REGULATORY RNAs AT THE HEART OF SUGAR METABOLISM:

NEW MECHANISMS AND NOVEL DISCOVERIES

APPROVED BY SUPERVISORY COMMITTEE

Wade C. Winkler, Ph.D.

Vanessa Sperandio, Ph.D.

Kevin H. Gardner, Ph.D.

Benjamin Tu, Ph.D.

REGULATORY RNAs AT THE HEART OF SUGAR METABOLISM: NEW MECHANISMS AND NOVEL DISCOVERIES

by

IRNOV

DISSERTATION

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

December, 2010

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my mentor, Dr. Wade Winkler, for all his guidance and support during the past five years. His passion for science is contagious and has been invaluable at times for moving me forward. I am grateful to have had the opportunity to grow both as a scientist and a person under his guidance.

I would like to thank Dr. Vanessa Sperandio, Dr. Kevin Gardner, and Dr. Benjamin Tu for being in my thesis committee. Their advice and encouragement have been crucial for directing my projects. I would also like to acknowledge the Sara and Frank McKnight Foundation for their pre-doctoral fellowship, which has allowed me to pursue many interesting avenues during my studies.

I need to extend my utmost gratitude to the past and present members of the Winkler laboratory for their friendship and tremendous helps throughout my graduate career. In particular, I would like to thank Jennifer Collins for getting me started with the *glmS* project. I am indebted to Dr. Alexis Kertsburg and Stephanie Baker for their helps when I first joined the lab as a new graduate student. I want to specially thank Dr. Arati Ramesh, Dr. Laura Motta-Mena, and Michael Dambach for their advices and for making the lab a fun place to work in. Most importantly, I would like to thank to Dr. Catherine Wakeman for her advices, friendship, and the many interesting discussions we have had over the years.

REGULATORY RNAs AT THE HEART OF SUGAR METABOLISM:

NEW MECHANISMS AND NOVEL DISCOVERIES

Irnov, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2010

Wade C. Winkler, Ph.D.

Bacteria are adept at using a variety of posttranscriptional strategies to regulate gene expression. Specifically, various RNA-mediated genetic control elements have been discovered in the past decade through a combination of genetics, bioinformatics, and transcriptomic approaches. Together, these RNA elements control the expression of many genes involved in diverse cellular processes such as energy metabolism, stress response, biofilm formation, and pathogenesis. In the Gram-positive bacterium *Bacillus subtilis*, several RNA elements have been shown to be required for the precise coordination of genes

involved in various sugar utilization pathways. These genetic switches typically regulate gene expression by modulating the formation of a transcription termination element in a ligand-dependent manner. Interestingly, two unique elements, the *glmS* ribozyme and the *eps*-associated RNA (EAR), are missing the signature elements required for control of transcription termination or translation initiation. The latter mechanism is more commonly found in Gram-negative bacteria. Our objective is to study the mechanisms by which these two RNAs control gene expression. Additionally, we would like to identify other regulatory RNAs that are important for sugar metabolism in *Bacillus subtilis*.

Both the *glmS* ribozyme and EAR are positioned at the center of the sugar metabolism pathways in *B. subtilis*. The *glmS* RNA is a glucosamine-6-phosphate responsive element that regulates the expression of the GlmS enzyme, which directs sugar precursors from glycolysis into the cell wall biosynthesis pathway. The EAR element resides within the 16-kb *eps* operon that is required for biofilm exopolysaccharide production. Our data demonstrates that both RNAs employ novel mechanisms: the *glmS* ribozyme utilizes a ligand-specific RNase-mediated degradation event, while EAR uses a processive antitermination mechanism for complete synthesis of the long operon. Furthermore, by using high-throughput sequencing approach we have successfully identified many new regulatory RNA candidates, including various long 5`-UTR, toxin-antitoxin systems, prophage-

encoded RNAs, and several developmentally regulated small RNAs. Their functions are still under investigation.

Collectively, our studies provide important insights into the different aspects of bacterial physiology, including RNA decay pathways, transcription of long operons and cellular differentiation. We argue that posttranscriptional regulation is of greater importance to *Bacillus subtilis* (and probably all bacteria) than previously realized.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Du, L., Schageman, J.J., **Irnov**, Girard, L., Hammond, S.M., Minna, J.D., Gazdar, A.F., and Persemlidis, A. 2010. MicroRNA expression distinguishes SCLC from NSCLC lung tumor cells and suggests a possible pathological relationship between SCLCs and NSCLCs. J. Exp. Clin. Cancer Res. 29: 75.

Irnov, I., Sharma, C. M., Vogel, J. and Winkler, W. C. 2010. Identification of regulatory RNAs in *Bacillus subtilis*. *Nucleic Acids Res* doi: 10.1093/nar/gkq454.

Irnov, I. and Winkler, W. C. 2010. A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. *Mol Microbiol* 76(3): 559-75.

Dann, C. E., 3rd, Wakeman, C. A., Sieling, C. L., Baker, S. C., **Irnov, I.** and Winkler, W. C. 2007. Structure and mechanism of a metal-sensing regulatory RNA. *Cell* 130(5): 878-92.

Irnov, Kertsburg, A. and Winkler, W. C. 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.

Roy, H., Ling, J., **Irnov, M.**, and Ibba, M. 2004. Post-transfer editing *in vitro* and *in vivo* by the beta subunitphenylalanyl-tRNA-synthetase. *EMBO J.* 23(23): 4639-48.

LIST OF FIGURES

LIST OF TABLES

LIST OF DEFINITIONS

A	Adenosine
asRNA	antisense RNA
ATP	Adenosine triphosphate
bp	Basepair
β−gal	Beta-galactosidase
BME	β−Mercaptoethanol
BSA	Bovine Serum Albumin
DNA	Deoxyribonucleic acid
DTT	Dithiotreitol
EAR	*eps*-Associated RNA
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharide
EtBr	Ethidium bromide
G	Guanosine
GlcN6P	Glucosamine-6-phosphate
IPTG	Isopropyl-β-D-thiogalactopyranoside
kbp	kilobasepair
kD	kilo Dalton
mRNA	Messenger RNA

nt	Nucleotides
NTP	Nucleotide triphosphate
oligo	Oligonucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
RBS	Ribosomal Binding Site
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNase	Ribonuclease
RNA-Seq	RNA Sequencing
rRNA	Ribosomal RNA
RT	Room temperature or reverse transcription
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulfate
sRNA	Small RNA
T	Thymidine
TA	Toxin-Antitoxin
TBAB	Tryptone Blood Base Agar
tRNA	Transfer RNA
TSS	Transcription Start Site
U	Uridine

UTP	Uridine triphosphate
UTR	Untranslated region
WT	Wild-type
X-gal	Bromo-chloro-indolyl-galactopyranoside

CHAPTER ONE

Introduction and Literature Review

The ability to respond to sudden changes in the environment is essential for microorganisms. For this purpose, microorganisms have evolved various gene regulation strategies to precisely regulate their gene expression profiles. Although the majority of these control mechanisms are thought to be at the level of transcription initiation, recent studies have found a growing number of RNA-mediated strategies that affect gene expression at post-initiation stages. In this introduction, I will review previous work on certain regulatory RNAs and the mechanisms by which they control gene expression. I will focus specifically on RNA-mediated control of the sugar metabolism pathways in the Gram-positive bacterium *Bacillus subtilis*.

Regulatory RNA and posttranscriptional gene regulation

Depending on their structural relationship to the target gene, regulatory RNAs can be divided into two broad categories: those that are co-transcribed (in *cis*) or transcribed independently (in *trans*) of the target mRNA (Fig. 1-1). *Cis*-acting regulatory RNAs are typically located within the untranslated region (UTR) of the target transcript. On the other hand, *trans*-acting regulatory RNAs, which can be encoded either in *cis* or in *trans* relative to their target transcript, are transcribed separately from the mRNA target. Both classes of regulatory RNAs are capable of activating or repressing gene expression.

Trans-acting regulatory RNA

In the past several decades, many small (~70-200 nt) *trans*-acting noncoding RNAs ('sRNA'), have been discovered in bacteria (Wassarman *et al.*, 1999; Waters and Storz, 2009). Overlapping methods have been used for their discovery, including directed cloning and sequencing of sRNA pools, computational approaches, deep sequencing, and hybridization of sRNA pools to genomic tiling arrays (Altuvia, 2007; Livny and Waldor, 2007; Landt *et al.*, 2008; Livny *et al.*, 2008; Sittka *et al.*, 2008; Sharma and Vogel, 2009). Although a few noncoding RNAs serve 'housekeeping' functions, such as processing of pre-tRNAs by RNase P, most sRNA candidates identified by these approaches are presumed to function as regulatory agents (Fig. 1-2). For example, one widespread sRNA class regulates gene expression by sequestering an RNA-binding protein (*e.g.,* CsrA) and preventing it from controlling translation of target mRNAs (*e.g.,* Babitzke *et al.*, 2009). Another widespread class, coined 6S RNAs, regulates expression patterns by structurally mimicking an open promoter and associating with RNA polymerase to prevent transcription initiation of target genes (Wassarman, 2007). However, most sRNAs are likely to affect gene expression by directly base-pairing to one or more target mRNAs (Waters and Storz, 2009; Gottesman *et al.*, 2006). Some interact with their target mRNA via antisense base-pairing and are *trans*-acting but *cis*-encoded relative to the target gene. Typically, these antisense sRNAs associate with the target transcript through formation of long base-paired regions of greater than 65 nucleotides in order to regulate mRNA stability, translation or transcription elongation (Waters and Storz, 2009). *Trans*-encoded sRNAs, in contrast, usually associate with their target transcripts through shorter, more imperfect, base pairing interactions (Gottesman, 2005). Dozens of such sRNAs have been

well characterized in *E. coli* and other Gram-negative bacteria. Indeed it has been estimated that 200-300 sRNAs will be present in the average bacterial genome, equivalent in numbers to the total complement of cellular transcription factors (*e.g.,* Hershberg *et al.*, 2003). Most are individually responsive to different stress conditions; they largely assist the adaptive responses of the microorganism as its local environment undergoes changes. Similar to transcription factors, many sRNA regulators are predicted to regulate multiple targets (*e.g.,* Bejerano-Sagie and Xavier, 2007; Sharma *et al.*, 2007; Valentin-Hansen *et al.*, 2007; Papenfort and Vogel, 2009); therefore, the full range of their regulatory complexities remains to be determined.

Small noncoding RNAs are capable of using multiple mechanisms to modulate gene expression. Many sRNAs bind near to the start of the coding region in order to affect translation initiation efficiency of the target gene by preventing ribosome binding (Gottesman, 2005; Bouvier *et al.*, 2008); however, some reduce translation by interacting further upstream within the 5' untranslated region (5' UTR) (Sharma *et al.*, 2007; Darfeuille *et al.*, 2007; Vecerek *et al.*, 2007). Yet others reduce expression by associating with the mRNA further downstream, within the coding region, or in an inter-cistronic region to promote mRNA decay (*e.g.,* Pfeiffer *et al.*, 2009). Interestingly, a small subset of sRNAs is also capable of activating gene expression either by preventing the formation of a ribosome binding site (RBS) inhibitory hairpin or protecting target mRNA from degradation (*e.g.*, Kalamorz *et al.*, 2007). In Gram-negative bacteria, the majority of the sRNAs studied so far require an RNA-binding protein, Hfq, for their regulatory functions. Hfq has been shown to be important for the intracellular stability of many if not most sRNAs and necessary for promoting productive interactions with target

transcripts (Jousselin *et al.*, 2009). In contrast, the role of Hfq in Gram-positive bacteria is still unclear. Several sRNAs that have been studied in different *Bacillus*, *Staphylococcus*, and *Listeria* species do not appear to require Hfq for their stability or activity (Silvaggi *et al.*, 2005; Gaballa *et al.*, 2008; Nielsen *et al.*, 2010).

Cis-acting regulatory RNA

Cis-acting regulatory RNAs are most commonly found at the 5` leader region of mRNAs, although they can also be positioned in the inter-cistronic region of multi-gene transcripts or even at the 3`-UTR of a given transcript (Irnov *et al.*, 2006; Toledo-Arana *et al.*, 2009). To date, the most successful approach to identify *cis*-acting regulatory RNAs has been through comparative sequence analyses looking for conserved RNA secondary structure arrangements *(e.g.,* Barrick *et al.*, 2004; Weinberg *et al.*, 2007; Weinberg *et al.*, 2010). However, current bioinformatics approaches oftentimes fail to detect poorly conserved or newly evolved regulatory RNAs. Thus more recently, many researchers have also started using less-biased approaches, such as whole-genome tiling microarrays and RNA-sequencing, to identify long untranslated regions that might harbor novel *cis*-acting regulatory RNAs (*e.g.,* Livny *et al.*, 2008; Irnov *et al.*, 2010).

Cis-acting regulatory RNAs, in general, can fold into very specific structures for sensing diverse cellular signals. Typically, these RNAs consist of two parts: a ligand-binding domain and a remaining portion that is involved in harnessing the ligand-induced conformational changes for genetic control. This class of regulatory RNAs can be grouped into several categories based on the nature of their effectors: those that respond to intracellular proteins, tRNAs, small metabolites, metal ion, temperatures, and a

collection of 'orphan' regulatory RNAs with unknown ligands (Gollnick and Babitzke, 2002; Stulke, 2002; Henkin, 2008; Dambach and Winkler, 2009; Gutierrez-Preciado *et al.*, 2009; Klinkert and Narberhaus, 2009; Roth and Breaker, 2009; Ramesh and Winkler, 2010). In most cases, bacterial *cis*-acting regulatory RNAs control gene expression via transcription attenuation or translation inhibition (Winkler and Breaker, 2005). Both mechanistic strategies rely upon the specific orchestration of alternate base-pairing schemes in the expression platform (Fig. 1-3).

For transcription attenuation, a metabolic signal is received by the aptamer domain that in turn stimulates formation of a transcription termination signal, usually in the form of an intrinsic terminator helix, thus turning the gene expression "OFF". It is the interchange between terminator and antiterminator helices that dictate expression levels. In the absence of the aptamer-bound signal, an alternate antiterminator helix will be formed as the default state configuration. In other instances, the default conformation is formation of a transcription terminator helix. For these RNAs, ligand association promotes formation of an antiterminator helical element, thereby preventing terminator formation and turning the gene expression "ON", often referred to as transcription antitermination. The outcome of these mechanisms is a single "decision" between terminator and antiterminator elements during the active process of transcription. Therefore, the "choice" of forming terminator or antiterminator elements derives from precise coordination of multiple processes including transcription kinetics, the nascent RNA folding pathway, and ligand-binding kinetics. Indeed, it has been shown in at least one example that harmonization between transcription and ligand-binding kinetics is

important for gene regulation by a number of different regulatory RNAs *in vivo* (Wickiser *et al.*, 2005).

The other major mode of regulation is through translation inhibition. Similar to transcription attenuation, ligand association influences the thermodynamics interplay between helical pairings. However, rather than controlling terminator formation, the ribosome binding site (RBS) is either occluded from ribosomal access or rendered more accessible upon ligand binding. Interestingly, there is a nonrandom phylogenetic distribution for transcription attenuation and translation inhibition mechanisms in eubacterial species. For reasons not yet revealed, Gram-positive bacteria preferentially utilize transcription attenuation mechanisms, whereas Gram-negative bacteria more often than not rely upon translation inhibition.

Regulatory RNAs in Bacillus subtilis

B. subtilis has been a model organism for studying *cis*-acting regulatory RNAs. At present, more than 70 different *cis*-acting RNAs have been characterized in this organism (Irnov *et al.*, 2006). The *B. subtilis* genome encodes for at least 26 protein-sensing RNAs, 19 tRNA-sensing RNAs, 35 metabolite-sensing RNAs, 1 metal-sensing RNA, and several other *cis*-acting regulatory RNAs that respond to unknown ligands. Collectively, these RNAs regulate about 4% of *B. subtilis* genome including diverse sets of genes involved in amino acid and nucleotide biosynthesis and transport, cofactor biosynthesis, metal transport, sugar metabolism, and protein translation. In this organism, most of the well-studied *cis*-acting regulatory RNAs use transcription attenuation as the mechanism to

control gene expression, with the exception of the *glmS* ribozyme and the *eps*-associated RNA (EAR), which will be discussed in other chapters of this document.

On the other hand, the role of *trans*-acting small RNAs in *B. subtilis* biology is still underappreciated. There are 14 "non-housekeeping" sRNAs identified previously in this organism, although only a handful of sRNAs that have been studied in detail. For example, one sRNA, FsrA, regulates several iron-responsive genes (*sdhCAB*, *citB*, *yvfW*, *leuCD*), while another, SR1, controls the expression of the regulator of arginine catabolism, AhrC (Gaballa *et al.*, 2008; Heidrich *et al.*, 2006). Two antisense RNAs have been known to regulate the expression of a toxin gene (*txpA*) and an unknown gene (*yabE*), respectively (Silvaggi *et al.*, 2005; Eiamphungporn and Helmann, 2009). Deletion of *hfq* has not been previously reported to affect stability or activity of these sRNAs. Furthermore, it has been hypothesized that *B. subtilis* Hfq might not play the traditional role in sRNA biogenesis due to a shorter C-terminal region as compared to Hfq in proteobacteria; the C-terminal extension has been shown to be important for mRNA binding by *E. coli* Hfq (Vecerek *et al.*, 2008).

Regardless, recent transcriptomic studies have uncovered many more examples of potential regulatory RNAs, both *cis* and *trans*-acting RNAs (Rasmussen *et al.*, 2009; Irnov *et al.*, 2010). Although the biological functions of these RNAs are still unknown, future studies will likely reveal the full extent of the RNA-based genetic regulation strategies in *B. subtilis*.

Genetic regulation of the sugar metabolism pathways in *Bacillus subtilis*

Bacteria spend most of their time under nutrient limiting conditions in nature. In order to increase their fitness, microorganisms have to develop flexible yet efficient mechanisms to utilize various sugars as sources of carbon and energy. For example, *Bacillus subtilis*, a soil-dwelling Gram-positive bacterium, possesses many different genes necessary for breaking down and metabolizing more than 20 different sugars (Steinmetz, 1993; Reizer *et al.*, 1999). Also, specific sugar molecules are required to build polysaccharides for the cell wall and the biofilm or capsular matrix. Although many of the relevant genes are controlled at the level of transcription initiation, a surprising number of genes involved in sugar metabolism are likely to be regulated posttranscriptionally by a variety of RNA elements (Fig. 1-4).

Genetic regulation of alternative sugars catabolism genes in Bacillus subtilis

Generally, genes required for the utilization of specific alternative sugar are only expressed in the absence of the preferred substrate (usually glucose) and in the presence of the corresponding sugar. This induction can be achieved through several different mechanisms. For example, *B. subtilis xyl*, *ara*, and *gnt* operons, which encode the enzymes necessary for the utilization of xylose, arabinose, and gluconate, respectively, are repressed by specific transcription factors. The presence of the appropriate sugars relieves this inhibition by interacting with the corresponding repressor proteins to induce derepression (Fujita and Fujita, 1987; Gartner *et al.*, 1992; Mota *et al.*, 1999). In contrast, the *gut* and *lev* operons require activator proteins, GutR and LevR, for their expression. Similar to the repressor proteins, GutR is only active in the presence of the appropriate

alternative sugar, glucitol (Ye *et al.*, 1994). LevR, on the other hand, requires a fructose-specific phosphorylation event in order to be able to activate gene expression (Martin-Verstraete *et al.*, 1998).

In addition to the abovementioned transcription factor-based regulation, several sugar catabolism operons are regulated by post-initiation mechanisms that involve RNA-mediated antitermination. Generally, the 5`-UTR of these transcripts contains a *cis*-acting RNA element which overlaps an intrinsic transcription terminator stem-loop. The RNA element can fold into a specific structure that controls the formation of the terminator site. In the absence of the appropriate signal, transcription is initiated but stops at the more stable terminator stem-loop. In the presence of the appropriate signal, binding of a specific protein factor stabilizes an alternative RNA structure that prevents terminator formation and thereby allows transcription elongation to proceed into the downstream genes. *Bacillus subtilis* uses five antiterminator proteins to regulate the expression of genes required for the utilization of sucrose, β-glucosides, glycerol and also glucose (Stulke and Hillen, 2000). In most cases, the RNA-binding activity of these proteins is controlled by an inducer-specific phosphorylation event, which promotes dimerization. In the case of the GlpP antiterminator, however, RNA-binding activity is dependent on the direct binding of glycerol-3-phosphate (Glatz *et al.*, 1996). Although each protein controls different sets of genes, some cross-activation have been previously observed. For example, the SacY and SacT antiterminator proteins can recognize multiple RNA elements located in the *sacB*, *sacXY*, *sacPA*, *licS*, and *bglPH* operons, which share some structure similarities (Steinmetz *et al.*, 1989).

