator at the end of FsrA which is qualitatively comparable to that at the end of *ykuI* results in significant read through (Figure 3-12A).

Hfq and ORFs

The largest class of RNAs we recovered in our Hfq pull down studies were those that mapped to known open reading frames. We further subdivided these Hfq associated RNAs into four groups that represented the location within the open reading frame that our recovered reads mapped to. These were reads that encompassed the entire ORF, those that overlapped the start codon or ribosome

binding site, those that overlapped or were adjacent to the stop codon, and peaks that mapped to the internal portion of the mRNA (Figure 3-13 A-D). Of the 17 ORFs in which the entire coding sequence was found to associate with Hfq, ten of these displayed decreased expression levels of twofold or greater in the Hfq mutant versus WT. In fact none of the recovered ORFs had a twofold or greater increase in the *Δhfq* strain. Thus Hfq appears to be stabilizing these RNAs, or deletion of Hfq results in secondary affects that lead to decreased transcription of these genes. One ORF in particular, *ywzA* was dramatically changed between the two strains. Deletion of Hfq resulted in a 62-fold decrease in expression in the mutant compared to WT. In fact the RPKM for *ywzA* decreased from over 1400 in the WT to just 23 in the mutant, yet the neighboring genes show minimal differences in expression between the two data sets (Figure 3-14). This putative ORF is a hypothetical protein of unknown function that is conserved in closely related *Bacillus* species. It is 81 amino acids in length and has a $_{p}Ka$ of 11 in addition to a large hydrophobic stretch and a basic C-terminal tail. In many respects it resembles the hydrophobic toxins of type I TAs discussed previously and may in fact be one; however we do not observe any evidence of a corresponding antitoxin on the adjacent strand.

Hfq and start codons

Coincidentally we also pulled down 17 ORFs whose Hfq associated reads clustered around the start codon or ribosome binding site (Figure 3-13B). This is an intriguing location to detect Hfq associated sequences as this location has been demonstrated to be the predominant mRNA target of sRNA regulators. Therefore, reads that map to this location within an ORF may potentially represent targets of sRNA action. Most of the mRNAs that we grouped in this category displayed negligible changes between the WT and *Δhfq* strains and many of the inherent problems described above can explain some of the changes we observe. However, one gene in particular, *csbA* displayed a nearly 90-fold down regulation in the *hfq* mutant versus WT. Closer, examination of neighboring genes clearly dismisses the possibility of read through of transcription terminators or expression from the opposite strand as possible sources of the difference we observe between the two samples. Interestingly, transcription start site mapping detected a faint signal 85 nucleotides upstream of the start codon suggesting that *csbA* possess a leader sequence (Figure 3-15). This leader appears to have the potential to adopt a complex secondary structure and possess a potential intrinsic transcription terminator within the 5' UTR. This putative terminator ends 2 bases before the initiation codon and the

helix adjacent to the run of 5Us is comprised of the ribosome binding site for *csbA* (Figure 3-15). Additionally, this structured leader has the capability of forming alternative secondary structures which would result in the formation of a putative antiterminator in front of the UUG start codon, in addition to freeing the sequestered Shine-Dalgarno sequence. This raises the possibility that a small RNA or a protein acts on this leader and induces a conformational change which could in turn regulate this ORF at the transcriptional and/ or translational level. However, again we cannot rule out secondary effects on transcription as the reason for the large changes in *csbA* expression between these two samples.

Internal peaks

The largest category of ORF associated peaks (45%) we detected associated with Hfq mapped to internal portions of mRNAs. Manual inspection of these 50 different genetic loci failed to produce any readily evident correlation between Hfq associated reads and changes in gene expression between the WT and *Δhfq* strains. It is not clear whether reads that map to the internal portions of these ORFs are physiologically relevant or not. In general sRNAs modulate their mRNA targets through affecting translation initiation; however in *Salmonella* the MicC sRNA functions via base-pairing within the coding sequence of its target OmpD, which leads to rapid degradation of this

mRNA via RNaseE (Pfeiffer et al. 2009). Therefore the internal peaks that we observe may in fact represent sites of sRNA action within the coding sequence. In support of this many of the internal regions that co-IP with Hfq contain repeated tracts of G's which may represent seed sequences for sRNAs in Gram-positive organisms, which often contain conserved C-rich repeats (Geissmann et al. 2009).

Alternatively, these peaks may be stable degradation intermediates that have no functional relevance within the cell. It is also possible that these peaks actually represent sRNAs that are embedded within a coding sequence or on the antisense strand from an ORF. This became readily evident when examining the *malP* locus which codes for a maltose specific phosphotransferase. Examination of this sequence revealed a weak intrinsic transcription terminator present on the opposite strand from *malP*; 112 base pairs downstream of the Hfq associated reads we observed a faint transcription start site signal corresponding to the antisense strand, taken together this suggests that the Hfq associated reads actually correspond to a sRNA and not to an internal portion of an ORF (Figure 3-16). It is plausible that many of the reads that map to the coding regions of ORFs may be sRNAs. We cannot be sure how much coverage of the *B. subtilis* transcriptional landscape our TSS data set actually maps and without a signal we

would not be able to differentiate between an ORF associated peak and new sRNAs.

Hfq and 5′ leader sequences

One surprising finding from our coIP data set was the number of *cis*-acting regulatory RNAs that associated with Hfq *in Vivo.* In fact, over 75 percent of the catalogued leader sequences at the time we under took these studies were pulled down with Hfq. As discussed in chapter 2 the physiological relevance of the association of these RNA elements with Hfq is unclear and may be an artifact of their abundant expression within the cell. As an initial test of the significance of this interaction, a putative Hfq-enriched aptamer corresponding to the orphan riboswitch *ykoY* was examined by northern blotting for wild-type and *ΔHfq* strains (Barrick et al. 2004) (Figure 3-17A). Interestingly, the signal for this sequence was reduced in the *Δhfq* strain, suggesting that Hfq may indeed affect its intracellular abundance. This finding was further corroborated upon examination of the *ykoY* locus in the global transcriptome analysis of the WT and *Δhfq* mutant strain. The top panel in Figure 3-17 depicts the mapped reads we recovered from our Hfq coIP which corresponded to the aptamer domain of the orphan riboswitch *ykoY.* In total we observed an over 35-fold enrichment in the Hfq co-IP versus the mock control. Deletion of *hfq*

resulted in an over 10-fold decrease in the number of reads that specifically mapped to the aptamer portion of the *ykoY* leader; and a threefold reduction in RPKM value for the coding portion of the *ykoY* ORF in comparison to WT. Inspection of the ORFs flanking this genomic locus in general possessed relatively low RPKM values that changed marginally between the two samples and there is no annotated ORF on the antisense strand of *ykoY* suggesting that the observed difference is not an artifact of sequencing as described above. Intriguingly, the region corresponding to the coding portion of *ykoY* displayed a threefold decrease in the RPKM value in the *Δhfq* strain suggesting that association with Hfq may be physiologically relevant. The rationale being that if Hfq was associated with terminated aptamer domains due to their intracellular abundance, or because they resemble physiologically relevant RNA substrates, one would not anticipate observing an actual change in the expression value for the *ykoY* coding region because the association presumably occurs after the aptamer domain has executed its regulatory function and terminated transcription. Alternatively, it's also possible that deletion of Hfq leads to indirect down regulation of transcription for the *ykoY* gene accounting for the decreased expression of the coding region. This is plausible because most intrinsic transcription terminators are not absolute, and noise in the system leads to read through of these elements on some basal level. Therefore it's

possible that the increase in RPKM for the coding portion of *ykoY* is a function of increased transcription of the leader and subsequent read through of the terminator.

Discussion and future directions

Our initial motivation for examining the global transcriptomic profile of the *Δhfq* deletion mutant was to assist in the validation of the suite of RNA elements we found to be associated with Hfq in our coIP studies. It has been well documented in Gram-negative organisms that Hfq dependant sRNAs exhibit a decreased half-life when this protein is lost from the cell (Vogel and Luisi 2011, Condon and Bechhofer 2011, Storz et al. 2011). Indeed, our preliminary results utilizing northern blotting clearly established a relationship between RNA abundance and deletion of Hfq for a select subset of RNAs we identified in our coIP studies (Figure 3-3, -4, -8 and -17). However, inherent limitations in the manner in which we prepared our samples for deep-sequencing prevented us from making global statements concerning the validity of the various RNA species we found to interact with Hfq in our co-immunoprecipitation studies. Most notably was the lack of information concerning the strand from which the RNA elements of interest originated from, and the inability to distinguish primary

and secondary effects of deletion of *hfq* in relation to changes in gene expression profiles.

The latter scenario is elegantly illustrated by the gene encoding the transporter csbA. We identified regions of this ORF that associated with Hfq corresponding to portions of the start codon as well as the stop codon. Closer examination of the transcription start site for this gene identified a putative 5'UTR that has the potential to form a complex secondary structure and may possibly contain sequences capable of influencing transcription attenuation or translation initiation (Figure 3-15). One could imagine a scenario in which sRNA pairing to this region could remodel this UTR thus allowing transcriptional read through of the putative terminator element or exposing the ribosome binding site thus promoting translation. Indeed, csbA exhibits a 90-fold decrease in RPKM value in the Hfq deletion mutant versus the wild type (Figure 3-15). However, computational analysis of our data set concerned with identifying changes in genetic regulons mediated by various transcription factors in *B. subtilis* identified csbA as being regulated by the general stress response sigma factor, σ^B. In fact, a majority of the validated ORFs containing σ^B promoters (including the sigma B operon) were up regulated in the WT but not *Δhfq* data sets and several of these were also detected in our co-IP experiments (3-10D).

This suggests that activation of the sigma B regulon requires Hfq either directly or indirectly. This is intriguing as *rpoS*, the major stress response sigma factor in *Salmonella* and *E. coli,* is positively regulated by multiple Hfq-dependant sRNAs (Battesti et al. 2011, Majdalani et al. 1998, Weber et al. 2005). However, we do not detect any putative 5'UTR by TSS mapping, or do we recover any Hfq associated reads that map to the operon coding for σ^B, thus regulation at the post-transcriptional level is not intuitively obvious based on our data sets. Regardless, the question then becomes are the changes we observed do to transcription factor specific increases in gene expression, loss of post-transcriptional regulatory function do to deletion of Hfq, or both? Clearly, further experimentation is required in order to de-convolute the complex regulatory interplay of these various networks. Repeating the total RNA-seq experiment utilizing strand-specific adaptors would be a relatively fast and effective way to begin to parse out some of the issues described above and would in all likely hood further validate or exclude many of the RNA elements we recovered in our Hfq co-IP.

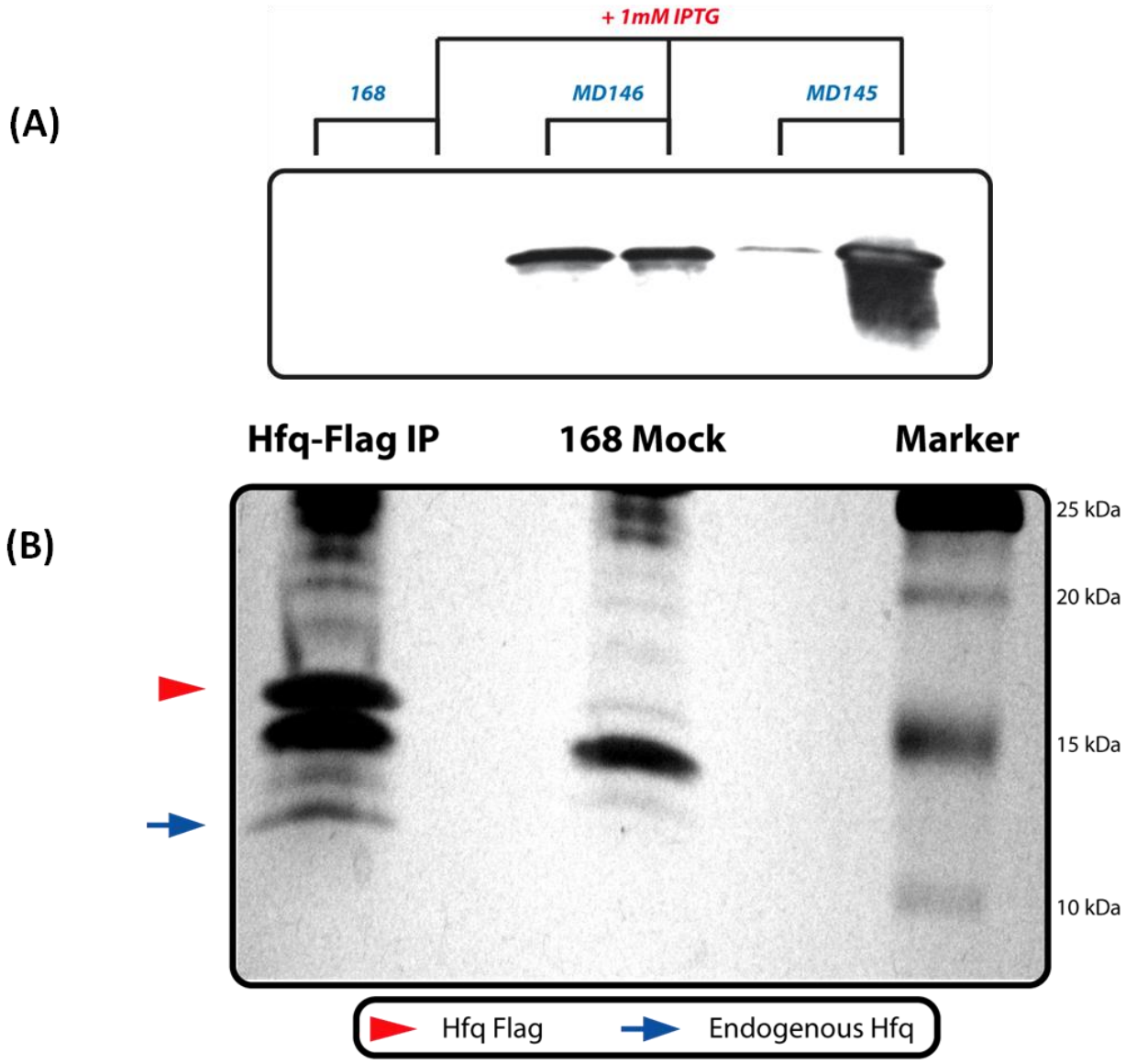


Figure 3-1. Analysis of Hfq over expression versus endogenous protein levels. (A) Western blot of FLAG-tagged Hfq at the endogenous locus (MD145) as well as after induction of an over expression construct integrated at an ectopic locus within the genome (MD146). Cells were grown in glucose minimal media under the same conditions at the Hfq pull down experiments. Hfq protein production was induced for five hours in the presence of 1mM IPTG. (B) FLAG-tagged Hfq physically interacts with endogenous untagged Hfq. Hfq was overexpressed under the same conditions as in (A) and the the protein was immunoprecipitated using anti-FLAG M2 monoclonal antibody followed by protein precipitation using cold acetone. Hfq associated proteins were then run on a 4-20% SDS-PAGE and silver stained. Bands of interest were then excised and identified using mass

spectrometry at the UT Southwestern protein identification core facility. The red arrow is the exogenous FLAG-tagged Hfq and the blue arrow represents the endogenous untagged Hfq which physically associated with the immunprecipitated Hfq.

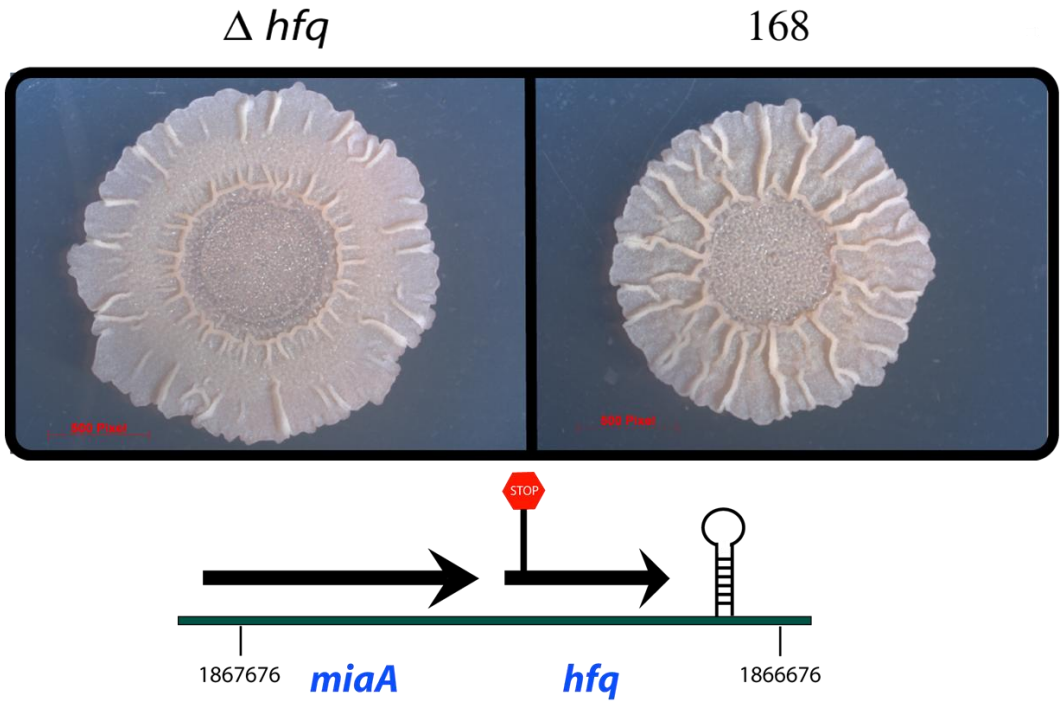


Figure 3-2. Complex colony architecture of wild type *Bacillus subtilis* Hfq deletion mutant. A markerless deletion of Hfq was constructed using the pMAD allelic replacement system. A stop codon was placed in the codeing sequence that corresponded to the second codon. A slight deficit in biofilm formation was observed in the Hfq knockout versus the wild-type as evident by the decreased aerial projections and the flatter smoother colony architecture.

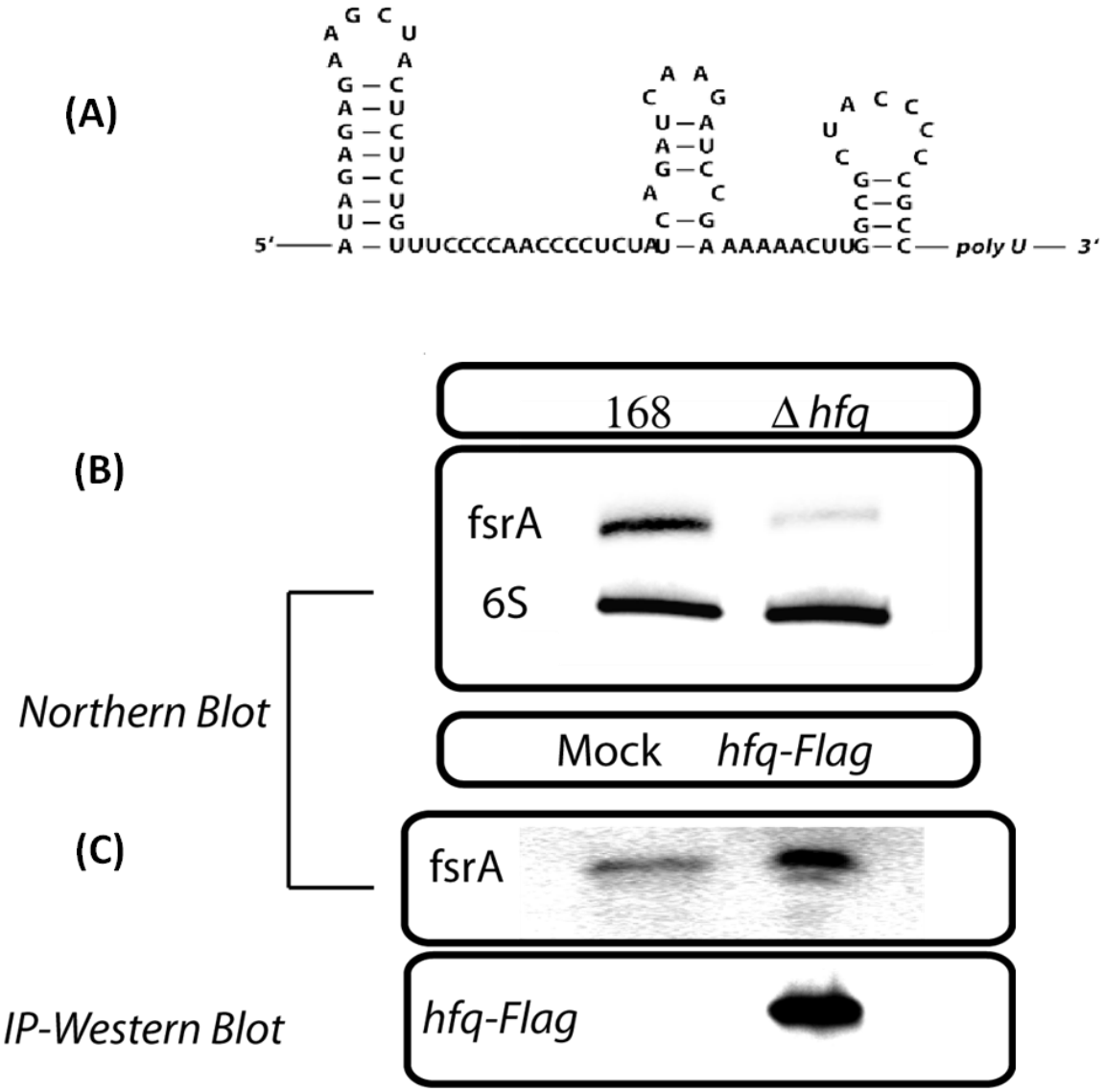


Figure 3-3. The iron-responsive small RNA FsrA associates with Hfq and deletion of this protein effects the abundance of this sRNA. (A) Depicts the secondary structure of *B. subtilis* FsrA. (B) Deletion of Hfq results in a 50% decrease in the abundance of FsrA by Northern Blotting. (C) FsrA associates with Hfq *in vivo* as this sRNA can be detected by Northern Blotting after immunoprecipitation of Hfq and phenol:chloroform extraction of RNAs that associate with this protein.

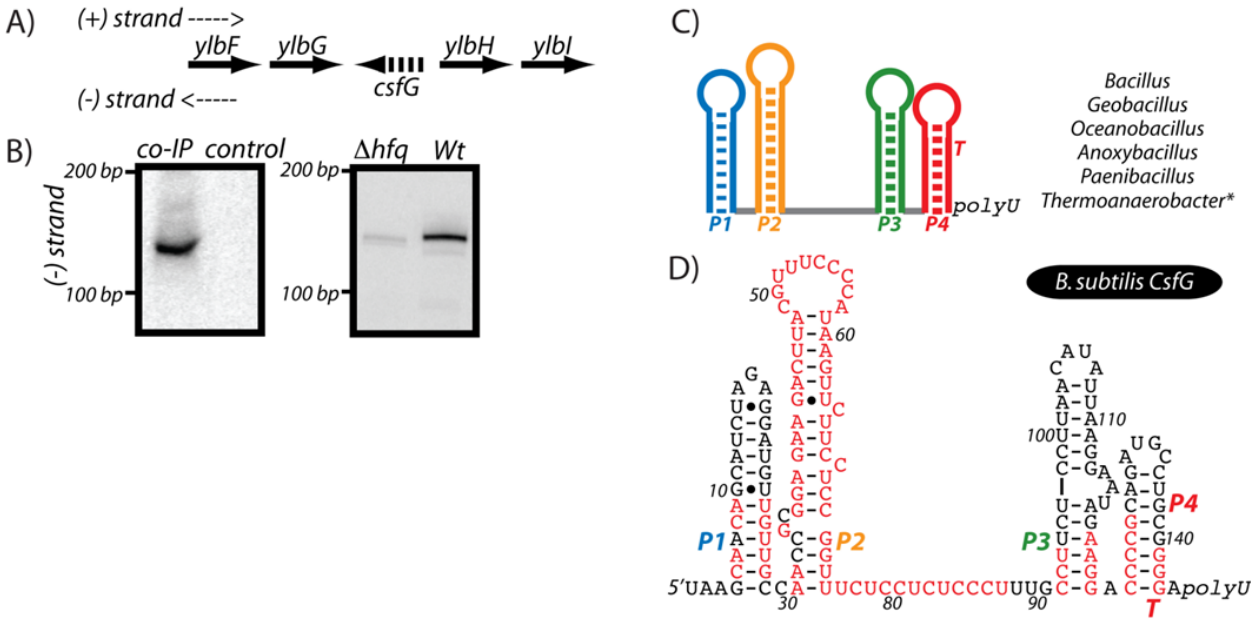


Figure 3-4. The conserved sRNA CsfG associates with Hfq and deletion of Hfq affects its abundance. (A) Depicts the genomic arrangement of CsfG within the *Bacillus subtilis* genome. (B) CsfG associates with Hfq *in vivo* as determined by Northern Blotting following pull down of Hfq and subsequent isolation of bound RNA species by phenol:chloroform extraction. Loss of Hfq results in a sevenfold reduction in CsfG abundance by Norther Blotting. (C) The general structure of CsfG and its phylogenetic distribution. (D) The consensus secondary structure of CsfG nucleotides depicted in red rpresent greater than 95% conservation, while those in black correspond to greater than 80%.

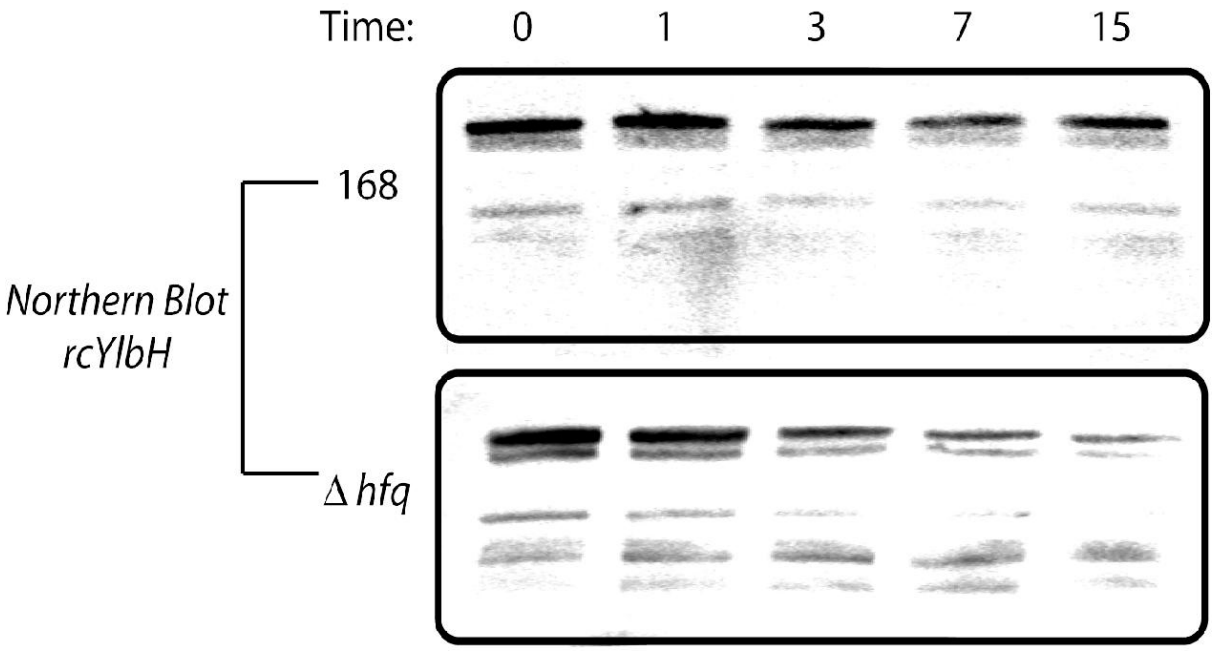


Figure 3-5. Deletion of Hfq does not affect mRNA half-life of CsfG (rcYlbH). A xylose inducible over expression construct of CsfG was integrated into the ectopic Sac locus in both 168 and *ΔHfq* strains. Expression of CsfG was induced for one hour via the addition of 1mM xylose to each culture respectively. Following induction of CsfG the antibiotic rifampicin was added to each culture and samples were collected at the indicated times. CsfG expression was monitored by northern blotting and normalized to a 5S rRNA control. We dod not observe any difference in half-life between the wild-type and the Hfq deletion mutant.

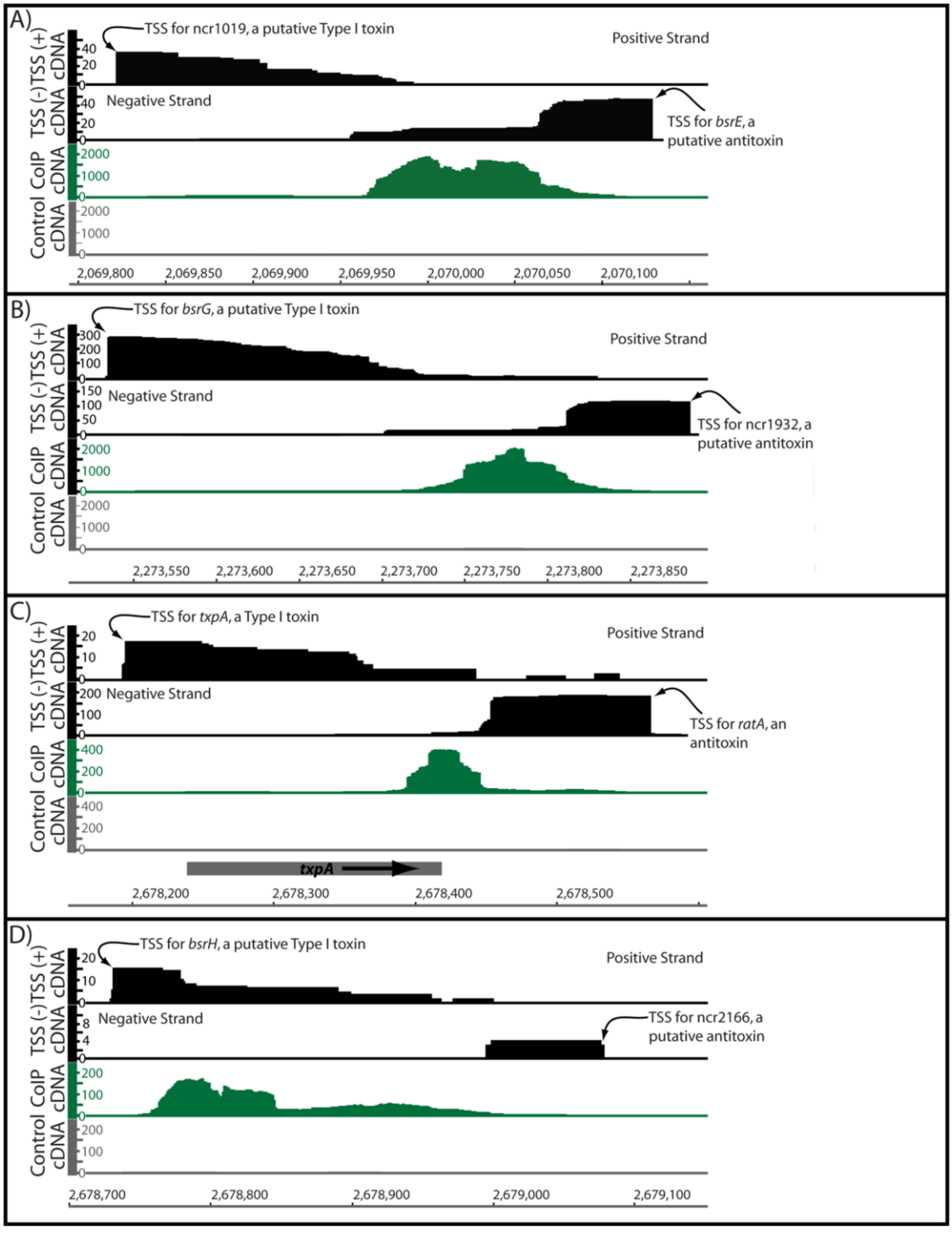

A)
TSS for ncr1019, a putative Type I toxin
Positive Strand
Negative Strand
TSS for bsrE, a putative antitoxin
TSS (+) cDNA
TSS (-) cDNA
Control cDNA
CoIP cDNA
2,069,800
2,069,850
2,069,900
2,069,950
2,070,000
2,070,050
2,070,100
B)
TSS for bsrG, a putative Type I toxin
Positive Strand
Negative Strand
TSS for ncr1932, a putative antitoxin
2,273,550
2,273,600
2,273,650
2,273,700
2,273,750
2,273,800
2,273,850
C)
TSS for txpA, a Type I toxin
Positive Strand
Negative Strand
TSS for ratA, an antitoxin
txpA
2,678,200
2,678,300
2,678,400
2,678,500
D)
TSS for bsrH, a putative Type I toxin
Positive Strand
Negative Strand
TSS for ncr2166, a putative antitoxin
2,678,700
2,678,800
2,678,900
2,679,000
2,679,100

FIGURE 3-6. Hfq associates with type I antitoxin transcripts. (A-D) A previous study of *B. subtilis* transcription start sites (TSS) (Irnov *et al*., 2010) revealed the presence of at least four putative type I toxin/ antitoxin systems. The TSS data from this study are shown in black for the purpose of indicating the orientation of these RNA transcripts, relative to regions of Hfq enrichment. In this study, Hfq appeared to preferentially associate with antisense transcripts. In particular, the sequences that appeared to coIP with Hfq corresponded most often to the portions of the antitoxin transcripts that are predicted to base pair with the toxin-encoding mRNAs. Throughout the figure, the y-axis shows the volume of cDNA reads as a function of genomic position (x-axis) as determined by Illumina-based sequencing of Hfq-associated RNA molecules. Data shown in gray correspond to the negative control reaction (co-immunoprecipitation of Hfq that lacked an epitope tag). Data shown in black correspond to a prior transcription start site (TSS) mapping study (Irnov *et al*., 2010).

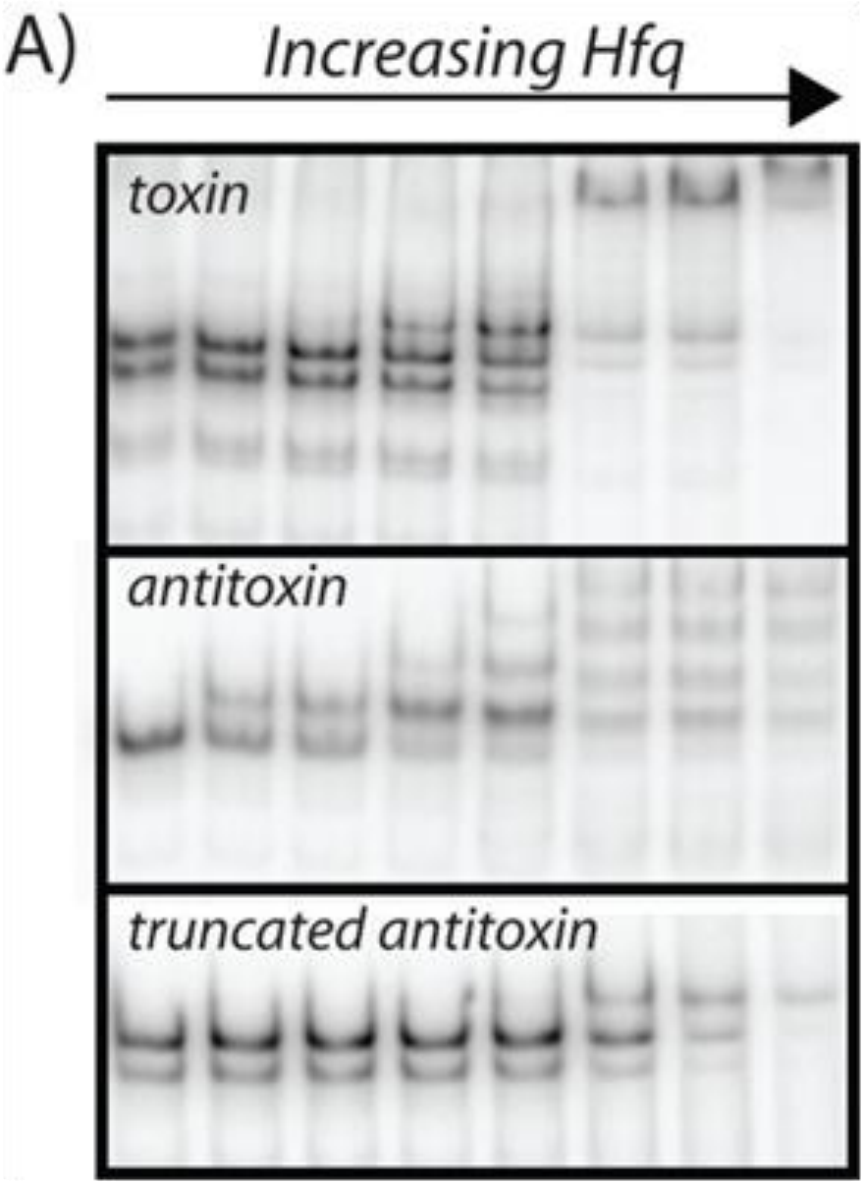


Figure 3-7. Hfq preferentially binds to A/U rich residues present in the RNA antitoxin anti-BsrG. EMSAs corresponding to varying amounts of recombinant Hfq ranging from 1nM to 500µM. Deletion of a 5' A/U rich tract from the anti-toxin (anti-BsrG) ameliorates binding suggesting Hfq may preferentially recognize this sequence.

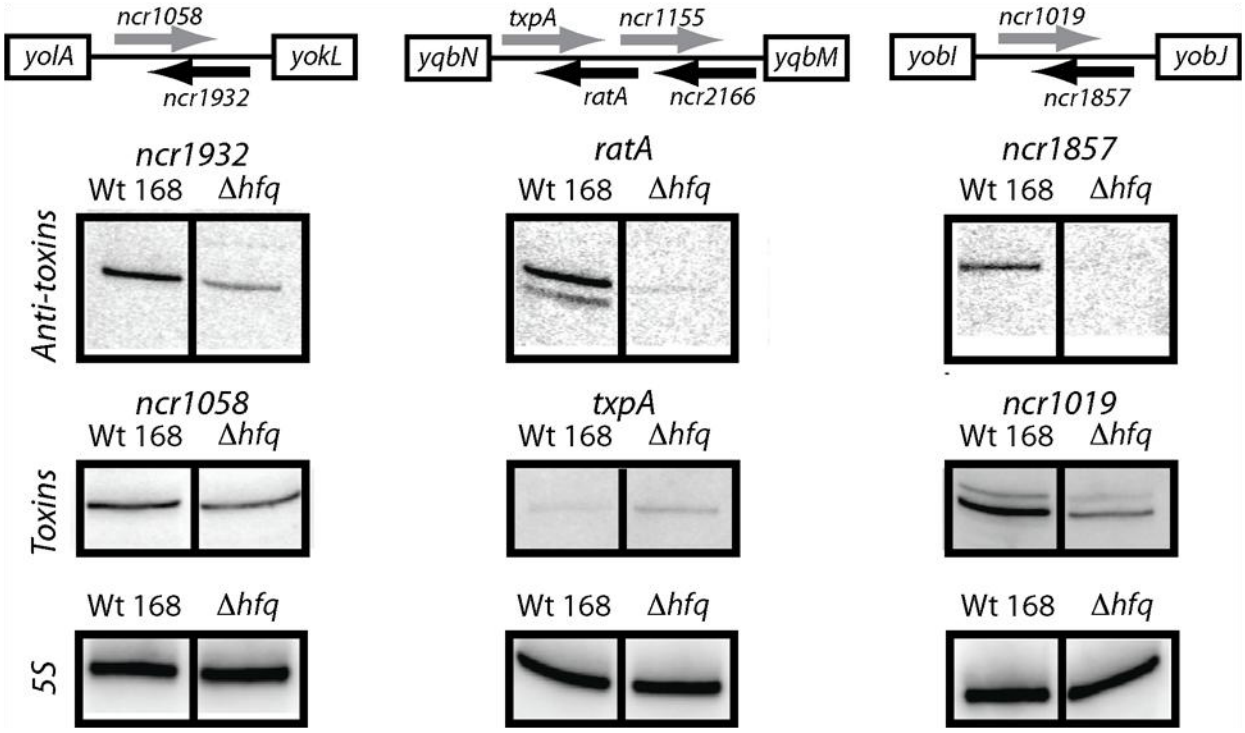


Figure 3-8. Loss of Hfq results in decreased abundance of antitoxin RNA levels. Northern Blots of wild-type or Hfq deletion mutant *B. subtilis* cells grown to stationary phase in glucose minimal media. Total RNA was extracted and subjected to Northern Blotting. Deletion of Hfq resulted in a roughly 50% decrease in each of the antitoxin RNA species while having a negligible effect on the corresponding toxin mRNAs. The genomic context of each TA pair is depicted as a cartoon above their corresponding blots.

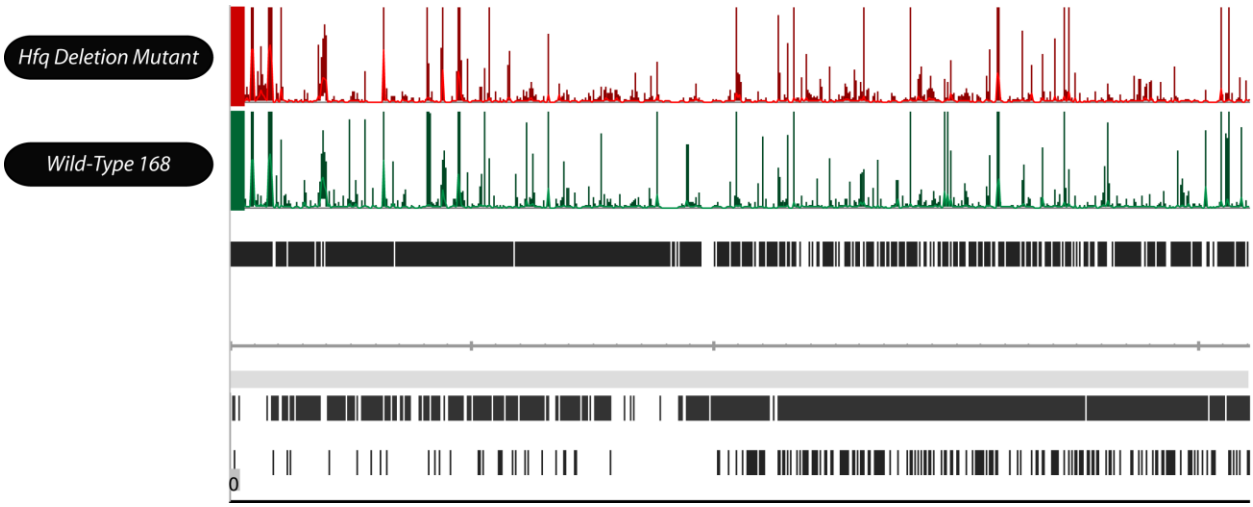


Figure 3-9. Global distribution of cDNA reads for Hfq eletion mutant and wild-type *Bacillus subtilis.* Total RNA was extracted from both wild-type *B. subtilis* 168 or Hfq deletion mutant and subjected to illumina total RNA-seq. The growth conditions used were identical to those for the Hfq pull down experiements. The distribution of reads over the entire *B. subtilis* chromosome are dipicted.

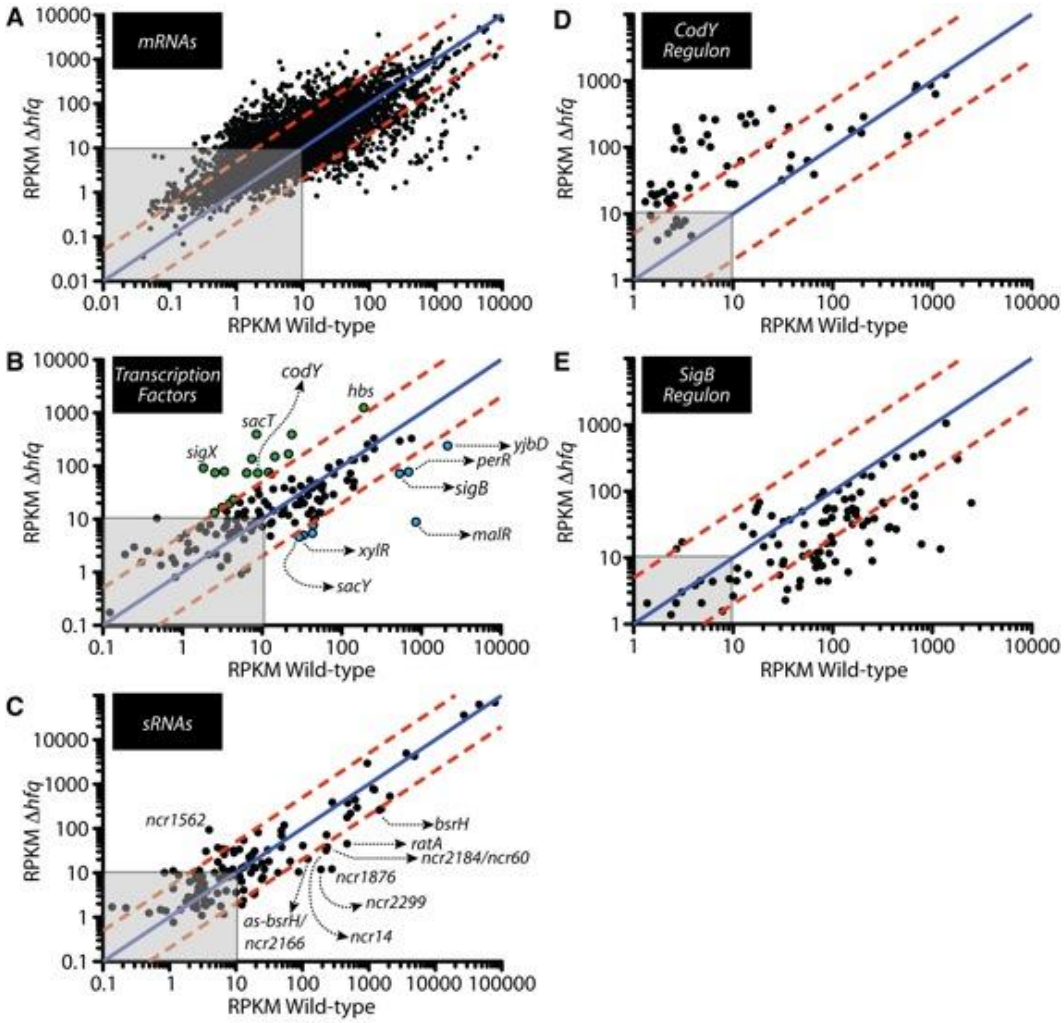


Figure 3-10. Global transcriptomic changes between wild-type *B. subtilis* 168 and the Hfq deletion mutant. In all plots the blue dashed lines represent fold enrichment values of 5 or greater. The grey box denotes RPKM values of ten or lower which is our arbitrary threshold for significant number of reads (A) Depicts the total changes we detect for the Hfq deletion mutant versus wild-type 168. In total nearly 60% of the genome changed in the Hfq KO. (B) sRNAs are capable of producing large changes in gene expression by modulating the expression of transcription factors, thus indirectly influencing gene expression patterns.

(C) Represents the changes in expression level of the cataloge of sRNAs in *Bacillus subtilis.* (D) The transcription factor codY is up regulated in the Hfq mutant, subsequent analysis of genes known to be activated by codY are also upregulated in the Hfq deletion cell line. (E) The stress responsive sigma factor, sigmaB appears up regulated in wild-type in comparison to the *∆hfq* mutant and genes known to be activated by this transcription factor are also up regulated. Suggesting Hfq is required for activation of sigmaB and subsequently its regulon.

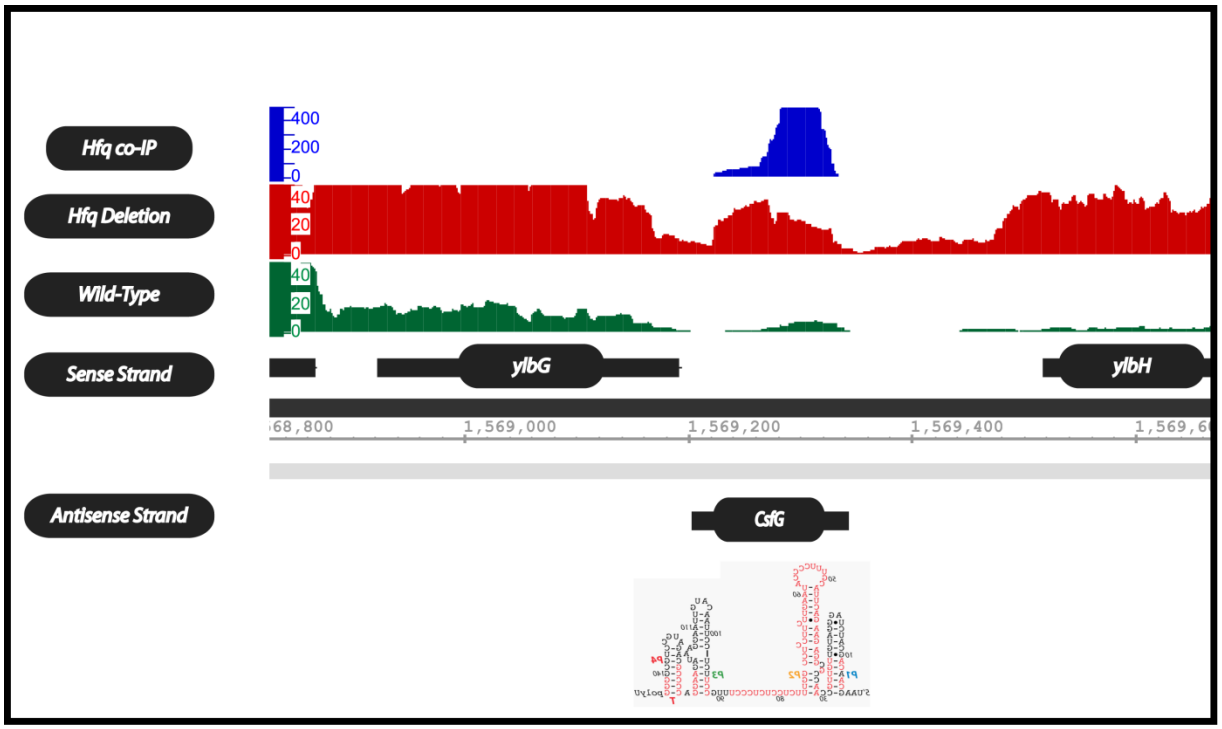


Figure 3-11. Screen shot of the sRNA CsfGs genomic location. The observed increase in CsfG expression observed for the Hfq deletion mutant in comparison to the wild-type is in direct contradiction with the results we obtained using Northern Blotting for this sRNA. Closer examination of this genetic locus offers a potential explanation for this incongruity. It appears that the *ylbG-ylbH* operon is up-regulated in the Hfq deletion mutant (red), and since directionality is lost upon completion of the Illumina sequencing protocol increased cDNA reads on the sense strand opposite of CsfG are incorrectly assigned to the CsfG locus, hence yielding a false positive.