Posttranscriptional regulation of the glmS gene: balancing glycolysis and cell wall biosynthesis

In *B. subtilis*, RNA-mediated genetic mechanisms are also important for balancing the flux of sugar intermediates required for glycolysis, cell wall biosynthesis, and exopolysaccharide production. The GlmS enzyme catalyzes the production of glucosamine-6-phosphate (GlcN6P) from fructose-6-phosphate, which is the first step in the cell wall biosynthesis pathway. Since fructose-6-phosphate is also an important intermediate in glycolysis, the activity of the GlmS protein has to be tightly regulated to maintain a balanced flow of sugar intermediates for both pathways. In *Bacillus subtilis*, the expression of the *glmS* gene is regulated posttranscriptionally by a unique self-cleaving RNA element located at its 5`-untranslated region (Winkler *et al.*, 2004). This conserved RNA element, often referred to as the *glmS* ribozyme, has been shown to site-specifically self-cleave *in vitro* upon binding of GlcN6P. Our data revealed that self-cleavage also occurs in vivo and resulted in decreased expression of the *glmS* gene (Collins *et al.*, 2007). In fact, the *glmS* ribozyme is the only example of a natural ligand-dependent self-cleaving RNA that can regulate gene expression. Other known small self-cleaving ribozymes, such as the hammerhead and hairpin ribozymes, are thought to be constitutively active once they are synthesized (Doudna and Cech, 2002; Fedor and Williamson, 2005). However, the *glmS* RNA does not contain the typical elements required for modulating transcription termination or translation initiation as decribed previously (Winkler, 2005). Instead, the data presented in this thesis shows that the *glmS* ribozyme control gene expression by modulating mRNA stability (Collins *et al.*, 2007). This novel genetic control mechanism is dependent on the ability of the RNA to undergo

self-cleavage in the presence of GlcN6P and the activity of a novel ribonuclease, RNase J1.

Posttranscriptional regulation of the exopolysaccharide genes: synthesis of long operons

In their natural environment, many bacteria exist as a part of complex multicellular communities or biofilms. In this setting, a population of cells is typically enclosed within a protective matrix, which in *B. subtilis* consists of a protein (TasA) and sugar polymers (exopolysaccharide/EPS) (Branda *et al.*, 2001; Sutherland, 2001; Branda *et al.*, 2005; Aguilar *et al.*, 2007). Although the exact identity of *B. subtilis* EPS is still unknown, the genes responsible for the assembly, modification, transport and polymerization of these molecules have been identified (Branda *et al.*, 2001). *B. subtilis* contains 15 *eps* genes (*epsA-O*) that are arranged into one 16-kilobases long operon. The transcription of this operon has been shown to be under the control of many different transcription factors and is activated only in a subset of cells during biofilm formation (Kearns *et al.*, 2005; Chu *et al.*, 2008; Vlamakis *et al.*, 2008; Chai *et al.*, 2009; Lopez *et al.*, 2009; Winkelman *et al.*, 2009). Interestingly, there is also evidence of posttranscriptional regulation by a *cis*-acting RNA element located in the intercistronic region between *epsB* and *epsC* (Irnov and Winkler, 2010). Similar to the *glmS* ribozyme, this *eps*-associated RNA (EAR) element does not seem to have the elements necessary for controlling transcription attenuation or translation of the downstream gene. Mutation in the RNA element negatively affects the expression of distal *eps* genes, but has no effect on the expression of proximal genes (Irnov and Winkler, 2010). Our work suggests that EAR might affect the expression of *eps* genes using a processive antitermination

mechanism similar to what has been described previously for ribosomal RNA operons, the RfaH regulon in *E. coli*, and for genomic transcription of lambda and lambda-like phages. It has been hypothesized that the EAR processive antitermination system causes RNA polymerase to be resistant to downstream termination signals and allows for the complete synthesis of the unusually long *eps* operon (Irnov and Winkler, 2010).

Comparing Bacillus subtilis to Escherichia coli: the role of trans-acting RNAs?

Regulation of the sugar metabolism genes by regulatory RNAs seems to be a common theme in bacteria. The Gram-negative bacterium *Escherichia coli* uses at least five different regulatory RNAs to control the expression of sugar metabolism genes (Fig. 1-5). As in *B. subtilis*, the β-glucoside utilization operon in *E. coli* is also regulated by an RNA-protein antitermination system (Houman *et al.*, 1990). In fact, the *E. coli* BglG antiterminator protein is a homologue of *B. subtilis* LicT and also requires an inducer-specific phosphorylation event to become active (Amster-Choder and Wright, 1990). However, the other *E. coli* RNA elements that regulate sugar acquisition and catabolism, are *trans*-acting small RNAs (sRNAs) (Gorke and Vogel, 2008). One of the sRNAs, Spot42, regulates the utilization of galactose (Moller *et al.*, 2002). Another small RNA, SgrS, is involved in controlling the expression of a glucose transporter gene, *ptsG* (Vanderpool and Gottesman, 2004; Kawamoto *et al.*, 2006). Yet two other sRNAs, GlmZ and GlmY, work in a hierarchical manner to modulate *glmS* expression (Urban and Vogel, 2008). In this case, the *glmS* mRNA by itself is poorly translated due to the presence of an inhibitory stem-loop that sequesters its RBS. GlmZ RNA, when it is expressed, pairs with the 5`-region of the inhibitory stem-loop, thereby permitting

ribosome access to the RBS and activation of *glmS* translation (Kalamorz *et al.*, 2007). However, GlmZ RNA is unstable *in vivo*. The expression of GlmY, which is induced under low GlcN6P conditions, antagonizes the degradation of GlmZ, presumably through structural mimicry that titrates the degradation factor(s) away from GlmZ (Urban and Vogel, 2008).

Interestingly, both *ptsG* and *glmS* genes are also regulated by RNA elements in *B. subtilis*; however, they are controlled by *cis*-acting regulatory RNAs rather than *trans*-acting sRNAs. It is unclear why *E. coli* prefers *trans*-acting RNAs while *B. subtilis* uses *cis*-acting RNAs to control the same set of genes. Nevertheless, the fact that *trans*-acting RNAs are used to regulate sugar metabolism genes in other bacteria brings up the possibility of finding the same mechanism in *B. subtilis*. A more systematic approach is needed to thoroughly analyze the contribution of *trans*-acting RNAs in *B. subtilis* in general, and sugar metabolism pathways in particular.

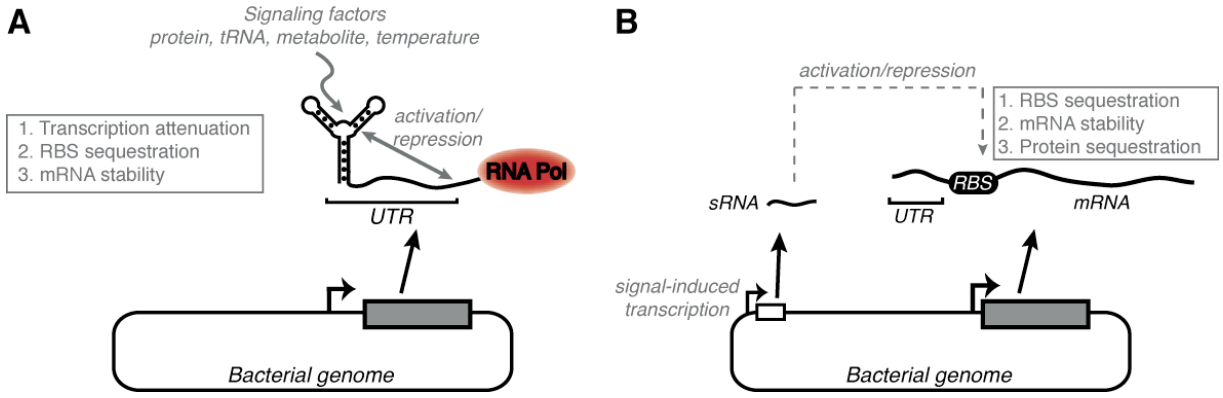


Figure 1-1. RNA-mediated genetic control mechanisms. (A) *Cis*-acting regulatory RNAs are usually located within the untranslated region (UTR) of target mRNAs. In most cases, these RNAs are capable of sensing various signaling factors to regulate the expression of proximal genes. (B) *Trans*-acting regulatory RNAs (also called 'sRNA') are transcribed as independent transcripts separate from their targets. The synthesis of these RNAs is oftentimes regulated by various cellular signals. Both types of regulatory RNAs are capable of activating and repressing gene expression.

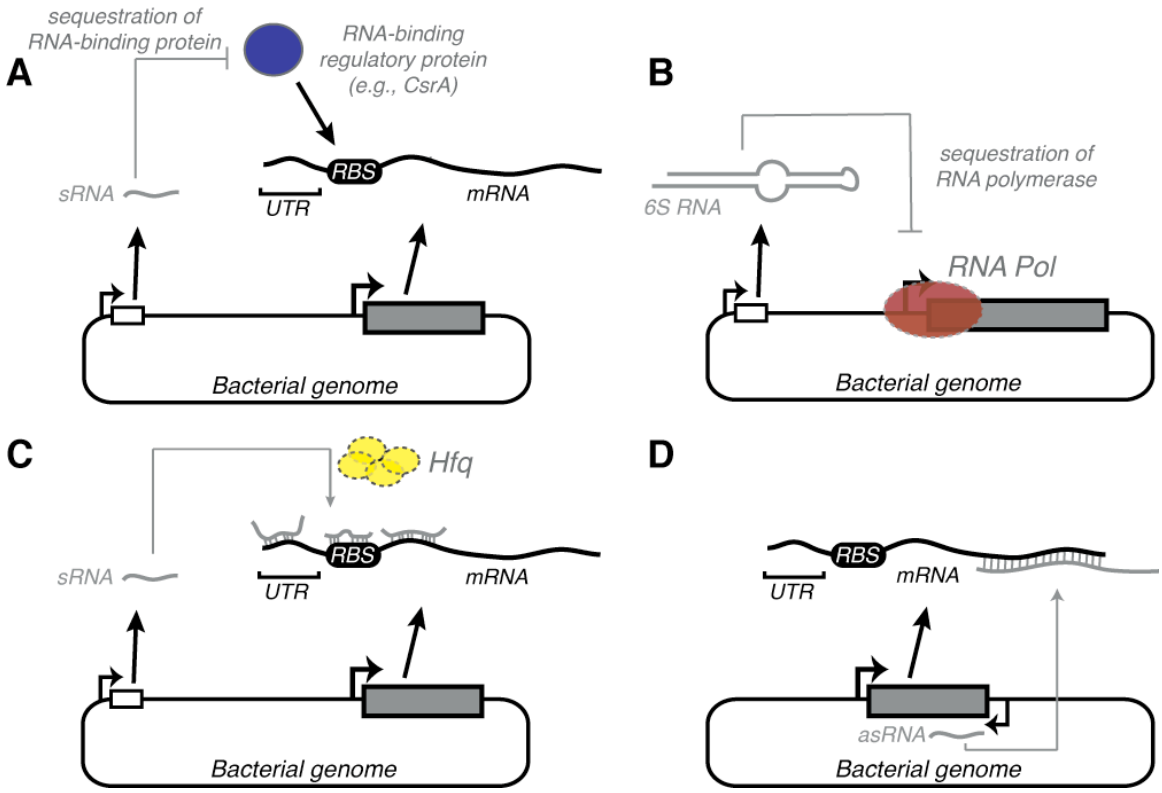


Figure 1-2. Different classes of *trans*-acting regulatory RNAs. (A-B) Example of protein-sequestering *trans*-acting RNAs. This particular class of RNAs can indirectly modulate gene expression by (A) sequestering specific RNA-binding regulatory protein normally required for genetic regulation (*e.g.*, CsrA and RsmA) or (B) direct binding to RNA polymerase (*e.g.,* under stationary phase condition), thereby preventing it from initiating transcription. (C-D) Example of *trans*-acting RNAs capable of base-pairing with its target genes. Shown in (C) is a more common example of *trans*-acting RNAs that form imperfect base-pairing with sequences within the untranslated region, near ribosomal binding site, or even within the coding region of transcripts. In most cases, a chaperone protein, Hfq, is required for stability and activity of these RNAs *in vivo*. Shown in (D) is an example of 'antisense RNA' which is typically encoded at the same genomic locus as its target, but is transcribed from the opposite DNA strand. Antisense RNAs form longer and more extensive base-pairing with its target mRNAs and do not usually require Hfq.

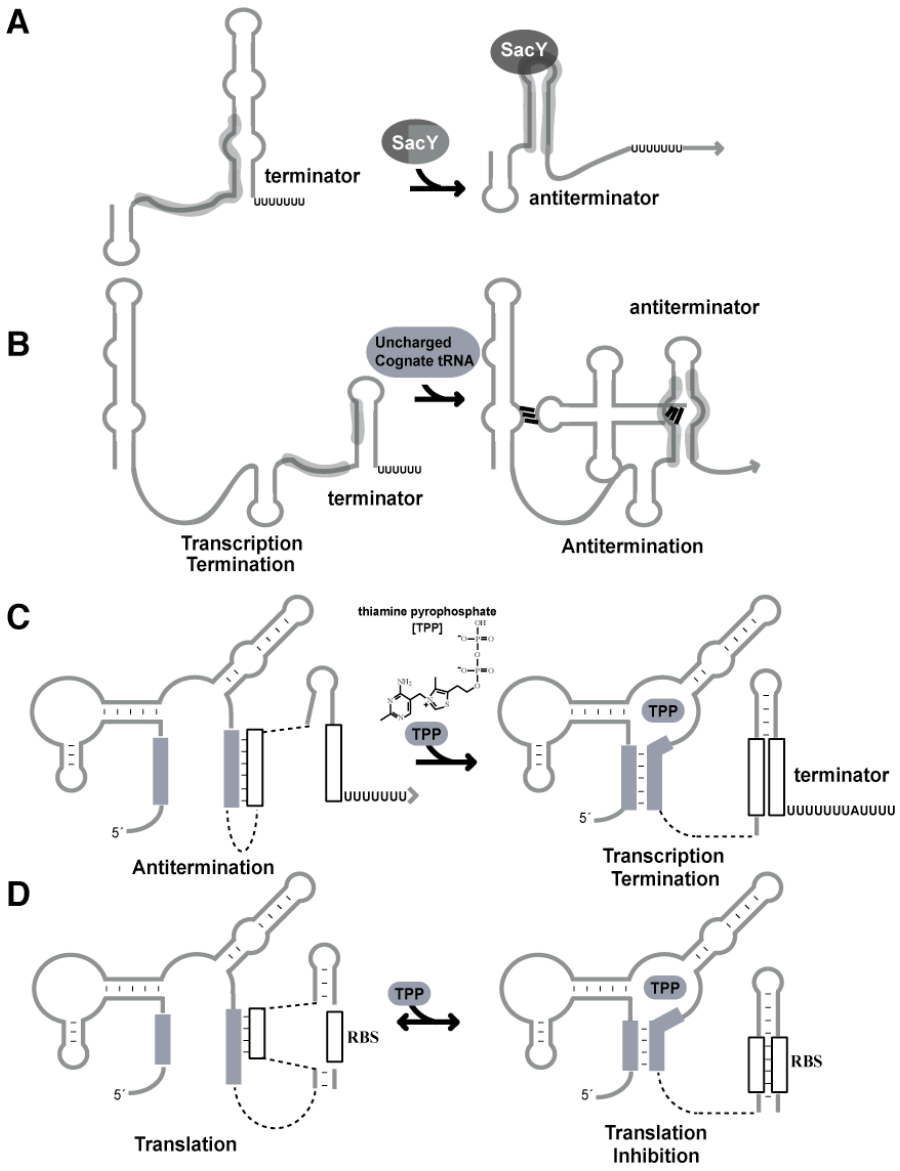


Figure 1-3. Different classes of *cis*-acting regulatory RNAs. (A) Example of a protein-responsive RNA. In the absence of sucrose, SacY is inactivated through phosphorylation by SacX. In its activated state, SacY stabilizes an antiterminator helix for the *sacB* and *sacXY* transcripts. (B) Example of tRNA-sensing RNAs. A decrease in amino acid levels alters intracellular tRNA charging ratios. Uncharged cognate tRNAs associate with the appropriate T-box RNA to promote transcription antitermination, thereby increasing expression of aminoacyl tRNA synthetases, biosynthesis genes and transporters. (C) Example of a metabolite-sensing regulatory RNA – transcription attenuation. Binding of thiamine pyrophosphate (TPP) stimulates formation of an anti-antiterminator helix, thereby allowing for formation of an intrinsic transcription terminator within the 5′ UTR of a *B. subtilis* thiamine biosynthetic cluster. D. Example of a metabolite-sensing regulatory RNA – translation inhibition. Binding of TPP stimulates formation of a helical element that sequesters the ribosome binding site, thereby reducing translation initiation efficiency for *Escherichia coli thiM* transcripts.

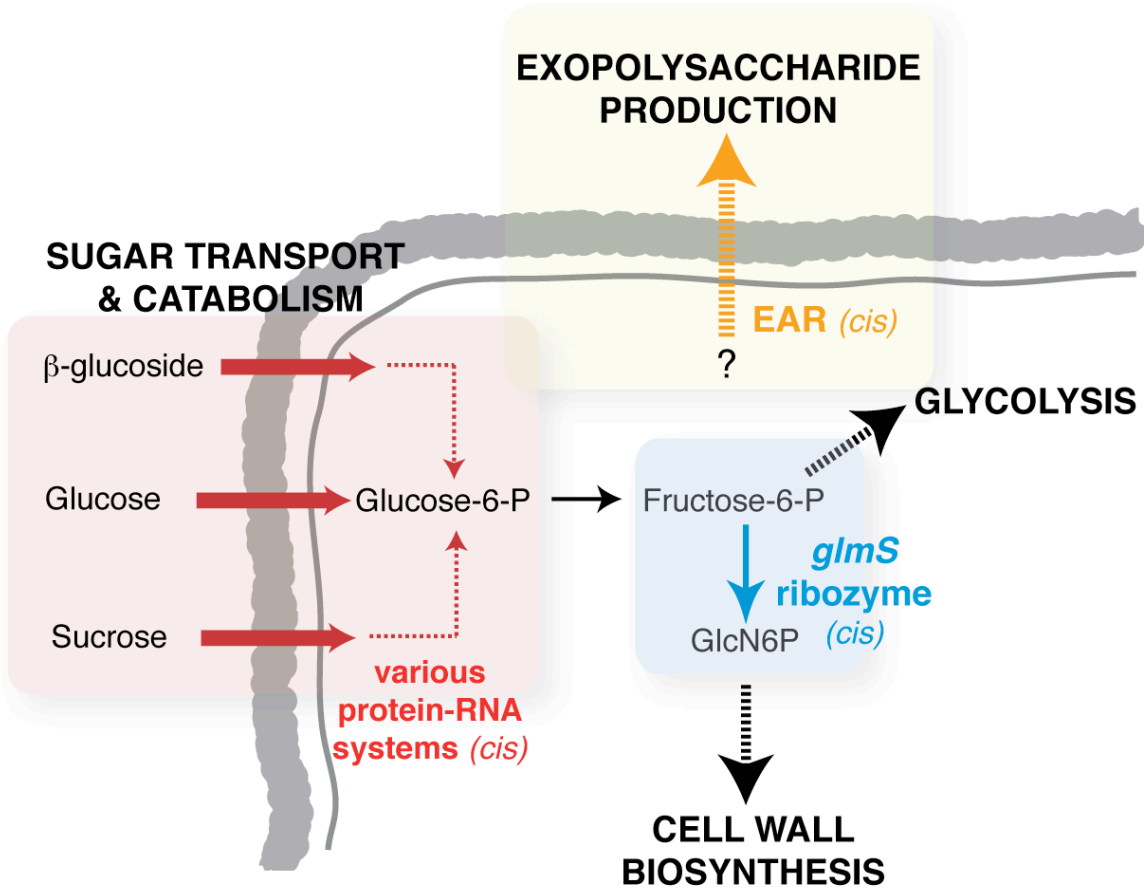


Figure 1-4. Posttranscriptional regulation of *Bacillus subtilis* sugar metabolism pathways. Various RNA elements are found to regulate different stages of sugar utilization processes in *B. subtilis*. For example, the transport and catabolism of glucose, sucrose, and β-glucoside are controlled by several RNA elements that interact with their respective protein partners (red box). Another RNA element, the *glmS* ribozyme, modulates the flux of sugar molecules required for glycolysis and cell wall biosynthesis (blue box). Also, the synthesis of extracellular polysaccharides is dependent on the *eps*-Associated RNA (EAR) element (yellow box). The identity of the sugar molecules required for exopolysaccharide productions in this organism is still unknown. It is important to note that all of these RNA elements are acting in *cis* (*i.e.*, they are located within the untranslated region of their target mRNAs). Each point of regulation is indicated with the appropriate colored arrows (*e.g.*, the *glmS* ribozyme regulates the conversion of fructose-6-P to GlcN6P, thus represented by blue arrow). Solid arrows indicate one-step process, while dashed arrows represent multi-step processes.