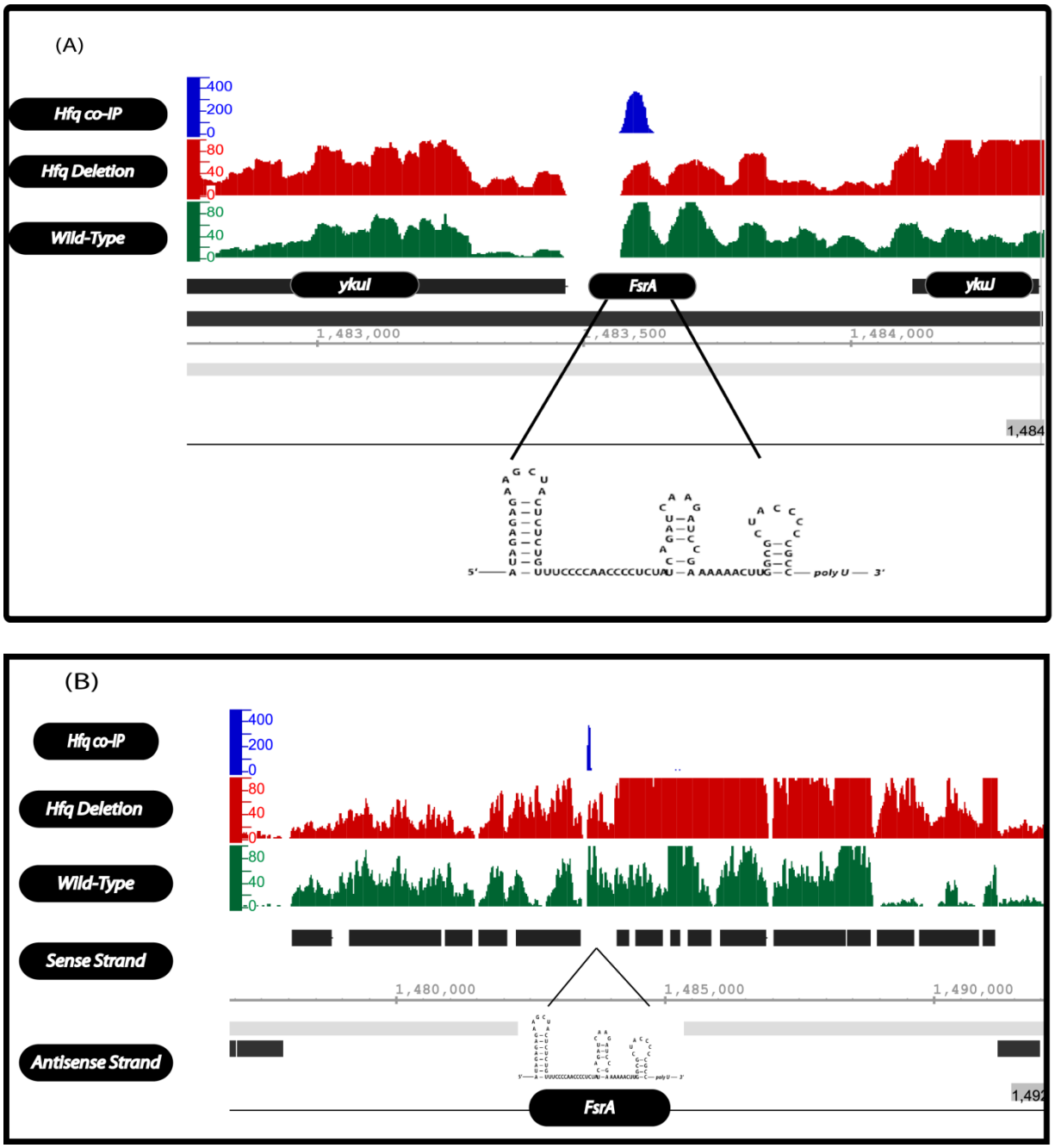


Figure 3-12. Screen shots for the FsrA genomic locus in *B. subtilis.* (A) A close-up screen shot of the FsrA locus. Reads depicted in blue are those that coIP with Hfq. We detected cDNA reads associated with Hfq that mapped over the entire

FsrA sRNA. Reads colored in red and green correspond to mapped cDNA reads derived from total RNA of the Hfq deletion mutant and wild-type 168 respectively. Quantification of Northern Blots specific for FsrA in Figure 3-3B yielded a 50% reduction in FsrA expression in the Hfq mutant versus wild-type and this reduction is recapitulated by the total RNA-seq results depicted above. Notice the lack of reads at the end of the *ykuI* gene which has an intrinsic transcription terminator; however a similar terminator present at the end of the FsrA sRNA results in significant read through of this regulatory element. (B) Is a zoomed out screen shot for the FsrA locus. Notice how this sRNA is in the middle of a 12KB genomic region that does not contain any genes on the antisense strand therefore changes in this region are probably not attributed to differences emanating from the antisense strand.

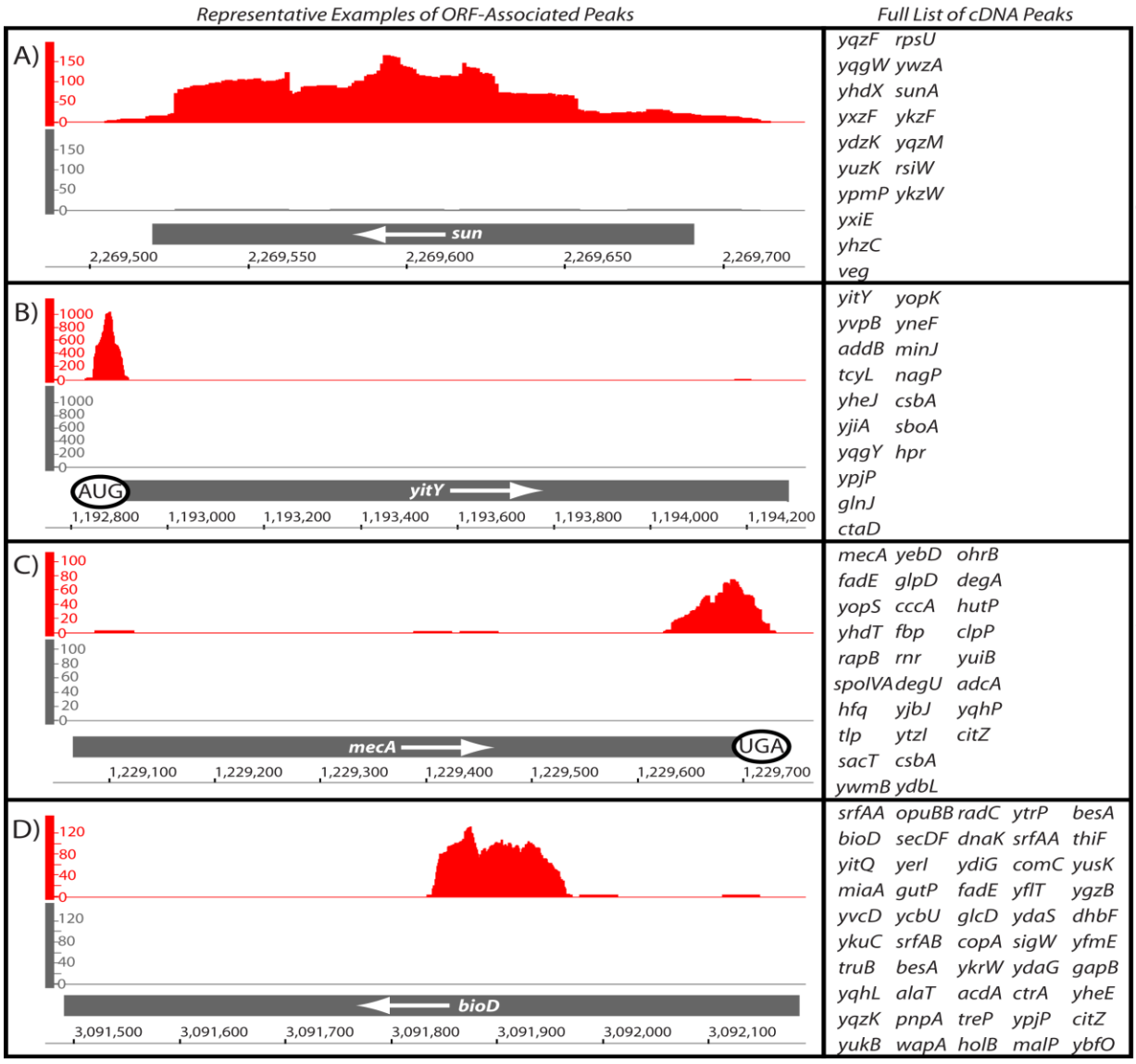


FIGURE 3-13. Representative Hfq-enriched peaks associated with open reading frames (ORF). (A) Representative data for a full mRNA sequence that was enriched by coIP of Hfq. A full list of genes sharing this pattern is shown to the right. (B) Representative data for an Hfq-enriched peak that overlaps with the translational start codon. Other genes sharing this arrangement are listed to the right. (C) Representative data for an Hfq-enriched peak that overlaps with the translation stop codon. Other genes sharing this arrangement are listed to the right. (D) Representative data for an Hfq-enriched peak located within an ORF.

Other genes sharing this arrangement are listed to the right. More details on these peaks are included in Table 2-4. The y-axis shows the volume of cDNA reads as a function of genomic position (x-axis) as determined by Illumina-based sequencing of Hfq-associated RNA molecules. Data shown in gray correspond to the negative control reaction (coimmunoprecipitation of Hfq that lacked an epitope tag).

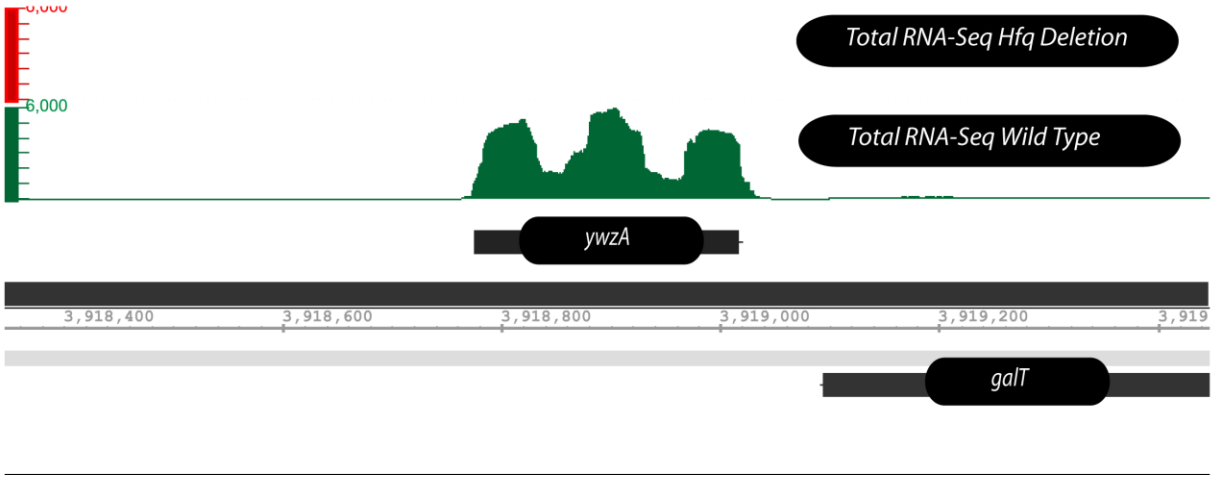
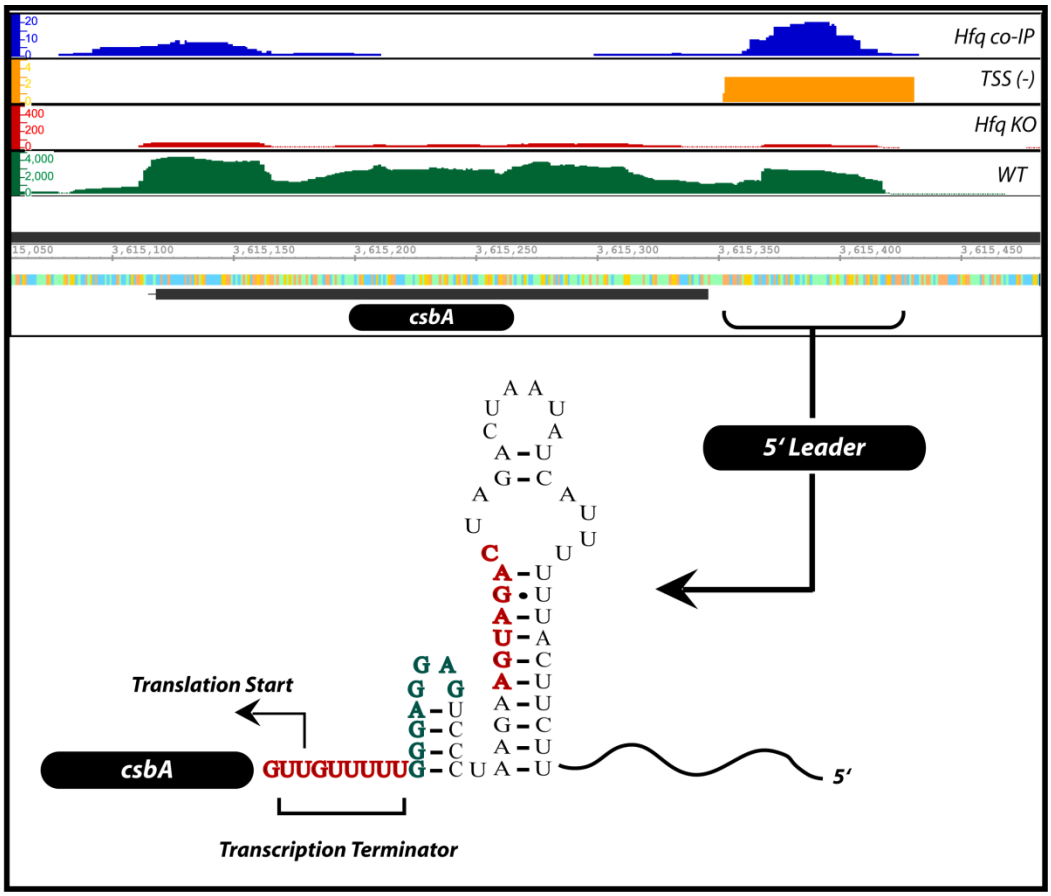


Figure 3-14. The small peptide *ywzA* is massively down-regulated in the Hfq deletion mutant. One class of ORF-associated reads we detect interacting with Hfq are those in which we recover reads spanning the entire ORF (see Figure 3-15A). In general these ORFs tended to code for small peptides of unknown function. One of these in particular, *ywzA* exhibited an over 62 fold-reduction in mapped cDNA reads in the Hfq mutant as compared to wild-type and is depicted above. Notice the low level of mapped reads around this locus, suggesting that the observed difference is not an artifact of sample preparation.



3-15. The gene encoding csbA may be regulated by a 5' leader in an Hfq dependant manner. The gene encoding for the putative membrane transporter csbA displayed Hfq associated reads that mapped to both the start codon as well as the stop codon. Further inspection of 5' region of this gene revealed a putative 5' leader sequence that overlapped with Hfq associated reads. This leader appears to contain a structured element resembeling an intrinsic transcription terminator (red residues) directly in front of the translation start site. Furthermore the putative RBS is sequestered within an inhibitory helix (green residues). This structure appears to be able to form alternative secondary structure that would resilt in the formation of an antiterminator (red residues in helix). Deletion of Hfq results in a 90 folod reduction in csbA mRNA suggesting that expression of this gene is Hfq dependant.

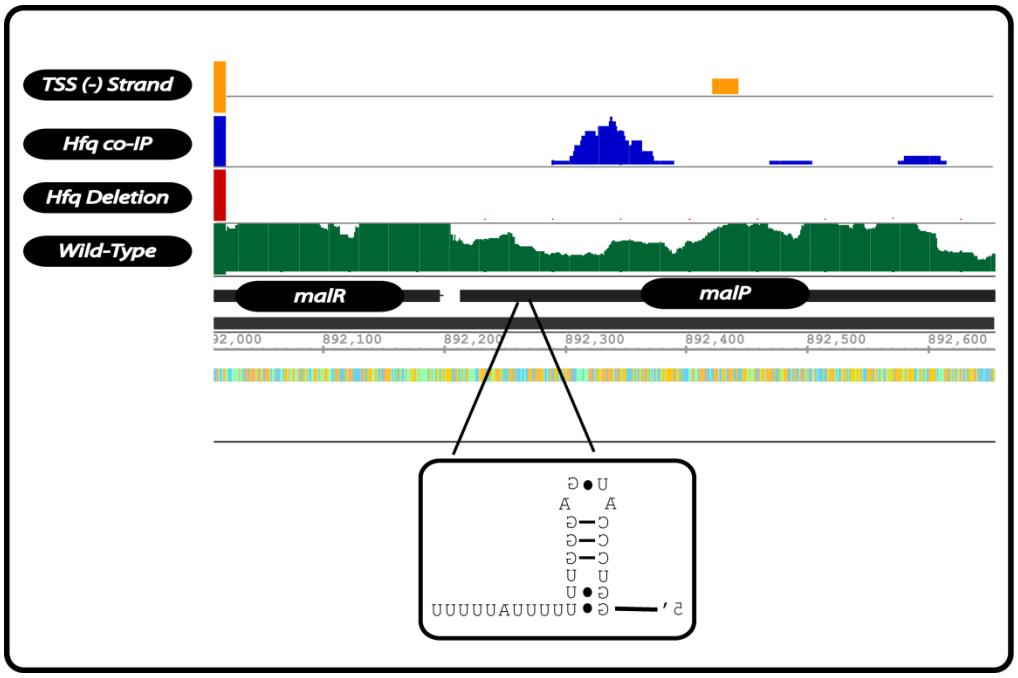


Figure 3-16. Depicts a screen shot of the internal ORF associated peak present in *malP* which may represent a new sRNA present on the antisense strand. We detect a faint TSS signal on the antisense strand of the *malP* locus (depicted in yellow). Reads shown in blue correspond to Hfq associated reads. Closer examination of the genomic region downstream of the reads recovered in the Hfq pull down has the potential to form an intrinsic transcription terminator raising the possibility that these reads may represent a sRNA resident on the antisense strand across from the *malP* locus.

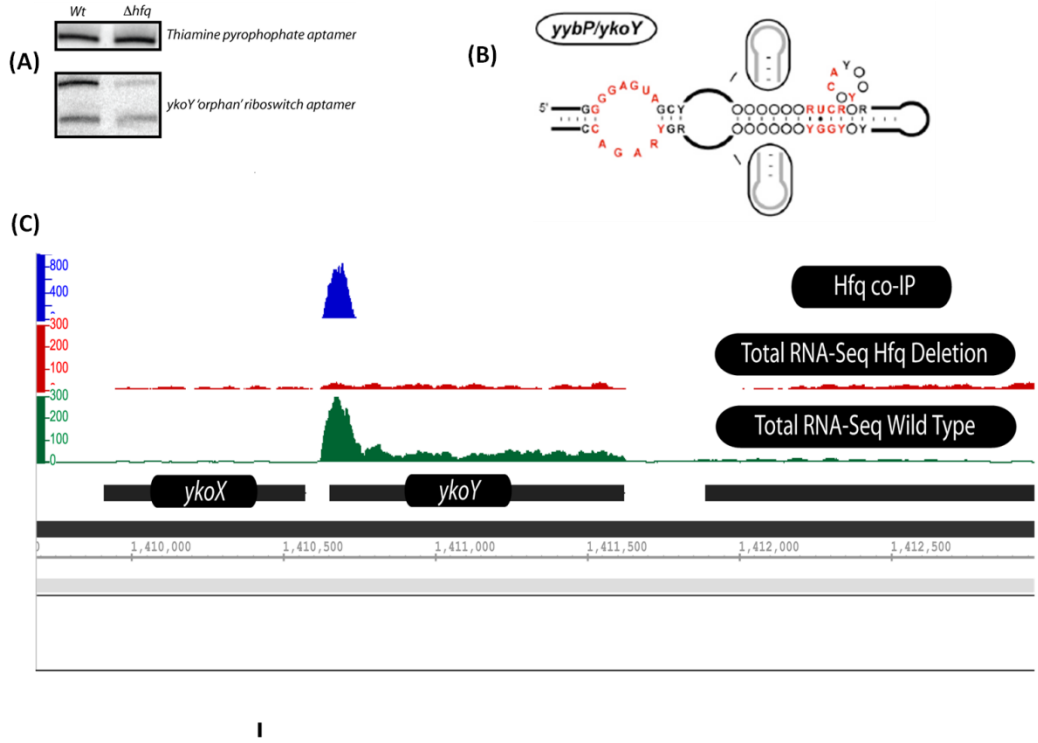


Figure 3-17. The orphan riboswitch *yybP/ykoY* coIPs with Hfq and its abundance is effected by deletion of Hfq. (A) Deletion of Hfq results in decreased abundance of the *ykoY* aptamer domain relative to wild-type. (B) The consensus secondary structure of the *yybP/ ykoY* aptamer domain predicted by comparative sequence alignment (Barrick et al. 2004). (C) Depicts in blue the apatamer domain of *ykoY* which pulls down with Hfq, whereas reads depicted in red correspond to the total RNA-seq of the Hfq mutant and green denotes wild-type.

CHAPTER FOUR

Identification of Sporulation Specific Small RNAs and the Examination of the Global Transcriptome of *Bacillus subtilis* Spores

Introduction on Sporulation

Our perception of bacterial cell biology and ecology has radically changed over the last twenty years. The idea that bacteria exist as simple independent metabolic machines has given way to the realization that many bacteria adopt complex life styles in nature. In order to persist in a highly competitive and constantly changing environment bacteria have evolved the capability of differentiating into phenotypically distinct alternative cell types. The best understood model of bacterial differentiation is the process of sporulation in *Bacillus subtilis.* Upon nutrient deprivation *Bacillus subtilis* executes a tightly cooridinated and highly regulated genetic program that leads to the formation of a stress resistant metabolically dormant endospore. Initiation of re-growth occurs via a process known as germination and is initiated in response to environmental cues such as amino acids and cell wall muropeptide fragments derived from growing cells (Higgins and Dworkin 2012, Shah et al. 2008).

The onset of sporulation is initiated via a complex regulatory cascade that is governed by the phosphorylation state of the master transcriptional regulator Spo0A. Phosphorylation of Spo0A is intitiated via several different PAS domain containing sensory histidine-kinases (KinA, KinB, KinC, KinD and KinE) the major one being KinA. The physiologic signals that these sensory kinases respond to are unknown with the exception of KinC, which is activated in response to changes in the intracellular concentration of potassium ions; regardless limited evidence suggests that each of these kinases probably respond to different cellular and environmental cues (Jiang et al. 2000, Errington 2003, Lopez et al. 2009, Higgins and Dworkin 2012). Phosphorylation of Spo0A does not occur directly but instead is mediated through a phosphotransfer system involving two intermediates Spo0F and Spo0B (Errington 2003). These proteins in turn are regulated by phosphatases, which are subjected to regulation via other proteins; illustrating the complex layers of regulation governing induction of sporulation (Errington 2003, Higgins and Dworkin 2012). Phosphorylation of Spo0A leads to transcription of over 120 genes, and results in the activation of σ^{A}-RNAP and σ^{H}-RNAP which triggers entry into sporulation (Errington 2003, Kroos 2007, Higgins and Dworkin 2012).

Entry into sporulation is characterized initially by asymmetric membrane septation, which is followed by an engulfment step that results in a double membrane-bound pre-forespore. Proteins are then expressed within the forespore and within the mother cell environments to assemble a unique cell wall called the cell cortex, which is surrounded by a protein-rich cell coat. The transcriptional mechanisms that govern these processes have been extensively studied (reviewed in (Piggot and Hilbert 2004, Kroos 2007). A hierarchical cascade of sigma factors, which communicate between the forespore and mother cell compartments are required. Other transcription regulatory factors also participate in the overall pathway, which also exhibit specific spatial and temporal restrictions. However, the potential role for post-initiation regulatory mechanisms has not been as well studied for the endospore developmental program, although several prior searches for small, trans-acting regulatory RNAs (sRNAs) have uncovered a few potential candidate sRNAs (Silvaggi *et al.,* 2006; (Schmalisch et al. 2010);(Marchais et al. 2011)).

Results

CsfG is a fore spore specific sRNA

Upon examination of the promoter sequences of the 22 sRNA candidates we recovered in our Hfq pull down experiment, we identified a putative σ^F dependant promoter for CsfG suggesting that this sRNA may be involved in sporulation. In support of this, covariance model searches for CsfG revealed occurrences in many different endospore-forming Bacillales genomes (Figure 4-1A) (Marchais et al. 2011). In addition, this RNA element is strongly conserved at both the primary sequence level as well as the overall seconday structural arrangment (Figure 4-1A-C). Interestingly, CsfG contains multiple tracts of 3 to 4 repeated C residues which are present in terminal loops or unstructured regions of the sRNA and are over 95% conserved. The C-rich motif was recently identified in the *S. aureus* sRNA RsaE, as a conserved element present in multiple Gram-positive species that contain this sRNA (Geissmann et al. 2009). The significance of this motif is believed to stem from its functional ability to base pair with G-rich ribosome binding sites and as such the C-rich motif is usually present in terminal loops and unstructured regions of the sRNAs which contain them (Figure 4-1B).

A role for this sRNA in sporulation has been proposed previously, based on its genomic location and the fact that a putative sporulation-specific promoter sequence is also conserved upstream of *csfG* (Figure 4-2A and B)

(Amaya, Khvorova and Piggot 2001, Marchais et al. 2011). Indeed, microscopy studies conducted in our lab by Vinetha Zacharia utilizing a plasmid containing this putative promoter region fused to an YFP cellular reporter gene revealed that it is transcriptionally active only within the forespore under conditions leading to activation of the sporulation developmental program (Figure 4-3C and 3D). Further supporting this is the fact that an σ^F deletion strain fails to produce YFP when placed in nutrient limiting conditions that induce sporulation. Interestingly cells in which chromosomal CsfG has been deleted display no defect in sporulation (Marchais et al. 2011). In fact deletion of Hfq has no discernible effect on sporulation either (Silvaggi et al. 2006). One possible explanation for a lack of phenotype associated with deletion of CsfG could be redundancy. For example, quorum sensing in *Vibrio cholera* is governed by four Hfq-dependant sRNAs, and deletion of all four are required before a deficit is observed (Lenz et al. 2004). Alternatively, CsfG may be important in regards to progression through sporulation. We did not conduct time course analysis of sporulation efficiency, instead we examined the effects of the strain deficient for CsfG with wild-type at 24 hours post-induction of sporulation; thus we conducted end-point assays which would not reveal any changes in the rate of sporulation. It is also plausible that deletion of CsfG may affect the rate of germination as compared to wild type *B. subtilis* for the same reason just mentioned. However,

a function for CsfG in germination seems unlikely since we were unable to detect this sRNA in fully differentiated spores (see below). Currently, we are undertaking experiments in the lab to try and determine the functional significance of this sRNA, as well as trying to identify putative targets. In order to address these questions I constructed a constitutively expressed CsfG that was integrated into an ectopic locus in the *B. subtilis* genome. Our aim is to initiate sporulation, collect cells every two hours, and then perform next-generation sequencing; comparing these expression profiles to wild-type and *csfG* deletion mutants grown under the same conditions. It is our expectation that we will observe many changes in RNA expression patterns between the over-expression strain and the deletion strain. The challenge then becomes determining primary targets for CsfG as compared to secondary changes in gene expression for reasons outlined in the introduction of this manuscript.

mRNA expression in spores

To gain further potential insight into sporulation-specific expression patterns, we analyzed the general transcriptome of spores by RNA-Seq. Part of the goal of this analysis was to discover the suite of transcripts that were significantly enriched in the endospore, after completion of the developmental

pathway. This would likely reveal the transcripts that are stored for the metabolic and anabolic needs during germination. In addition, transcriptomic profiling of terminally differentiated spores may reveal new, novel sRNAs that may be important for this developmental program. Since the spore represents the final product of the overall pathway, we compared its transcriptome to cells deleted of σ^F (*ΔsigF*). σ^F is the first forespore specific sigma factor activated upon entry into sporulation, thus cell lacking this sigma factor are blocked at the earliest committed stage of the developmental pathway and subsequently cannot differentiate into spores.