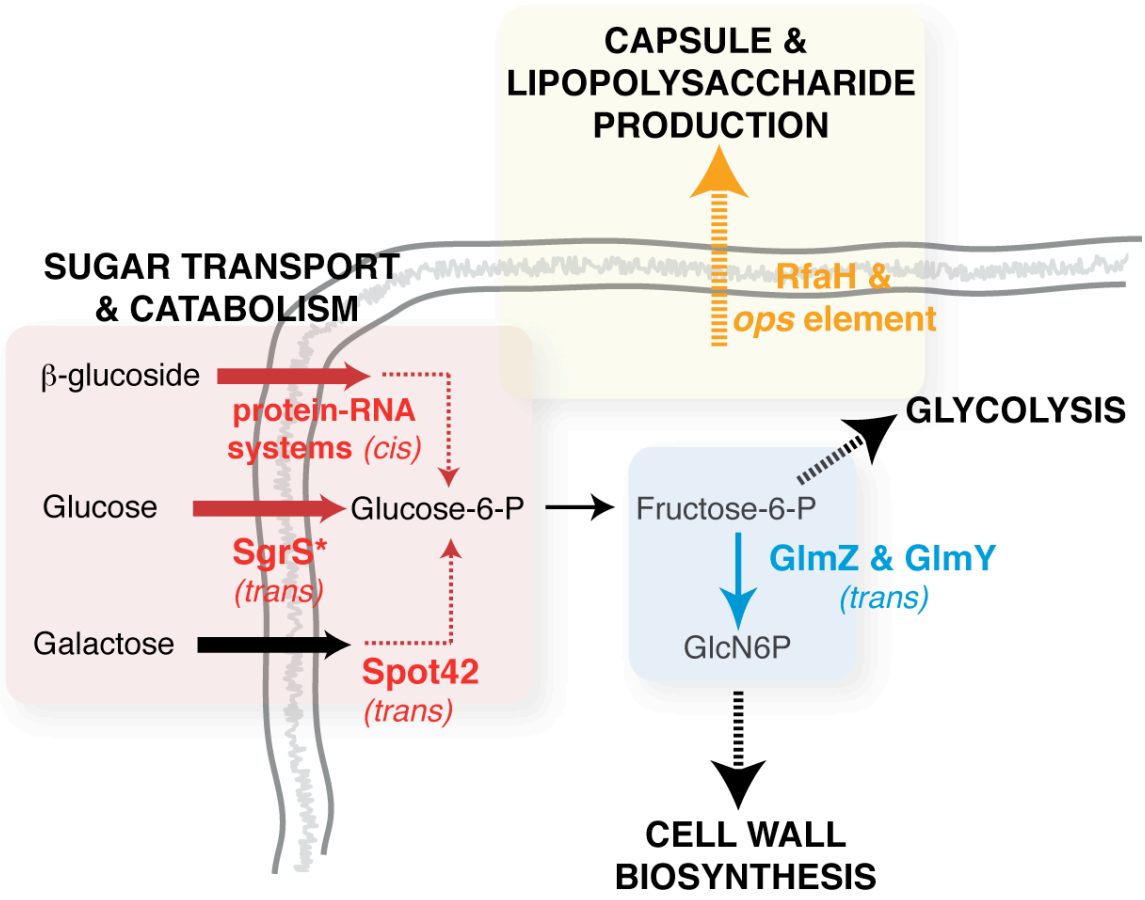


Figure 1-5. Posttranscriptional regulation of *Escherichia coli* sugar metabolism pathways. Similar to *B. subtilis*, the transport and catabolism of several sugar substrates are also controlled by multiple RNA elements in *E. coli* (red box). In this case, the utilization of β-glucosides is controlled by a protein-sensing RNA element (acting in *cis*). The utilization of glucose and galactose is regulated by two independent *trans*-acting RNAs, SgrS and Spot42. In contrast to *B. subtilis*, the synthesis of glucosamine-6-phosphate (GlcN6P) in *E. coli* is regulated by *trans*-acting RNAs (GlmZ and GlmY; blue box). Additionally, polysaccharide production is mediated by a NusG-like protein factor, RfaH, which uses a similar mechanism as the EAR element (yellow box). Each point of regulation is indicated with the appropriate colored arrows (*e.g.*, GlmZ and GlmY regulate the conversion of fructose-6-P to GlcN6P, thus represented by blue arrow). Solid arrows indicate one-step process, while dashed arrows represent multi-step processes.

CHAPTER TWO

Mechanism of genetic control by the *glmS* ribozyme

Introduction

Microbes employ a diverse assortment of *cis*-acting regulatory RNA elements, which are almost exclusively located within the 5′ untranslated region (UTR) of the gene(s) that they regulate. They have been demonstrated to function as sensors for RNA-binding proteins, unrelated RNAs, small molecule metabolites, or specific metal ions (reviewed in Winkler, 2005; Cromie *et al.*, 2006; Dann *et al.*, 2007). C*is*-acting regulatory RNAs that directly respond to intracellular metabolites in order to control expression of downstream genes are commonly referred to as riboswitches (Winkler and Breaker, 2005; Gilbert and Batey, 2006; Schwalbe *et al.*, 2007). For these RNAs, binding of the correct ligand stabilizes conformational changes within the UTR that in turn dictate the availability of nucleotides required for regulation of downstream gene expression. As a reflection of their overall importance to microbial genetic circuitry, a minimum of 4.1% of the *Bacillus subtilis* genome is predicted to be controlled by *cis*-acting regulatory RNAs (Irnov *et al.*, 2006).

Riboswitches are typically comprised of two portions: a ligand-binding region referred to as the aptamer domain and a downstream portion that includes nucleotides directly involved in genetic control. Although most metabolite-sensing RNAs inhibit downstream gene expression once the aptamer domain adopts a ligand-bound conformation, some exhibit ligand-induced gene activation (Mandal and Breaker, 2004). Binding of the appropriate ligand is typically harnessed for regulating formation of a

transcription terminator helix or structural elements near the ribosome binding site (RBS) that affect translation initiation. There is considerable precedence for control of transcription terminator formation (transcription attenuation) by regulatory RNAs (Landick and Yanofsky, 1987). Additionally, there are numerous examples of bacterial RNA elements that regulate translation initiation. However, an entirely different mechanistic strategy was proposed upon the discovery of a unique RNA element located within the 5′ leader region of the *B. subtilis glmS* gene (Fig. 2-1A), which encodes for glucosamine-6-phosphate (GlcN6P) synthase, an enzyme that catalyzes formation of GlcN6P and glutamate from fructose-6-phosphate and glutamine (Barrick *et al.*, 2004; Winkler *et al.*, 2004).

The *glmS* regulatory RNA was discovered through a bioinformatics-based search for sequence conservation within bacterial intergenic regions, an experimental approach that has also led to the discovery of several different proven metabolite-sensing riboswitches (Barrick *et al.*, 2004; Mandal *et al.*, 2004; Corbino *et al.*, 2005; Dann *et al.*, 2007; Roth *et al.*, 2007; Weinberg *et al.*, 2007). Many of the candidate *cis*-acting regulatory RNAs that emerged from these studies contained sequence elements required for transcription attenuation or control of translation initiation. The *glmS* RNA appeared to lack these features, therefore it was not immediately obvious how it would offer control over gene expression. However, *in vitro* characterization of the *glmS* UTR revealed that GlcN6P directly associated with the RNA, adding it to the list of known metabolite-sensing riboswitches (Winkler *et al.*, 2004). Surprisingly, GlcN6P binding was found to stimulate site-specific self-cleavage near the 5′ end of the transcript, demonstrating that the RNA is a metabolite-responsive ribozyme (Fig. 2-1A). Subsequent

structural analyses have revealed the three-dimensional architecture of the *glmS* ribozyme in metabolite-bound and unbound conformations and suggested structural features that are likely to participate in promoting self-cleavage (Klein and Ferre-D'Amare, 2006; Cochrane *et al.*, 2007; Klein *et al.*, 2007). These and other biochemical analyses (Hampel and Tinsley, 2006; Roth *et al.*, 2006; Tinsley *et al.*, 2007) have revealed that ligand-induced cleavage does not correlate with dramatic structural rearrangements and therefore the *glmS* ribozyme is pre-organized for ligand binding. The latter observations agree well with growing biochemical evidence suggesting that GlcN6P functions as a cofactor for the phosphodiester cleavage reaction (McCarthy *et al.*, 2005; Cochrane *et al.*, 2007).

Other autolytic ribozymes are required for processing of viral or viroid RNA genomes; the *glmS* ribozyme is therefore unique for its catalytic dependency on GlcN6P and for its presumed role in microbial gene control. Despite rapid improvements in understanding the mechanism of GlcN6P-induced self-cleavage *in vitro*, the correlation between cleavage and genetic control has not been studied. Furthermore, this relationship is not intuitively obvious. There are no candidate transcription terminators within the *glmS* 5′ UTR and the ribosome binding site is 235 nucleotides downstream from the site of self-cleavage. Adding further intrigue, GlcN6P-stimulated cleavage of the *glmS* 5′ UTR has been shown to produce RNA products with a 2′-3′ cyclic phosphate and a 5′-hydroxyl (Winkler *et al.*, 2004; Klein *et al.*, 2007), similar to other small self-cleaving ribozymes (Fedor and Williamson, 2005). Studies of mRNA turnover in *E. coli* suggest that transcripts containing a 5′- hydroxyl are actually poor substrates for degradation by the RNase E-dependent degradosome complex (Carpousis, 2002; Celesnik *et al.*, 2007).

Specifically, RNase E, which fulfills a global role in *E. coli* mRNA turnover, has been demonstrated to be a 5′-end dependent endoribonuclease that exhibits greatest affinity for transcripts containing a 5′-monophosphate and decreasing affinity for transcripts containing diphosphate, triphosphate or hydroxyl groups at the 5′-terminus (Mackie, 1998; Jiang and Belasco, 2004; Celesnik *et al.*, 2007). Based upon these observations, one would predict that self-cleavage by the *glmS* ribozyme would lead to stabilization rather than destabilization of the overall transcript, in contrast with the GlcN6P-responsive feedback repression model that has been proposed for the *glmS* ribozyme (Winkler *et al.*, 2004). For all of these reasons, we chose to investigate the relationship between ribozyme self-cleavage and *glmS* expression and to directly test whether molecular identity at the 5′ terminus is an important feature of this mechanism.

Results

The glmS UTR responds to fluctuations in intracellular GlcN6P

As a preliminary test of the importance of the *glmS* ribozyme in maintaining cell wall homeostasis, the *glmS* ribozyme was mutated such that it was incapable of self-cleavage ("M9"; Fig. 2-1A); this site-directed mutation was not accompanied by any other genomic alterations. In contrast to a wild-type strain, the M9 mutant strain was incapable of sporulation, demonstrating that over-expression of *glmS* is itself deleterious and that glucosamine pools are likely to be strictly maintained during *B. subtilis* development (Fig. 2-1B). This strain also exhibited a defect in biofilm production (data not shown), most likely due to imbalances in glucosamine precursors simultaneously required for synthesis of peptidoglycan and extracellular biofilm polysaccharides.

In order to study the *glmS* regulatory mechanism in greater detail *Bacillus subtilis* strains were engineered to exhibit controllable production of GlcN6P (Fig. 2-2A). Specifically, the *glmS* coding region was replaced through allelic exchange with a gene encoding for erythromycin resistance (*erm*). The upstream ribozyme was not deleted and was therefore transcriptionally fused to *erm* upon creation of this strain. The *glmS* coding sequence was instead positioned under control of an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoter on a self-replicating plasmid (pDG148). For some strains an additional copy of the *glmS* 5′ UTR was fused to a *lacZ* reporter gene and ectopically integrated into the genome. These strains were all dependent upon addition of IPTG for continued growth, confirming that *glmS* is an essential gene (Fig. 2-2B) (Kobayashi *et al.*, 2003). Cellular growth ceased approximately 2 hours after withdrawal of IPTG. This observation corresponded with a significant change in morphology from rod-shaped bacilli to round-shaped entities of variable size, likely to be protoplasts (Fig. 2-2B). Measurements of *glmS* transcripts by reverse transcription/quantitative real-time PCR (qPCR) demonstrated a ~7-fold decrease in abundance as extracellular IPTG was raised from 1 μM to 5 mM (data not shown). These data together confirm that the *B. subtilis glmS* gene is indeed essential for peptidoglycan biosynthesis and that intracellular GlcN6P pools can be manipulated through control of the GlmS enzyme. Furthermore, these data together argue that the *glmS* regulatory ribozyme is likely to fulfill a critically important role in maintaining cell wall homeostasis.

Posttranscriptional regulation of Bacillus subtilis glmS

In response to a decrease in GlcN6P *B. subtilis* strains containing a ribozyme-*lacZ* fusion exhibited a modest (~3-fold) but reproducible increase in reporter expression (Fig. 2-2B; compare data from cells cultured with 1 μM to those with 5 mM IPTG). This responsiveness of the *glmS* ribozyme to fluctuations in intracellular GlcN6P agrees well with prior biochemical results that indicated the ribozyme binds directly and selectively to GlcN6P (Winkler *et al.*, 2004; Klein and Ferre-D'Amare, 2006; Cochrane *et al.*, 2007). Therefore, these data support the hypothesis that the *glmS* ribozyme responds to *in vivo* GlcN6P levels for regulation of downstream gene expression. To delineate whether this regulatory influence is wielded over translation or transcript abundance different ribozyme-*lacZ* constructs were examined during conditions of high and low GlcN6P (Fig. 2-2C). A fusion of *lacZ* to the *glmS* promoter region, which was identified through primer-extension mapping of the 5′ terminus (data not shown) exhibited increased expression overall relative to a wild-type ribozyme sequence, suggesting that the ribozyme limits downstream expression. However, the promoter-*lacZ* fusion lacked GlcN6P-responsiveness, indicating that GlcN6P-responsive regulation occurs post-transcription initiation (Fig. 2-2C). A translational reporter fusion was constructed through in-frame fusion of the *glmS* N-terminus with *lacZ*. This construct demonstrated ~3-fold responsiveness to GlcN6P, identical to the transcriptional ribozyme-*lacZ* fusion. That additional GlcN6P-responsive regulation was not observed for the translational fusion relative to the transcriptional fusion argues against metabolite-responsive control over translation initiation. Site-directed mutagenesis of the ribozyme self-cleavage site from 5′-AG-3′ to 5′-CC-3′, referred to as the M9 mutant, has been previously

demonstrated to eliminate self-cleavage activity *in vitro* (Winkler *et al.*, 2004). Structural analyses of *glmS* ribozymes have revealed that these nucleotides participate in direct hydrogen bonding interactions to the GlcN6P ligand (Klein and Ferre-D'Amare, 2006; Cochrane *et al.*, 2007; Klein *et al.*, 2007), suggesting that mutagenesis of these resides is likely to eliminate the ability of GlcN6P to bind to the ribozyme. Given that the M9 ribozyme is inactive for self-cleavage *in vitro*, expression of the M9-*lacZ* fusion would be expected to correspond to the upper expression limit for the dynamic range exhibited by the regulatory ribozyme *in vivo*. Analysis of the M9-*lacZ* fusion during low and high GlcN6P demonstrated that indeed expression was significantly increased and unresponsive to fluctuations in GlcN6P (Fig. 2-2C). However, M9-*lacZ* expression was still ~8-fold higher than that of the wild-type sequence during conditions of GlcN6P deprivation, conditions when wild-type expression would be expected to be maximally engaged. This observation suggests that either intracellular glucosamine pools are not thoroughly depleted under our experimental conditions or that ribozyme self-cleavage is in part stimulated even in the absence of GlcN6P *in vivo*. It is also possible that other metabolites bearing structural resemblances to portions of GlcN6P could stimulate ribozyme self-cleavage *in vivo*.

The glmS ribozyme controls mRNA stability in response to intracellular glucosamine-6-phosphate pools

Ribozyme-containing transcripts were monitored by qPCR in order to directly measure the impact of GlcN6P variations on transcript abundance (Fig. 2-3A). Ribozyme-containing transcripts increased ~37-fold in response to GlcN6P depletion

whereas the M9 mutant exhibited almost no change. Similar to the analysis of ribozyme-*lacZ* reporters, M9 transcripts were found to be substantially higher (~13-fold) than that of wild-type transcripts during conditions of high GlcN6P (data not shown). It is not immediately clear why measurements of transcript abundance by qPCR revealed a greater fold-change upon GlcN6P depletion than with measurements of LacZ expression (*i.e.,* 37-fold versus 3-fold, respectively). The amplicons chosen for qPCR analyses consisted of approximately 100 nucleotide portions of the overall transcript(s), whereas LacZ expression could only result from full-length, translation-competent mRNAs. One potential explanation therefore is that there are mRNA degradation intermediates that are detected by qPCR analysis but that are not functional *lacZ* templates. Ribozyme-containing transcripts were further examined by Northern blot analyses via an antisense RNA probe to the *glmS* UTR (Fig. 2-3B). These tests revealed that ribozyme-*erm* transcripts were undetectable during conditions of high GlcN6P. However, GlcN6P depletion led to significant transcript accumulation. In contrast, M9-*erm* transcripts were highly expressed regardless of GlcN6P levels. To ensure that these results were general for all ribozyme-containing transcripts we monitored levels of ribozyme-*lacZ* transcripts, which were also found to be detectable only during low GlcN6P conditions.

Interestingly, in addition to observing correctly-sized full-length transcripts we also noted a smaller RNA species that hybridized well with an antisense probe to the *glmS* UTR (Fig. 2-3B,C; Fig. 2-4A). The length of this RNA species was found to be approximately ~250 nucleotides, roughly the size of the *glmS* ribozyme, with the 3` end mapped to a GC-rich stem loop located ~40 nucleotides upstream of the *glmS* start codon (Fig. 2-4B). Similar to full-length ribozyme-containing mRNAs the presence of the

isolated *glmS* UTR RNA was inversely proportional to GlcN6P levels, remaining undetectable during high GlcN6P but becoming abundant during GlcN6P depletion. These observations together suggest that the 5′ UTR becomes physically separated from *glmS* transcripts for a subpopulation of the ribozyme-containing mRNAs. Furthermore, stability of the ribozyme-containing UTR is controlled identically to the full-length transcripts, in agreement with the hypothesis that regulatory control emanates entirely from the metabolite-sensing ribozyme. We tested whether the well-studied RNA-binding protein Hfq was required for stabilization of the isolated *glmS* 5′ UTR under conditions that favored the pre-cleaved state (Fig. 2-3C). Deletion of Hfq, which is required for stabilization of many small *trans*-acting regulatory RNAs in other bacteria (Storz *et al.*, 2005; Gottesman *et al.*, 2006; Vogel and Papenfort, 2006), did not noticeably affect abundance of the *glmS* UTR species, suggesting that other factors are likely to be required for stabilization of pre-cleaved *glmS* transcripts.

As highlighted in Fig. 2-1A, there are 3 distinct regions of the *glmS* UTR. The most 5′ portion is a short 62mer region predicted to be released upon ribozyme self-cleavage. This is followed by the catalytic ribozyme domain and a nonconserved region between the ribozyme and downstream *glmS* coding sequence. Sequences corresponding to the latter region vary considerably between UTRs that contain the GlcN6P-sensing ribozyme (Barrick *et al.*, 2004; Griffiths-Jones *et al.*, 2005; data not shown). To test whether the UTR was released from full-length transcripts as a result of RNase-mediated cleavage within this region a mutant (“M11”) was constructed that was deleted for this portion (Fig. 2-1A). Indeed, when the M9 ribozyme-inactivating mutation was combined with the M11 deletion mutant, full-length transcripts were observed by Northern blotting,

but not the ~250 nucleotide *glmS* UTR fragment (Fig. 2-3C). Consistent with the 3` end mapping experiment, these data suggest that there might be a processing event within the region between the ribozyme and downstream coding portion for a subpopulation of *glmS* transcripts. However, deletion of several known endoribonucleases (*i.e.* RNase J1, RNase J2, RNase P, and RNase III) did not affect the accumulation of this RNA species (data not shown). Nevertheless, the fact that full-length transcripts are still stabilized by the M9 mutation, even in the context of the M11 deletion, suggests that the RNase-mediated release of the UTR is not likely to be an essential feature of *glmS* genetic mechanism.

The widespread ribonuclease, RNase J1, is required for selective degradation of self-cleaved transcripts

Our data demonstrate that self-cleaved *glmS* transcripts are destabilized relative to their pre-cleaved counterparts. This may be a consequence of active stabilization of pre-cleaved mRNAs by protein factors, active destabilization of the 3′ cleavage product by RNase enzymes, or a combination of both mechanisms. Degradation of bacterial mRNAs has been best studied in *E. coli* (Kushner, 2004; Deutscher, 2006; Condon, 2007) wherein the endoribonuclease enzyme, RNase E, performs a global role in initiating mRNA decay. It is still unclear what features of mRNA degradation paradigm established for *E. coli* will be generally upheld by other bacteria given that RNase E does not appear to be present in many species (Condon and Putzer, 2002; Even *et al.*, 2005). In organisms that lack RNase E, a counterpart for RNase E activity has not been conclusively identified.

Akin to many Gram-positive bacteria, the *B. subtilis* genome lacks an RNase E homolog. However, recent data have suggested that an essential ribonuclease, RNase J1,

may fulfill the functional requirements of this enzyme (Even *et al.*, 2005). RNase J1 has been reported to be an endoribonuclease that cleaves within the UTRs of tRNA-sensing *B. subtilis* regulatory RNAs (Even *et al.*, 2005). Moreover, it has also been reported to exhibit 5′ to 3′ exoribonuclease activity, an activity previously unobserved for bacteria (Mathy *et al.*, 2007). Other reports have suggested that RNase J1 is required for processing of 16S rRNA and the signal recognition particle RNA component (Britton *et al.*, 2007; Yao *et al.*, 2007). More recently, RNase J1 has been shown to interact with other ribonucleases, a helicase, and an enolase to form a multiprotein complex similar to the degradosome complex in *E. coli* (Commichau *et al.*, 2009; Mathy *et al.*, 2010). The sum of these reports hints at a broad role for RNase J1 activity in *B. subtilis* and other microbes.

To investigate a role for RNase J1 in *glmS* regulation, ribozyme-containing transcripts were analyzed in the presence of RNase J1 or after its withdrawal (Fig. 2-5A). Depletion of RNase J1 led to accumulation of ribozyme-containing transcripts, demonstrating that the enzyme is required for *glmS* degradation. In agreement with these data, expression of a ribozyme-*lacZ* fusion increased significantly upon removal of RNase J1 (Fig. 2-5B). Denaturing polyacrylamide gel electrophoresis of M9 transcripts revealed the presence of pre-cleaved transcripts during the presence and absence of RNase J1 (Fig. 2-5C). However, wild-type sequences, which were undetectable in the presence of RNase J1, accumulated in response to RNase J1 depletion, indicating that RNase J1 is specifically required for degradation of the 3′ cleavage product. Notably, the size of these transcripts diminished by approximately 60 nucleotides, roughly the same size as the 5′ cleavage fragment that was predicted to be released upon ribozyme self-

cleavage (Fig. 2-1A). These data therefore offer direct evidence for intracellular self-cleavage by the *glmS* ribozyme and suggest that the 3′ cleavage product is specifically targeted for degradation by RNase J1. Indeed, wild-type transcripts that accumulated during RNase J1 depletion were found to exhibit a 5′ terminus exactly at the site of ribozyme self-cleavage, as ascertained by 5′ primer extension analyses ('G1') (Fig. 2-5D). In total, these data indicate that RNase J1 is specifically required for degradation of *glmS* transcripts in response to ribozyme self-cleavage.