Specifically, in this analysis the expression patterns of the *sigF* mutant were compared to that of free spores, which were accumulated after 24 hours of incubation in liquid sporulation medium. Quantification of colony forming units for total cells and chloroform-treated cells revealed that, after 24 hours, sporulation had progressed to completion for the wild-type strain. The *sigF* mutant was cultured in identical medium as compared to the final, wild-type spores; therefore, changes in gene expression would not result from changes in medium composition between the strains. In general, the overall expression level of all genes was lower than for wild-type cells cultured under non-sporulating conditions (data not shown), which is likely to reflect lowered overall

mRNA abundance in cells undergoing sporulation (Moeller et al. 2006). The observed decrease in the amount of global RNA as compared to the *sigF* mutant may also be a result of technical limitations of RNA isolation from spores (Moeller et al. 2006). However, we are unable to make any quantitative statements to this effect since we focused our analysis on the narrow comparison between the *sigF* mutant and free spores.

A total of 607 genes were enriched in spores by at least 2-fold with 187 of these genes displaying enrichment values greater than fivefold relative to *sigF* (Figure 4-4 and 4-5). Of which 32% (2-fold enrichment) and 31% (5-fold enrichment) were predicted to function during sporulation, the remaining up regulated genes belonged predominantly to gene categories predicted to function in transport and anabolic biosynthetic processes (Figure 4-5 and 4-6). The fact that the majority of genes up regulated 2-fold or more in comparison to the *ΔsigF* mutant belonged to sporulation and biosynthetic processes is not surprising. Since the *sigF* mutant is blocked at the earliest committed stage of sporulation, one would expect large differences between spores and the mutant. The large increase in biosynthetic genes is also logical for two reasons. First, sporulation is induced by culturing cells in a nutrient deficient media, however the *sigF* mutant cannot sporulate in order to deal with these limiting conditions.

Instead, the cell activates a catabolic stress response in an attempt to stave of starvation. This is supported by the gene categories that are up regulated in the *sigF* deletion mutant which consist largely of nutrient transporters for amino acids and carbohydrates, genes involved in proteolysis, and pathways involved in catabolic metabolism (Figures 4-7 and 4-8). Conversely, the increase in mRNAs involved in anabolic metabolism within the spore is logical in the context of germination. The intracellular build up of mRNAs involved in anabolic processes within the dormant spore effectively "primes" the cell to grow and divide once a signal for germination is received.

Of these, 45 had been previously implicated as members of sporulation sigma factor regulons (Figure 4-4B) (Eichenberger *et al.,* 2004; Wang *et al.,* 2006). However, upon manual examination of the expression data, many of the genes exhibiting RPKM values <10 were expressed at levels low enough to result in incomplete coverage of the full gene. Therefore, we analyzed the data more closely using an arbitrary RPKM cut-off of 10. This resulted in a small catalog of genes (N=48) that were expressed at detectable levels and that were enriched in spores (from 5-fold for *cysH* to 290-fold for *ykzP*) (Figure 4-4C). Manual examination of these transcripts revealed the expression data to cover the full length for each of the respective genes. Of these genes, 15 have been predicted

to be members of either the *sigF* or *sigG* regulons (Wang *et al.,* 2006). Interestingly, the remaining 34 have not been predicted to have a specific role in sporulation gene expression patterns, which makes them an intriguing set for future analyses by researchers studying bacterial endospore formation. Most of these genes corresponded to genes of unknown function; however, putative metabolic functions have been attributed to some, including maltose utilization (*malA*, *malR*, *malP*), phosphate transport (*pstS*), arginine synthesis and transport (*argF*, *argH*, *artR*), the mannitol phosphotransferase system (*mtlF*), an RNase P RNA subunit (*rnpB*), lactate transport (*lctP*), and cysteine synthesis (*cysK*, *cysH*). It is not immediately obvious why these respective gene categories might be enriched in spores.

sRNA expression in spores

Our goal in collecting transcriptomic measurements of free spores and the *ΔsigF* control strain was primarily to investigate sRNA populations at the start and end of the sporulation pathway. In particular we were interested in examining the expression profiles of four new sporulation specific sRNA candidates recently discovered in our lab termed CsfG, SurG, SurE, and SurD (Irnov, Dambach, Winkler unpublished results). These four putative sRNA regulators were identified through deep-sequencing of a *B. subtilis* strain mutant

for a regulatory RNA required for biofilm formation, termed Exopolysaccharide Associated RNA (EAR) (Irnov and Winkler 2010). The EAR element is required for antitermination of the exopolysaccharide operon (EPS), cells mutant for EAR lead to premature transcription termination of this operon and subsequently fail to produce EPS. Production of EPS is a hallmark of biofilm formation in all bacteria and serves as a matrix that functions to both protect and provide structural integrity to the bacterial community (Aguilar et al. 2007, Lopez et al. 2009, Straight and Kolter 2009). Cells mutant for EAR fail to produce a complex colony architecture (Figure 4-9A), which is characteristic of *B. subtilis* biofilms and display an over 1000-fold reduction in spores in comparison to WT cells capable of forming biofilms (Irnov and Winkler 2010). In a biofilm, only a small number of cells undergo sporulation, which takes place preferentially at vertical extensions called fruiting bodies (Branda *et al*., 2001). Previously, it was determined that a subpopulation of the extracellular biofilm matrix-producing cells further differentiates into endospore-forming cells (Vlamakis *et al.,* 2008). Despite the fact that none of the *eps* genes directly participate in sporulation, their deletion led to elimination of fruiting body formation and reduced the degree of sporulation for biofilms on solid medium; sporulation was not influenced for the mutant strains when grown in liquid medium (Vlamakis *et al.,* 2008 (Irnov and Winkler 2010)).

The lack of EPS production severely affected the formation of endospores in a biofilm community more than any other cell types (Figure 4-9C). Therefore, it is very likely that the full catalog of sporulation-specific sRNA regulators also exhibited decreased expression under these experimental conditions. In other words, we reasoned that the comparison of the transcriptomic profiles of wild-type and EPS-deficient communities should offer a functional means for identifying the catalog of sRNAs specifically expressed during sporulation. Also, by extracting total RNA from the biofilm community, we reasoned that we should be able to detect sRNAs that are expressed at any temporal stage during sporulation, given that sporulating cells were not synchronized within the community. Therefore, we chose to investigate whether the small catalog of sRNAs that were specifically decreased in the EPS-deficient community exhibited sporulation-specific expression. To investigate the potential for sporulation-related expression of these sRNAs Vinetha Zacharia fused the DNA sequences containing putative promoter elements upstream of SurD, SurE, and SurG to an YFP reporter, in an analogous manner as had been done for CsfG, and monitored YFP expression at various time intervals during sporulation (Figure 4-10-12).

Interestingly, the *surD* promoter-YFP fusion revealed a transient YFP signal within the pre-forespore compartment at earlier time points followed by a

stronger fluorescent signal occurring specifically within the mother cell compartment at later points (Figure 4-10D and E). Consequently, a closer inspection of the region upstream of *surD* revealed the presence of a potential promoter region exhibiting weak homology to the forespore-specific sigma factors, *sigF*, *sigG*, as well as the mother cell-specific *sigK* factor. Again, the *sigF* mutant strain showed no observable fluorescent signal, supporting an expression pattern that was specific to the sporulation pathway-dependent transcription factors. Fusion of the region upstream of *surE* to YFP revealed promoter-active expression specifically from within the mother cell environment, presumably from reliance upon *sigE* and/or *sigK* (Figure 11D and E). Finally, the region upstream of *surG* was fused to YFP and monitored for spore-forming cells (Figure 12B and C). This promoter fusion was specifically active in the developing forespore relative to the *sigF* control strain. Therefore, all four of the significantly (>30-fold) decreased sRNA candidates appeared to be specifically expressed during sporulation in a compartment-specific manner. Given that the transcription factors that exhibited the most significantly decreased expression also correlated with sporulation, it is clear from the aggregate data that the functional loss of sporulation within the biofilm-deficient community could be exploited as a basis for discovering new elements that are specifically expressed during sporulation. Moreover, just as the key transcription

factors for sporulation exhibited the greatest decreased expression, we anticipate that the four sporulation-specific sRNAs observed herein are likely to constitute the key post-transcriptional factors for sporulation.

Most of the ~100 sRNAs that have been identified previously (Rasmussen *et al*., 2009; Irnov *et al.,* 2010 and reference within) were not detectable in either free spores or the *ΔsigF* strain cultured in sporulation medium. In fact, only 17 of the greater than 100 putative sRNAs predicted for *B. subtilis* meet our threshold criteria of having an RPKM value of at least 10 (Table 4-1). Among this group only three exhibited a fold enrichment value greater than 2. The *sigF* mutant also had 17 sRNAs with RPKM values greater than 10, with 15 sRNAs being common to the sRNAs present in spores which meet our cutoff; nine of these sRNAs had a fold enrichment greater than 2 in comparison to spores (Table 4-2). Given the mother cell-specific expression patterns of SurD and SurE, we also did not anticipate them to be present in spores. Indeed, expression of these sRNAs was absent in both *ΔsigF* and wild-type spores (Figure 13A-B). Interestingly, expression of the forespore-specific CsfG was also virtually undetectable from both strains. This appears to conflict with the observation that the putative promoter region of CsfG activates YFP production in the forespore, including at later time points (Figure 4-3). However, as we examined

cells that had completely differentiated into spores it is likely that CsfG functions during the sporulation developmental program in a transient fashion and thus does not accumulate. It is also feasible that the fluorescent protein accumulated in the forespore to such a degree that its sensitivity as a reporter was insufficient to detect CsfG mRNA repression at later time points.

The most dramatically enriched sRNA we detected in spores was the SurG sRNA identified in our transcriptome studies of EPS deficient *B. subtilis* biofilms. The SurG sRNA was surprisingly enriched in free spores, exhibiting a 330-fold increase in expression relative to the *ΔsigF* strain (Figure 13A-C). The expression levels of SurG rivaled that of tmRNA and 4.5 S RNA, and it was more abundant than 6S-1 or 6S-2 sRNAs. This observation suggests that SurG may accumulate in spores for a potential functional role during latency, or, alternatively, during germination. Finally, we searched all of the intergenic regions of the *B. subtilis* genome for RNA peaks that might correspond to new sRNAs; however, new sRNA candidates were not identified in these particular data. Taken together, these results suggest that CsfG, and other putative forespore-specific sRNA candidates, are most likely to be important during intermediate stages of sporulation while, in contrast, SurG is likely to accrue to high concentrations in spores for a role that is important to free spores.

Further manual inspection of SurG revealed that it exhibited features unique from most antisense-dependent sRNAs. Specifically, we found a repeated sequence motif, 5'-UGAGGUG-3', within the SurG sequence that exhibited the potential to participate in formation of a small stem-loop. Also, a remarkably similar sRNA could be identified for other *Bacillus* species, suggesting that this particular sRNA element is widely conserved in a subset of endospore-forming Firmicutes species (Figure 14A and B). Intriguingly, this conserved sequence motif is similar but not identical to the consensus-binding site of the RNA-binding protein, CsrA (Figure 14C). In *E. coli* and other Gram-negative bacteria, CsrA has been shown to activate or repress gene expression by binding to hairpin structures with a 5'-GGA-3' sequence located within the terminal loop region, typically in close proximity to the ribosome binding site (Dubey *et al.,* 2004). In general, CsrA is believed to activate genes involved in motility and glycolysis while repressing genes involved in glycogen synthesis, peptide transport and biofilm formation (reviewed in Timmermans and Van Melderen, 2010). CsrA activity is usually antagonized by small noncoding RNAs that contain a tandem array of individual CsrA-binding sites (*e.g.,* CsrB and CsrC in *E. coli* and RsmY in *Pseudomonas* species; reviewed in Babitzke and Romeo, 2007). The expression of these CsrA-sequestering sRNAs is presumed to titrate CsrA away from its mRNA targets. In contrast, the role of CsrA in *B. subtilis* and other Gram-positive

bacteria is still not well understood. In *B. subtilis*, CsrA has been shown to repress translation of the highly abundant flagellin protein, encoded by *hag* (Yakhnin *et al.,* 2007). More recently, a protein antagonist for *B. subtilis* CsrA was identified. Specifically, secretion of flagellin during flagellar assembly results in the freeing of FliW protein from a complex of FliW and Hag proteins (Mukherjee *et al.,* 2011). FliW then associates tightly with CsrA and prevents its repression of *hag*. However, no bona fide 'CsrB-like' sRNA has been identified in *B. subtilis*, or other Gram-positive microorganisms. Therefore, from our aggregate data, we hypothesize that the SurG sRNA may constitute a CsrA-sequestering sRNA for *B. subtilis* and other *Bacillus* species. Also, based on our observation of a forespore-specific promoter for the SurG sRNA and the striking accumulation of SurG molecules within spores, we hypothesize that this sRNA may specifically antagonize an RNA-binding protein such as CsrA within the forespore. However, it is of course also possible that an RNA-binding protein other than CsrA is the target for binding to the SurG sRNA; future experimentation will be required to differentiate these scenarios.

Discussion and future directions

A subset of the sRNA genes characterized herein was active in the forespore, while another subset was specifically expressed later in the mother cell after

engulfment of the forespore. Surprisingly, SurD showed a dual expression pattern switching from an early forespore-specific expression into a mother cell-specific expression later in the developmental pathway. Similar dual expression patterns have been observed for the CtpB protease required for SigK activation (Camp and Rudner, 2007). However, in this case, *ctpB* is synthesized in both the forespore and mother cell compartments at the same time as opposed to a spatiotemporal switching that was observed for *surD*. Therefore, just as an ordered, hierarchical progression of transcription factors transpires between the two compartments for transcriptional control of target transcripts, we speculate that this overall catalog of sporulation-committed sRNAs may offer a similar progression, albeit via a post-initiation layer of genetic control (Figure 4-15).

Future experimentation will be required for determination of the molecular functions of these sporulation-specific sRNAs. The functional characterization of bacterial sRNAs can oftentimes be challenging, as their regulatory functions can sometimes be redundant (*e.g.,* Lenz *et al.,* 2004), and in other instances they may be required for "fine tuning" of gene expression patterns. Therefore, the functions of these sRNAs remain to be elucidated; however, several (CsfG, SurD, SurE) exhibit a secondary structure organization that is consistent with known *trans*-acting regulatory sRNAs (Figures 4-3 and 4-10-11). Therefore, we speculate

that they are likely to associate with target mRNAs to affect translation or mRNA stability (Figure 4-15B). Prediction of mRNA targets for sRNAs is a notoriously difficult task. Nonetheless, we searched the conserved regions of the sRNAs using RNAPredator (Eggenhofer *et al.,* 2011) for the highest scoring mRNA target matches (Tables 4-3-5). Although sRNAs have been also shown to target coding regions as well as 5' untranslated regions (Waters and Storz, 2009), we restricted our search for CsfG to the 5' leader regions of potential target mRNAs in order to limit the list of potential targets, which was surprisingly large for this particular sRNA. The 15 top scoring hits to CsfG that satisfied this criterion including genes for diverse metabolic functions, such as uncharacterized permeases and efflux proteins, terpenoid synthesis, the cytrochrome c550 *cccA*, membrane-associated proteins, glutamate synthase, citrate synthase and dihydrodipicolinate reductase. Again, these predictions must be cautiously interpreted and numerous lower scoring hits involve functions that are also directly pertinent to sporulation, such as cortex assembly; only dedicated experimentation will resolve the actual targets. Similar analyses with SurD and SurE revealed surprisingly few top scoring candidate targets with base-pairing interactions that were proximal to the ribosome binding site. Instead, the majority of these target loci involved excerpts of protein coding regions. Assessing the top 15 candidates overall revealed a few potentially intriguing metabolic connections between

putative targets. For example, the top 15 hits for SurD included several secondary metabolite genes, including hits to three separate genes involved in plipastatin synthesis (*ppsA*, *ppsC*, *ppsD*), and the *pks* polyketide synthesis pathway (*pksJ*). The SurD target list also included the endopeptidase, *spoIIR*, that is expressed in the forespore and that activates *sigE* within the mother cell. It is worth noting that SurD appeared by our analysis to be expressed in the forespore at early time points, where it could theoretically impact the expression dynamics of *spoIIR*, and then was increased in expression in the mother cell, where it could impact expression patterns of other, diverse mother cell genes. Other potential targets included a DNA ligase, DNA helicase, RNA helicase, a 2' phosphodiesterase, and a few sporulation genes (*spoIVA*, *spoVAA*). The 15 top scoring hits for SurE included multiple genes for potassium transport (*ktrB*, *khtT*), a proline iminopeptidase (*ybaC*), a sensor histidine kinase for membrane fluidity (*desk*), a response regulator phosphatase (*rapD*), a metalloprotease (*mpr*) and multiple genes for citrate transport (*citZ*, *cimH*). However, it is again important to note that these predictions must be cautiously interpreted and will only be meaningful after direct experimentation.

In general, we hypothesize that the sRNA list presented herein is likely to comprise the fundamental sRNA catalog involved in sporulation for *B. subtilis*

and, possibly, for other *Bacillus* species. These data also revealed their compartment-specific expression patterns, including their presence or absence in fully formed spores. This study therefore reveals new features of the genetic circuitry that are likely to underlie the endospore developmental pathway. Also, it highlights how experimental conditions that functionally alter the proportion of a specific cellular class can be used to discover the sRNA catalog that is specifically expressed from within those cells. From the total catalog of putative sRNAs in *B. subtilis*, we anticipate that individual subsets will be specifically expressed by other cell types, such as those expressing competence machinery, motility organelles, etc. Therefore, variations to the methods used herein should ultimately prove useful in classifying the many functional roles of putative sRNA regulatory molecules in *B. subtilis*.

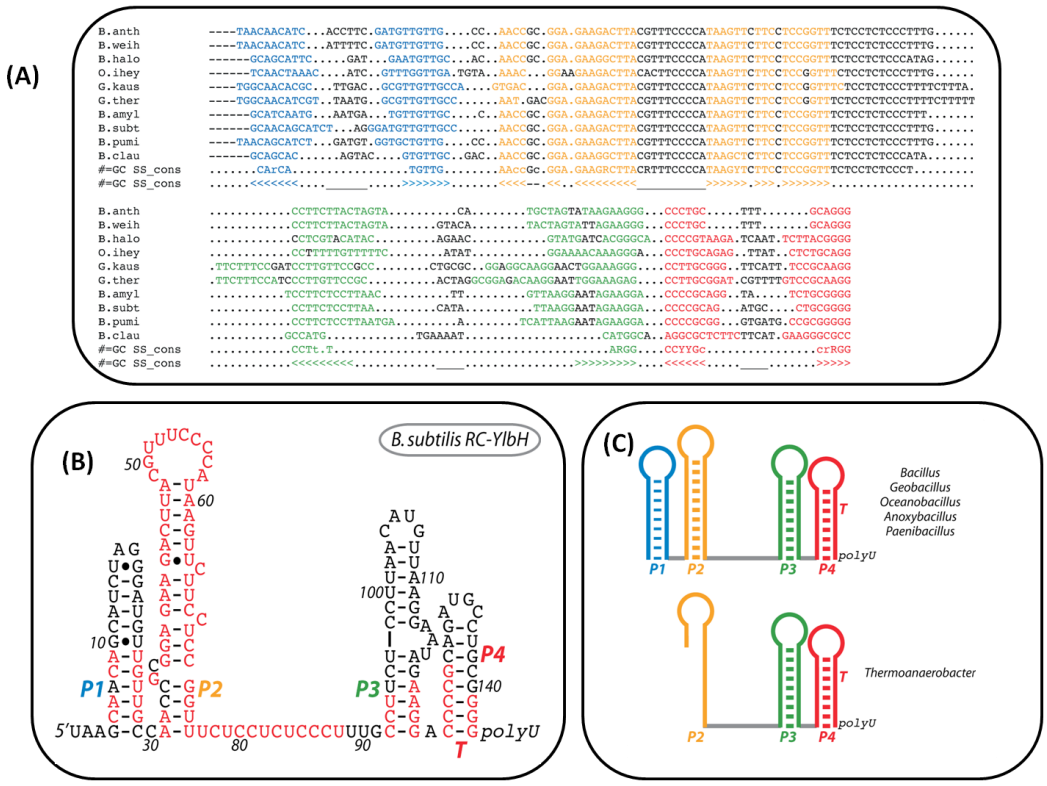
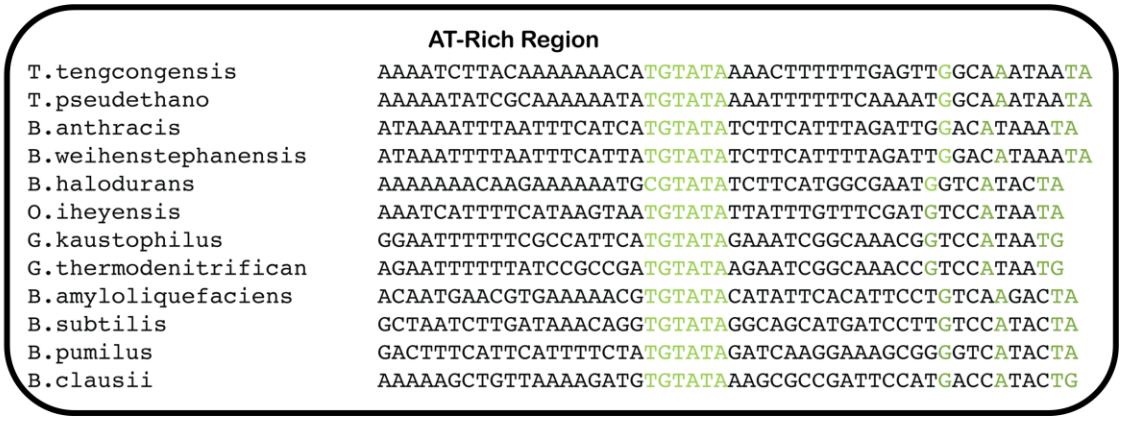


Figure 4-1. The sRNA CsfG is conserved in endospore-forming *Bacillales* genomes. (A) CsfG is conserved at both the primary sequence level as well as the overall secondary structural arrangment in the genomes of endospore-forming *Bacillales.* The colored residues correspond to pairing which result in formation of helical elements. (B) Depicts the consensus secondary structure of CsfG. Residues colored red are 95% conserved where as those colored black are 80% conserved in organisms containing this sRNA. The color of the "P" label corresponds to the colored residues in A which comprise the respective helix. (C) Represents the various Genus' containing CsfG. Covariance computational searches identified a truncated form of this sRNA in *Thermoanaerobacter.*

(A)



(B)

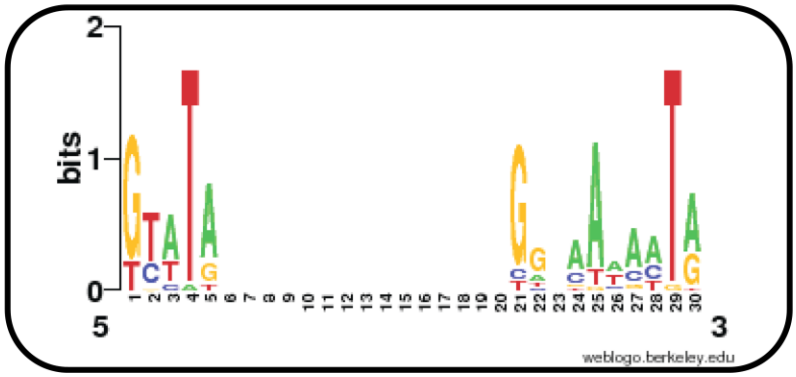


Figure 4-2. CsfG contains a strong fore-spore specific σ^F promoter. (A) Comparative sequence alignment of putative promoter regions upstream of CsfG in organisms containing this sRNA identifies potential σ^F binding sites which are colored in green. (B) Weblogo representation of the nucleotide composition for the σ^F promoter element upstream of CsfG.

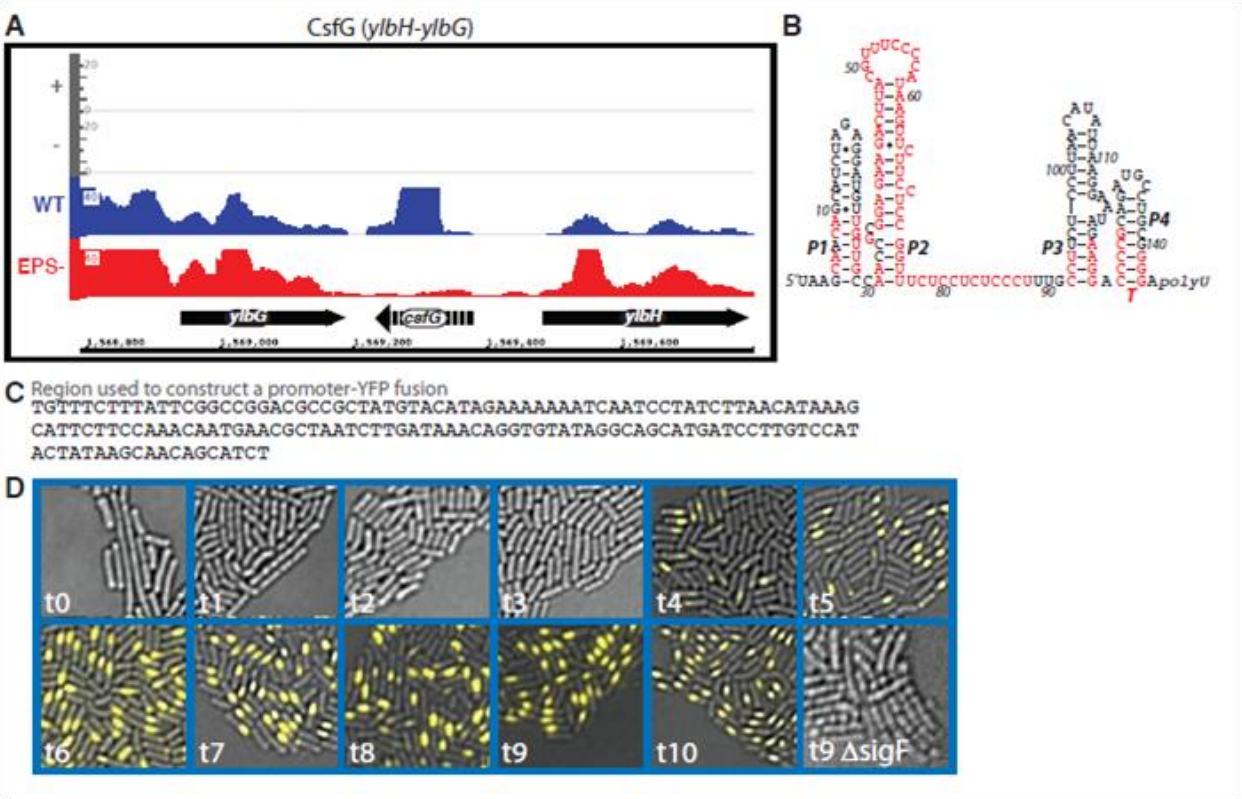


Figure 4-3. Analysis of the CsfG sRNA under sporulation inducing conditions. (A) The genomic locus for CsfG is shown at the bottom with the two flanking genes. The distribution of cDNA reads obtained from the RNA-Seq analysis of the wild-type and EPS-deficient biofilm colonies are represented by blue and red graphs, respectively. Also shown are 5'-transcription start site mapping data from stationary phase cells for both the positive and negative genomic strands, as shown in grey (as reported by Irnov *et al.,* 2010); no such transcription start sites were observed in early stationary phase cells. The secondary structure of the *B. subtilis* CsfG sRNA is shown in (B); red lettering denotes positions that are >95 % conserved in the sequence alignment. The genomic sequence located immediately upstream of the CsfG sRNA gene is shown in (C). This sequence was subcloned into an *amyE* integration vector (pDG1662) to control transcription of a promoter-less copy of yellow fluorescent protein (YFP) that was dependent on an *rpsD* ribosome binding site. Expression of the CsfG promoter-YFP fusion appeared specifically in the forespore compartment (D), but was absent from a Δ*sigF* control strain.