Molecular identity at the 5′ terminus is a signal for RNase J1-mediated mRNA degradation

Analyses of RNase E-mediated mRNA decay in *E. coli* have demonstrated that transcripts containing a 5′ hydroxyl are poorly degraded while those containing a 5′ monophosphate are subjected to rapid turnover (Mackie, 1998; Jiang *et al.*, 2000; Tock *et al.*, 2000; Jiang and Belasco, 2004; Celesnik *et al.*, 2007). Our data on the mechanism of *glmS* regulation therefore stand in contrast to features of the *E. coli* mRNA decay paradigm. We sought to directly investigate whether molecular identity at the 5′ terminus is important for RNase J1-mediated mRNA degradation. To validate that the 3′ cleavage product retains a 5′ hydroxyl group *in vivo*, we extracted total RNA and confirmed that the 3′ cleavage product could only ligate to an upstream DNA oligonucleotide if the total RNA sample was first phosphorylated by T4 polynucleotide kinase (Fig. 2-6). An altered *glmS* ("M12") was then constructed such that transcription initiated at the +1 G residue (Fig. 2-1; Fig. 2-7A, B), which for wild-type ribozyme sequence is rendered the 5′ terminus upon self-cleavage. Therefore, the M12 transcript is identical in sequence to the

wild-type 3′ cleavage product. However, the transcripts differ in that M12 RNAs contain a triphosphate at their 5′ terminus while wild-type transcripts exhibit a 5′-hydroxyl group upon self-cleavage. Strikingly, M12 RNAs accumulated in the presence and absence of RNase J1, in contrast to the specific accumulation of self-cleaved wild-type transcripts upon RNase J1 depletion (Fig. 2-7). These data suggest that transcripts containing a 5′-hydroxyl can be targeted for degradation by *B. subtilis* RNase J1, which agrees well with the previous *in vitro* experiments (Mathy *et al.*, 2007).

In addition to RNase J1, the *B. subtilis* genome also contains a separate ribonuclease exhibiting high sequence identity (49%) to the RNase J1 protein, which has been renamed RNase J2 (Even *et al.*, 2005). RNase J1 and RNase J2 have been demonstrated to cleave tRNA-sensing RNA elements *in vitro* at similar locations (Even *et al.*, 2005). The role of RNase J2 in *B. subtilis* mRNA degradation is unclear, although unlike RNase J1 it is not an essential gene. To investigate the potential role for RNase J2 in *glmS* regulation, wild-type and M12 transcripts were monitored during the presence and absence of both enzymes (Fig. 2-7D). M12 transcripts accumulated regardless of the presence or absence of RNase J1 and RNase J2 (compare lanes 2, 4, and 6 in Fig. 2-7D). However, depletion of only RNase J1 resulted in accumulation of the 3′ cleavage product for wild-type transcripts (compare lanes 1, 3, and 5). These experiments were conducted with strains either expressing wild-type *glmS*-*lacZ* or M12-*lacZ* fusions, but which still also retained the wild-type ribozyme sequence within the endogenous *glmS* locus. Therefore, the apparent increase in signal intensity for M12 transcripts upon RNase J1 depletion (as shown in Fig. 2-7D) can be explained by the additive effects of M12 RNAs and accumulation of wild-type self-cleaved ribozymes. In total, these data argue that

RNase J2 is not required for regulation of *glmS* stability by the GlcN6P-sensing ribozyme.

These data together predict that the addition of RNase J1 to *E. coli* would result in degradation of transcripts containing a 5′-hydroxyl. To test this hypothesis the hammerhead ribozyme, which also generates products containing 2′, 3′-cyclic phosphate and 5′-hydroxyl termini, was fused upstream of the *rpsT* transcript (Fig. 2-8A). This construct, which was previously used to demonstrate the longevity of transcripts containing a 5′ hydroxyl in *E. coli* (Celesnik *et al.*, 2007), also retains a second promoter (P2) upstream of *rpsT* but downstream of the hammerhead ribozyme. Northern blots reveal the presence of the pre-cleaved transcript, the 3′ cleavage product, and the P2 transcript (Fig. 2-8B). However, only the 3′ cleavage product is specifically destabilized upon concurrent expression of RNase J1 in *E. coli*. These data further establish RNase J1 in the molecular recognition and degradation of 5′ hydroxyl-containing transcripts, including those that are unrelated to the *glmS* ribozyme.

Discussion and Future Directions

The intracellular stability of mRNAs is intimately linked to gene expression through restriction of the amount of protein produced from individual mRNA templates. Therefore, in addition to genetic strategies that control transcription initiation or translation initiation, mechanisms that determine mRNA stability are essential in establishing intracellular expression levels. There are many factors that can potentially influence mRNA stability (Grunberg-Manago, 1999; Condon, 2003). For example, the overall degree of translation can affect mRNA stability, presumably through ribosome-

mediated protection from RNase enzymes for mRNA transcripts that are heavily translated (Arnold *et al.*, 1998; Braun *et al.*, 1998; Deana and Belasco, 2005). The phosphorylation state at the 5′ terminus also significantly influences RNA half-lives (Jiang *et al.*, 2000; Tock *et al.*, 2000; Celesnik *et al.*, 2007). Furthermore, polyadenylation of mRNA 3′ termini can have a significant impact on intracellular stability for a subset of the total mRNA population (Xu and Cohen, 1995; Li *et al.*, 1998; Dreyfus and Regnier, 2002; Kushner, 2004; Joanny *et al.*, 2007). Structural elements proximal to the 5′ or 3′ termini, such as 5′ stem-loops or 3′ intrinsic transcription terminators, can also significantly influence mRNA stability (Belasco *et al.*, 1985; McLaren *et al.*, 1991; Emory *et al.*, 1992; McDowall *et al.*, 1995; Hambraeus *et al.*, 2000; Sharp and Bechhofer, 2005). Finally, regulatory protein factors or ribosomal initiation complexes may act to specifically stabilize or destabilize mRNAs (Liu *et al.*, 1995; Glatz *et al.*, 1996; Folichon *et al.*, 2003; Sharp and Bechhofer, 2005). It is likely that the sum of many such features, acquired as the result of evolutionary selective pressure, are needed to establish appropriate expression levels. Therefore, mechanisms that alter these parameters can have profound effects upon gene expression.

A plethora of bacterial *cis*-acting regulatory RNAs have been discovered in the past several decades. The majority of these RNA elements exert their regulatory influence over the processes of transcription elongation or translation initiation. Specifically, these RNAs typically offer sensitive control over a transcription terminator helix within the 5′ UTR or structured regions near to or encompassing the ribosome binding site. Bioinformatics-searches for bacterial regulatory RNAs have uncovered numerous examples where these regulatory elements are either absent or simply difficult

to identify. This and other observations suggest that regulatory mechanisms other than transcription attenuation and SD-sequestration may still await detection.

In imagining the potential regulatory strategies employed by these heretofore unidentified RNAs, the potential control of mRNA stability via action of *cis*-acting regulatory RNAs would be predicted to provide an efficient means for affecting gene expression for the reasons outlined in the preceding paragraphs. Indeed, there are established examples of *cis*-acting regulatory RNAs that specifically interact with RNase enzymes in order to establish expression levels of downstream genes (*e.g.,* Diwa *et al.*, 2000). Also, a variety of *trans*-acting regulatory RNAs have been identified that are required for specific destabilization of target mRNAs (Storz *et al.*, 2005; Gottesman *et al.*, 2006; Vogel and Papenfort, 2006). An accruement of scientific literature in the past few decades have demonstrated that RNAs are capable of the structural sophistication required for performing autocatalytic reactions, such as self-cleavage, an activity that could be conceivably coupled to control of RNA stability. Several classes of self-cleaving RNAs have been intensively studied, including the hammerhead, hepatitis delta virus (HDV), hairpin, and Varkud satellite (VS) ribozymes (Doudna and Cech, 2002; Fedor and Williamson, 2005). None of these RNAs are employed for control of gene expression or respond to metabolic changes through alterations in their catalytic behavior. However, a wide variety of chemical-responsive, self-cleaving ribozymes have been designed through molecular engineering approaches (Breaker, 2004), demonstrating proof-in-principle that signal-responsive RNA cleavage should also be considered as a legitimate genetic strategy for biological RNAs. As detailed herein, the latter prediction was confirmed with the discovery of the *glmS* ribozyme in *B. subtilis* and other Gram-positive

bacteria, which was demonstrated to site-specifically self-cleave *in vitro* in response to association with the *glmS* metabolic product, GlcN6P (Winkler *et al.*, 2004). This observation established the seemingly simple hypothesis that signal-induced self-cleavage might be coupled to intracellular stability of the *glmS* transcript. A challenge for this mechanistic model lies in the extensive data regarding mRNA degradation pathways in *E. coli*, the primary organism employed for study of bacterial mRNA decay (Grunberg-Manago, 1999; Coburn and Mackie, 1999; Carpousis, 2002; Condon, 2003; Kushner, 2004; Deutscher, 2006). In this microbe the RNase E endoribonuclease associates with the 5′ terminus of mRNAs and cleaves internally at AU-rich regions, resulting in RNA products with 3′ hydroxyl and 5′ monophosphate termini. Since RNase E exhibits the greatest affinity for substrates containing 5′ monophosphate groups it then rapidly subjects the latter RNAs to further cleavage reactions. Products of RNase E cleavage are further degraded through the combined action of multiple 3′-5′ exoribonucleases. Based upon this logic one might expect an RNA element that entrained self-cleavage activity for genetic purposes to produce a downstream product containing a 5′ monophosphate group, thereby signaling the downstream transcript for degradation by RNase E. However, the *glmS* ribozyme has been clearly established to produce products containing 2′, 3′-cyclic phosphate and 5′ hydroxyl groups on the upstream and downstream RNAs, respectively (Winkler *et al.*, 2004; Klein and Ferre-D'Amare, 2006; Cochrane *et al.*, 2007; Klein *et al.*, 2007). RNA substrates containing 5′ hydroxyl groups in fact exhibit the poorest affinity for RNase E-mediated degradation (Jiang *et al.*, 2000; Tock *et al.*, 2000; Celesnik *et al.*, 2007). Furthermore, although RNase E has been established as a core component governing mRNA decay in *E. coli,* it is absent in many

bacterial lineages and it is still unknown how well data from *E. coli* will extrapolate to other microorganisms. The *glmS* ribozyme therefore provides a window into mRNA degradation pathways in Gram-positive bacteria and allows for investigations into the basic principles of ribozyme-mediated control of gene expression.

Our data herein reveal that the GlcN6P-sensing *B. subtilis glmS* ribozyme indeed wields its regulatory influence over mRNA stability, rather than affecting transcriptional elongation or translation initiation (Fig. 2-9). This genetic activity relies upon the specific destabilization of the 3′ cleavage product upon GlcN6P stimulation of self-cleavage. Moreover, the recently discovered RNase J1 protein is explicitly required for degradation of the 3′ cleavage product for hammerhead and *glmS* ribozymes. The fact that targeting of transcripts by RNase J1 occurs with RNAs containing a 5′ hydroxyl, but not in response to those with a 5′ triphosphate group, establishes two hypotheses. First, the RNase J1 protein is likely to behave, at least in part, similar to RNase E with regards to its reliance upon molecular identity at the 5′ terminus. Therefore, recognition of the 5′ terminus may be a rate-limiting step in mRNA turnover for bacteria in general, as opposed to being a unique requirement by *E. coli.* However, these data also suggest that there are likely to be fundamental biochemical differences between the RNase E and RNase J1 ribonucleases. Given the widespread biological distribution of RNase J in bacteria (Even *et al.*, 2005) comparative analyses of these proteins will be necessary before generalizations into bacterial mRNA turnover can be established.

However, there are still several features of the *glmS* ribozyme that remain to be investigated in order to fully understand the ribozyme-mediated gene regulation. First of all, what is the function of the cleavage event in the region between the core ribozyme

and the *glmS* coding region? We have tested most of the known endoribonucleases in *B. subtilis*, however the RNase enzyme responsible for this activity remains to be discovered. A more recently identified *B. subtilis* ribonuclease, RNase Y, has been shown to cleave several SAM-responsive riboswitches *in vivo* (Shahbabian *et al.*, 2009). Thus it is very likely that RNase Y might also be the enzyme responsible for processing *glmS* UTR from the coding region. Additionally, the functional outcome of the RNase-mediated processing is also unclear. The separated *glmS* coding transcript, while supposedly still available for translation, will probably have a shorter half-life without the protection from the structured RNA at its 5` end. Perhaps, this processing event provides a way to avoid disproportionate overexpression of the GlmS protein, especially under low GlcN6P condition when full-length *glmS* mRNA can accumulate to a high level *in vivo* (see Fig. 2-3B).

Secondly, it has been shown previously that only one nucleotide upstream of the cleavage site is required for GlcN6P-dependent ribozyme cleavage *in vitro*, albeit at a slower rate (Winkler *et al.*, 2004). *B. subtilis glmS* ribozyme contains 62 residues upstream of the cleavage site (Fig. 2-1A). In fact, *glmS* ribozymes from different Gram-positive bacteria are predicted to maintain ~30-60 nucleotides between the transcription start site and ribozyme cleavage site (data not shown). What is the role, if any, of this upstream region? While it might directly affect the kinetics of ribozyme cleavage, one interesting feature of this region is the presence of an evolutionarily conserved purine-rich region (5`-AGGAAAAGAGA-3`) that somewhat resembles a consensus ribosomal binding site (Fig. 2-1A). In some cases, for example in *B. subtilis*, the conserved motif is also part of a stem-loop structure at the 5`-end of *glmS* transcript (data not shown). Both

of these elements (ribosome binding sites and 5` stem-loops), are individually known to be positive-acting determinants for mRNA stability *in vivo* (*e.g.,* Hambraeus *et al.*, 2000; Sharp and Bechhofer, 2005). Hence, it would be interesting to look into the contribution of these elements on the stability of *glmS* mRNA *in vivo*. One can imagine that a stabilizing element might be needed to prevent 5`→3` RNA degradation process, thus allowing the *glmS* ribozyme time to be transcribed, folded, ready for ligand-sensing and gene regulation. Otherwise, the ribozyme might be degraded before it becomes active.

Overall, these studies help elucidate some of the basic processes involved in bacterial mRNA decay pathways and highlight many similarities and differences between the *B. subtilis* and the *E. coli* paradigms. In addition to shedding light on mRNA degradation mechanisms in Gram-positive bacteria, the study of the *glmS* ribozyme mechanism will assist in developing an experimental framework from which the molecular engineering of 'designer' chemical-responsive ribozymes can be built. Furthermore, the demonstration that the RNase J1 protein, which is broadly distributed in eubacterial organisms, can be selectively targeted to RNAs containing a 5′ hydroxyl terminus suggests that other signal-responsive self-cleaving regulatory ribozymes may still await discovery.

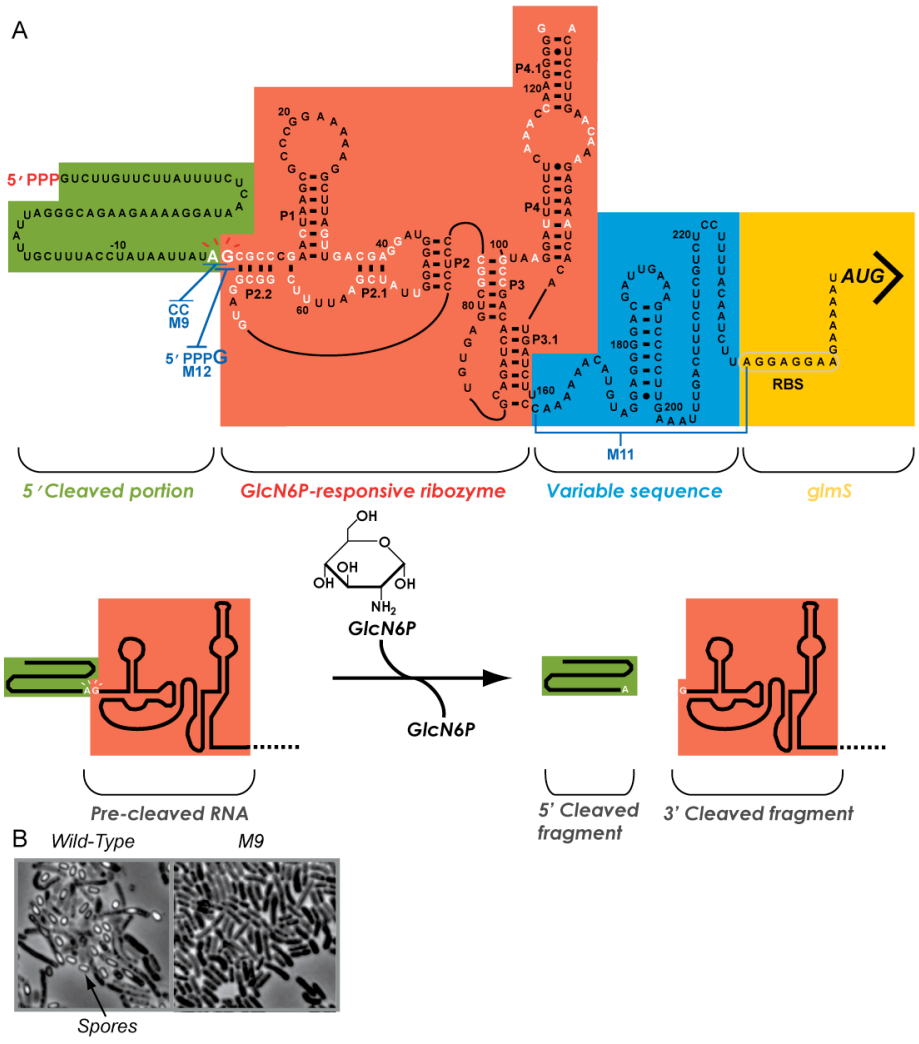
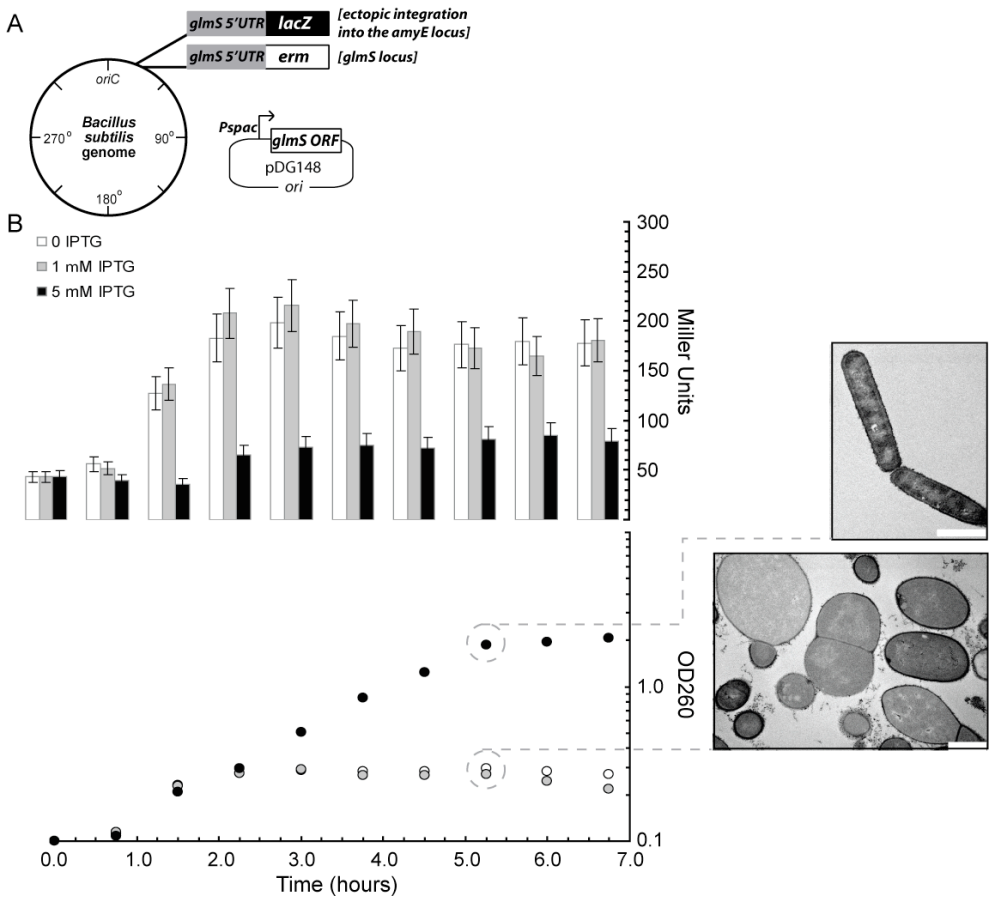


Figure 2-1. The *Bacillus subtilis* 5′ untranslated region (5′ UTR) contains a glucosamine-6-phosphate (GlcN6P)-sensing ribozyme. (*A*) The sequence of the *B. subtilis glmS* 5′ UTR is shown. Positions that exhibit greater than 90% conservation are indicated in white (Griffiths-Jones *et al.*, 2005; data not shown). This RNA element has been previously demonstrated to site-specifically cleave *in vitro* in response to GlcN6P (Winkler *et al.*, 2004). Nucleotides are numbered relative to the site of self-cleavage, with the first nucleotide of the downstream cleavage product corresponding to +1. The green shaded region denotes the portion of the sequence that is released upon self-cleavage by the ribozyme. The red shaded region denotes the region corresponding to the GlcN6P-sensing ribozyme (Winkler *et al.*, 2004). The GlcN6P-induced self-cleavage reaction is depicted in the lower cartoon. The blue shaded region denotes the nucleotide stretch located between the ribozyme and the downstream coding segment, which is variable between UTRs that contain the *glmS* ribozyme. Yellow shading denotes the downstream *glmS* ribosome binding site leading into the coding region. (*B*) A strain containing the M9 mutation (Figure 1; Winkler *et al.*, 2004) within the endogenous *glmS* 5′ UTR was assayed for its ability to form spores, a developmental program that is likely to require proper coordination of amino sugar pools. *B. subtilis* strains were induced into sporulation by the resuspension method as described elsewhere (Sterlini and Mandelstam, 1969; Nicholson and Setlow, 1990). Wild-type *B. subtilis* cells were capable of sporulation (*i.e.*, phase-bright objects) in contrast to M9, which exhibited a loss in spore formation.