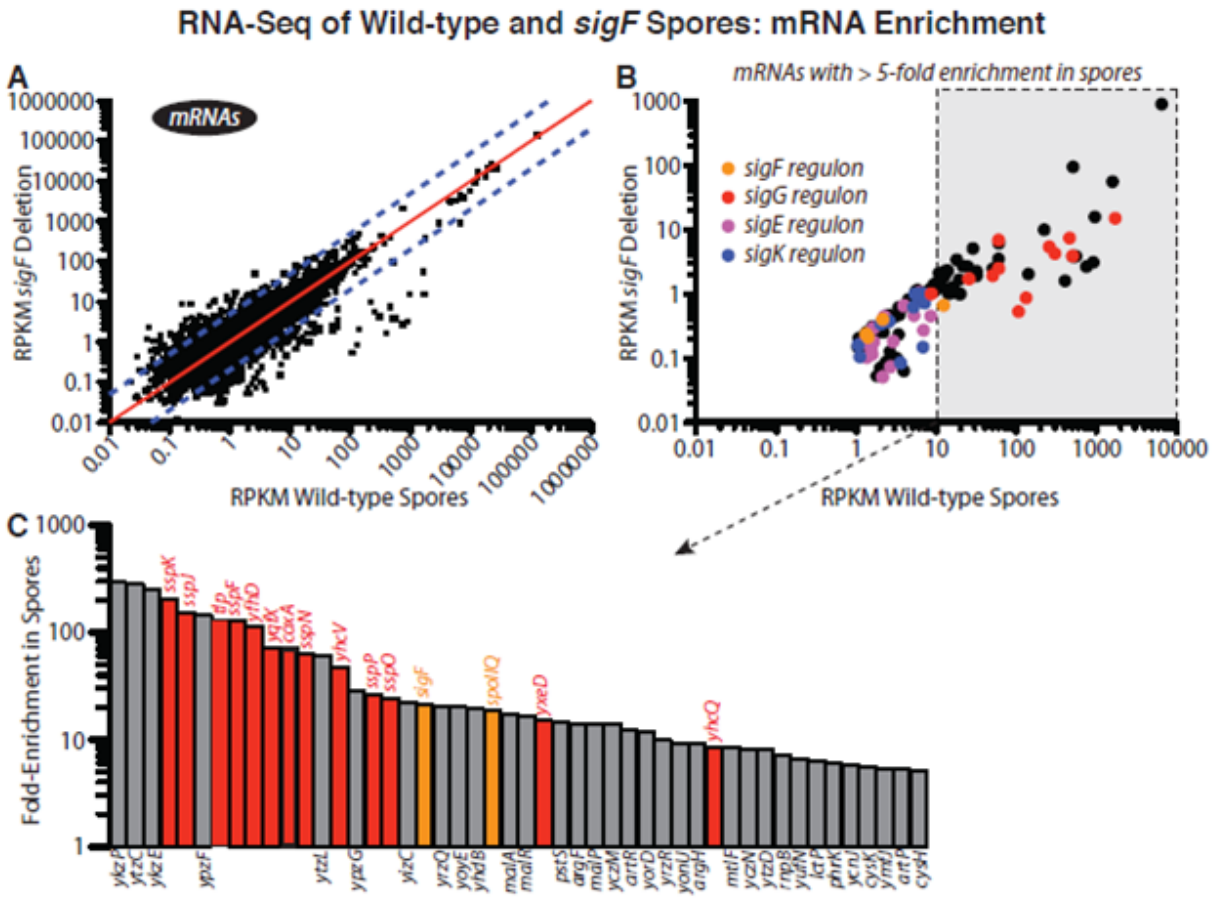
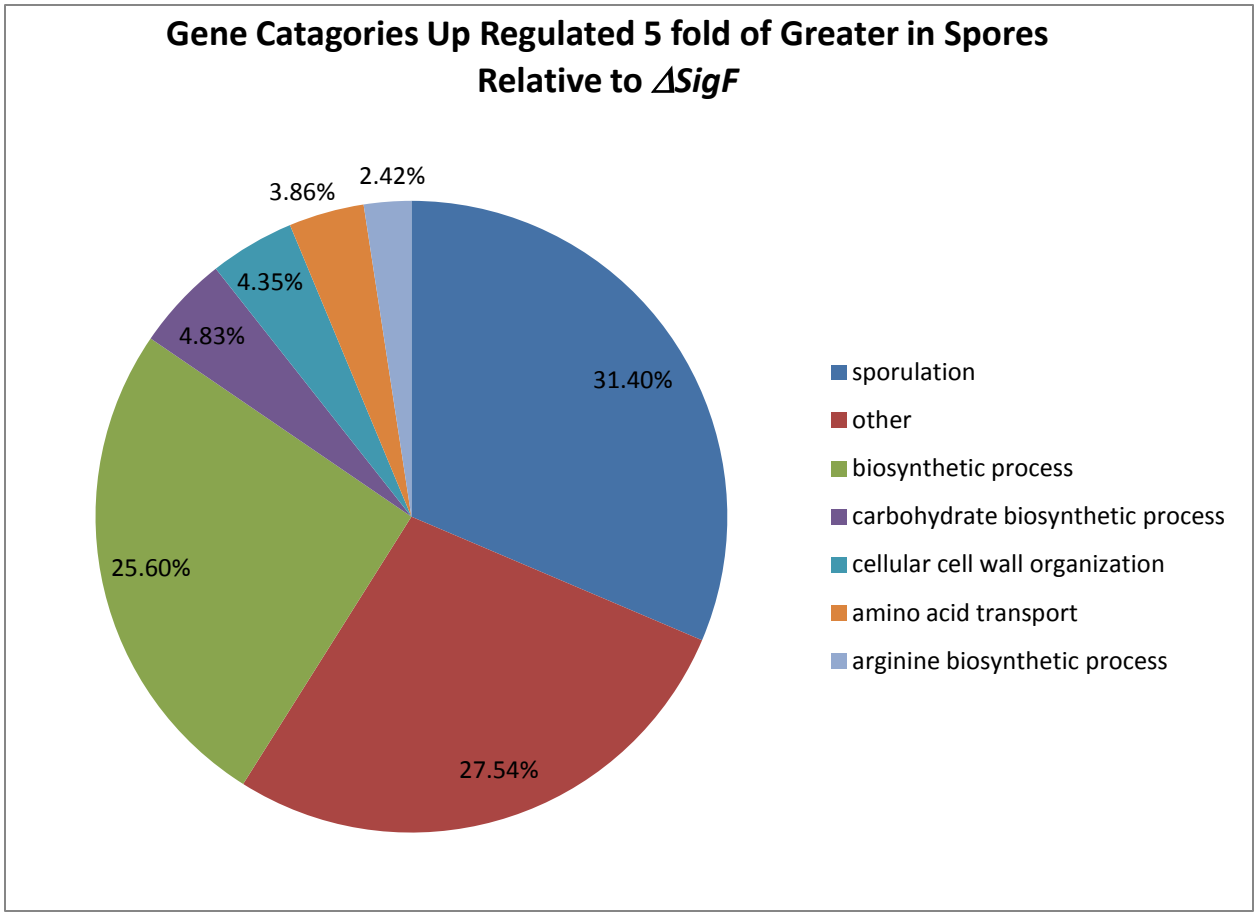


Figure 4-4. Transcriptomic analysis of wild-type and *ΔsigF* spores: mRNA enrichment. Wild-type NCIB3610 and NCIB3610 Δ*sigF* strains were cultured in sporulation medium (Harwood and Cutting, 1991) for 24 hours, whereupon chloroform-based killing assays revealed that the wild-type culture had completely sporulated. In contrast, no chlorofom-resistant colonies formed with the Δ*sigF* strain, consisted with a block in the sporulation pathway. Total RNA was extracted from both cultures, converted to cDNA and subjected to high-throughput sequencing on an Illumina GAXII sequencer. (A) A representative graphical depiction of these RNA-Seq data is shown. Reads per kilobase of genomic sequence were normalized to the total number of mapped cDNA reads ('RPKM') for RNA-Seq datasets. RPKM values for Δ*sigF* cells were then plotted against RPKM values for the wild-type spores. Many genes exhibited significantly increased (>5-fold) expression for wild-type spores (see Supplementary Table S4

for all RPKM values). The blue dashed lines indicate 5-fold changes in gene expression. These genes are again shown in (B), where they are color-coded according to their inclusion in sporulation sigma factor regulons (as cataloged by Eichenberger *et al.,* 2004; Wang *et al.,* 2006). Manual inspection of genes exhibiting RPKM values <10 revealed that many of these lowly expressed genes did not have cDNA reads covering the full length of the gene. Therefore, genes that exhibited RPKM values >10 and that were enriched in spores by at least 5-fold are tabulated in (C). Red bars denote that the genes have been implicated in the *sigG* regulon whereas orange indicates that the genes have been attributed to the *sigF* regulon.



COG Functional Category	Number of Genes
Sporulation	65
Other	62
Biosynthetic Process	53
Carbohydrate Biosynthetic Process	10
Cell Wall Organization	9
Amino Acid Transport	8
Arginine Biosynthetic Process	5

Figure 4-5. Functional gene categories up-regulated fivefold or greater in spores compared to *ΔsigF*. A total of 212 genes were up-regulated by fivefold or greater in spores as compared to the σ^F mutant. Up-regulated genes were grouped based on their category of gene (COG) function. Genes categories that changed fivefold or greater but represented less than 2% of the observed 212 changes were collectively grouped together as "other".

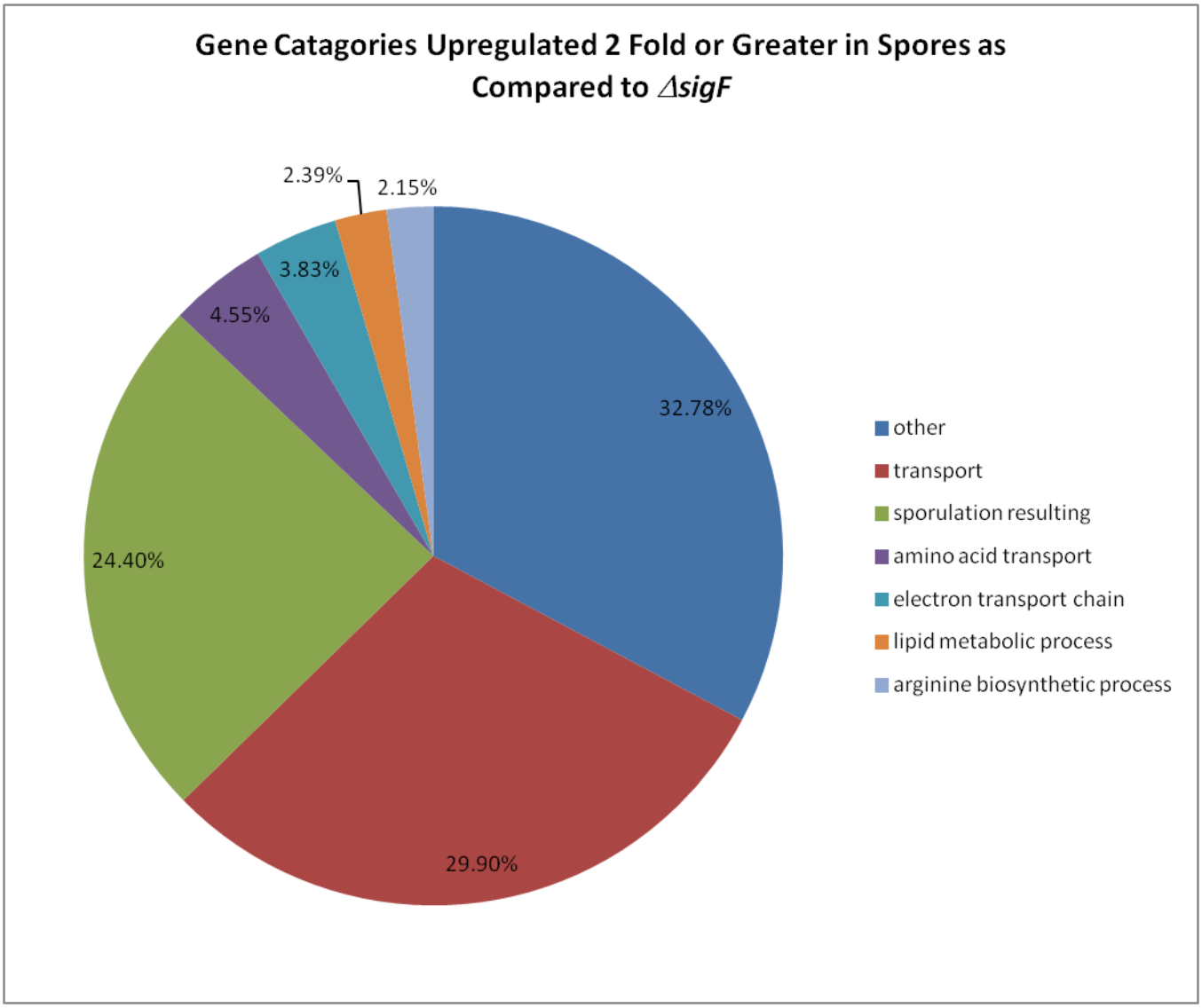


Figure 4-6. Functional gene categories up-regulated twofold or greater in spores compared to *ΔsigF*. Up-regulated genes were grouped based on their category of gene (COG) function. Genes categories that changed twofold or greater but represented less than 2% of the observed changes were collectively grouped together as "other".

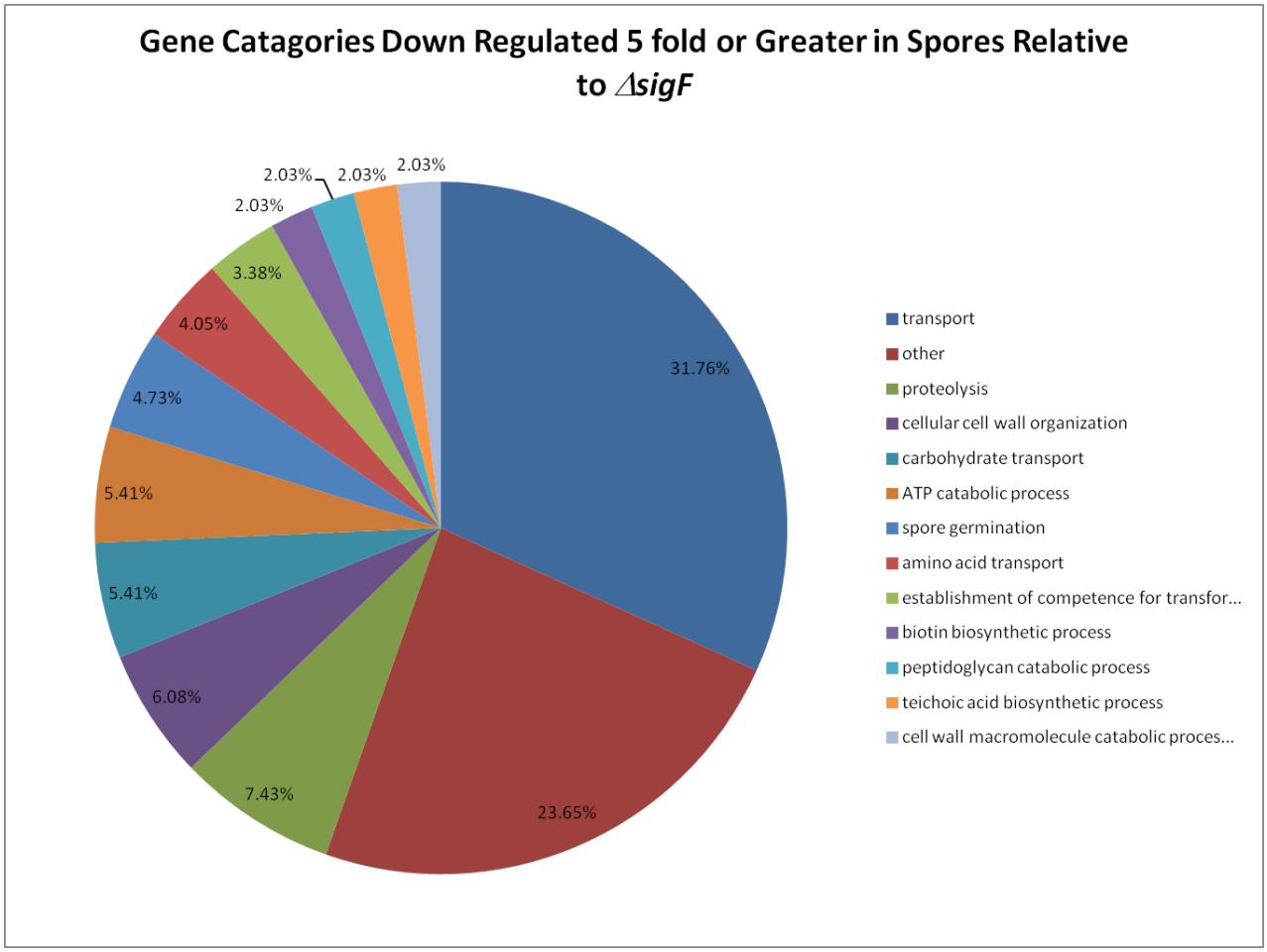


Figure 4-7. Functional gene categories down-regulated fivefold or greater in spores compared to *ΔsigF*. down-regulated genes were grouped based on their category of gene (COG) function. Genes categories that changed fivefold or greater but represented less than 2% of the observed changes were collectively grouped together as "other".

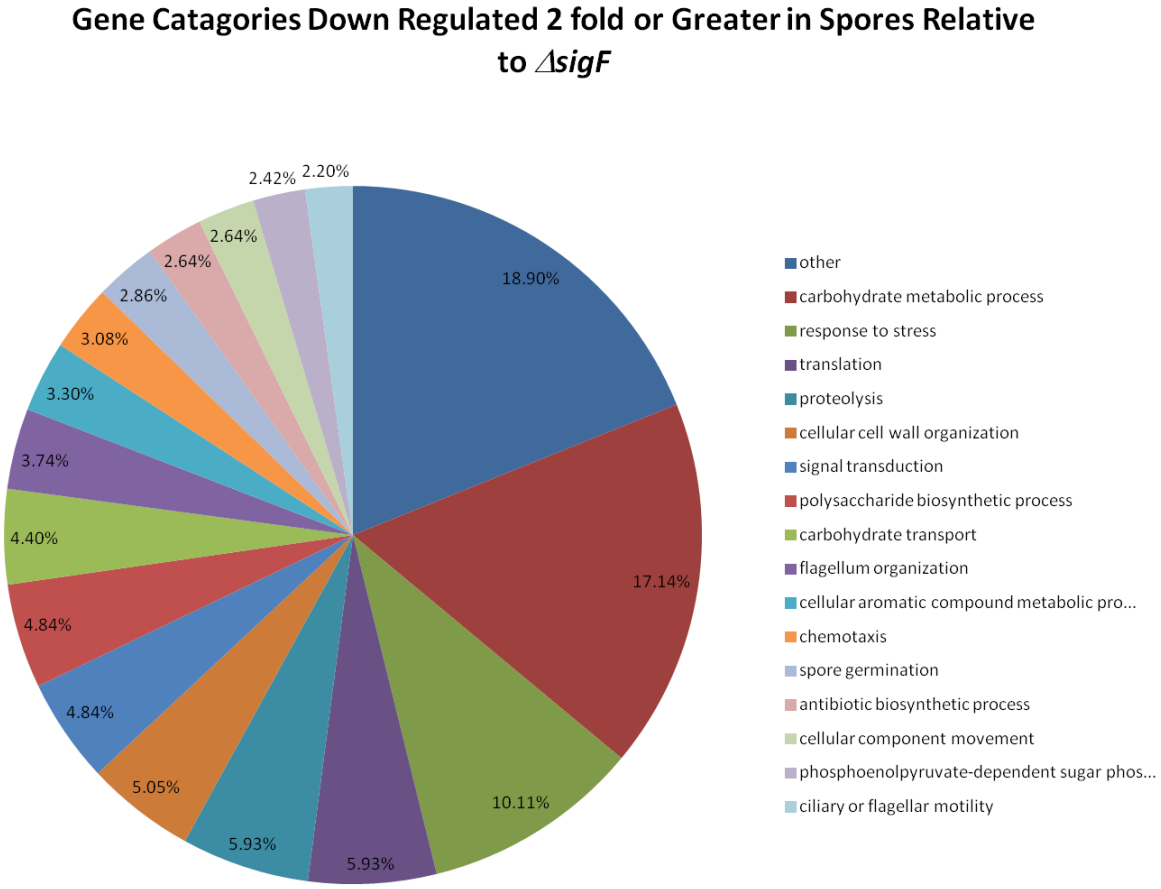


Figure 4-8. Functional gene categories down-regulated twofold or greater in spores compared to *ΔsigF.* Down-regulated genes were grouped based on their category of gene (COG) function. Genes categories that changed twofold or greater but represented less than 2% of the observed changes were collectively grouped together as "other".

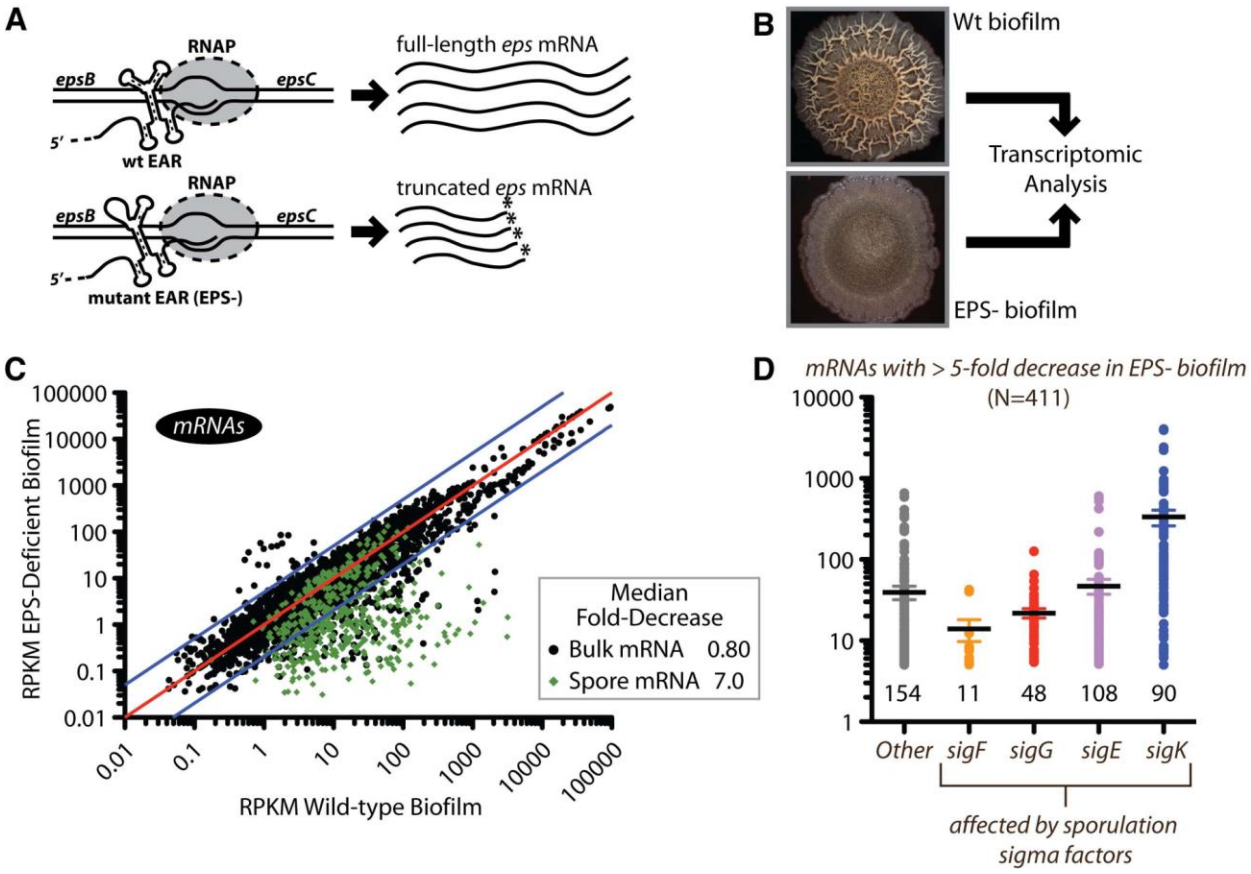


Figure 4-9. Transcriptomic analysis of wild-type and exopolysaccharide (EPS)-deficient *Bacillus subtilis* communities. (A) The EAR element is a cis-acting RNA element that promotes antitermination within the *B. subtilis eps* operon (Irnov and Winkler, 2010). Deletion or mutation of the EAR element results in a biofilm-deficient phenotype, shown in (B) as a marked decrease in colony topology on MSgg agar medium. In this study, we extracted total RNA from wild-type and EAR-minus colonies for transcriptomic analysis using microarrays and high-throughput sequencing (RNA-Seq) approaches. The two different transcriptomic approaches complemented one another (discussed more in Figure S1), although RNA-Seq experimentation permitted more detailed inspection of gene expression patterns, particularly with transcripts emanating from intergenic regions. (C) Representative graphical depictions of RNA-Seq data. Reads per kilobase of genomic sequence were normalized to the total number of mapped cDNA reads ('RPKM') for RNA-Seq datasets. RPKM values for wild-type genes were then plotted against RPKM values for the EPS-deficient mutant strain.