C

lacZ Fusion	*'High' GlmS (5 mM IPTG)*	*'Low' GlmS (1mM IPTG)*	*Fold Difference*
Promoter	610 +/-85	410 +/-45	0.67
Wild-type Transcriptional	81 +/-13	270 +/-43	3.3
Wild-type Translational	150 +/-33	460 +/-52	3.0
M9 Transcriptional	1900 +/-190	2300 +/-250	1.2

Figure 2-2. GlcN6P-responsive regulation *in vivo* by the *B. subtilis glmS* ribozyme. (*A*) *B. subtilis* strains (all derived from strain BR151) containing inducible control of *glmS* were constructed in the following manner. The *glmS* coding sequence was placed under an IPTG-inducible promoter on a low-copy plasmid (pDG148). The chromosomal *glmS* coding region was exchanged for an *erm* cassette, thus creating a transcriptional fusion of *erm* to the *glmS* 5′ UTR. Finally, the *glmS* 5′ UTR was fused to a *lacZ* reporter (pDG1661, BGSC, Ohio) and the resulting sequences were integrated into the *amyE* gene. (*B*) These cells were cultured in glucose minimal medium and IPTG was either supplied in excess or was depleted at OD_{600} = 0.1. The growth of the strains was monitored by measurements of OD_{600} and samples were withdrawn for analysis by electron microscopy. Samples were withdrawn every 45 minutes for measurements of β-galactosidase activity. (*C*) A variety of different *glmS* 5′ UTR-*lacZ* fusions were assayed for β-galactosidase activity in strains exhibiting inducible control of GlmS. For these experiments, which were conducted identical to experimentation described in (*B*), cells were harvested for β-galactosidase activity measurements 5 hours after OD_{600} = 0.1.

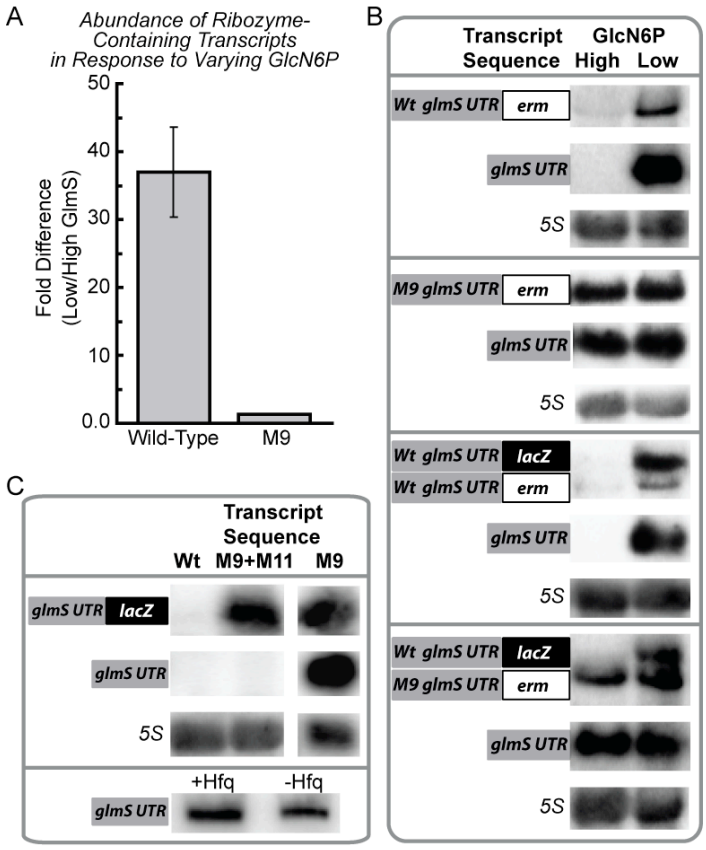


Figure 2-3. GlcN6P-induced ribozyme self-cleavage controls intracellular abundance of ribozyme-containing transcripts. (*A*) qPCR analyses of *glmS* 5′ UTR regions during conditions of high or low GlmS reveal significant differences in abundances of wild-type transcripts in response to GlmS reduction; however, virtually no change for M9 transcripts occurs under these conditions. The bacterial strains and growth conditions for this experiment, which was performed in triplicate, were identical to those described in Figure 2-2. Specifically, cells were harvested for qPCR measurements 5 hours after OD_{600} = 0.1. (*B*) Transcripts containing either wild-type or M9 *glmS* 5′ UTRs were assayed by Northern blot analyses during conditions of high and low GlmS. Antisense RNA probes, which constituted the reverse complementary sequence of the *glmS* 5′ UTR, were employed for detection of ribozyme-containing transcripts. The latter transcripts, which were resolved by 1 % agarose gel electrophoresis (see Methods for further details), were found to accumulate during conditions of GlmS deprivation. An RNA species, roughly the size of the *glmS* 5′ UTR and resolved by 2 % agarose gel electrophoresis (see Methods) was also found to accumulate during conditions of GlmS deprivation. A representative Northern blot showing simultaneous resolution of these different RNA species within the same gel is shown in Supplementary Materials. (*C*) Deletion of the 74 nucleotide intervening region between the 3′ end of the ribozyme and the downstream *glmS* ribosome binding site (M11) resulted in the absence of the smaller *glmS* 5′ UTR species, suggesting that an RNase enzyme endonucleolytically cleaves within this region. The M9 construct was also assayed by Northern blot analyses in a wild-type background strain as well as a strain deleted for Hfq, an RNA-binding protein commonly observed to stabilize small non-coding RNAs. Loss of Hfq appeared to have minimal influence on M9 stability.

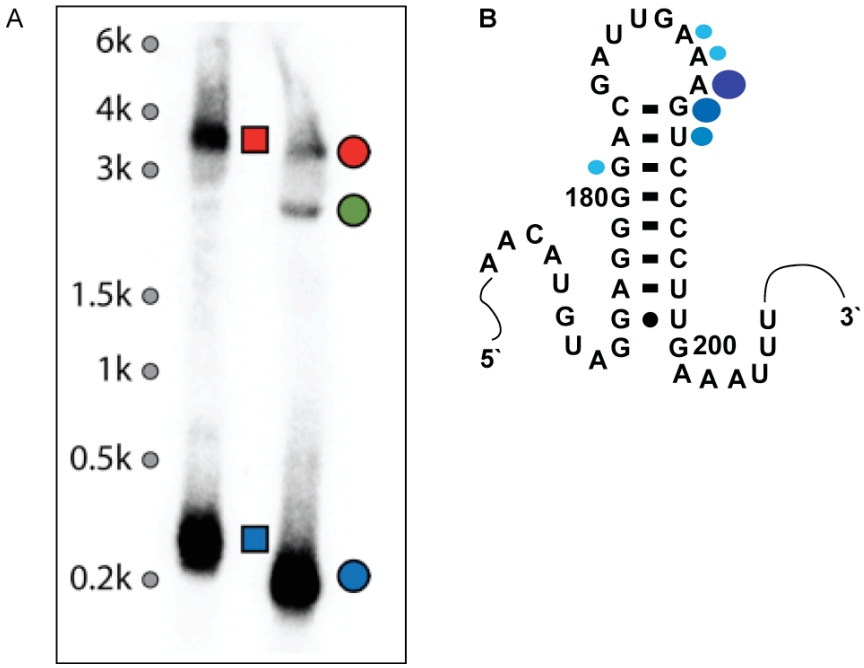


Figure 2-4. The *glmS* ribozyme also exist as a small independent transcript *in vivo*. (A) Northern blot analyses of wild-type *glmS* and M9 containing transcripts. RNA was extracted from *B. subtilis* strain BR151 containing M9-*lacZ* fusion (left lane) and GP41 containing *glmS* ribozyme-*lacZ* fusion grown in the absence of 1 % xylose (right lane). Total RNA samples were resolved by 1.5% formaldehyde denaturing agarose gel and probed with antisense RNA probes to the *glmS* UTR. M9-containing transcripts are indicated with square boxes, while transcripts containing the wild-type ribozyme are indicated with circles. Red denotes transcripts containing *lacZ*, blue denotes small RNA species which correspond to the *glmS* 5`-UTR, and green denotes the endogenous *glmS* transcripts. The migrations of RNA size standards are shown on the far left (indicated with grey circles). (B) The 3`-end of the small *glmS*-containing RNA species was mapped to within a GC-rich stem loop in the region between the core ribozyme and the *glmS* coding region (shaded in blue in Fig. 2-1). Blue circles indicated various 3`-end obtained from our sequencing. The size and color of each circle correspond to the number of times that particular nucleotide was found to be at the 3`-end of the *glmS* RNA (*i.e.*, bluer, bigger circles indicate higher occurrences of the 3`-end at that position).

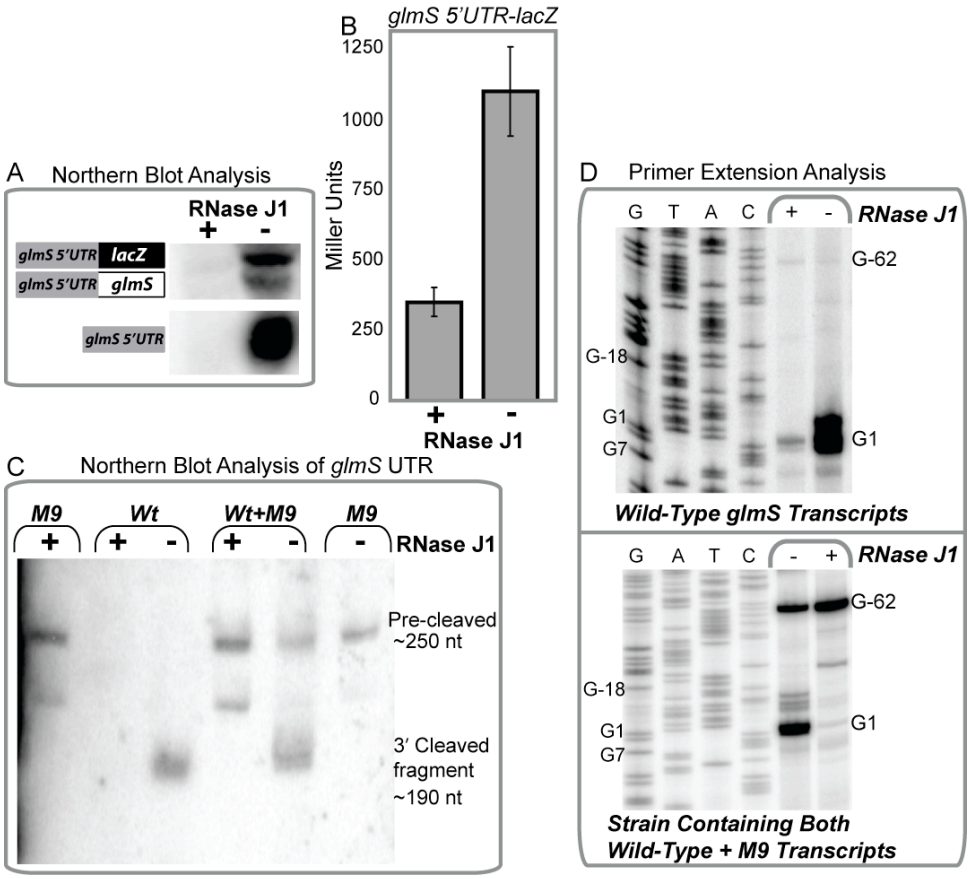


Figure 2-5. Degradation of self-cleaved transcripts is dependent upon action by RNase J1. (*A*) Ribozyme-containing transcripts were analyzed by Northern blot analyses using an antisense radiolabeled probe that could hybridize to the *glmS* 5′ UTR. RNA was extracted from strain GP41 (Jörg Stülke, University of Göttingen, Germany), which exhibits xylose-inducible control of RNase J1. For these experiments, cells were cultured in rich media (2XYT) in the presence or absence of 1.5 % xylose to mid-exponential phase at which point they were harvested for OD_{600} and experimental assays (β-galactosidase activity assays, RNA extraction). (*B*) The observation that depletion of RNase J1 led to accumulation of ribozyme-containing transcripts was further supported by measurements of activity for *lacZ* fusions to the *glmS* 5′ UTR, which exhibited an increase in expression in response to RNase J1 depletion. (*C*) Northern blotting of the *glmS* 5′ UTR when resolved by denaturing 6% PAGE revealed that wild-type transcripts accumulated during RNase J1-deprived conditions. Furthermore, the size of these RNAs was approximately 60 nucleotides smaller than that of M9 transcripts, consistent with ribozyme self-cleavage *in vivo*. (*D*) Direct support of the latter was obtained through primer extension analyses of transcripts containing the *glmS* ribozyme. Wild-type transcripts were again found to accumulate in response to withdrawal of RNase J1. However, the site of self-cleavage was found to constitute the 5′ terminus for these transcripts.

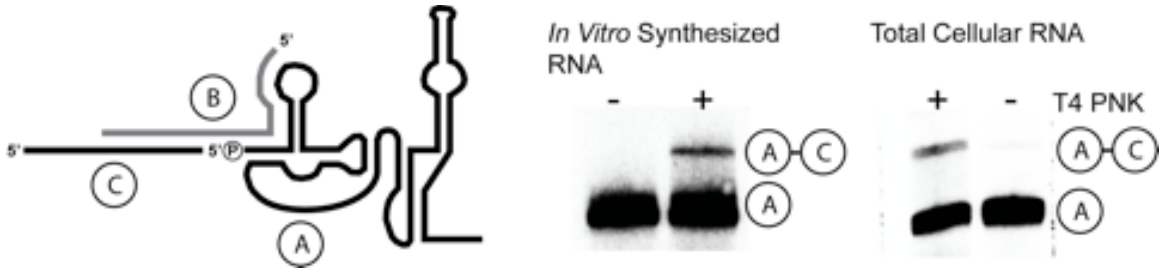


Figure 2-6. The 3`-cleaved product of the *glmS* ribozyme contains a 5`-hydroxyl *in vivo*. The 5`-phosphorylation state of the transcripts was analyzed using PABLO (Phosphorylation Assay By Ligation of Oligonucleotides) as described in Celesnik *et al.* (2007), which involves ligating a DNA oligonucleotide (denoted as C in the schematic above) to an RNA transcript containing 5`-monophosphate using a "splint" DNA (denoted as B above). The products were then analyzed by Northern blotting using an antisense RNA probe to the *glmS* ribozyme. Previous *in vitro* analysis has shown that the *glmS* ribozyme produces a 3`-cleaved product containing 5`-hydroxyl in the presence of GlcN6P (Winkler *et al.*, 2004). Thus, the ligated product (denoted as 'A-C' in the schematic above) was present only after the addition of a phosphate group by T4 PNK (middle panel). Similar pattern was observed with total RNA extracted from cells grown under RNase J1-depleted conditions (right panel).

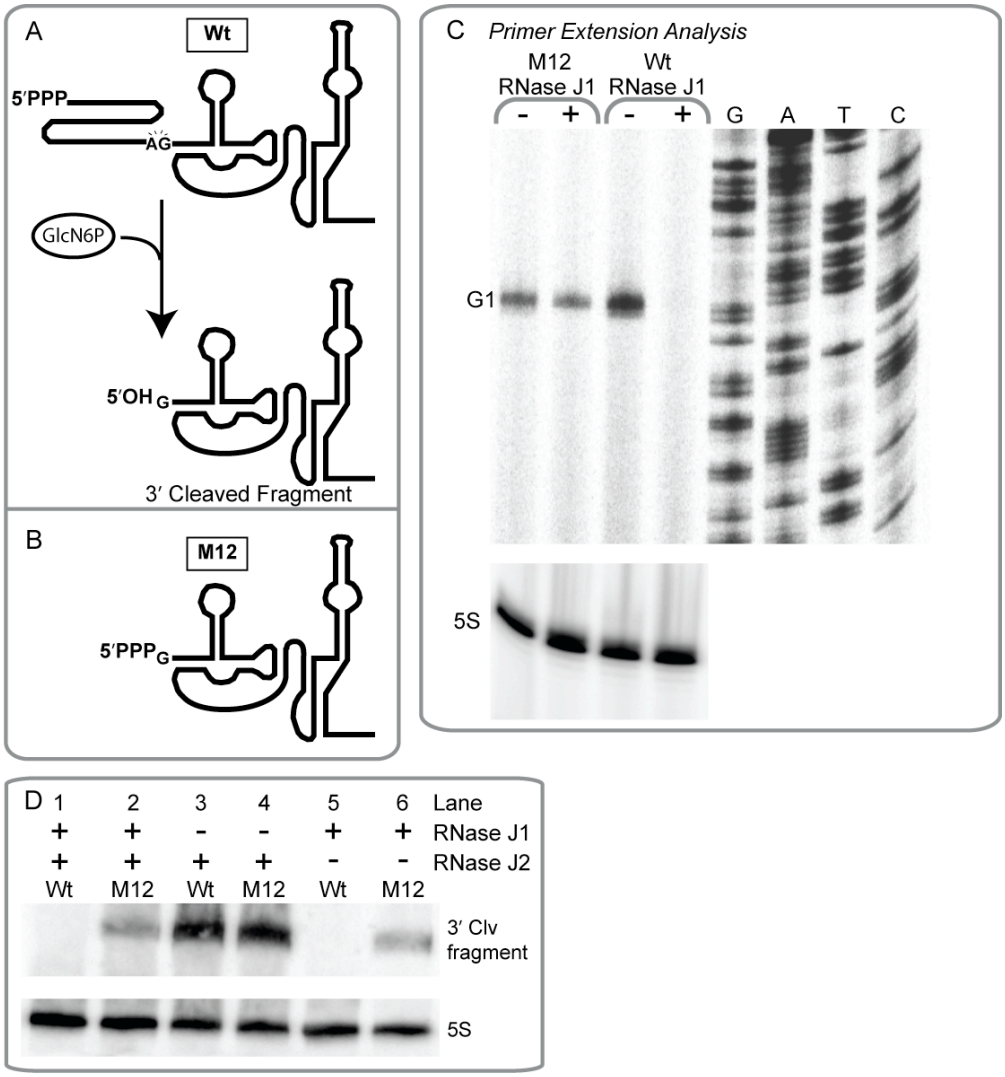


Figure 2-7. Molecular identity at the 5′ terminus of self-cleaved transcripts is important for RNase J1-mediated degradation. (*A*) A schematic highlighting molecular identity at the 5′ terminus upon ribozyme self-cleavage. Wild-type sequences contain a hydroxyl group at the 5′ terminus after self-cleavage. (*B*) In contrast, the M12 mutant was created such that transcription initiated at G+1 (the site of self-cleavage) and therefore contained a triphosphate group at the 5′ terminus (see Methods for further description). (*C*) Primer extension analyses of total RNA samples during conditions of replete or depleted RNase J1 revealed that M12 transcripts accumulated regardless of RNase J1 status whereas wild-type, hydroxyl-terminated transcripts were degraded in an RNase J1-dependent manner. For these experiments cells were cultured as described in Figure 4 and in Methods. (*D*) Depletion of RNase J1 but not RNase J2 led to accumulation of wild-type transcripts. Similarly, M12 transcripts accumulated regardless of the presence or absence of RNase J1 and RNase J2.