Many genes exhibited significantly decreased expression for the EPS-deficient strain (see Supplementary Materials for more detailed discussion). Genes that have been previously implicated as being partially or fully affected by expression of sporulation sigma factors, *sigF*, *sigG*, *sigE*, and *sigK* are denoted in green; this list was compiled from a combination of primary literature (Eichenberger et al, 2004; Wang et al, 2006) and data presented on-line (www.http://bsubcyc.org/). This allowed for the widest possible definition of 'sporulation regulons', which included many central metabolism genes, biosynthesis genes, and other 'housekeeping' genes. In general, the majority of sporulation genes are significantly decreased while bulk transcripts are less decreased overall for the EPS-deficient strain. Specifically, the median fold-decrease was almost 10-fold greater for sporulation-related genes. This finding was supported in particular by significantly decreased expression of sporulation transcription factors. (D) Of the 411 genes exhibiting significantly (>5-fold) decreased expression, only 154 have not been previously correlated with regulons for sporulation sigma factors.

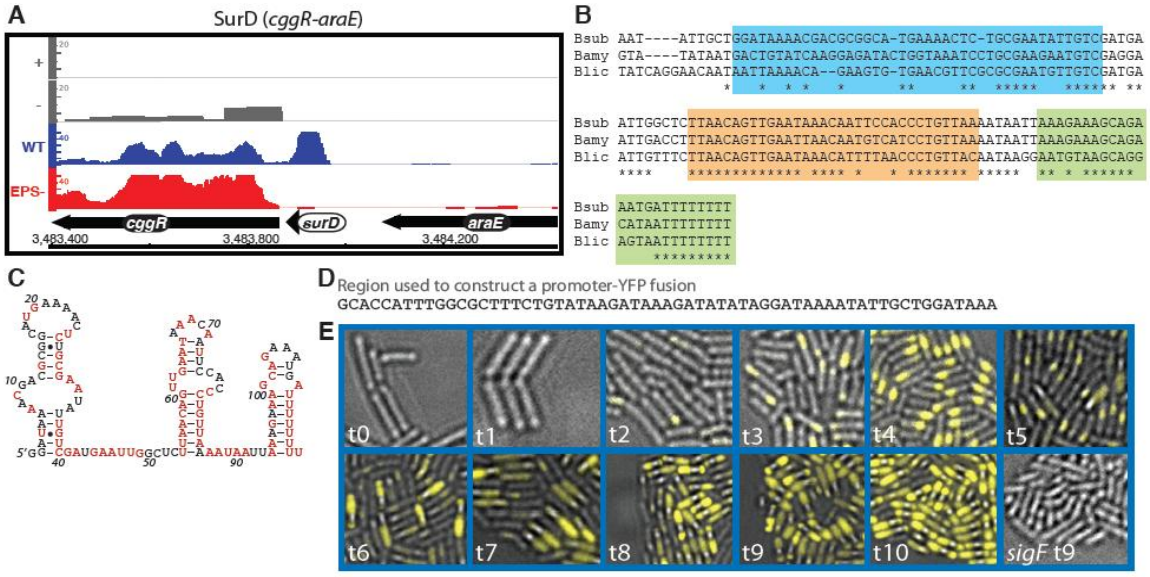


Figure 4-10. Discovery and analysis of the SurD sRNA. (A) The genomic locus is shown at the bottom with the two flanking genes. The distribution of cDNA reads obtained from the RNA-Seq analysis of the wild-type and EPS-deficient biofilm colonies are represented by the blue and red graphs, respectively. Also shown are 5'-transcription start site mapping data from stationary phase cells for both the positive and negative genomic strands, shown in grey (as reported by Irnov *et al.,* 2010). A 5'-transcription start site was not previously identified for the SurD sRNA under stationary phase conditions; however, a start site was identified for the downstream gene, *cggR*. (B) An abbreviated comparative sequence alignment is shown for SurD from a few closely related *Bacillus* species. Colored blocks denote regions of putative secondary structure. Regions of RNA that exhibited secondary structure potential were supported by observation of residues that co-varied while maintaining base-pairing potential. The green block denotes what appears by sequence context to be an intrinsic transcription terminator. The predicted secondary structure of the *B. subtilis* SurD sRNA is shown in (C); red lettering denotes positions that are >95 % conserved in the sequence alignment. The genomic sequence located immediately upstream of the SurD gene, which presumably contains the promoter sequence, is shown in (D). This sequence was subcloned into an *amyE* integration vector (pDG1662) to control transcription of a promoter-less copy of yellow fluorescent protein (YFP) that was dependent on a *rpsD* ribosome binding

site. (E) Expression of the SurD promoter-YFP fusion appeared at early time intervals during sporulation in the forespore, but then increased at later time intervals specifically within the mother cell compartment. No YFP expression was observed for a Δ*sigF* control strain at either early or late time points.

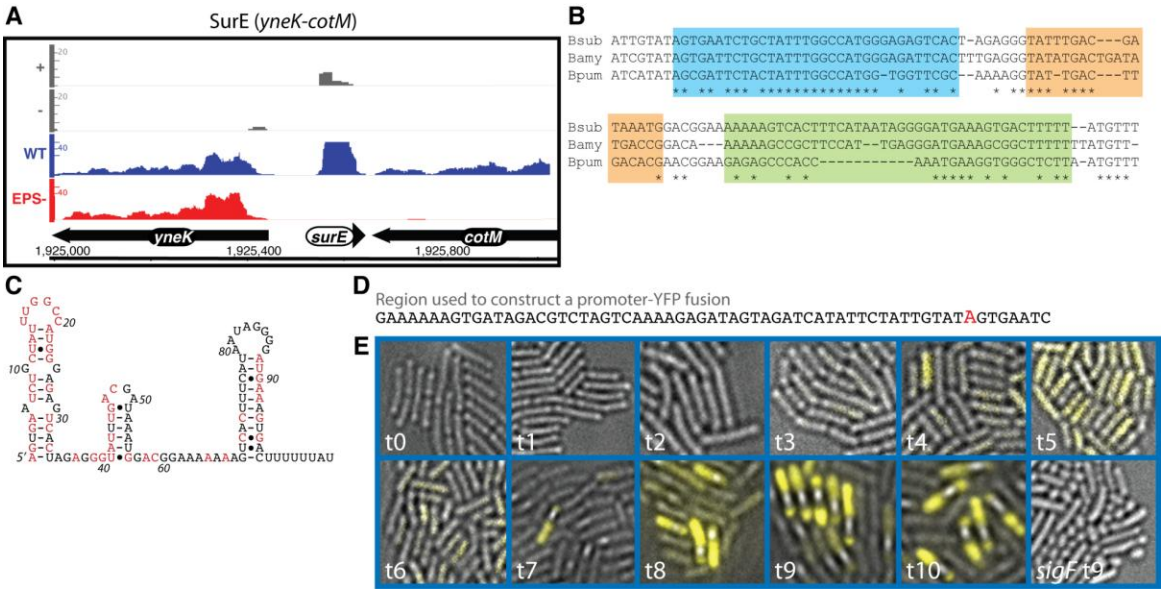


Figure 4-11. Analysis of the SurE sRNA. (A) The genomic locus for SurE (previously discovered by our prior mapping of global transcription start sites and temporarily annotated as ncr992; Irnov *et al*., 2010) is shown at the bottom with the two flanking genes. The distribution of cDNA reads obtained from the RNA-Seq analysis of the wild-type and EPS-deficient biofilm colonies are represented by blue and red graphs, respectively. Also shown are 5'-transcription start site mapping data from stationary phase cells for both the positive and negative genomic strands, as shown in grey (Irnov *et al.,* 2010). (B) An abbreviated comparative sequence alignment is shown for SurE from a few closely related *Bacillus* species. Colored blocks denote regions of putative secondary structure. Regions of RNA that exhibited secondary structure potential were supported by observation of residues that co-varied while maintaining base-pairing potential. The green block denotes what appeared by sequence context to be an intrinsic transcription terminator. The secondary structure of the *B. subtilis* SurE sRNA is shown in (C); red lettering denotes positions that are >95 % conserved in the sequence alignment. The genomic sequence located immediately upstream of the SurE gene, which presumably contains the promoter sequence, is shown in (D). This sequence was subcloned into an *amyE* integration vector (pDG1662) to control transcription of a

promoter-less copy of yellow fluorescent protein (YFP) that was dependent on an *rpsD* ribosome binding site. Expression of the SurE promoter-YFP fusion appeared specifically in the mother cell at later time intervals during sporulation (E), but was absent in a Δ*sigF* control strain.

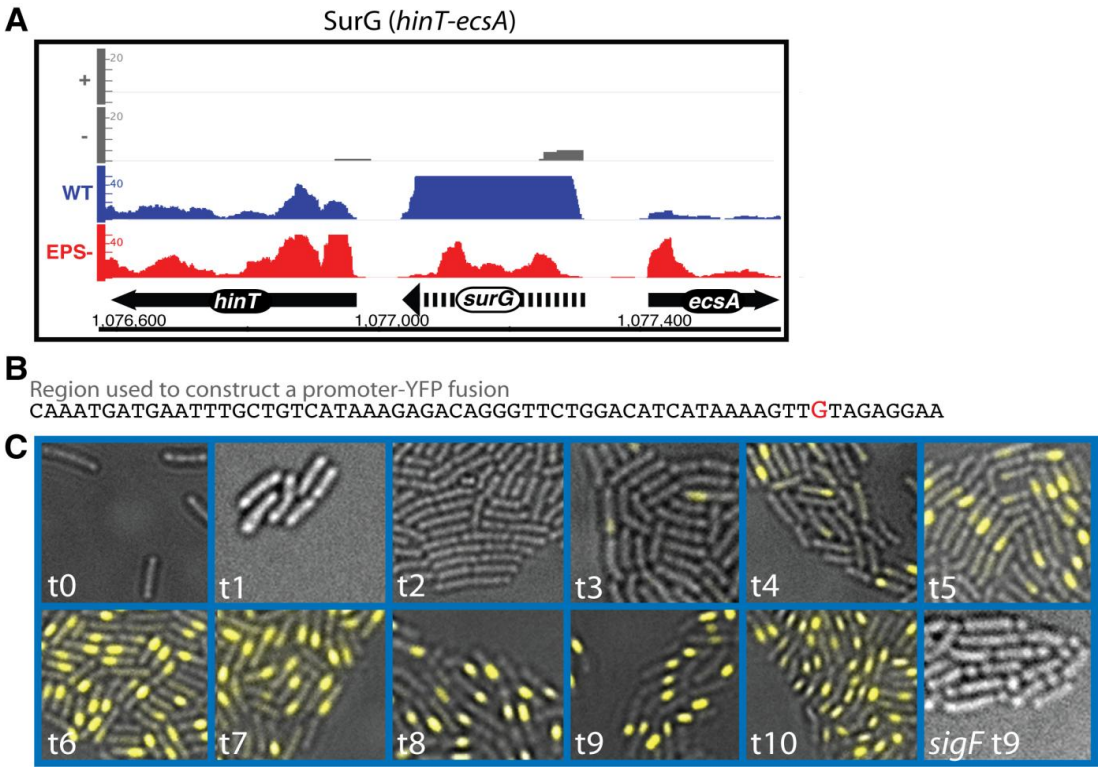


Figure 4-12. Analysis of the SurG sRNA. (A) The genomic locus for SurG (previously discovered by our prior mapping of global transcription start sites and temporarily annotated as ncr1670; Irnov *et al*., 2010) is shown at the bottom with its two flanking genes. The distribution of cDNA reads obtained from the RNA-Seq analysis of the wild-type and EPS-deficient biofilm colonies are represented by blue and red graphs, respectively. Also shown are 5'-transcription start site mapping data from stationary phase cells for both the positive and negative genomic strands, as shown in grey (as reported by Irnov *et al.,* 2010). The genomic sequence located immediately upstream of the SurG sRNA gene is shown in (B). This sequence was subcloned into an *amyE* integration vector (pDG1662) to control transcription of a promoter-less copy of yellow fluorescent protein (YFP) that was dependent on an *rpsD* ribosome binding site. (C) Expression of the SurG promoter-YFP fusion appeared specifically in the forespore compartment but was absent from a Δ*sigF* control strain.

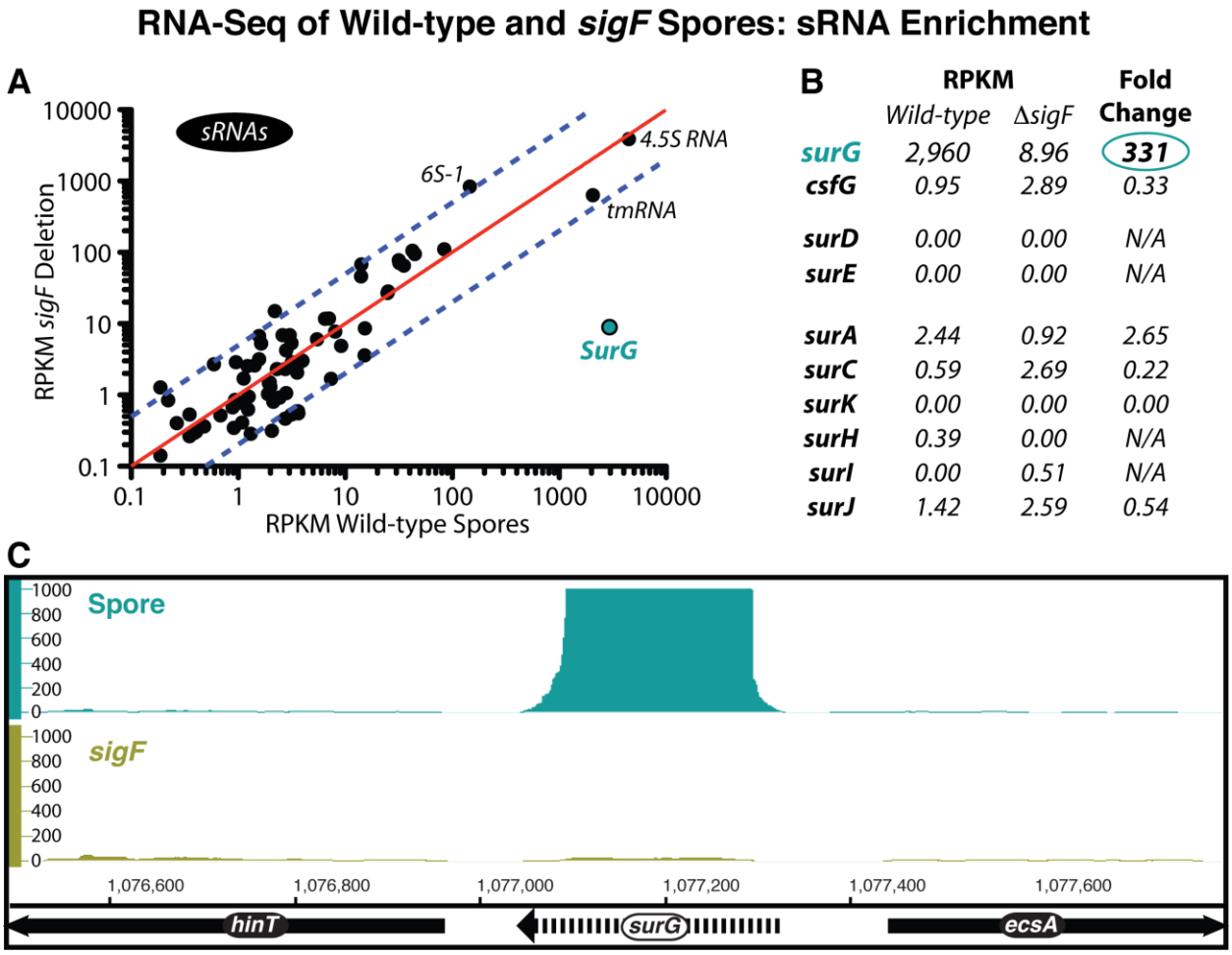


Figure 4-13. Transcriptomic analysis of wild-type and Δ*sigF* spores: sRNA enrichment. Total RNA was extracted from NCIB3610 and NCIB3610 Δ*sigF* strains after 24 hours in sporulation medium (Harwood and Cutting, 1991), converted to cDNA and subjected to high-throughput sequencing on an Illumina GAXII sequencer. We compiled a comprehensive sRNA catalog that included the genomic coordinates for all putative sRNAs that have been discovered (Rasmussen *et al.,* 2009; Irnov *et al.,* 2010; and references within). We calculated the reads per kilobase of genomic sequence as normalized to the total number of mapped cDNA reads ('RPKM') for each of these respective coordinate sets. The RPKM values for these genomic coordinates were then plotted for expression data of Δ*sigF* cells and wild-type spores, shown in (A). The blue dashed lines denote 5-fold changes in gene expression. Many sRNAs that have

been identified previously are not shown in this plot as they exhibited an RPKM value of 0.0. Also, manual examination of the sRNAs that exhibited RPKM values <10 revealed that in many instances the expression data did not cover the full length of the putative sRNA gene. Therefore, we consider these sRNAs to be categorized as exhibiting particularly low expression levels, as denoted by the grey shaded box. Notably, only one sRNA, SurG, was both highly expressed and significantly enriched in wild-type spores. The other sporulation-specific sRNA molecules characterized in this study were essentially not expressed in either wild-type spores or the Δ*sigF* cells, as tabulated in (B). The strikingly increased RPKM value for SurG in spores resulted from increased expression over the entire length of the *surG* gene and was not the result of any transcriptional overlap from neighboring genes, as shown in (C).

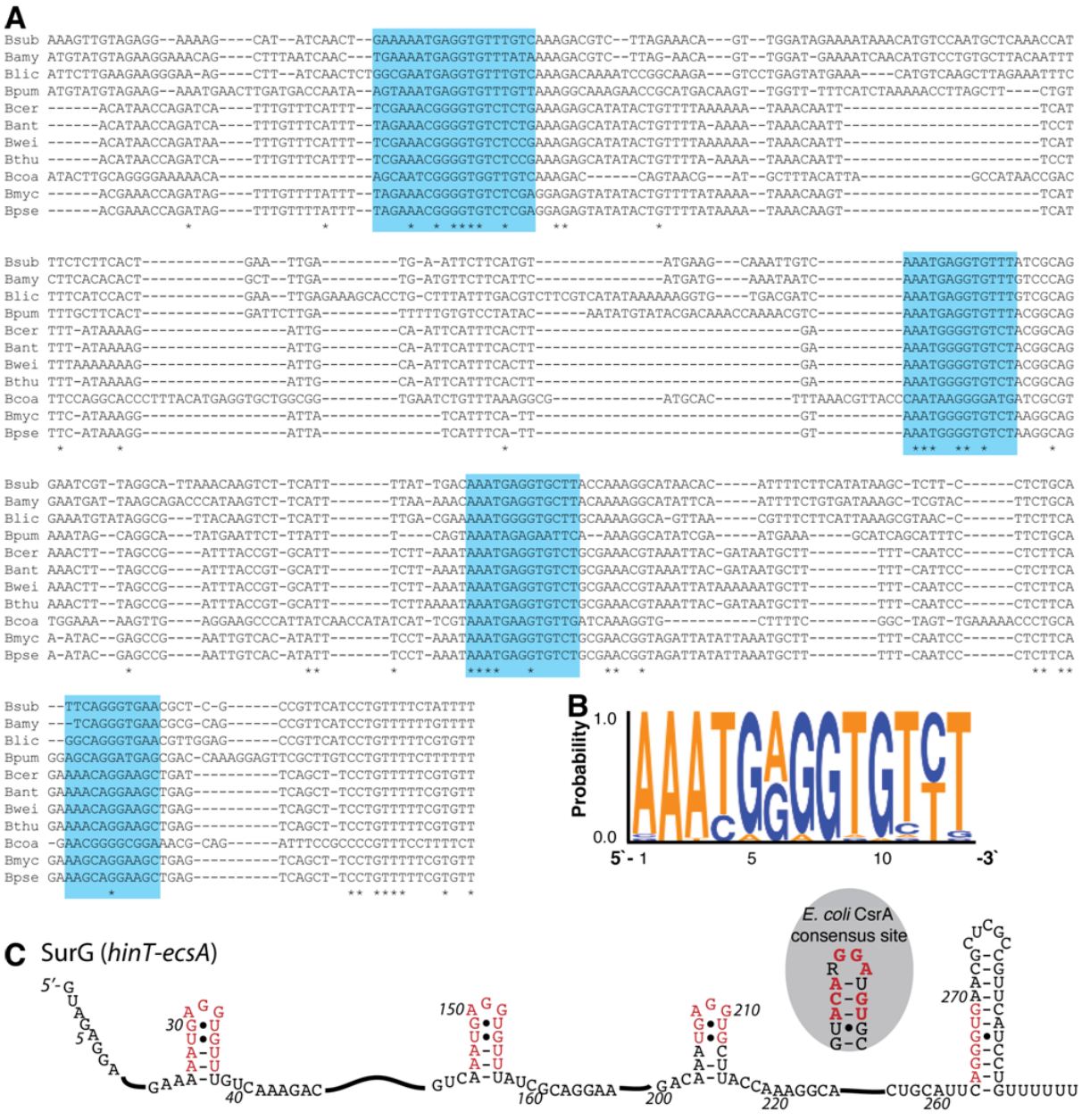


Figure 4-14. Comparative sequence alignment of SurG. (A) Representative comparative sequence alignment of SurG from several *Bacillus* species. A repeated sequence motif is shaded in blue (see Discussion). A weblogo depiction of the repeated sequence motif is shown in (B). The consensus secondary structure of the *B. subtilis* SurG sRNA is shown in (C), which was derived from an expanded version

of the sequence alignment in (A). Red lettering denotes positions that are >95 % conserved in the sequence alignment. This analysis revealed the presence of a repeated stem-loop element, which includes a purine-rich small terminal loop. We note the similarity, but not identity, between this stem-loop motif and the consensus-binding site for *E. coli* CsrA, shown in the inset (Babitzke and Romeo, 2007; Sonnleitner and Haas, 2011). We speculate from these aggregate data that SurG may correspond to a protein-sequestering sRNA, akin to sRNAs that sequester CsrA.

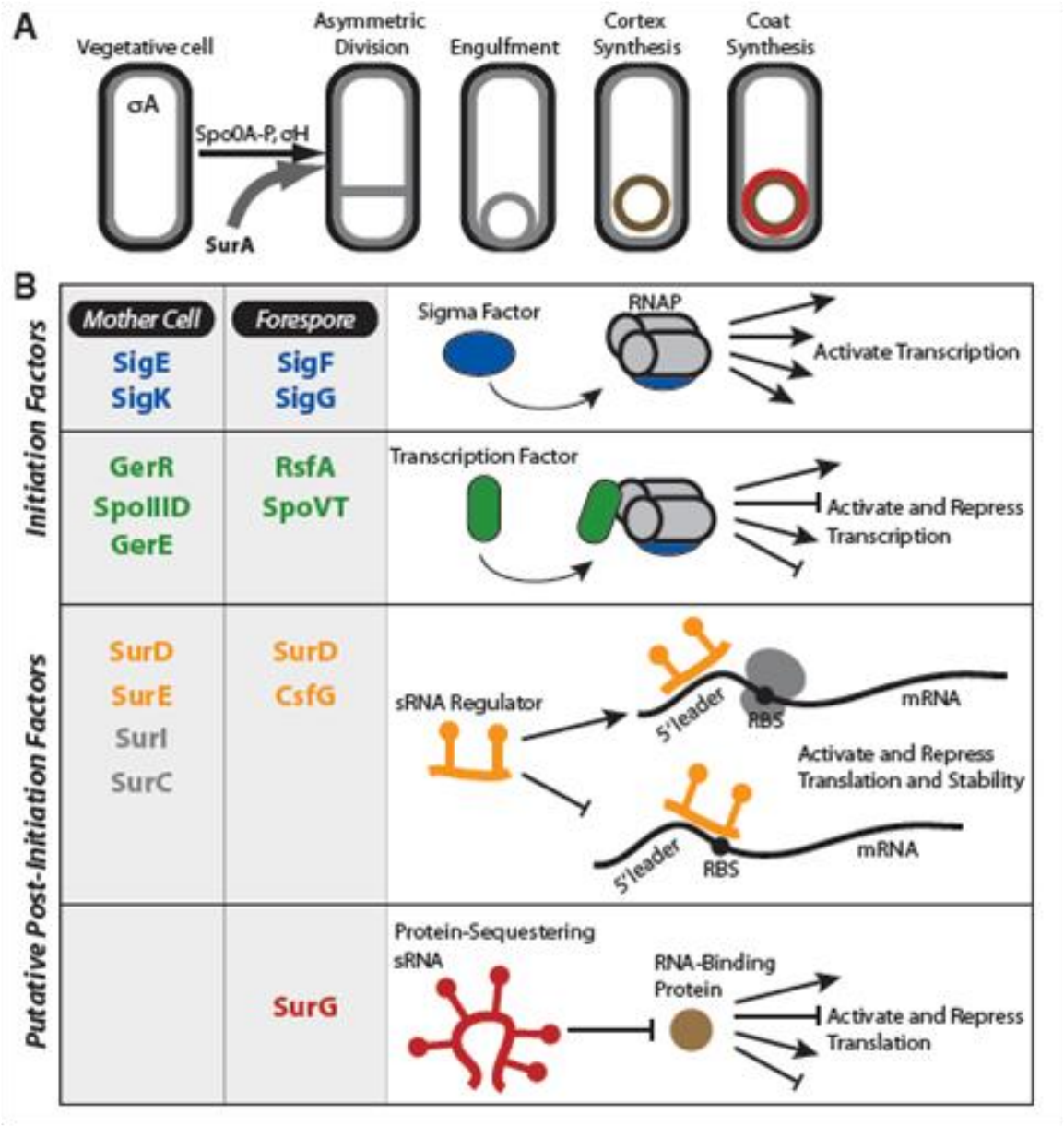


Figure 4-15. Summary model for *Bacillus subtilis* sporulation regulatory factors. Two successive sigma factors, SigF and SigG, control gene expression within the developing forespore (Wang *et al.,* 2006; Kroos, 2007). Similarly, two successive sigma factors, SigE and SigK are specifically produced and activated within the mother cell compartment (Eichenberger *et al.,* 2004; Kroos, 2007). For many genes, activation of their expression by these sigma factors (shown as blue

arrows) is also influenced by a variety of transcription regulatory factors (shown as green arrows). Many genes are specifically activated and repressed by these transcription regulatory factors. We observed significantly decreased expression of these various sporulation regulatory factors and their target genes specifically for the EPS-deficient strain; therefore, we reasoned that sRNAs specifically expressed during sporulation should also exhibit decreased expression in our data set. Our test of this hypothesis led to discovery of compartment-specific expression patterns for a small subset of sRNA regulatory molecules (shown in orange). A few putative sRNAs that were previously implicated in sporulation were not appreciably detected in this study, shown in grey. The SurG sRNA (red letters) was specifically enriched in spores and exhibited a general architecture resembling that of protein-sequestering sRNAs. We determined a speculative temporal expression pattern for these sRNAs based on appearance of engulfment relative to activity of their respective promoter regions. However, further experimentation will be required to determine the basis of transcriptional control of the sRNA regulatory molecules. As discussed in the text, we hypothesize that these RNA-based regulators are responsible for a second, post-initiation layer of genetic regulation during sporulation, in addition to the previously established catalog of transcription factors.

sRNA	spore RPKM	ΔsigF-RPKM	Spore/ ΔsigF ratio
SurG/ncr1670/hinT-ecsA	2961.5	8.965461	330.3228914
ncr2339	15.2257	3.608352	4.219563391
ssrA/tmRNA	2063.84	635.7212	3.246446713
bsrC	15.3109	8.595417	1.781285306
scRNA/4.5S RNA	4478.96	3872.7	1.156546854
ncr2	24.892	27.11961	0.917858701
ncr1155/ncr58/bsrH	25.2667	28.2198	0.895355034
ncr560/ncr18	84.4726	110.1225	0.767078118
ncr1058/ncr46/bsrG	35.4666	65.39894	0.54231093
6S-2	44.8386	94.89171	0.472523891
ncr1932/as-bsrG	32.0322	71.11016	0.450458134
rnpB	32.0322	71.11016	0.450458134
ncr1857/bsrE	31.6859	77.55031	0.408584569
ncr1019/ncr39/as-bsrE	42.6319	105.5507	0.403899548
ncr2184/ncr60	14.0722	46.1869	0.304679682
ncr82	14.1348	68.01303	0.20782503
6S-1	146.674	840.7941	0.17444699

Table 4-1. Expression levels of sRNA regulators in spores. Of the over 120 putative small RNAs identified to date, only 17 of these exhibited exceeded our threshold RPKM value of 10. Only 3 of these 17 exhibited a 2 fold or greater enrichment when compared to the *ΔsigF*. With the forespore-specific sRNA showing an enrichment of over 330 fold versus the σ^F mutant.

sRNA	ΔsigF-RPKM	spore RPKM	ΔsigF/Spore ratio
bsrI	15.016	2.2	6.825472727
6S-1	840.794	146.674	5.732400425
ncr82	68.013	14.1348	4.811739953
ncr2184/ncr60	46.1869	14.0722	3.2821355
ncr1019/ncr39/as-bsrE	105.551	42.63188	2.475863133
ncr1857/bsrE	77.5503	31.6859	2.447473731
ncr1932/as-bsrG	71.1102	32.0322	2.219962132
rnpB	71.1102	32.0322	2.219962132
6S-2	94.8917	44.8386	2.116295112
ncr1058/ncr46/bsrG	65.3989	35.4666	1.843960621
ncr214	11.7086	6.474468	1.808426577
ncr1733/ncr26	11.8191	7.038282	1.679252124
ncr560/ncr18	110.123	84.4726	1.303648191
ncr1155/ncr58/bsrH	28.2198	25.2667	1.116875386
ncr2	27.1196	24.892	1.089492314
ssrA/tmRNA	635.721	635.7212	1
scRNA/4.5S RNA	3872.7	4478.96	0.864642878

Table 4-2. Expression levels of sRNA regulators in *ΔsigF* deletion strain. Of the over 120 putative small RNAs identified to date, only 17 of these exhibited exceeded our threshold RPKM value of 10 in the *ΔsigF* mutant strain; nine of which exhibited a 2 fold or greater enrichment when compared to the spores. The sRNA BsrI exhibiting the greatest enrichment in the σ^F mutant, corresponding to a nearly 7 fold increase versus spores.

Accession Number for Putative Target Gene	Function	Target Site Info[a]	
		mRNA Start End[b]	Energy [kJ/mol]; z-Score
CsfG			
1. BSU40860	Putative lyase	-21 to 2	-29.16; -4.87
2. BSU10300	Serine alkaline protease (subtilisin E)	-15 to 9	-27.30; -4.41
3. BSU25190	Cytochrome c550	-27 to 2	-26.26; -4.15
4. BSU02570	Putative ABC efflux transporter	-28 to -5	-25.83; -4.04
5. BSU23750	Putative lyase	-52 to -25	-25.71; -4.01
6. BSU00900	2 C methyl D erythritol 4 phosphate cytidylyltransferase	-30 to -4	-22.41; -3.94
7. BSU27100	Putative efflux transporter	-29 to -2	-25.09; -3.86
8. BSU19640	Hypothetical protein	-30 to -5	-24.81; -3.79
9. BSU01530	N acetylmuramoyl L alanine amidase	-17 to 6	-24.32; -3.67
10. BSU34010	Putative amino acid permease	-17 to 6	-23.74; -3.52
11. BSU24510	Hypothetical protein	-21 to 1	-23.73; -3.52
12. BSU28970	Putative integral inner membrane protein	-19 to 7	-23.04; -3.35
13. BSU01720	Hypothetical protein	-31 to -3	-22.96; -3.33
14. BSU22490	Dihydrodipicolinate reductase	-27 to -5	-22.81; -3.29
15. BSU18440	Glutamate synthase subunit	-20 to -2	-22.78; -3.28
16. BSU24140	Citrate synthase 3	-12 to 13	-22.42; -3.2

[a]Many good candidates were also observed for potential association with coding regions; however, due to the large number of candidate targets, only the 16 top scoring hits with targeted regions upstream of the ribosome binding site were considered herein.
[b]Coordinates are shown as reported by RNApredator for the accession genome sequence NC_000964.

Table 4-3. Target prediction for CsfG using RNA Predeator (Eggenhofer *et al.,* 2011). The top 15 potential mRNA targets for CsfG base pairing near the ribosome binding site.

Accession Number for Putative Target Gene	Function	Target Site Info[a]	
		mRNA Start End[b]	Energy [kJ/mol]; z-Score
SurD			
1. BSU04780	Putative RNA helicase	1170-1197	-19.70; -3.84
2. BSU36970	Pro sigma(E) endopeptidase	476-502	-16.96; -3.07
3. BSU18320	Plipastatin synthetase	7332-7348	-16.75; -3.01
4. BSU13400	ATP dependent DNA ligase	1487-1513	-16.59; -2.97
5. BSU07840	Bifunctional 2'-3' cyclic nucleotide 2' phosphodiesterase	3706-3732	-16.35; -2.9
6. BSU34790	Thioredoxin reductase	80-99	-16.23; -2.87
7. BSU14900	Cytochrome caa3 oxidase subunit	81-107	-15.80; -2.75
8. BSU18310	Plipastatin synthetase subunit	7130-7153	-15.60; -2.69
9. BSU17180	Polyketide synthase (*pks*)	10204-10226	-15.02; -2.53
10. BSU18340	Plipastatin synthetase subunit	4022-4045	-15.02; -2.53
11. BSU16590	Hypothetical protein	-61 to -47	-14.98; -2.51
12. BSU22800	Morphogenetic stage IV sporulation protein	635-661	-14.97; -2.51
13. BSU23020	ATP dependent DNA helicase	193-206	-14.85; -2.48
14. BSU13980	Cell wall protein	812-824	-14.78; -2.46
15. BSU23440	Stage V sporulation protein	191-207	-14.71; -2.44

[a]Top scoring hits in this analysis only corresponded to coding regions. The 15 top scoring hits are shown in the table. A focused study on upstream regions was not conducted herein.
[b]Coordinates are shown as reported by RNApredator for the accession genome sequence NC_000964.

Table 4-4. Target prediction for SurD using RNA Predeator (Eggenhofer *et al.*, 2011). The top 15 potential targets for CsfG base pairing near the ribosome binding site.

Accession Number for Putative Target Gene	Function	Target Site Info[a]	
		mRNA Start End[b]	Energy [kJ/mol]; z-Score
SurE			
1. BSU11549	Hypothetical protein	107-131	-22.1; -3.63
2. BSU31100	Potassium transporter	1012-1036	-18.21; -2.67
3. BSU09860	Potassium efflux	-103 to -80	-18.12; -2.65
4. BSU39230	Cell wall associated protein precursor	250-267	-18.01; -2.63
5. BSU01140	Putative proline iminopeptidase	602-620	-18.00; -2.62
6. BSU19190	Histidine kinase *desK* – sensor of membrane fluidity	401-422	-17.81; -2.58
7. BSU05760	Hypothetical protein	2-26	-17.81; -2.58
8. BSU03620	LysR-family regulator	770-794	-17.73; -2.56
9. BSU30490	Putative RNA methylase	537-551	-17.33; -2.46
10. BSU36380	Response regulator aspartate phosphatase	442-461	-17.3; -2.45
11. BSU38550	Branched chain biosynthesis	412-425	-17.05; -2.39
12. BSU02240	Extracellular metalloprotease	343-368	-16.70; -2.3
13. BSU13050	Hypothetical protein	15-29	-16.70; -2.3
14. BSU29140	Citrate synthase	251-276	-16.15; -2.17
15. BSU38770	Citrate/malate/H+ symporter	292-307	-15.69; -2.06

[a]Top scoring hits in this analysis only corresponded to coding regions. The 15 top scoring hits are shown in the table. A focused study on upstream regions was not conducted herein.
[b]Coordinates are shown as reported by RNApredator for the accession genome sequence NC_000964.

Table 4-5. Target prediction for SurE using RNA Predeator (Eggenhofer *et al.*, 2011). The top 15 potential targets for CsfG base pairing near the ribosome binding site.

CHAPTER 5

Material and Methods

Co-immunoprecipitation of Hfq

Five ml of glucose minimal media (Irnov and Winkler, 2010) with 50 µg/ml tryptophan was inoculated by a single colony of either wild-type 168 or MD145 from a freshly streaked plate and incubated without shaking overnight at 37˚C. 50 ml minimal medium was inoculated with 0.5 ml of overnight culture (OD_{600} ~1.5) and incubated shaking at 37˚C overnight (approximately 20 hours). These cells were induced then induced with isopropyl □-D-2-thiogalactopyranoside (IPTG) to a final concentration of 1 mM and incubated shaking for five hours at 37˚C. The cultures were transferred to 50 mL conical tubes and centrifuged at 2,900 x g for seven minutes; the resulting cell pellet was flash frozen in liquid nitrogen and stored at -80˚C. Lysates were prepared using a method reported elsewhere (Sittka *et al*., 2008).

Briefly, cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 150 mM KCl, 1 mM $MgCl_2$, 1 mM DTT and 10 mg/ml lyzozyme) at placed on ice for ten minutes. The cell suspension was then flash frozen by liquid nitrogen and thawed for two minutes in a 55˚C heat block. Approximately 400 µl of sterile

glass beads were added and the suspension was subjected to five rounds of vortexing for 30 seconds followed by 30 seconds on ice (5x). The lysate was then centrifuged at 20,000 x g for 30 minutes. The supernatant was then transferred to a fresh tube and incubated with 35 µl of anti-FLAG M2 monoclonal antibody (Sigma) for 30 minutes while rocking at 4˚C. Subsequently, 75 µl of protein A sepharose (Sigma), pre-washed in 1 ml of lysis buffer lacking lyzozyme, was added to each tube and rocked for an additional 30 minutes at 4˚C. Each sample was then centrifuged for two minutes at 500 x g and washed 5x with 500 µl lysis buffer. The mixture was then resuspended in 500 µl wash buffer with 300 mM sodium acetate, mixed with 500 µl of phenol:chloroform:isoamyl alcohol (25:24:1 pH 7.8) and centrifuged at 20,400 x g for ten minutes. The aqueous phase was collected from each separation and placed in a new tube with 1.5 ml of 100% isopropyl alcohol and placed at -20˚C overnight. To the organic phase (containing the beads), 1 ml of cold acetone was added for protein precipitation. The precipitated RNA was next centrifuged at 20,400 x g for 30 minutes, the supernatant was discarded and the pellet was washed with 1 ml of 70% ethanol followed by centrifugation at 20,400 x g for 10 minutes. The supernatant was carefully removed and the pellet was air dried for ten minutes at room temperature followed by re-suspension in 20 µl of sterile distilled water. Roughly

1 µg of RNA for the coIP and mock control samples was used for generation of Illumina-compatible cDNAs.

Preparation of samples for RNA-seq

The samples were prepared for sequencing by following the manufacturer's instructions of the Illumina mRNA sequencing kit, except that the initial polyA-enrichment and final gel purification steps were omitted. Instead of the latter, the cDNA was purified using Qiagen PCR cleanup kit after adaptor ligation. The processed samples were sequenced using an Illumina Genome Analyzer (GAIIx) housed within the DNA Sequencing Core Facility at the UT Southwestern Medical Center. The resulting cDNA sequences were mapped onto the *B. subtilis* genome (NC_000964.3) using 'Burrows-Wheeler Aligner' (BWA) software (Li and Durbin, 2009). Subsequent data processing were done using SAMtools (Li *et al*., 2009) and custom-made Python scripts. Mapped reads were visualized using Integrated Genome Browser (IGB).

RNA-Seq analysis

For the biofilm samples, we obtained 28,845,542 and 28,433,082 cDNA reads from the wild-type and M3 biofilm samples, respectively. These cDNA reads were then mapped onto the latest *B. subtilis* genome (NC_000964.3) using

"Burrows-Wheeler Aligner" software (version 0.5.8a; Li and Durbin, 2009; freely available at http://bio-bwa.sourceforge.net/) with the following parameters: -t 2 -l 20 -k 2 -n 4. These parameters will allow up to 4 mismatches with a maximum of 2 mismatches allowed for the first 20 nucleotides (seed sequence). For the spore samples, we obtained 20,163,832 and 26,782,380 cDNA reads from the spore and *DsigF* samples, respectively. These cDNA reads were then mapped onto the latest *B. subtilis* genome (NC_000964.3) using "Bowtie2" software (version 2.0.0-beta5; Langmead *et al.*, 2009; freely available at http://bowtie-bio.sourceforge.net/bowtie2/) with the --very-sensitive preset.

The output from BWA and Bowtie2 were then converted into SAM format using "SAMtools" (Li *et al.*, 2009; freely available at http://samtools.sourceforge.net/) and further analyzed with custom-made Python scripts. For cDNA reads with multiple hits in the genome, only the first mapped coordinate will be used. In total, 24,210,581 and 24,408,252 cDNA reads from the biofilm samples were successfully mapped onto the genome for the wild-type and M3 samples, respectively. In the case of the spore samples, 18,765,300 and 24,744,182 cDNA reads were successfully mapped onto the genome for the spore and *DsigF* datasets, respectively. However, despite rRNA depletion steps, ~80-90% of cDNA reads still originated from rRNA loci.

Quantification of gene expression

The expression level of each gene was quantified in reads per kilobase per million mapped reads or 'RPKM' (Mortazavi *et al.*, 2008).

R = (10^9 x C) / N x L

where C = number of mappable reads to certain region of interest

N = total number of mappable reads in the experiments

L = length of the region of interest (nt)

For this analysis, the list of *B. subtilis* genes was obtained from NCBI (NC_000964_3.gff). cDNA reads mapped to each gene were extracted using SAMtools. The subsequent RPKM analyses were automatically calculated using a python script. Finally, the fold-change for each gene was calculated as: fold-change = RPKM (M3)/RPKM (wild-type) or fold-change=RPKM (spore)/RPKM (D*sigF*) for the biofilm and spore samples, respectively.

Visualization of sequencing analyses

Integrated Genome Browser (IGB; freely available at http://www.bioviz.org/igb/) was used to visualize all of the Illumina sequencing

analysis. To generate the input file for IGB (with ".gr" extension), we converted the BWA-generated SAM file into the appropriate format using a custom-made python script. Briefly, the IGB format consists of two columns per line. The first column denotes the genomic coordinate. The second column represents the number of cDNA hits for that particular nucleotide position.

SDS-PAGE and western blot analysis

Aliquots of bacterial culture (5 ml) were harvested at various growth stages by centrifugation at 2,900 for five minutes followed by storage at -80˚C. Total cell lysates were prepared by thawing the cells on ice for ten minutes followed by resuspension in 500 µl of lysis buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM MgCl, 1 mM DTT, and Sigma Protease Cocktail (P 8849) diluted 1:100). The samples were boiled for 20 minutes followed by centrifugation for ten minutes at 20,400 x g with subsequent collection of the supernatant. Total protein concentration was determined by the Bradford method and 2 mg of each sample was mixed with 5 µl of 6x SDS-loading buffer in a total volume 25 µl. The samples were boiled an additional five minutes and resolved on a 10-20% SDS-PAGE gel followed by transfer to a nitrocellulose membrane. Each blot was probed with a 1:1000 dilution of anti-FLAG M2 monoclonal antibody (Sigma 1804) overnight at 4˚C in 5% (w/v) non-fat milk/TBST followed by a 1:5,000

dilution of horseradish peroxidase-conjugated sheep-anti-mouse IgG (GE Healthcare NA931V) for 30 minutes at room temperature.

The blots were subsequently developed with ECL development reagent for one minute (GE Healthcare RPN 2106V) and exposed to autoradiography film. After exposure each blot was stripped in mild stripping buffer (200 mM glycine, 0.1% SDS, 1.0% TWEEN 20 at a final pH of 2.2) at room temperature for ten minutes with gentle agitation, followed by two ten minute washes with PBS and two ten minute washes with TBST. Each blot was then blocked for 30 minutes at room temperature with 5% non-fat milk/TBST and incubated overnight 1:2000 with rabbit serum raised against *B. subtilis* sigma A (a generous gift from Dr. Masaya Fujita), followed by a 1:5,000 dilution of horseradish peroxidase-conjugated donkey anti-rabbit IgG (GE Healthcare NA934V) and developed as described above.

RNA Purification

Bacterial cells were centrifuged at 2,900 x g for five minutes and pellets were stored at -80˚C. The cells were thawed for ten minutes on ice followed by resuspension in 750 µl LETS buffer (0.1 M LiCl, 10 mM EDTA, 10 mM Tris-HCl pH 7.4 and 1% SDS) and disrupted by continuous vortexing with 400 µl of sterile glass beads for four minutes followed by incubation at 55˚C for five minutes. This

suspension was centrifuged for ten minutes at 20,400 x g and the supernatant was collected and mixed with 1 mL of TRI reagent (Ambion AM9738) and incubated at room temperature for 5 minutes. Subsequently, 200 µl chloroform was added to each sample and vigorously mixed for 15 seconds followed by incubation at room temperature for three minutes. The samples were centrifuged at 20,400 x g for 15 minutes and the top 600 µl of the phase-separated mixture was collected and precipitated with 1 ml of isopropyl alcohol overnight at -80˚C. Precipitated RNA was then pelleted and the supernatant was discarded. The pellet was washed with 200 µl of 70% ethanol, air-dried for 10 minutes at room temperature and resuspended in 50 µl purified water, followed by incubation at 55˚C for five minutes.

Quantitative real-time RT-PCR

Total RNA (4 µg) was incubated with 1 µl of RQ1 RNase-free DNase (Promega MG10A) in 0.5 mM $MgCl_2$ at a total volume of 20 µl for 30 minutes at 37˚C followed by 10 minutes at 75˚C for enzymatic inactivation. 1 µg of this RNA was converted to cDNA using the iScript cDNA Synthesis Kit (Bio-Rad 170-8891) per the manufacturer's instructions. Quantitative PCR amplification was then performed using specific primer pairs (1.25 µM) and the appropriate template nucleic acids in the presence of iTaq SYBR Green with ROX (Bio-Rad 172-5850) on

an ABI 7900 HT Fast Real Time PCR System. Subsequent data analysis was executed using the ABI SDS 2.2.2 software package. The primer pairs used were as follows MD399/400 (*hfq*) and MD421/422 (Sigma A).

Northern Blot analyses

Total RNA samples (15–20 μg) were heated at 65°C for ten minutes in gel loading buffer (45 mM Tris–borate, 4 M urea, 10% sucrose [w/v], 5 mM EDTA, 0.05% SDS, 0.025% xylene cyanol FF, 0.025% bromophenol blue) and resolved by 6% denaturing (8 M urea) polyacrylamide electrophoresis. RNAs were transferred to BrightStar-Plus nylon membranes (Ambion) using a semi-dry electroblotting apparatus (Owl Scientific) according to manufacturer instructions. The blots were UV-crosslinked and hybridized overnight at 42°C in UltraHyb-Oligo buffer (Ambion) with the appropriate 5′ -radiolabeled (^{32}P) DNA oligonucleotide (Table S4). The blots were then washed twice for 15 min using low stringency wash buffer (1 × SSC, 0.1% SDS, 1 mM EDTA). Radioactive bands were visualized using ImageQuant or ImageJ software and a Typhoon PhosphorImager (Molecular Dynamics).

Genetic transformation of Bacillus subtilis

All strains were transformed with 1-10 µg plasmid DNA using the one-step method. Briefly, cells were inoculated from a single colony into transformation media consisting of 12.5g of K_2HPO_4, 3 g of KH_2PO_4, 0.5g trisodium citrate, 0.1g $MgSO_4$, 1g Na_2SO_4, 50µM $FeCl_3$ 2µM $MnSO_4$, 0.4% glucose and 0.2% glutamate in 500 mL of water, and placed in a 37˚ incubator standing over night. The following day they were placed in a shaking incubator and the OD_{600} was monitored until the cultures reached exponential phase (between 0.4 and 0.8) and were then inoculated with plasmid DNA. The cultures were subsequently replaced in the incubator shaking for one hour and then plated on an LB plate with the appropriate antibiotic.

Deletion of Hfq using the pMAD system

To generate the markerles Δ*hfq* deletion construct, approximately 400 bp upstream of *hfq* gene, including the first codon, was amplified from *B. subtilis* 168 chromosomal DNA using primer pairs MD324/325. These oligonucleotides were engineered to contain an in-frame stop codon following the ATG start codon. They also introduced a *Not I* restriction site into the *hfq* gene. The second half of the construct was generated via PCR amplification using primer pairs MD326/MD327. This forward oligonucleotide (MD327) begins at the ATG of *hfq*

and is directly complimentary to MD325, while the reverse oligonucleotide MD327 hybridizes with sequences located approximately 100 bases downstream of the *hfq* coding sequence. The two PCR products were mixed in equal amounts and sewn together through amplification with primers MD324/327; the resulting PCR product was then checked for digestion with *Not I*. This DNA was then sub-cloned into pMAD (Arnaud et al., 2004) via *EcoR1* and *BamHI* sites, which were added by the oligonucleotide primers and the correct clones were confirmed by DNA sequencing. The pMAD plasmid carries a temperature-sensitive origin of replication, an erythromycin resistance cassette and a constitutively active *lacZ* gene. Correspondingly, the pMAD-based plasmids were transformed into *B. subtilis* strain 168 at the permissive temperature for plasmid replication (30°C) with selection for resistance to erythromycin (1 μg ml^{-1} and lincomycin 25 μg ml^{-1}) on plates containing bromo-chloro-indolyl-galactopyranoside (X gal). To stimulate integration of the plasmids via Campbell recombination the cells were cultured overnight in 2xYT broth at the restrictive temperature (37°C) and then incubated on solid medium at 37°C with selection for resistance to erythromycin and lincomycin. The resulting isolates were blue from pMAD-encoded *lacZ* expression. To screen for recombination-based loss of the integrated plasmids, the strains were incubated overnight in 2xYT broth without shaking and without antibiotics at 30°C, followed by shaking incubation for 5 h at

30°C and then 3 h at 37°C. The cells were then serially diluted and plated on tryptose blood agar base (TBAB, Difco) at 37°C in the absence of antibiotic. Individual colonies were patched onto TBAB plates with and without erythromycin and lincomycin. Isolates that were sensitive to antibiotics, and were white on X gal-containing medium, were presumed to result from recombination-loss of the integrated plasmid. Chromosomal DNA was isolated from these strains and used as templates for diagnostic PCR reactions and subsequent DNA sequencing reactions to confirm mutagenesis of the targeted genomic locus.

Construction of an ectopic, inducible, epitope-tagged hfq Allele.

The *hfq* coding sequence was amplified using primers MD381/382 and sub-cloned via *HindIII* and *NheI* into pHyper-SPANK (a gift from David Rudner), an IPTG-inducible expression vector that integrates into the *amyE* gene. To ensure efficient translation the forward oligo used to amplify *hfq* added the ribosome binding site from the *rpsD* gene. Generation of a DNA template encoding the FLAG peptide sequence was accomplished via primer extension of overlapping oligonucleotides MD383/384, which regenerated the 3x FLAG repeat as encoded by the yeast vector p417-CYC-NTAP (a gift from Benjamin Tu, Sunil Laxman). This DNA fragment was then subcloned into pHyper-SPANK and fused to the 3'

terminus of the *hfq* coding sequence, resulting in plasmid pMD-145. Transformants were screened by diagnostic PCR, verified by DNA sequencing, and used for the co-immunoprecipitation experiments described herein.

Construction of a Strain Expressing an Epitope-Tagged hfq Allele.

The *hfq* gene containing a C-terminal 3x FLAG tag was amplified from pMD-145 using the oligonucleotide primers MD395/396 and sub-cloned into the pMAD vector via *BglII* and *EcoRI*. In order to preserve the C-terminal FLAG sequence during allelic exchange, 400 bp immediately downstream of the *hfq* coding sequence was amplified using MD397/MD398 and also sub-cloned into the above pMAD vector, downstream of the epitope-tagged copy of *hfq*. Positive clones were subsequently screened by diagnostic PCR and confirmed by sequencing, resulting in plasmid pMD-131. Transformation and allelic exchange using the pMAD-based plasmids are described above.

Construction of promoter-YFP fusions

To generate each of the sRNA promoter-YFP fusions, the targeted promoter regions of the *surD*, *surE*, *csfG*, and *surG* sRNA genes were PCR amplified using the following primers, respectively: VMZ007/VMZ008, VMZ005/VMZ006, VMZ013/VMZ014, and VMZ009/VMZ010. The amplified promoter regions were digested with BamHI and HindIII and subcloned into

pVMZ006, a pDG1622-derived vector containing a constitutive *rpsD* promoter-YFP fusion. For subcloning of the sRNA promoter regions, the pVMZ006 plasmid was first digested with BamHI and HindIII for excision of the existing *rpsD* promoter region and then purified by gel electrophoresis. The digested vector was then ligated with sRNA promoter DNA fragments that had been digested with BamHI and HindIII. The resulting vectors (pVMZ010, pVMZ011, pVMZ012, and pVMZ013) were transformed into *B. subtilis* NCIB3610 to integrate via double crossover recombination into the *amyE* locus using standard techniques (Irnov *et al.,* 2010). Correct isolates were confirmed by diagnostic PCR reactions and Sanger sequencing.