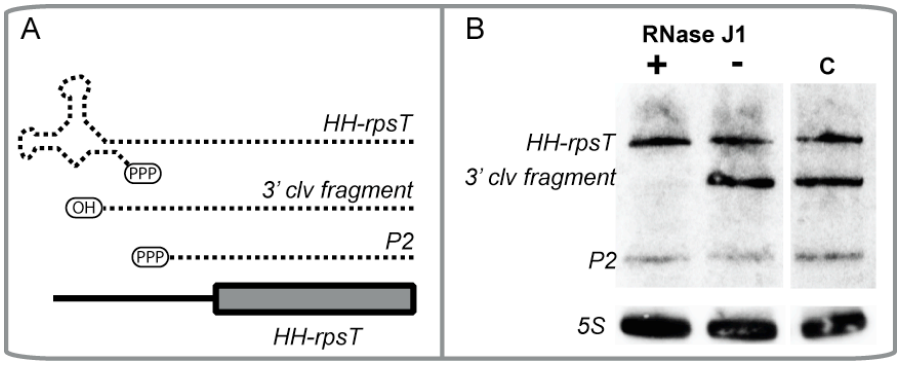


Figure 2-8. Degradation of 5′ hydroxyl-containing transcripts in *E. coli* is dependent on RNase J1. (*A*) A schematic is shown of the hammerhead-*rpsT* construct (*HH-rpsT*) used to generate transcripts with 5`-hydroxyl groups *in vivo*. Different transcripts produced by this transcriptional unit are shown in dotted lines with the corresponding 5`-phosphorylation state. The full-length mRNA containing a hammerhead ribozyme fused to *rpsT* (denoted as *HH-rpsT*) is shown at the top. The 3`-product of the ribozyme self-cleavage *in vivo* (denoted as 3` clv fragment) exhibits identical sequence to the endogenous, triphosphorylated *rpsT* P1 mRNA, except for the presence of a 5`-hydroxyl. P2 denotes the second *rpsT* mRNA which is identical to the native *rpsT* P2 transcripts. The plasmid-encoded *rpsT* transcripts contain a 19 nucleotide sequence tag within the 3`-untranslated region of the mRNA, used for hybridization of a radiolabeled probe during Northern blotting. (*B*) *rpsT* transcripts were resolved by 6 % PAGE and assayed by Northern blot analyses. For these experiments, RNA was extracted from *E. coli* strain DH5α containing arabinose-inducible RNase J1 and the hammerhead-*rpsT* construct. These cells were cultured to mid-exponential phase in the presence or absence of 0.2 % arabinose. "C" denotes a control experiment with RNA extracted from *E. coli* strain DH5α containing only the *HH-rpsT* construct. The 3` clv fragment is diminished in the presence of RNase J1 while triphosphorylated transcripts (*HH-rpsT* and P2) remain unaffected.

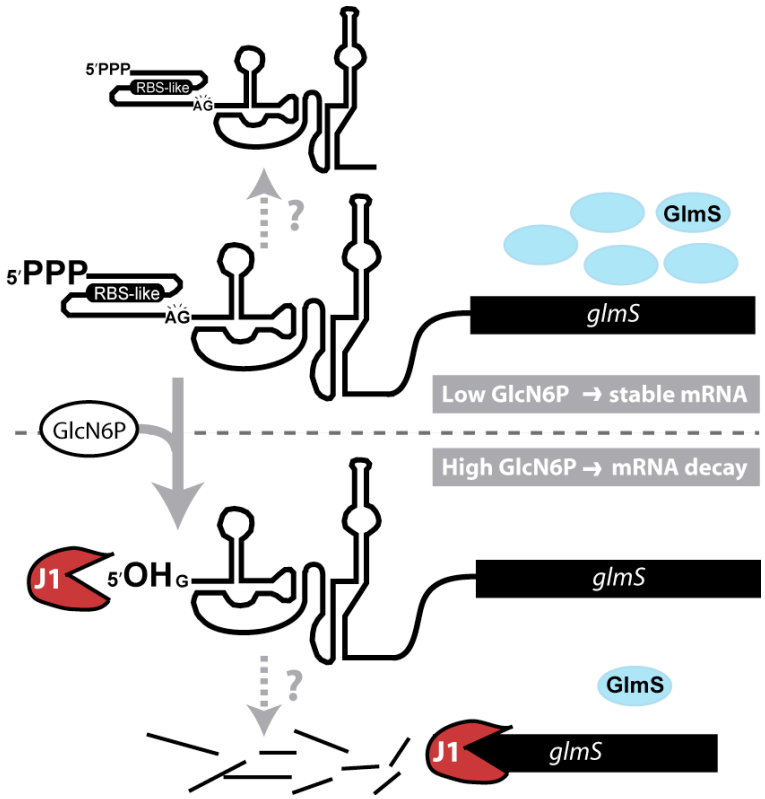


Figure 2-9. The model for the *glmS* ribozyme-mediated genetic control mechanism. Under low GlcN6P condition, the *glmS* ribozyme does not undergo self-cleavage. A combination of a 5`-triphosphate and an RBS-like motif (see Discussion) stabilizes the full-length *glmS* transcripts, leading to increased protein production (blue circles). Higher level of the GlmS proteins, in turn, catalyzes the production of more GlcN6P *in vivo*. Upon sensing GlcN6P, the *glmS* ribozyme undergoes a self-cleavage reaction producing a new RNA transcript with 5`-hydroxyl terminus. In *B. subtilis*, transcripts with 5`-hydroxyl are recognized by RNase J1 and promptly degraded. RNase J1-mediated decay of the *glmS* transcripts will eventually lead to a lower protein level and less GlcN6P inside the cells.

CHAPTER THREE

Characterization of a novel noncoding RNA required for antitermination of the exopolysaccharide operon in *Bacillus subtilis*

Introduction

Many, if not most, bacteria can form tightly associated multicellular communities called biofilms that adhere with surfaces or are formed at liquid-air interfaces. Key to formation of a bacterial biofilm is an extracellular matrix, typically comprised of exopolysaccharides (EPS), specialized proteins and occasionally DNA, which together encases the organisms and permits a surface-associated lifestyle (Branda *et al.*, 2001; Sutherland, 2001; Branda *et al.*, 2005; Aguilar *et al.*, 2007). In a *Bacillus subtilis* biofilm community, only a subset of cells is able to turn on the expression of genes necessary for production of exopolysaccharides due to a paracrine signaling pathway mediated by a cyclic lipopeptide, surfactin (Lopez *et al.*, 2009d). Other cells are likely to differentiate into distinct cell types (*e.g.,* spore formers, competent cells, motile cells, cells that produce exoproteases, *etc*) that together contribute to the survival of the community (Lopez *et al.*, 2009b). In this organism, a single operon (*eps*) encodes for synthesis of exopolysaccharides, although the individual chemical constituents of the polymer have not yet been identified. This operon, along with a separate transcript (*yqxM-sipW-tasA*) that encodes for synthesis and transport of a matrix protein, TasA, are subject to temporal and spatial control within the biofilm community (Branda *et al.*, 2006; Vlamakis *et al.*, 2008; Lopez *et al.*, 2009b). Multiple overlapping signaling pathways are coordinated for control of transcription initiation of these operons (Kearns *et al.*, 2005; Chu *et al.*, 2008; Chai *et al.*, 2009; Winkelman *et al.*, 2009). The master regulator for biofilm formation,

SinR, directly represses both operons. Under the appropriate conditions for initiating biofilm formation, SinR repression is relieved by expression of multiple antagonist proteins *(e.g*., SinI and SlrA) that sequester SinR, thereby resulting in the production of exopolysaccharide and TasA. *sinI* expression is triggered by low levels of phosphorylated Spo0A, a global regulator of gene expression that is responsive to nutrient availability, quorum sensing, and the production of secondary metabolites (Bai *et al.*, 1993; Perego and Hoch, 1996; Ratnayake-Lecamwasam *et al.*, 2001; Fujita and Losick, 2005; Kearns *et al.*, 2005; Lopez *et al.*, 2009d). SlrA is part of a minor pathway stimulating matrix production and its expression is negatively controlled by a separate factor, YwcC (Kobayashi, 2008; Chai *et al.*, 2009). A separate repressor, AbrB, also affects *eps* expression (Chu *et al.*, 2008; Hamon *et al.*, 2004) by binding directly to the promoter region (Murray *et al.*, 2009). Phosphorylated Spo0A represses AbrB expression, thereby relieving AbrB repression of *eps* under biofilm conditions.

In this chapter, we show that an additional layer of co-transcriptional genetic control is required for expression of *eps* genes. We report the discovery of a conserved RNA element located between the second and third genes of the *eps* operon, which we coin 'EAR' (*eps*-associated RNA). The EAR element is unique from other conserved RNA motifs that have been discovered (*e.g.,* Gardner *et al.*, 2009) and exhibits a conserved overall secondary structure architecture interspersed with conserved primary sequence determinants. It can be identified in other *Bacillus* species as well as a few additional genera of the order Bacillales, and is always located within biofilm or capsular polysaccharide operons. Our data demonstrate that the EAR element is required for *eps* expression in *B. subtilis*.

Bacteria are known to employ both small noncoding *trans*-acting regulatory RNAs (Gottesman *et al.*, 2006) as well as *cis*-acting regulatory RNAs for transcriptional and posttranscriptional gene regulation (Winkler, 2005). The latter are typically located in the 5' leader region of mRNAs or within intergenic regions of multigene transcripts, where they can fold into two mutually exclusive structural configurations that differentially affect downstream expression. The equilibrium between the conformational states is usually influenced by a cellular signal, such as a specific protein, unrelated RNA, small molecule metabolite, metal ions, or a change in temperature (Gollnick and Babitzke, 2002; Stulke, 2002; Henkin, 2008; Dambach and Winkler, 2009; Gutierrez-Preciado *et al.*, 2009; Klinkert and Narberhaus, 2009; Roth and Breaker, 2009; Ramesh and Winkler, 2010). Most often, *cis*-acting regulatory RNAs control expression of a proximally located downstream gene by affecting accessibility of the ribosome binding site or by influencing formation of a single adjacent transcription termination signal (Winkler, 2005), although one class of metabolite-sensing RNAs (*i.e.,* 'riboswitch') regulates mRNA stability (Collins *et al.*, 2007). We report herein that the 'EAR' element is similar with these signal-responsive *cis*-acting regulatory RNAs in its overall structural complexity, but employs an uncommon mechanism for control of transcription termination.

There are two general mechanisms for control of transcription termination. For the first class, typically referred to as 'attenuation', formation of an individual terminator signal is controlled via mutually exclusive secondary structure arrangements. The 'equilibrium' between these secondary structures is generally influenced by the presence or absence of an effector ligand. For the second class, first discovered for regulation of

bacteriophage λ genes (Roberts, 1969; Condon *et al.*, 1995; Roberts *et al.*, 2008), RNAP is modified to become resistant to downstream terminator and pause sites (Weisberg and Gottesman, 1999; Nudler and Gottesman, 2002; Greive and von Hippel, 2005; Roberts *et al.*, 2008). Since this results in read-through of multiple terminator sites over long stretches of DNA it is generally referred to as 'processive antitermination'. For these mechanisms, short DNA or RNA determinants assist assembly of multiple protein accessory factors with the transcription elongation complex that together promote read-through of termination sites (Fig. 3-1). The most thoroughly analyzed examples affect transcription of the phage λ genome and rRNA genes (*rrn*) in *E. coli*. Such mechanisms have only been poorly characterized in non-proteobacterial species. Indeed, only five processive antitermination classes have been described overall and more than a decade has passed since the last new example. Interestingly, a λ-like phage, HK022, employs a two-stem RNA motif of approximately 65 nucleotides in length ('*put*') that alone is capable of inducing antitermination activity (King *et al.*, 1996; Weisberg and Gottesman, 1999; Fig. 3-1). Also, a DNA-binding protein, RfaH, has been demonstrated to cause processive antitermination of certain operons in proteobacterial species, including genes required for lipopolysaccharide synthesis (Artsimovitch and Landick, 2002). The EAR element is responsible for a new processive antitermination mechanism class, which shares features with the latter two classes (*put* and RfaH antitermination) in that it is an RNA element that promotes antitermination of polysaccharide genes. Specifically, we find that the EAR element provides read-through of distally located termination sites within the *eps* operon and is also capable of promoting read-through of heterologous terminators. In the absence of this antitermination mechanism *eps* transcription is

truncated at intermediary points of the operon, resulting in an exopolysaccharide-deficient phenotype. Therefore, we hypothesize that a key function of this mechanism is to ensure completed synthesis of the long *eps* transcript, where each gene is likely to be required for EPS production. Given that the EAR element can be found in other species within the order Bacillales, it is likely that EAR-mediated antitermination is a general requirement for synthesis of biofilm and capsular polysaccharide operons in these microorganisms. Furthermore, we also show that the EAR element plays an important for the development of distinct cell-types within multicellular communities.

Results

Identification of a conserved RNA element located within exopolysaccharide operons

During the course of a bioinformatics-based search for *B. subtilis* intergenic regions with high G+C content, which in many cases is correlated with the presence of structured noncoding RNA (see Irnov *et al.*, 2006), we identified a particularly intriguing candidate region. It consists of a ~120-nucleotide region that can be found in a subset of Bacillales, mostly restricted to the Bacillaceae with a few examples in Paenibacillaceae (Fig. 3-2A). It does not resemble any previously established RNA motifs, such as those cataloged by the Rfam database (Gardner *et al.*, 2009). The motif is nonrandom with regards to its genomic distribution; each individual example is located within or upstream of an operon encoding for biofilm or capsular polysaccharide genes (Fig. 3-2C). When the motif is located within an intergenic region of a polysaccharide operon, it is always upstream of the biosynthesis genes but downstream of a few uncharacterized putative regulatory genes. Such is the arrangement for *B. subtilis*, where it is situated within a

~250 nt intergenic region between the 2nd and 3rd gene (*epsB-C*) of the 15-gene *eps* operon (Fig. 3-2C). For these reasons we designate the motif as the 'EAR' element, for *eps*-associated RNA.

The EAR element adopts a complex secondary structure

Comparative sequence analysis revealed many residues that are likely to be conserved at the level of primary sequence, mostly located within helical portions (Fig. 3-2). Sequence analyses also revealed many co-varying residues that when considered together provide strong evidence for the presence of base-paired segments. From these data we proposed a secondary structure arrangement consisting of five helical segments (P1-P5) and a pseudoknot at the 5' terminus (Fig. 3-2B).

To further test this secondary structure model, we performed in-line probing analyses of the *B. subtilis* EAR element *in vitro*. This structural probing assay takes advantage of the natural instability of RNA wherein single-stranded or flexible regions are more susceptible to spontaneous cleavage compared to base-paired regions (Soukup and Breaker, 1999; Nahvi *et al.*, 2002; Wakeman and Winkler, 2009). The analysis resulted in probing patterns that closely supported the secondary structural model (Fig. 3-3). One notable exception is the conserved pseudoknot (PK) region, which appeared to exhibit overall internucleotide 'flexibility' in a manner similar to bulged residues and single-stranded regions, suggesting that it is poorly formed *in vitro* (Fig. 3-3A). Also, the L1 terminal loop contains a stretch of five nucleotides (nt 24-28), adjacent to the pseudoknot, that appear to be highly conserved at the primary sequence level (Fig. 3-2). However, it is unlikely that these conserved residues participate in formation of an

intrinsic secondary structural element for RNA molecules *in vitro* given that their spontaneous cleavage frequencies resembled those of bulged or otherwise 'unstructured' residues. From this we speculate that the conserved L1 residues may be required *in vivo* for a functional role other than simple secondary structure formation, such as assisting intermolecular interactions with other cellular components. Given that the putatively pseudoknotted region appears to be required for EAR function *in vivo* (*e.g.,* M8 mutant in Fig. 3-4B), but was poorly formed *in vitro*, we speculate that its formation may also be enhanced *in vivo*, perhaps in response to additional factors. We also note from our *in vitro* structural probing data that the P3 and P4 helices appear to be modestly less structured than the P1 and P5 helices (Fig. 3-3A). Nonetheless, the probing data overall support the secondary structure as proposed by comparative sequence alignments. Finally, these data also demonstrated that mutation of the P3 helix (M3) does not perturb the global secondary structure arrangement in vitro, and instead locally alters the P3 and L3 components. Similarly, mutation of the L3 terminal loop (M4) has essentially no effect on the overall secondary structure arrangement *in vitro*, including the P3 helix (Fig. 3-3B,C). We therefore considered these particular mutants to be particularly useful (as described below) in that they reflect mutagenic alteration of conserved residues (M3) and nonconserved residues (M4), while maintaining an overall common secondary structure arrangement.

The EAR element is important for EPS production

To test whether the EAR element is functionally required for expression of downstream genes, we dissected the 126 nt region encompassing the EAR element from

the *B. subtilis epsB-C* intercistronic region. For this experiment we utilized a genetic approach that would not incorporate any additional genomic changes, such as introduction of antibiotic resistance cassettes. Also, since common laboratory strains of *B. subtilis* are unable to form biofilms (Aguilar *et al.*, 2007; Earl *et al.*, 2008), all genetic changes were introduced into the most frequently utilized 'undomesticated' strain, NCIB3610. This revealed that deletion of the EAR element resulted in decreased biofilm formation, as indicated by loss of pellicles (floating biofilm communities) and colony rugosity (Fig. 3-4A,B,C). Similarly, site-directed mutagenesis (*e.g.,* M3, M5, M8) of conserved nucleotides and helical portions also resulted in a biofilm-deficient phenotype. In contrast, mutagenesis (M4) of a nonconserved terminal loop (L3) had no effect. Compensatory mutations (M6, M7) were then introduced into the RNA element in an attempt to restore base-pairing integrity to helices that had been disrupted by mutagenesis while still maintaining an altered primary sequence. Restoration of helical integrity appeared to be sufficient to restore biofilm formation at one location (P1) but not another (P3), suggesting that a combination of secondary structure and primary sequence is likely to be important for EAR function. Scanning electron microscopy of biofilm communities (*e.g.,* Branda *et al.*, 2006) was also used to qualitatively assess whether mutation of the EAR element appeared to affect exopolysaccharide (EPS) production. Electron micrographs of wild-type and the neutral mutant, M4, revealed the presence of a coating consistent with biofilm EPS, whereas the deleterious M3 mutant exhibited an apparent lack of this material (Fig. 3-4D). Thus, the EAR element is functionally required for EPS production, presumably through control of downstream gene expression.

The EAR element exerts regulatory control over distally located genes

It is generally expected that *cis*-acting regulatory RNAs alter expression levels of an immediately downstream gene in response to specific stimuli (Irnov *et al.*, 2006). To assess the importance of the EAR element on expression of the most proximal downstream gene, *epsC*, we measured the relative abundances of *eps* genes for wild-type, ΔEAR, M3 and M4 strains (Fig. 3-5A,B). Surprisingly, mutation or deletion of the EAR element resulted in virtually unchanged levels of the most proximal downstream gene, *epsC*. Furthermore, levels of *epsD* and *epsE* remain unchanged for all strains. In contrast, more distally located genes (*epsF*, *epsH*, and *epsM*) exhibited substantially decreased abundance for ΔEAR and M3 strains but not for the neutral mutant, M4. Therefore, the EAR element affects expression levels of distally located genes within the *eps* operon, without affecting proximal genes. From these data we hypothesized that the EAR-mediated regulatory mechanism specifically affects the efficiency of transcription elongation at distal locations within the *eps* locus, most likely through an influence over transcription termination.

EAR-mediated read-through of intrinsic terminators within the eps operon

Except for rRNA antitermination, the known antitermination systems all promote read-through at intrinsic terminators (Roberts *et al.*, 2008). Furthermore, previous observation had revealed that intrinsic terminators were the dominant mode of termination in *Firmicutes* (de Hoon *et al.*, 2005; Sierro *et al.*, 2008). Assuming the EAR element is also proficient in this activity, our data suggested that the target termination sites were likely to be located between *epsE* and *epsF*, based upon the drop in expression

that occurs at this juncture for EAR mutants (Fig. 3-5). Consistent with this hypothesis, a search for intrinsic terminator hairpins (G-C rich stem loop followed by a run of U residues) within this region revealed several reasonable candidates, although we note that one of the putative terminators (‘1’) is a particularly poor candidate (Fig. 3-6; Fig. 3-7A). To test for EAR antitermination at this locus, we fused the EAR element to the upstream half of *epsF*, which was in turn fused to a *lacZ* reporter gene. We then measured LacZ activity for constructs containing wild-type or M3 EAR elements. Expression levels of *lacZ* were substantially higher (~10-fold) for constructs containing a wild-type EAR (Fig. 3-7B). The decreased *lacZ* expression that occurs upon introduction of the M3 mutation suggests that, indeed, EAR is likely to promote read-through of intrinsic terminators in *epsF*. This experiment was repeated with truncated versions of the *epsF* putative terminator cluster. Only the region that included the strongest terminator candidate (‘2’ in Fig. 3-7A) was required for antitermination activity in this assay. The very weak terminator candidate (‘1’ in Fig. 3-7A) was also just upstream of this terminator in our constructs; therefore, we cannot rule out the possibility that this terminator is also a target for EAR antitermination. However, it is also possible that this weak terminator candidate is not relevant in vivo and that an individual intrinsic terminator in *epsF* (‘2’) is a primary target for EAR antitermination.

As a preliminary assessment for whether cellular conditions may affect EAR antitermination, expression of wild-type and M3 EAR-*epsF-lacZ* reporter fusions were monitored under varying growth conditions (Fig. 3-7C). As a control for this experiment, we also measured expression of *lacZ* reporter constructs lacking the *epsF* terminator sites but containing either a wild-type or M3 EAR element. Modest antitermination activity

(2.6-fold) was observed during exponential phase growth in rich medium. The apparent overall antitermination activity notably increased in stationary phase (17-fold). In contrast, no appreciable change in expression was observed between exponential and stationary growth phases for the wild-type and M3 constructs lacking *epsF* terminator sites. When cultured in the defined minimal medium typically used to promote biofilm formation (MSgg), the apparent antitermination activity was increased during exponential phase growth (10-fold) and modestly further increased during stationary phase growth (18-fold). However, it is important to note that the increase in apparent antitermination activity was due primarily to decreased expression of the M3 construct, rather than increased expression for the wild-type construct. Therefore, these data do not necessarily prove that EAR-mediated antitermination activity is increased under these conditions but may instead suggest that intrinsic termination activity is generally improved during stationary phase or in growth in minimal medium. In any case, the inclusion of wild-type EAR sequence always promotes high level of LacZ expression despite the presence of terminator elements, further highlighting its antitermination function. Since *B. subtilis* populations are known to be heterogenous, especially in stationary phase and in minimal medium, these data also do not address whether antitermination activity might vary more significantly for certain cellular subpopulations during growth and development. It is possible that some cells (*e.g.,* EPS producing bacteria) exhibit a greater overall change in antitermination activity than detected by our reporter assays, which is masked by the averaging of *lacZ* activity during bulk culture conditions. Future experimentation at the single-cell level will be required to address these points.