Fluorescence microscopy

PY79 cells were induced to enter sporulation using a modified version of a published protocol (Harwood and Cutting, 1991). Briefly, one colony of PY79 was inoculated into 5 mL LB broth and was incubated without shaking overnight at 37 °C to reach mid-log phase by morning, whereupon the cells were diluted into 20 mL growth media (CH1, CH2, CH3, and CH4 media; Harwood and Cutting, 1991) without tryptophan. Cells were subsequently incubated at 37 °C, shaking to reach an OD_{600} = 0.5-0.8, at which point the cells were centrifuged for 5

minutes at low speed. These pellets were resuspended in an equal volume of resuspension media (Mandelstam sporulation salts, 0.1% L-glutamate, 1 mM $CaCl_2$, 10 mM $MgSO_4$; Harwood and Cutting, 1991). The point at which cells were resuspended in resuspension media was defined as T_0. Cells were incubated at 37 °C with shaking and aliquots of sporulating cultures were removed every hour until T_{10}. Specifically, a 500 μL aliquot of sporulating cell samples was removed from the culture at each time point, pelleted, and resuspended in 25 μL of phosphate-buffered saline solution (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 2 mM KH_2PO_4). A 5 μL aliquot was pipetted onto a 1.3 % agarose pad on a glass slide to immobilize the cells. Samples were then imaged using an Applied Precision DeltaVision RT Deconvolution microscope equipped with a CoolSNAP-HQ digital CCD camera (Photometrics), courtesy of the UT Southwestern Molecular and Cellular Imaging Facility. The cells were viewed using an Olympus PlanApo 60x oil objective under brightfield channel and YFP channel (excitation 492 +/- 18 nm, emission 535 +/- 30 nm). Images were acquired with DeltaVision softWoRx (Applied Precision) and processed using ImageJ software.

Sporulation assays

Sporulation of *Bacillus subtilis* was induced by the resuspension method detailed above in the "Fluorescence Microscopy" section. After 24 hours

sporulation efficiency was determined by spinning down 500 µl of cells and resuspending in 500 µl of saline EDTA plus or minus 10% chloroform (v/v). The cell suspensions were then vortexed for 10 seconds, and incubated at room temperature for 10 minutes. Cells were then pelleted, washed with saline-EDTA, and serially diluted onto TBAB plates. Sporulation efficiency was calculated as the proportion of chloroform-resistant colony forming unit (CFU) compared to the saline-EDTA treated control.

Growth conditions and media composition

All *B. subtilis* strains were typically cultured at 37 ^{0}C in either defined glucose minimal medium (0.5% glucose, 0.5 mM $CaCl_2$, 5 µM $MnCl_2$, 15 mM $(NH_4)_2SO_4$, 80 mM K_2HPO_4, 44 mM KH_2PO_4, 3.9 mM sodium citrate, 8.1 mM $MgSO_4$), 2xYT (16 g tryptone, 10 g yeast extract, 5 g NaCl per liter), or on Tryptone Blood Agar Base (TBAB, Difco) plates.

All *E. coli* strains were cultured at 37 ^{0}C in LB (10 g tryptone, 5 g yeast extract, 5 g NaCl per liter). To analyze *B. subtilis* colony architecture, 5 µl of overnight culture in 2xYT was spotted onto MSgg plates (100 mM MOPS (pH 7), 5 mM potassium phosphate (pH 7), 0.5% glycerol, 0.5% glutamate, 2 mM $MgCl_2$, 700 µM $CaCl_2$, 50 µM $MnCl_2$, 100 µM $FeCl_3$, 1 µM $ZnCl_2$, 2 µM thiamine; Branda *et al.,* 2001) supplemented with 1.5% Bacto agar (Difco) and incubated at 30 ^{o}C

for 48-72 hours. Images of *B. subtilis* colonies were captured at 6-10X magnification using a Zeiss AxioCam Mrc 5 camera equipped with a 0.63X objective lens (owned by the UT Southwestern Medical Center Live Cell Imaging Core Facility). When appropriate, antibiotics were included at the following concentrations: 50 µg ml^{-1} ampicillin (for *E. coli*) and 100 µg ml^{-1} spectinomycin for *B. subtilis*.

Oligonucelotides used in these studies

Table 5-1. Oligonucleotides used in this study.

Name	Description	Sequence (written 5' to 3')
MD324	Forward oligonucleotide for amplification of *hfq* upstream region	ATCATCGGATCCCAAGAAGAGTCATTCGCGCACTGG
MD325	Reverse oligonucleotide for amplification of *hfq* upstream region. Adds a stop codon and *Not I* site to *hfq* coding sequence.	CGGGCGGCCGCCTACATGTTTCGTCCTCCTTGATTCTC
MD326	Forward oligonucleotide for amplification of *hfq* and its downstream region. Adds stop codon and *Not I* site to *hfq* coding sequence.	ATGTAGGCGGCCGCCCGATTAATATTCAGGATCAGTTT
MD336	Reverse oligonucleotide for amplification of *hfq* and its downstream region. Adds an *EcoRI* site.	AAAAAAGAATTCGAGTTCAAGCTGGACGTTTTTTTG
MD381	Forward oligonucleotide for amplification of *hfq*. Adds the *rpsD* ribosome binding site and a *Hind III* site.	AAAAAAAAGCTTCCAAAGGAGGAGTCACATTATGAAACCGATTAA TATTC

MD382	Reverse oligonucleotide for amplification of *hfq*.Adds *Nhe I* site.	TTTTGCTAGCTTCGAGTTCAAGCTGGACGTTTTTTTG
MD383	Forward oligonucleotide for creation of the sequence encoding for the FLAG epitope. Adds *Nhe I* site.	AAAAGCTAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCG
MD384	Forward oligonucleotide for creation of the sequence encoding for the FLAG epitope. Adds *Sph I* site.	TTTTGCATGCTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCG
MD395	Forward oligonucleotide for fusion of the FLAG-encoding sequence (3x) to *hfq*. Adds *BamHI* site.	AAAAAAGGATCCATGAAACCGATTAATATTCAG
MD396	Forward oligonucleotide for fusion of the FLAG-encoding sequence (3x) to *hfq*. Adds *EcoRI* site.	TTTTTTGAATTCGCGGCCGCTTACTTGTCATCGTCATCC
MD397	Forward oligonucleotide for amplification of *hfq* downstream region. Adds *Eco*RI site.	AAAAAAGAATTCGCCATGTCAAGACATGAG
MD398	Reverse oligonucleotide for amplification of *hfq* downstream region. Adds *Bgl* II site.	TTTTTTAGATCTGATATAGCCGTCCTCC
MD399	Forward oligonucleotide for amplification of an internal region of *hfq* for the purposes of qPCR quantification.	TCAAATCCGGAAAGAAAATACGTAT
MD400	Reverse oligonucleotide for amplification of an internal region of *hfq* for the purposes of qPCR quantification.	GCCCCGCAACTGAAAGC
MD414	Oligonucleotide used as probe in Northern analysis of the *ykoY* orphan riboswitch region.	GCACCTTGGTGATTTCATTGAA
MD421	Forward oligonucleotide for amplification of an internal region of *sigA* for the purposes of qPCR quantification.	TGCGGCGTGGTCAGAAG
MD422	Reverse oligonucleotide for amplification of an internal region of *sigA* for the purposes of qPCR quantification.	GACGACTTTGTTTCCGGAATACATTCCGGACGCTACTCCCC
MD425	Oligonucleotide used as a probe during Northern blotting analysis of CsfG.	GGAGAAACCGGAGGAAGAACTTATGGGGAAACG
MD426	Oligonucleotide used as a probe for Northern blotting analysis of the orphan riboswitch leader region, *ylbH*.	CGTTTCCCCATAAGTTCTTCCTCCGGTTTCTCC
MD427	Oligonucleotide used as a probe for Northern blotting analysis of *txpA*.	CAAAGCCGATCATGACCATTAGAGATTCATAGGTCG

MD428	Oligonucleotide used as a probe for Northern blotting analysis of *ratA*.	CAAGTGGTAATGTGGTAATGTGGTACCAACTATAAGCTTAC
MD429	Oligonucleotide used as a probe for *ncr1019*.	GCATTAATGCCTGGAATGTTGACATAGCATCACCCC
MD430	Oligonucleotide used as a probe for Northern blotting analysis of *bsrE*.	GTACAGAGCCGGGGTGTTGGTAGCACCTCGGTC
MD433	Oligonucleotide used as a probe for Northern blotting analysis of *bsrG*.	CATTATCATTAATGATTCGTAAACAGTCATTTTCCCACCCC
MD434	Oligonucleotide used as a probe for Northern blotting analysis of *ncr1932*.	GAGCCAGGGTGCTACCAACACCCTGGTCTTTTTATTTTATG

Table 5-2. Oligonucleotides used for construction of YFP-promoter fusions (V. Zacharia)

Name	Oligonucleotide sequence (written 5'-3')	
VMZ005	AATAATGGATCCGAAAAAAGTGATAGACGTCTAGTC	Forward oligo for *surE* promoter
VMZ006	GGCTGCAAGCTTGCAGATTCACTATACAATAGAATATG	Reverse oligo for *surE* promoter
VMZ007	AATAATGGATCCGCACCATTTGGCGCTTTCTG	Forward oligo for *surD* promoter
VMZ008	GGCTGCAAGCTTTTATCCAGCAATATTTTATCCTATATATC	Reverse oligo for *surD* promoter
VMZ009	AATAATGGATCCCAAATGATGAATTTGCTGTC	Forward oligo for *surG* promoter
VMZ010	GGCTGCAAGCTTTTCCTCTACAACTTTTATGATGTCC	Reverse oligo for *surG* promoter
VMZ013	TGTTTCTGGATCCGGCCGGACGCCGCTATGTAC	Forward oligo for *csfG* promoter
VMZ014	TTTTTTAAGCTTAGATGCTGTTGCTTATAGTATGG	Reverse oligo for *csfG* promoter

BIBLIOGRAPHY

Aguilar, C., H. Vlamakis, R. Losick & R. Kolter (2007) Thinking about Bacillus subtilis as a multicellular organism. *Curr Opin Microbiol,* 10**,** 638-43.

Amaya, E., A. Khvorova & P. J. Piggot (2001) Analysis of promoter recognition in vivo directed by sigma(F) of Bacillus subtilis by using random-sequence oligonucleotides. *J Bacteriol,* 183**,** 3623-30.

Arnaud, M., A. Chastanet & M. Débarbouillé (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl Environ Microbiol,* 70**,** 6887-91.

Babitzke, P. & T. Romeo (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current Opinion in Microbiology,* 10**,** 156-163.

Baker, J. L., N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge & R. R. Breaker (2012) Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. *Science,* 335**,** 233-235.

Barrick, J. E., K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser & R. R. Breaker (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A,* 101**,** 6421-6.

Barrick, J. E., N. Sudarsan, Z. Weinberg, W. L. Ruzzo & R. R. Breaker (2005) 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *Rna-a Publication of the Rna Society,* 11**,** 774-784.

Bartel, D. P. (2009) MicroRNAs: Target Recognition and Regulatory Functions. *Cell,* 136**,** 215-233.

Battesti, A., N. Majdalani & S. Gottesman (2011) The RpoS-Mediated General Stress Response in Escherichia coli. *Annual Review of Microbiology, Vol 65,* 65**,** 189-213.

Blencowe, B. J. (2006) Alternative splicing: new insights from global analyses. *Cell,* 126**,** 37-47.

Blumenthal, T. & G. G. Carmichael (1979) RNA replication: function and structure of Qbeta-replicase. *Annu Rev Biochem,* 48**,** 525-48.

Bocobza, S., A. Adato, T. Mandel, M. Shapira, E. Nudler & A. Aharoni (2007) Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes Dev,* 21**,** 2874-9.

Bouvier, M., C. M. Sharma, F. Mika, K. H. Nierhaus & J. Vogel (2008) Small RNA Binding to 5 ' mRNA Coding Region Inhibits Translational Initiation. *Molecular Cell,* 32**,** 827-837.

Breaker, R. R. (2012) Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology,* 4.

Brennan, R. G. & T. M. Link (2007) Hfq structure, function and ligand binding. *Current Opinion in Microbiology,* 10**,** 125-133.

Cavanagh, A. T., A. D. Klocko, X. C. Liu & K. M. Wassarman (2008) Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of sigma(70). *Molecular Microbiology,* 67**,** 1242-1256.

Chao, Y. & J. Vogel (2010) The role of Hfq in bacterial pathogens. *Curr Opin Microbiol,* 13**,** 24-33.

Cheah, M. T., A. Wachter, N. Sudarsan & R. R. Breaker (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature,* 447**,** 497-500.

Coller, J. (2008) Methods to determine mRNA half-life in Saccharomyces cerevisiae. *Methods Enzymol,* 448**,** 267-84.

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

Condon, C. & D. H. Bechhofer (2011) Regulated RNA stability in the Gram positives. *Current Opinion in Microbiology,* 14**,** 148-154.

Croft, M. T., M. Moulin, M. E. Webb & A. G. Smith (2007) Thiamine biosynthesis in algae is regulated by riboswitches. *Proc Natl Acad Sci U S A,* 104**,** 20770-5.

Dambach, M. D. & W. C. Winkler (2009a) Expanding roles for metabolite-sensing regulatory RNAs. *Current Opinion in Microbiology,* 12**,** 161-169.

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

De Hoon, M. J., S. Imoto, K. Kobayashi, N. Ogasawara & S. Miyano (2004) Predicting the operon structure of Bacillus subtilis using operon length, intergene distance, and gene expression information. *Pac Symp Biocomput***,** 276-87.

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

Errington, J. (2003) Regulation of endospore formation in Bacillus subtilis. *Nat Rev Microbiol,* 1**,** 117-26.

Fozo, E. M., M. R. Hemm & G. Storz (2008) Small toxic proteins and the antisense RNAs that repress them. *Microbiol Mol Biol Rev,* 72**,** 579-89, Table of Contents.

Fozo, E. M., K. S. Makarova, S. A. Shabalina, N. Yutin, E. V. Koonin & G. Storz (2010) Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic Acids Res,* 38**,** 3743-59.

Gaballa, A., H. Antelmann, C. Aguilar, S. K. Khakh, K. B. Song, G. T. Smaldone & J. D. Helmann (2008) The Bacillus subtilis iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci U S A,* 105**,** 11927-32.

Garst, A. D., A. L. Edwards & R. T. Batey (2011) Riboswitches: Structures and Mechanisms. *Cold Spring Harbor Perspectives in Biology,* 3.

Geissmann, T., C. Chevalier, M. J. Cros, S. Boisset, P. Fechter, C. Noirot, J. Schrenzel, P. François, F. Vandenesch, C. Gaspin & P. Romby (2009) A search for small noncoding RNAs in Staphylococcus aureus reveals a conserved sequence motif for regulation. *Nucleic Acids Res,* 37**,** 7239-57.

Gelfand, M. S., A. A. Mironov, J. Jomantas, Y. I. Kozlov & D. A. Perumov (1999) A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends in Genetics,* 15**,** 439-442.

Gerdes, K. & E. G. H. Wagner (2007) RNA antitoxins. *Current Opinion in Microbiology,* 10**,** 117-124.

Gottesman, S., C. A. McCullen, M. Guillier, C. K. Vanderpool, N. Majdalani, J. Benhammou, K. M. Thompson, P. C. FitzGerald, N. A. Sowa & D. J. FitzGerald (2006) Small RNA regulators and the bacterial response to stress. *Cold Spring Harb Symp Quant Biol,* 71**,** 1-11.

Gottesman, S. & G. Storz (2011a) Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harbor Perspectives in Biology,* 3.

--- (2011b) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol,* 3.

Henkin, T. M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev,* 22**,** 3383-90.

Higgins, D. & J. Dworkin (2012) Recent progress in Bacillus subtilis sporulation. *FEMS Microbiol Rev,* 36**,** 131-48.

Holmqvist, E., J. Reimegard, M. Sterk, N. Grantcharova, U. Romling & E. G. H. Wagner (2010) Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *Embo Journal,* 29**,** 1840-1850.

Ikeda, Y., M. Yagi, T. Morita & H. Aiba (2011) 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,* 79**,** 419-432.

Irnov, A. Kertsburg & W. C. Winkler (2006) Genetic control by cis-acting regulatory RNAs in Bacillus subtilis: general principles and prospects for discovery. *Cold Spring Harb Symp Quant Biol,* 71**,** 239-49.

Irnov, I., C. M. Sharma, J. Vogel & W. C. Winkler (2010) Identification of regulatory RNAs in Bacillus subtilis. *Nucleic Acids Res,* 38**,** 6637-51.

Irnov, I. & W. C. Winkler (2010) A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. *Mol Microbiol,* 76**,** 559-75.

Ishikawa, H., H. Otaka, K. Maki, T. Morita & H. Aiba (2012) The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3' poly(U) tail. *RNA,* 18**,** 1062-74.

Jacques, J. F., S. Jang, K. Prevost, G. Desnoyers, M. Desmarais, J. Imlay & E. Masse (2006) RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in Escherichia coli. *Molecular Microbiology,* 62**,** 1181-1190.

Jiang, M., W. Shao, M. Perego & J. A. Hoch (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis. *Mol Microbiol,* 38**,** 535-42.

Klein, D. J., M. D. Been & A. R. Ferre-D'Amare (2007) Essential role of an active-site guanine in glmS ribozyme catalysis. *Journal of the American Chemical Society,* 129**,** 14858-+.

Klein, D. J. & A. R. Ferre-D'Amare (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science,* 313**,** 1752-1756.

Kroos, L. (2007) The Bacillus and Myxococcus developmental networks and their transcriptional regulators. *Annu Rev Genet,* 41**,** 13-39.

Lenz, D. H., K. C. Mok, B. N. Lilley, R. V. Kulkarni, N. S. Wingreen & B. L. Bassler (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. *Cell,* 118**,** 69-82.

Link, T. M., P. Valentin-Hansen & R. G. Brennan (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 19286-19291.

Lopez, D., M. A. Fischbach, F. Chu, R. Losick & R. Kolter (2009) Structurally diverse natural products that cause potassium leakage trigger multicellularity in Bacillus subtilis. *Proc Natl Acad Sci U S A,* 106**,** 280-5.

Lorenz, C., T. Gesell, B. Zimmermann, U. Schoeberl, I. Bilusic, L. Rajkowitsch, C. Waldsich, A. von Haeseler & R. Schroeder (2010) Genomic SELEX for Hfq-binding RNAs

identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Research,* 38**,** 3794-3808.

Majdalani, N., C. Cunning, D. Sledjeski, T. Elliott & S. Gottesman (1998) 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,* 95**,** 12462-12467.

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

Mandal, M., B. Boese, J. E. Barrick, W. C. Winkler & R. R. Breaker (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell,* 113**,** 577-586.

Mandin, P. & S. Gottesman (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *Embo Journal,* 29**,** 3094-3107.

Marchais, A., S. Duperrier, S. Durand, D. Gautheret & P. Stragier (2011) CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers. *RNA Biol,* 8**,** 358-64.

Masse, E., F. E. Escorcia & S. Gottesman (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes & Development,* 17**,** 2374-2383.

Masse, E., H. Salvail, G. Desnoyers & M. Arguin (2007) Small RNAs controlling iron metabolism. *Current Opinion in Microbiology,* 10**,** 140-145.

Masse, E., C. K. Vanderpool & S. Gottesman (2005) Effect of RyhB small RNA on global iron use in Escherichia coli. *Journal of Bacteriology,* 187**,** 6962-6971.

McCullen, C. A., J. N. Benhammou, N. Majdalani & S. Gottesman (2010) Mechanism of Positive Regulation by DsrA and RprA Small Noncoding RNAs: Pairing Increases Translation and Protects rpoS mRNA from Degradation. *Journal of Bacteriology,* 192**,** 5559-5571.

Moeller, R., G. Horneck, P. Rettberg, H. J. Mollenkopf, E. Stackebrandt & W. L. Nicholson (2006) A method for extracting RNA from dormant and germinating Bacillus subtilis strain 168 endospores. *Current Microbiology,* 53**,** 227-231.

Moller, T., T. Franch, P. Hojrup, D. R. Keene, H. P. Bachinger, R. G. Brennan & P. Valentin-Hansen (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Molecular Cell,* 9**,** 23-30.

Moreno, R., S. Marzi, P. Romby & F. Rojo (2009) The Crc global regulator binds to an unpaired A-rich motif at the Pseudomonas putida alkS mRNA coding sequence and inhibits translation initiation. *Nucleic Acids Research,* 37**,** 7678-7690.

Muffler, A., D. Fischer & R. Hengge-Aronis (1996) The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for rpoS translation in Escherichia coli. *Genes Dev,* 10**,** 1143-51.

Nahvi, A., J. E. Barrick & R. R. Breaker (2004) Coenzyme B-12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Research,* 32**,** 143-150.

Narberhaus, F., T. Waldminghaus & S. Chowdhury (2006) RNA thermometers. *FEMS Microbiol Rev,* 30**,** 3-16.

Neusser, T., T. Polen, R. Geissen & R. Wagner (2010) Depletion of the non-coding regulatory 6S RNA in E. coli causes a surprising reduction in the expression of the translation machinery. *Bmc Genomics,* 11.

Otaka, H., H. Ishikawa, T. Morita & H. Aiba (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc Natl Acad Sci U S A,* 108**,** 13059-64.

Papenfort, K., N. Said, T. Welsink, S. Lucchini, J. C. D. Hinton & J. Vogel (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Molecular Microbiology,* 74**,** 139-158.

Pfeiffer, V., K. Papenfort, S. Lucchini, J. C. D. Hinton & J. Vogel (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nature Structural & Molecular Biology,* 16**,** 840-U63.

Piggot, P. J. & D. W. Hilbert (2004) Sporulation of Bacillus subtilis. *Curr Opin Microbiol,* 7**,** 579-86.

Prevost, K., H. Salvail, G. Desnoyers, J. F. Jacques, E. Phaneuf & E. Masse (2007) The small RNA RyhB activates the translation of shiA mRNA encoding a permease of

shikimate, a compound involved in siderophore synthesis. *Molecular Microbiology,* 64**,** 1260-1273.

Repoila, F. & F. Darfeuille (2009) Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biology of the Cell,* 101**,** 117-131.

Richards, G. R. & C. K. Vanderpool (2011) Molecular call and response: The physiology of bacterial small RNAs. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms,* 1809**,** 525-531.

Saito, S., H. Kakeshita & K. Nakamura (2009) Novel small RNA-encoding genes in the intergenic regions of Bacillus subtilis. *Gene,* 428**,** 2-8.

Salvail, H., P. Lanthier-Bourbonnais, J. M. Sobota, M. Caza, J. A. M. Benjamin, M. E. S. Mendieta, F. Lepine, C. M. Dozois, J. Imlay & E. Masse (2010) A small RNA promotes siderophore production through transcriptional and metabolic remodeling. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 15223-15228.

Schmalisch, M., E. Maiques, L. Nikolov, A. H. Camp, B. Chevreux, A. Muffler, S. Rodriguez, J. Perkins & R. Losick (2010) Small genes under sporulation control in the Bacillus subtilis genome. *J Bacteriol,* 192**,** 5402-12.

Shah, I. M., M. H. Laaberki, D. L. Popham & J. Dworkin (2008) A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell,* 135**,** 486-96.

Silvaggi, J. M., J. B. Perkins & R. Losick (2005) Small untranslated RNA antitoxin in Bacillus subtilis. *J Bacteriol,* 187**,** 6641-50.

--- (2006) Genes for small, noncoding RNAs under sporulation control in Bacillus subtilis. *J Bacteriol,* 188**,** 532-41.

Sittka, A., S. Lucchini, K. Papenfort, C. M. Sharma, K. Rolle, T. T. Binnewies, J. C. D. Hinton & J. Vogel (2008) Deep Sequencing Analysis of Small Noncoding RNA and mRNA Targets of the Global Post-Transcriptional Regulator, Hfq. *Plos Genetics,* 4.

Someya, T., S. Baba, M. Fujimoto, G. Kawai, T. Kumasaka & K. Nakamura (2012) Crystal structure of Hfq from Bacillus subtilis in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Res,* 40**,** 1856-67.

Sonnleitner, E., L. Abdou & D. Haas (2009) Small RNA as global regulator of carbon catabolite repression in Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 21866-21871.

Storz, G., J. A. Opdyke & K. M. Wassarman (2006) Regulating bacterial transcription with small RNAs. *Cold Spring Harb Symp Quant Biol,* 71**,** 269-73.

Storz, G., J. Vogel & K. M. Wassarman (2011) Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell,* 43**,** 880-891.

Straight, P. D. & R. Kolter (2009) Interspecies chemical communication in bacterial development. *Annu Rev Microbiol,* 63**,** 99-118.

Sudarsan, N., J. E. Barrick & R. R. Breaker (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *Rna-a Publication of the Rna Society,* 9**,** 644-647.

Sun, X. G., I. Zhulin & R. M. Wartell (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Research,* 30**,** 3662-3671.

Takyar, S., R. P. Hickerson & H. F. Noller (2005) MRNA helicase activity of the ribosome. *Cell,* 120**,** 49-58.

Thomason, M. K. & G. Storz (2010) Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annual Review of Genetics, Vol 44,* 44**,** 167-188.

Timmermans, J. & L. Van Melderen (2010) Post-transcriptional global regulation by CsrA in bacteria. *Cellular and Molecular Life Sciences,* 67**,** 2897-2908.

Tsui, H. C., H. C. Leung & M. E. Winkler (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. *Mol Microbiol,* 13**,** 35-49.

Vlamakis, H., C. Aguilar, R. Losick & R. Kolter (2008) Control of cell fate by the formation of an architecturally complex bacterial community. *Genes Dev,* 22**,** 945-53.

Vogel, J. (2009) A rough guide to the non-coding RNA world of Salmonella. *Molecular Microbiology,* 71**,** 1-11.

Vogel, J. & B. F. Luisi (2011) Hfq and its constellation of RNA. *Nature Reviews Microbiology,* 9**,** 578-589.

Vogel, J. & K. Papenfort (2006) Small non-coding RNAs and the bacterial outer membrane. *Current opinion in microbiology,* 9**,** 605-11.

Wachter, A., M. Tunc-Ozdemir, B. C. Grove, P. J. Green, D. K. Shintani & R. R. Breaker (2007) Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell,* 19**,** 3437-50.

Wassarman, K. M. (2007) 6S RNA: a regulator of transcription. *Molecular Microbiology,* 65**,** 1425-1431.

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

Weber, H., T. Polen, J. Heuveling, V. F. Wendisch & R. Hengge (2005) Genome-wide analysis of the general stress response network in Escherichia coli: sigma(S)-dependent genes, promoters, and sigma factor selectivity. *Journal of Bacteriology,* 187**,** 1591-1603.

Wilusz, C. J. & J. Wilusz (2005) Eukaryotic Lsm proteins: lessons from bacteria. *Nature Structural & Molecular Biology,* 12**,** 1031-1036.

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

Zhang, A., S. Altuvia, A. Tiwari, L. Argaman, R. Hengge-Aronis & G. Storz (1998) The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J,* 17**,** 6061-8.

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