EAR-mediated read-through of heterologous intrinsic terminators

A hallmark of processive antitermination mechanisms is the read-through of heterologous termination sites (Weisberg and Gottesman, 1999; Nudler and Gottesman, 2002; Roberts *et al.*, 2008). To test this possibility for the EAR element, it was placed upstream of a well-characterized, strong intrinsic terminator ('T7t'; Reynolds *et al.*, 1992) followed by *lacZ* (Fig. 3-8). Introduction of the M3 mutation into this construct lowered *lacZ* expression approximately five-fold as compared to the wild-type version. This experiment was then repeated for constructs containing instead a tandem array of three different strong, heterologous, transcription terminators ('T7t-λtR2-rrnB T1'; Artsimovitch and Landick, 2000) (Fig. 3-8). The M3 mutant again resulted in significantly decreased *lacZ* expression relative to wild-type (~30-fold), suggesting that the EAR element is capable of promoting read-through of multiple, heterologous terminators, akin to other processive antitermination systems.

EAR-mediated read-through of a heterologous rho-dependent terminator

All five of the previously described processive antitermination systems (Fig. 3-1) also promote readthrough at Rho-dependent termination sites (Weisberg and Gottesman, 1999; Roberts *et al.*, 2008). Rho is a hexameric RNA-binding protein best characterized in *E. coli,* where it is essential and is responsible for 3' end formation of most mRNAs (Richardson, 1990; Ciampi, 2006; Skordalakes and Berger, 2006). Rho binds to and 'pulls' upon the nascent transcript, destabilizing the RNAP active site. Rho is also responsible for causing operon polarity by gaining mRNA access and causing termination when ribosomes are absent. To investigate whether EAR-mediated antitermination causes

readthrough of Rho termination, a well-characterized Rho termination site ('λtR1') (Lau *et al.*, 1982) was inserted downstream of EAR and upstream of *lacZ*. LacZ activity remained roughly equal between wild-type and M3, arguing that EAR is unlikely to instigate read-through at Rho sites (Fig. 3-9A). Also, deletion of Rho in *B. subtilis* did not alter this result. As further investigation of a possible role for Rho termination/antitermination in *eps* genes, we monitored expression of *eps* genes for wild-type and *Δrho* strains (Fig. 3-9B). Deletion of Rho did not significantly impair *eps* expression or biofilm formation (Fig. 3-9C), although modestly increased expression was observed for *epsM* in the *Δrho* background. These results do not completely rule out the possibility that Rho termination could affect *eps* expression; however, our data in total suggest that intrinsic terminators are most likely to be the primary targets for EAR antitermination.

A lack of an obvious role for elongation factors in EAR antitermination

Protein accessory factors that tweak the molecular events during elongation can profoundly affect gene expression (Weisberg and Gottesman, 1999; Nudler and Gottesman, 2002; Borukhov *et al.*, 2005). In particular, several widespread classes of elongation factors, including NusA, NusG, NusB, and NusE (ribosomal protein S10), all of which participate in λ and *rrn* antitermination systems (Fig. 3-1), specifically affect elongation Borukhov *et al.*, 2005; Roberts *et al.*, 2008). *nusA* and *nusE* are essential in *B. subtilis* but *nusG* is not (Ingham *et al.*, 1999; Ingham and Furneaux, 2000), which contrasts with *E. coli*, where *nusG* is essential and *nusA* is dispensable in a *rho* mutant background. As a preliminary investigation of the potential involvement of the

nonessential elongation factors we monitored biofilm formation and *eps* expression for strains containing nonsense mutations in *nusG* or *nusB*.

NusG (Li *et al.*, 1992; Sullivan and Gottesman, 1992) may function as a bridge between RNAP and Rho (Cardinale *et al.*, 2008; Li *et al.*, 1992; Li *et al.*, 1993; Mooney *et al.*, 2009). It increases elongation by inhibiting certain pause sites (Artsimovitch and Landick, 2000; Burova *et al.*, 1995; Pasman and von Hippel, 2000) but can also enhance Rho-dependent termination and RNA release from stalled RNAP complexes (Burns *et al.*, 1998; Burns *et al.*, 1999; Nehrke and Platt, 1994). More recently, NusG has been shown to be important for coupling transcription and translation through its interaction with NusE (Burmann *et al.*, 2010). Deletion of *nusG* resulted in no obvious phenotype for the domesticated strains of *B. subtilis* (Ingham *et al.*, 1999). In contrast, deletion of *nusG* in NCIB3610 resulted in strikingly altered colony morphology, including decreased biofilm formation (Fig. 3-9C). However, this phenotype does not result from a defect in EAR-mediated antitermination of *eps* genes, as there was little change in *eps* expression in the mutant background (Fig. 3-9B). Therefore, the defect in biofilm formation for *ΔnusG* results from reasons other than a loss of EAR antitermination. NusB is an important component for both lambda and *rrn* antitermination system in *E. coli* (Sharrock *et al.*, 1985; Squires *et al.*, 1993). Together with NusE, it binds to 'boxA' region to promote antitermination activity (Mason *et al.*, 1992; Nodwell and Greenblatt, 1993; Luo *et al.*, 2008). Deletion of *nusB* in *B. subtilis*, however, resulted in no obvious biofilm deficiency (Fig. 3-9C), indicating that it too is unlikely to be required for EAR antitermination. These data suggest that EAR antitermination is likely to be mechanistically distinct from λ and *rrn* antitermination, which require both NusG and

NusB transcription elongation factors. Whether NusA and NusE play any role in EAR antitermination mechanism remains to be determined.

Investigation of EAR antitermination in vitro and within a heterologous host

In general, processive antitermination systems have not been studied in Firmicutes and other Gram-positive bacteria. Transcription elongation factors, such as the one described above, have also been only well-studied in *E. coli*. It is possible, if not likely, that some of their features and requirements will significantly vary in Gram-positive microorganisms. However, it is reasonable to hypothesize that *put* antitermination, which does not require these elongation factors, would function in a *B. subtilis* host, although it has not been directly tested. By corollary, one could assume that EAR antitermination would function in *E. coli* if, akin to *put* antitermination, no unique cellular factors were required. As a preliminary test of this hypothesis, we examined expression levels of *lacZ* reporter constructs containing wild-type and M3 EAR elements (Fig. 3-10A). Expression was nearly identical between the constructs in *E. coli*, suggesting that EAR antitermination does not function within this organism. This result may indicate the absence of a *Bacillus*-specific signaling factor within the heterologous host. However, it could also result from a missing or incompatible transcription cofactor. For example, *B. subtilis* RNA polymerase has been shown to interact with delta factor (RpoE), which is found exclusively in Gram-positive bacteria (Lopez de Saro *et al.*, 1999; Doherty *et al.*, 2010). Finally, it is also possible that the failure of EAR antitermination within *E. coli* results from key structural differences between the bacterial RNA polymerase enzymes. Assuming that the EAR element directly associates with

RNA polymerase, akin to the *put* antitermination element, it is possible that minor structural differences at this RNA-protein interface could disrupt EAR function.

To further test whether the EAR element alone is sufficient for antitermination activity, as in the case for the *put* RNA antitermination system, we assessed terminator read-through activity *in vitro*. Specifically, we performed *in vitro* transcription reactions that included purified *B. subtilis* RNAP, nucleotides, and PCR-generated DNA templates containing the EAR element followed by intrinsic termination sites (Fig. 3-10B). We examined the degree of run-off transcription (*i.e.,* corresponding to read-through of the intrinsic termination sites) for wild-type and M3 constructs. Using single-round transcription assays, we observed essentially no differences in the amount of run-off transcription between wild-type and M3. In combination with the lack of an obvious role for the traditional transcription elongation factors, this data suggests that EAR antitermination is mechanistically distinct from the *put* antitermination mechanism and is likely to require a novel cellular co-factor.

EAR is important for spore formation within biofilms

Under severe nutrient starvation, *B. subtilis* is capable of differentiating into spores, dormant cells that are resistant to many environmental stresses. Spore formation is an elaborate process that requires substantial changes in gene expression and uses enormous amount of energy (reviewed in Piggot and Hilbert, 2004). In a biofilm, only a small number of cells undergo sporulation, which takes place preferentially at the vertical extensions called fruiting bodies (Branda *et al.*, 2001). Interestingly, the ability to produce the extracellular biofilm matrix has been shown to affect sporulation within the

biofilm community. Specifically, a subpopulation of the 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). Similar results were observed upon deletion of *tasA*, which encodes for the matrix protein. From these results it was hypothesized that TasA and EPS production are so metabolically expensive that nutrient levels within the microniche that surrounds matrix producers plunge to very low levels. Presumably, these nutrient levels become decreased to such an extent that endospore formation is triggered (Vlamakis *et al.*, 2008). Perhaps the reason that a similar drop in sporulation does not occur for liquid cultures is due to the constant shaking of the culture medium, which is likely to prevent formation of microenvironments that are selectively nutrient-poor.

Consistent with the role of EAR in ensuring efficient EPS production, biofilm colonies formed by the ΔEAR strain were found to contain ~ten-fold fewer spores as compared to the wild-type biofilm community (Fig. 3-11). Furthermore this difference in sporulation frequency was not observed when both cells were grown in MSgg liquid culture (Fig. 3-11). Given the fact that EAR deletion does not affect the expression of *epsA-E* (Fig. 3-5), our results indicate that the genes downstream of *epsF* might be the limiting steps for efficient EPS production. Indeed, homology searches indicate that *epsF-O* are likely to encode for proteins that polymerize, transport, and assemble exopolysaccharide building blocks. In contrast, *epsA-E* are likely to encode for genes with regulatory roles.

Inactivation of the EAR element causes specific changes in the gene expression profiles of biofilm cells

To better understand the underlying mechanism by which EAR-mediated EPS production affects the formation of multicellular communities, we performed a microarray analysis of cells harvested from NCIB360 and ΔEAR biofilms. The typical *B. subtilis* biofilm community consists of multiple distinct cellular sub-populations, including but not limited to: motile cells, matrix producers, endospore-forming cells, competent cells (proficient for DNA uptake and recombination), and toxin-producing "cannibals" (reviewed Lopez *et al.*, 2009b). Microarray-based transcriptomic analysis of wild-type and ΔEAR biofilms revealed considerable differences in gene expression. Specifically, 124 genes were upregulated by at least 3-fold and 651 genes were similarly downregulated for the ΔEAR strain as compared to wild-type (Fig. 3-12; Fig. 3-13). As expected, our analysis revealed a marked and specific decrease in expression of *epsF-O* genes for the ΔEAR strain, further supporting a regulatory relationship between the EAR element and expression of the second half of the *eps* operon (Fig. 3-5C). Interestingly, the analysis also showed a modest increase in expression of *epsA-E* genes and a more significant increase (~8-fold) for the genes involved in the production of matrix protein (*i.e., yqxM*, *sipW*, *tasA*) (Table 3-1; Appendix I). While the molecular basis for this upregulation remains to be identified, it is also still unclear whether this result represents a specific increase in the total number of matrix-producers in the biofilm or a uniform increase in the expression of abovementioned genes in all cells.

Although the upregulated genes encompassed many different categories (Fig. 3-12), higher expression levels were detected for a variety of genes involved in amino acid metabolism, energy production, and protein translation. This profile might indicate a higher availability of nutrients in ΔEAR biofilm, perhaps due to the defect in EPS production.

A moderate increase in expression was observed for several genes involved in secondary metabolite biosynthesis (*ppsA*, *ppsD*, and *dhbB*). In *B. subtilis*, genes involved in secondary metabolites biosynthesis are arranged in several long operons (*e.g., ppsA-E* is ~38 kb and *dhbA-F* is ~12 kb); each operon is responsible for the synthesis of unique metabolite. Although only several genes within these secondary metabolite pathways exhibited more than 3-fold higher expressions for ΔEAR, the remaining genes in these operons were still moderately increased (~2-fold) (Appendix I). To date, not much is known regarding the functional roles of secondary metabolites during biofilm formation. One of the best characterized *B. subtilis* secondary metabolite, surfactin (encoded by *srf* operon), has recently been shown to stimulate extracellular matrix production by inducing potassium leakage (Lopez *et al.*, 2009a). However, the relative transcript abundances of genes responsible for surfactin production did not change under our experimental conditions (Table 3-1; Appendix I)

We also observed a moderate (2-3 fold) increase in the expression of *sdpABC* (previously annotated as *yvaWXY*) (Table 3-1). The *sdp* operon encodes for one of the two toxic peptides that are important for cannibalism by *B. subtilis* (Gonzalez-Pastor *et al.*, 2003). Intriguingly, the transcript abundance of the other toxin, Skf (encoded by *skfA-H*; previously annotated as *ybcOPST-ybdABDE*), did not change for the ΔEAR strain as

compared to wild-type (Table 3-1). A recent report shows that both of these cannibalism toxins are produced by a subpopulation of the cells that also produce extracellular matrix components (Lopez *et al.*, 2009c). However, our bulk transcriptomic analysis has suggested that their production may, in fact, be more nuanced within the biofilm community.

Similarly, the downregulated genes also consisted of a variety of different gene function categories (Fig. 3-13). However, a closer inspection revealed that a majority of the downregulated genes are predicted to be involved in the developmental pathway for endospore formation (data not shown). This is in agreement with the strikingly decreased population of spores that we observed for the ΔEAR biofilm community (Fig. 3-11). In *B. subtilis*, the endospore developmental pathway is regulated by the transcription factor Spo0A (reviewed in Hoch, 1993; Sonenshein, 2000). A high-level of phosphorylated Spo0A (Spo0A~P), induced by severe nutrient deprivation, has been shown to influence expression of ~500 genes, including a cascade of sigma factors that control expression of genes required for spore formation in a compartment-specific manner. The ΔEAR strain exhibited significantly decreased levels of the early sporulation sigma factors, SigE and SigF. This observation is likely to provide an explanation for the decreased expression of the other sporulation-related genes (Table 3-2).

Reduced expression levels were also observed for two genes encoding extracellular sigma factors, *sigY* and *sigV* (Table 3-2). SigY is known thought to regulate its own operon, a toxic peptide and its transporter, and a gene (*ybgB*) that encodes for a potential immunity protein (Cao & Helmann, 2003; Mascher *et al.*, 2007). It has also been reported that *sigY* expression is induced by nitrogen starvation (Tojo *et al.*, 2003).

SigV, on the other hand, is expressed during germination (Horsburgh *et al.*, 2001) and regulates the expression of genes involved in the cellular responses to high salt, ethanol, acid, and cell-wall targeting antibiotics (Zellmeier *et al.*, 2005). Deletion of both *sigY* and *sigV* does not affect biofilm formation (Mascher *et al.*, 2007).

We also observed decreased expression for several transcription factors controlling purine catabolism (*pucR*), fatty acid degradation (*fadR*), acetoin catabolism (*acoR*), gluconate utilization (*gntR*), and nitrogen metabolism (*tnrA*) (Table 3-2). Consistent with the reduced expression of these regulators, lower expression levels were also observed for the majority of genes known to be controlled by the abovementioned transcription factors. It is interesting to note that both *pucR* and *tnrA* are usually activated only under nitrogen-limited growth condition (*e.g.,* growth in media containing glutamate as the sole nitrogen source - Fisher, 1999; Beier *et al.*, 2002). Therefore, the reduced expression of both regulators (and also *sigY*) suggests that the ΔEAR biofilm might be more proficient in nitrogen stores than the wild-type biofilm.

The fact that the EAR deletion affected the expression of genes typically turned on in spores and matrix-producers/cannibals caused us to question whether the lack of EPS production also influenced expression of genes required for other cell types. Indeed, decreased expression was observed for *comK*, a known marker for competent cells (*e.g.,* Maamar *et al.*, 2007; Suel *et al.*, 2007) (Table 3-1). However, some but not all of ComK target genes showed similarly decreased abundance. For example, several known ComK targets, such as *comC*, *comEA*, *comER*, *comGA* and *yvzC*, exhibited reduced expression when the EAR element was deleted, while others did not show significant changes (Appendix I). It is possible that the competence genes were expressed at a very low level

under our culture conditions, thus making it difficult to get accurate measurements of gene expression. This is, in fact, consistent with the low rate of competence observed for strain NCIB3610 (data not shown). Another specialized cellular sub-population, called 'miners' consists of cells that function as dedicated producers and exporters of extracellular proteases (Lopez *et al.*, 2009b). We observed decreased expression for several genes encoding extracellular proteases (*i.e., aprE* and *bpr*) in the ΔEAR strain. The expressions of both genes were 4-5-fold lower in ΔEAR strain compared to wild-type (Table 3-1). In contrast, expression levels were largely unchanged for *hag* and *srfAA-AD*, which are the hallmark genes for motile cells and surfactin-producing cells, respectively (Table 3-1). However, from our transcriptomic data alone, we cannot exclude the possibility that the different cell types are also regulated through translational or posttranslational mechanisms.

The formation of distinct cell types within the biofilm proceeds through several distinct pathways and is governed by multiple master regulators (Lopez and Kolter, 2010). For example, sporulation, matrix production, and cannibalism are coordinated by Spo0A. Production of various exoproteases has been shown to be under the control of DegU. Finally, competence and surfactin production are regulated by yet another regulator, ComA. The activity of these master regulators is dependent upon phosphorylation by their cognate sensor kinases (*i.e., KinA-E*, *DegS*, *ComP*, respectively) (Hoch, 1993; Kunst *et al.*, 1994; Lopez and Kolter, 2010). Consistent with this post-translational mechanism, our data showed very little changes in the transcript abundance of these master regulators in both wild-type and ΔEAR biofilms, despite the considerable decrease in the expression of their downstream targets (Table 3-2). Furthermore, this

suggests that the lack of cellular differentiation seen in ΔEAR biofilms is caused not by the reduced expression of each master regulator, but perhaps due to the lack of signals that are required to activate each developmental pathway. Further experiments will be required to confirm this hypothesis.

In addition to our analysis of the in ΔEAR transcriptome, we also performed a high-throughput RNA sequencing (RNA-Seq) experiment to extend the range of our study. For this experiment, the M3 mutant strain was used rather than ΔEAR because it contains very minimal genomic changes from the wild-type NCIB3610 strain (see Fig. 3-4). Total RNA was extracted from the biofilm community and converted to cDNA with the appropriate oligonucleotide linkers and sequenced using an Illumina GAXII genome analyzer instrument. This analysis yielded approximately ~24 million cDNA reads with an average length of 36 nucleotides for each dataset (NCIB3610 and M3) that could be successfully mapped to *B. subtilis* genome (see Appendix I for more details on data analysis). This RNA-Seq analysis revealed 141 genes that exhibited higher expression and 653 genes exhibiting lowered expression (by at least 3-fold) in the M3 strain as compared to wild-type (data not shown), which is similar in magnitude to the number of gene changes observed in the microarray analysis. Moreover, the amplitudes of gene expression changes between the two datasets exhibited a reasonable correlation (Pearson coefficient = 0.79; Fig. 3-14A). Comparing both datasets, our analysis showed at least 60% overlap for genes with lower expression in the EAR mutants (based on a minimum 3-fold change). On the other hand, only about 15% of the genes detected with higher expression in the microarray analysis were confirmed by RNA-Seq (Fig. 3-14B). It is important to note that the Affymetrix microarray used here was designed and analyzed

based on the older version of *B. subtilis* genome (NC_000964.2; released in 2004), while the RNA-Seq experiment was analyzed using the latest genomic sequence (NC_000964.3; released in 2010). Hence, the modest overlaps seen here might be due to different gene annotations between the two sequences (for instance, name changes and discoveries of new genes). It is also possible that some of the discrepancies are due to the inherent differences between ΔEAR and M3 strains, which were not detected in our previous assays (see above). Regardless, the overall expression changes observed for various transcription factors and cell-type specific markers described above were confirmed by both approaches (Table 3-1; 3-2).

Discovery of biofilm-specific small RNAs

In recent years RNA-Seq has been successfully used to find novel regulatory RNAs in different organisms (*e.g.*, Sittka *et al.*, 2008; Liu *et al.*, 2009; Sharma *et al.*, 2010). In contrast to oligonucleotide microarray analyses, RNA-Seq is unbiased and does not rely on prior sequence information. Additionally, it also provides single-nucleotide resolution that would enable a more accurate determination of the 5` and 3` termini of various transcripts *in vivo* (reviewed in Wang *et al.*, 2009). However, most studies focus only on conditions where cells are mostly homogeneous (*i.e.*, during exponential growth phase). As stated above, *B. subtilis* cells differentiate into several distinct cell types during biofilm formation. We hypothesized that regulatory RNAs in bacteria would be specifically important for such complex developmental pathways, similar to the role of microRNAs in eukaryotic cell differentiation (*e.g.,* Chen *et al.*, 2004; Xiao *et al.*, 2007; Johnnidis *et al.*, 2008; Liu *et al.*, 2008; Xu *et al.*, 2009a).

To find RNA elements that are expressed only during biofilm formation, we took advantage of the fact that our EPS deficient strain appeared to exhibit decreased expression of genes corresponding to the development of multiple cell-types. We searched the mRNA-Seq data for transcription units of at least 50 nucleotides in length that were located exclusively within intergenic regions, and that were represented by at least 5 cDNA reads. We analyzed the data for peaks that fit these criteria and that exhibited higher expression in the wild-type biofilm compared to the EAR mutant. Based these criteria, we identified 9 candidate RNAs that exhibited decreased expression in the EAR mutant ranging from 3-fold to 100-fold in overall magnitude (Table 3-3). For simplicity, we will temporarily refer to these RNAs using the names of their neighboring genes. Although our RNA-Seq provided quantitative transcriptomic data it does not provide strand information. Hence, we determined the direction of transcription for these 9 candidate sRNAs using a combination of our previous 5`-end mapping data (Irnov *et al.,* 2010; see Chapter 4) and predictions of promoters and transcription terminator elements (DBTBS - Sierro *et al.*, 2008).

Among these candidates, we noticed two previously identified orphan riboswitches, *yxkD-yxkC* and *ylbG-ylbH* (Barrick *et al.*, 2004). Riboswitches are *cis*-acting metabolite-binding RNAs typically located in the 5`-UTR of mRNAs. These RNA elements are capable of sensing a variety of small metabolites and can activate or repress downstream gene expression upon ligand binding (Winkler and Breaker, 2005). Based on its genomic and structural arrangements, the *yxkD-yxkC* RNA would be predicted to activate the expression of *yxkD* gene, which encodes a putative membrane protein, upon binding of a still-unknown ligand (Barrick *et al.*, 2004). In other organism, this class of

metabolite-sensing RNAs (*i.e., ykkC/yxkD*-type riboswitches) appears to regulate the expression of multidrug resistance pumps, amino acid transporters, and multiple genes involved in purine metabolisms (Barrick *et al.*, 2004). Moreover, *ykkC/yxkD*-type riboswitches have been also found in tandem with another class of riboswitches that binds purine metabolites (Sudarsan *et al.*, 2006), raising a possibility that this RNA might sense a purine-related molecule. In our analysis, we noted almost 5-fold lower expression of *yxkD-yxkC* RNA in the EPS deficient strain. Interestingly, a similar fold change was also observed for *yxkD* gene, albeit expressed at a low level (Fig. 3-15). This might indicate that the reduced expressions for both the UTR and the coding region are due to a specific control at the level of transcription initiation. Thus, the weak expression observed for *yxkD* might represent the basal activity of this particular riboswitch. Alternatively, it is equally possible that the observed changes were due to the absence of a riboswitch ligand in the EPS deficient biofilm. Perhaps the *yxkD-yxkC* RNA is unstable *in vivo* in its unbound state (ligand-free), which would lead to a lower abundance of the UTR in addition to the coding region. The latter explanation will imply that the ligand for *ykkC/yxkD*-type riboswitches is present under biofilm conditions. However, *B. subtilis* also contains another example of this class of orphan riboswitches, which is located upstream of the *ykkC* gene (Barrick *et al.*, 2004). Given that there was no observable change in the expression of *ykkC* (both with the 5' UTR and the downstream genes; Fig. 3-15), we hypothesize that the former explanation is more likely to be true, *i.e.,* transcription initiation of the *yxkD-yxkC* transcript is affected by biofilm conditions. Moreover, although the metabolite ligand might not be present under our experimental conditions, our data showed that the *yxkD-yxkC* RNA was expressed and suggested that it

might play a role in the biofilm community. Hence, finding the regulators of *yxkD* might ultimately help identify the potential ligand for this class of riboswitches. Another candidate riboswitch identified by our analysis was the *ylbG-ylbH* RNA (Fig. 3-16A). This particular RNA element is found exclusively in Firmicutes and always located upstream of gene involved in DNA methylation (Barrick *et al.*, 2004). However, a more careful analysis of this genomic locus reveals that the *ylbG-ylbH* RNA is a small *trans*-acting RNA transcribed in the opposite strand (M. Dambach and W. Winkler, unpublished observation). Indeed while the expression of *ylbG-ylbH* RNA was ~35-fold lower in EPS deficient biofilm, there was no change in the expression of both the upstream and downstream genes (*ylbG* and *ylbH*, respectively; Fig. 3-16A). Furthermore, we note the presence of a strong transcription terminator element downstream of the putative small RNA supporting the idea of *ylbG-ylbH* RNA as an independent transcription unit (data not shown). Interestingly, there is a potential promoter sequence that resembles that of the σ^F and σ^G consensus sites (predicted by DBTBS - Sierro *et al.*, 2008), suggesting that expression of the *ylbG-ylbH* RNA might be limited to sporulating cells, specifically in the forespore region.

A third biofilm-specific sRNA candidate is the *hinT-ecsA* RNA, which we had previously discovered from late stationary phase cultures ('ncr1670'; Irnov *et al.,* 2010). This RNA was readily abundant in the wild-type biofilm, accumulating to approximately ~25% of the level of expression for the highly abundant sRNA, 6S (data not shown). However, its expression level was decreased by ~40-fold in the EPS deficient mutant biofilm (Fig. 3-16B). Similar to the *ylbG-ylbH* RNA, we also observed a potential σ^F or σ^G-dependent promoter for this RNA element, suggesting that it, too, is likely to be

expressed exclusively within the forespore. After close inspection, we found a repeated sequence motif, 5`-UGAGGUG-3`, within the *hint-escA* sRNA with the potential to participate in formation of a small stem-loop (Fig. 3-17A). Also, a remarkably similar sRNA could be identified for several other *Bacillus* species, suggesting that this particular sRNA element could be conserved in a subset of Firmicutes species (Fig. 3-17B). Intriguingly, this sequence motif is very similar to the consensus binding site of an RNA-binding protein, CsrA (Dubey *et al.*, 2005; Fig. 3-17A). In *E. coli* and other Gram-negative bacteria, CsrA has been shown to be able to activate or repress gene expression by binding to hairpin structures with 5`-GGA-3` sequence located within the loop region, typically near the RBS. In general, CsrA has been known 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*, the only known function of CsrA is to repress translation of the highly abundant flagellin protein, encoded by *hag* (Yakhnin *et al.*, 2007). Moreover, there has been no bona fide 'CsrB-like' small RNA identified in this, or other Gram-positive microorganism. Therefore, from our data, we hypothesize that the *hinT-ecsA* sRNA constitutes the CsrA-sequestering noncoding RNA for *B. subtilis*, and other *Bacillus* species. Also, if our prediction of a forespore-specific promoter for the *hinT-*

ecsA sRNA is correct, we hypothesize that this sRNA will antagonize CsrA's mRNA-binding ability only within the forespore during endospore development. Thus, future investigation of this RNA also is likely to reveal a role for CsrA in *B. subtilis*, especially during sporulation.

In addition to the abovementioned sRNAs, we also identified several more candidates with less obvious functions that were decreased in the EAR mutant strain (Fig. 3-18). For example, we predict from our data that a *cggR-araE* sRNA is expressed in the biofilm community, and, similar to *hinT-ecsA* and *ylbG-ylbH*, is predicted to be dependent on the forespore-specific sporulation sigma factors, σ^F and σ^G. The *desR-yocH* candidate RNA, which was also identified in a previous microarray analysis (Schmalisch *et al.*, 2010), is likely to be expressed from the mother cell during sporulation due to the presence of a putative σ^E-dependent promoter. Two more sRNAs, *rpsD-tyrS* and *yckB-yckC*, are predicted to be under the control of the stress-related sigma factor, σ^B. In addition, two of our candidates (*i.e., yckB-yckC* and *yrpD-yrpE*) also exhibit the potential to encode for small peptides (data not shown). This, however, does not rule out their possible roles as regulatory sRNAs. It has been shown previously that at least one of *E. coli* small RNA, SgrS, is both a regulatory RNA and a polypeptide-encoding mRNA (Wadler and Vanderpool, 2007). Finally, while there were 9 sRNA candidates with lowered expression levels in the EPS deficient mutant, we only observed one candidate small RNA exhibiting modestly increased expression (*adaB-ndhF*; Fig 3-19). Based on its expression pattern, this putative sRNA is more likely to play a role prior to the formation of multicellular communities. Indeed the *adaB-ndhF* RNA has also been detected in mid-exponential phase cultures (grown in M9 medium; Rasmussen *et al.*,

2009) and stationary phase cultures (Irnov *et al.*, 2010). Experiments to define the roles of these different sRNAs during biofilm formation are ongoing.

It is important to note that the analysis described above only focused on sRNAs that are differentially expressed between the wild-type and EPS deficient biofilms. We hypothesize that these sRNAs will play important roles during development of multicellullar communities. However, based on their predicted promoter sequences (Sierro *et al*., 2008), it appears that the majority of these sRNAs are likely to be involved exclusively in sporulation. We note that this is likely to be consistent with the fact the sporulation pathway appeared to be the most cellular sub-class that was affected by EAR mutations. Inspection of other previously identified sRNAs (described further in Chapter 4), revealed that they were not influenced by the lack of exopolysaccharide production (Table 3-4). It is very likely that these latter set of sRNAs are not involved in a particular differentiation pathway, and are instead important in regulating how an individual cell responds to various stress conditions.

Discussion and Future Directions

Characterization of systems in *E. coli* and λ have revealed some of the events that occur during processive antitermination (Weisberg and Gottesman, 1999; Roberts *et al.*, 2008; Fig. 3-1). First, RNA polymerase successfully transitions from an initiation complex into a TEC. For some processive antitermination systems, the TEC then encounters a promoter-proximal pause site, which allows time for accessory factors to associate with discrete elements in the nascent transcript and the TEC. In the event that binding of these accessory antitermination factors is unsuccessful within the lifetime of

the paused complex, elongation is likely to cease upon encountering downstream termination sites. However, when successfully bound by accessory factors, modified TEC complexes transcribe faster than a typical TEC and are more resistant to pause and termination sites. Presumably, these properties are maintained for as long as the TEC-binding antitermination factor(s) remain associated. However, the antitermination-proficient TEC eventually encounters a terminator site strong enough to cause dissociation, presumably located at the operon terminus. An interesting exception to this overall pathway is the *put* antitermination element, which does not require additional factors. The *put* RNA element (Fig. 3-1) interacts directly with a portion of RNA polymerase near to the RNA exit channel and is alone sufficient for inducing readthrough of pause and terminator sites *in vitro* (King *et al.*, 1996; King *et al.*, 2004; Weisberg and Gottesman, 1999; Sen *et al.*, 2002; Komissarova *et al.*, 2008).

In this study, we present the discovery of a new processive antitermination mechanism class, which is required for biofilm exopolysaccharide expression in *B. subtilis*. We find that the EAR element, located between *epsB* and *epsC*, can promote transcriptional read-through of endogenous and heterologous intrinsic termination sites. This antitermination activity is even maintained for a tandem arrangement of three 'strong' heterologous intrinsic terminator sites. From these data, we predict that the EAR antitermination mechanism will share certain features with the previously described processive antitermination mechanism classes. However, it is also clear that the EAR antitermination mechanism exhibits unique features from these other systems. For example, the EAR antitermination mechanism is one of only a few processive antitermination mechanism classes that are bacterially encoded. Also, EAR

antitermination requires an RNA element that is both unusual in its phylogenetic distribution and structural sophistication. Antitermination by the bacteriophage *put* element also requires a structured RNA element, which is comprised of a ~65 nucleotide dual-hairpin motif. However, antitermination by EAR appears to require a greater overall degree of information content as the ~120 nucleotide EAR element exhibits extensive conserved secondary structure and primary sequence features. Also, unlike *put*, the EAR element alone is not sufficient to promote antitermination activity *in vitro*. In addition to its function regulating the expression of *eps* genes, the EAR element also seems to play a role in the development of distinct cell types within biofilms. Whether cellular differentiation during biofilm formation requires the EAR element itself or EAR-mediated EPS production is still unclear. Moreover, it remains to be investigated whether the transcriptomic changes observed for the cell-type specific markers are due to changes in all biofilm cells or only the corresponding cell types. A more detailed single-cell level analysis will be needed to differentiate the two possibilities. Finally, our transcriptomic data also revealed several small RNAs expressed during biofilm formation. Most of them are predicted to be under the control of various sporulation sigma factors. Future investigations on the potential roles of these small RNAs might reveal novel posttranscriptional mechanisms for regulating an already complex development pathway.

What could be the functional role(s) of EAR antitermination? For the λ bacteriophage, processive antitermination by λN and λQ proteins is required for high-level expression of early and late phage transcripts, respectively. These mechanisms ensure both robustness of λ gene expression and completed synthesis of these unusually long operons (Weisberg and Gottesman, 1999). For example, estimates based upon prior

in vitro quantitative analyses suggest that a 'normal' length operon (<5000 nt) is likely to result in 80-90% of transcripts being fully transcribed. This contrasts with only 32-63% of the λ late operon (~23,000 nt) being fully synthesized without aid of the λQ processive antitermination system, which restores full transcript synthesis to ~90% (von Hippel and Yager, 1991; von Hippel and Yager, 1992). Similarly, *rrn* antitermination increases transcriptional efficiency and overall abundance of rRNAs to accommodate fast-growing bacteria (Condon *et al.*, 1995). Indeed, recent quantitative analyses of *rrn* antitermination demonstrate that the high rate of *rrn* transcription required for rapidly dividing cells results in dense RNAP 'traffic' along the transcripts (Klumpp and Hwa, 2008). Specific incorporation of multiple Rho-dependent termination sites within these transcripts helps to rapidly remove RNAP complexes that have not been altered into the terminator-resistant form, resulting in increased transcription elongation efficiency overall.

Based upon these observations on the λ and *rrn* antitermination systems, it can be presumed that EAR antitermination may have evolved for similar reasons: (1) full synthesis of long operons and (2) improved efficiency of transcription where robust expression and strict stoichiometry is required. At approximately 16 kb, the *eps* operon in *B. subtilis* is indeed substantially longer than the average bacterial transcripts (Dam *et al.*, 2007). Furthermore, ensuring that individual cells exhibit particularly high levels of *eps* expression may also be particularly important during *B. subtilis* biofilm development. Recent data have suggested that EPS production is important for endospore formation within the *B. subtilis* biofilm community, possible by creating microenvironments that are depleted of nutrients (Vlamakis *et al.*, 2008). Thus it is possible that the robust *eps* expression afforded by EAR antitermination may have evolved in part to ensure

metabolic 'specialization' of EPS-producing bacteria within the biofilm community, although further experimentation is required to test this hypothesis.

Therefore, one possible model is that the EAR element functions in 'constitutive antitermination' (Fig. 3-20B). However, the fact that this activity is not maintained *in vitro* using purified components suggests that additional cellular factors are required. It is possible that the EAR antitermination system may require cellular elongation factors (*e.g.,* 'Nus-like' factors), similar to other antitermination systems. This overall mechanistic model would suggest that EAR antitermination functions essentially as an extension of the regulatory mechanisms controlling transcription initiation. In other words, once transcription of *eps* has been initiated the EAR element may act as a constitutive 'gas pedal' to cause read-through of downstream termination and pause sites.

Alternatively, it is also possible that EAR antitermination requires participation of a particular physiological or population-based signal that occurs during biofilm formation (Fig. 3-20B). It may be particularly relevant for this model that we are only able to identify the EAR element within functionally related operons. This fact might suggest that the EAR element shares a specific regulatory relationship with exopolysaccharide gene clusters, perhaps integrating a signal that is commonly important during biofilm and capsule polysaccharide formation. However, this model raises an obvious question – why might it be useful to employ signal-responsive control of *eps* transcription elongation? Our data revealed that, in the absence of EAR antitermination, a major block for the transcription of *eps* operon is located inside the *epsF* coding region. It is possible that, for *B. subtilis,* EAR antitermination may be coordinated with the functional roles of the genes located upstream and downstream of *epsF*, respectively. For example, the

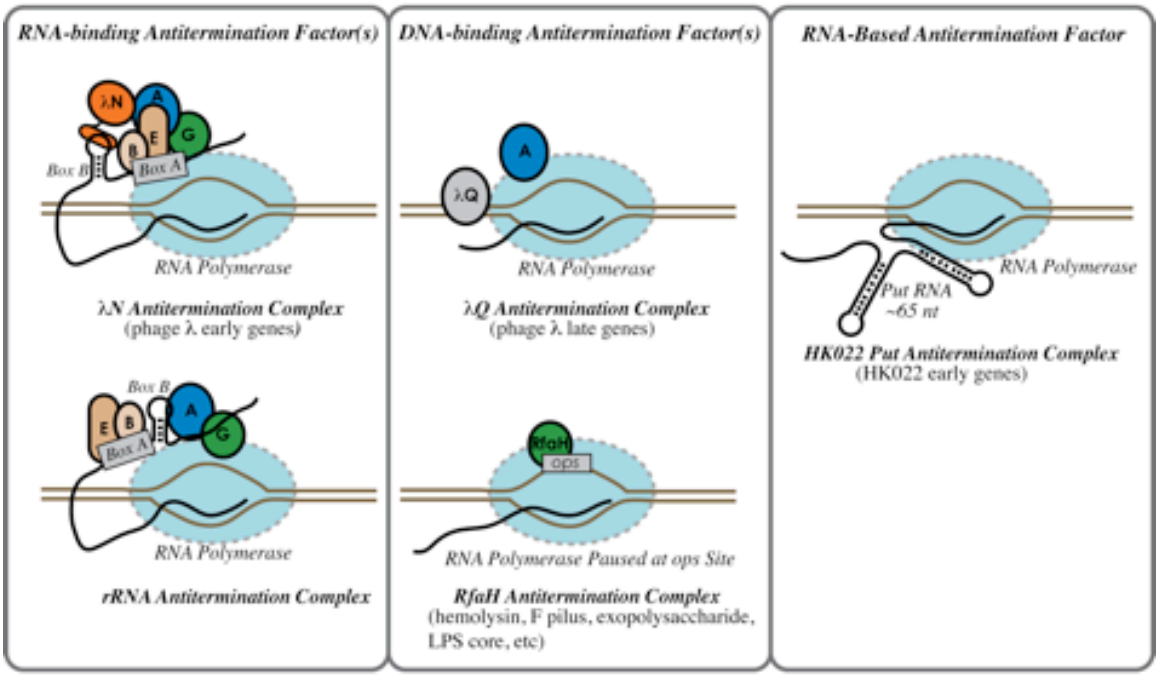
immediately upstream gene, *epsE,* is a glycosyltransferase that has also been shown to act as a "molecular clutch" by halting flagellar movement upon its expression (Blair *et al.*, 2008). Downstream of *epsF* are mostly genes required for polymerization and extracellular assembly of the exopolysaccharides. Therefore, upon SinR derepression, transcription of the *eps* operon would be initiated but would prematurely terminate within the *epsF* locus (Fig. 3-20A). These cells would be unable to complete exopolysaccharide production, but are likely to still be nonmotile due to *epsE* expression. Perhaps these cells remain in this nonmotile state while they await an unknown physiological or quorum signal that is perceived by the EAR antitermination system, which would then permit completed *eps* synthesis and full EPS polymerization. These hypotheses remain to be investigated in the future.

The regulation of biofilm formation in *B. subtilis* has been well studied at the transcriptional level (Aguilar *et al.*, 2007). Previous data revealed a cascade of transcription factors (*e.g.,* Spo0A, DegU, SinR, SlrR, RemA, RemB) that together contribute to the transcriptional activation of *eps* and *yqxM-sipW-tasA* operons (Branda *et al.,* 2001; Kearns *et al.,* 2005; Verhamme *et al.*, 2007; Chu *et al.,* 2008; Chai *et al.,* 2009; Winkelmann *et al.,* 2009). Our data together introduce a new layer of post-transcriptional regulation that is mediated by the EAR element and is essential for biofilm formation. Our data also suggest that structurally complex, 'riboswitch-like' RNA elements may also be employed for processive antitermination mechanisms (see Fig. 3-2). Given the rapidly expanding collection of uncharacterized but putative bacterial regulatory RNAs (*e.g.,* Barrick *et al.,* 2004; Corbino *et al.,* 2005; Weinberg *et al.,* 2007; Gardner *et al.,* 2009; Xu *et al.*, 2009b), this discovery may indicate that other RNA-mediated processive

antitermination mechanisms also await discovery. It has already been established that regulatory RNAs are widespread for regulation of transcription initiation (*e.g.,* 6S RNA) (Wassarman, 2007); we hypothesize that it remains to be determined whether regulatory RNAs controlling transcriptional elongation processivity are just as widespread.

Furthermore, the reliance of the *eps* operon on the EAR element for full mRNA synthesis is consistent with the hypothesis that processive antitermination might be a common mechanism through which long bacterial operons are synthesized. Indeed, development of the biofilm community is reliant upon coordination and expression of multiple long operons, including synthesis of exopolysaccharides and secondary metabolites. Consistent with the hypothesis, we have identified long 5`-leader regions for *srf* (~26 kb) and *dhb* (~12 kb) operons (see Chapter 4). A preliminary experiment on the *srf* leader region indicated that it might be necessary for synthesizing high level of *srf* mRNAs and efficient production of surfactin (data not shown). Based upon our observations with the EAR element, we speculate that processive antitermination as an altered mode of transcription elongation may be generally required for full expression of these operons and may be of greater fundamental importance to bacterial multicellular communities than previously realized. If this is all true, then there will be many more antitermination systems waiting to be discovered, which may or may not conform to the classical λ and *E. coli* processive antitermination systems.

A


RNA-binding Antitermination Factor(s)
DNA-binding Antitermination Factor(s)
RNA-Based Antitermination Factor
λN
A
B
E
G
Box B
Box A
RNA Polymerase
λN Antitermination Complex
(phage λ early genes)
λQ
λQ Antitermination Complex
(phage λ late genes)
Put RNA
~65 nt
HK022 Put Antitermination Complex
(HK022 early genes)
rRNA Antitermination Complex
RfaH
ops
RNA Polymerase Paused at ops Site
RfaH Antitermination Complex
(hemolysin, F pilus, exopolysaccharide,
LPS core, etc)

B

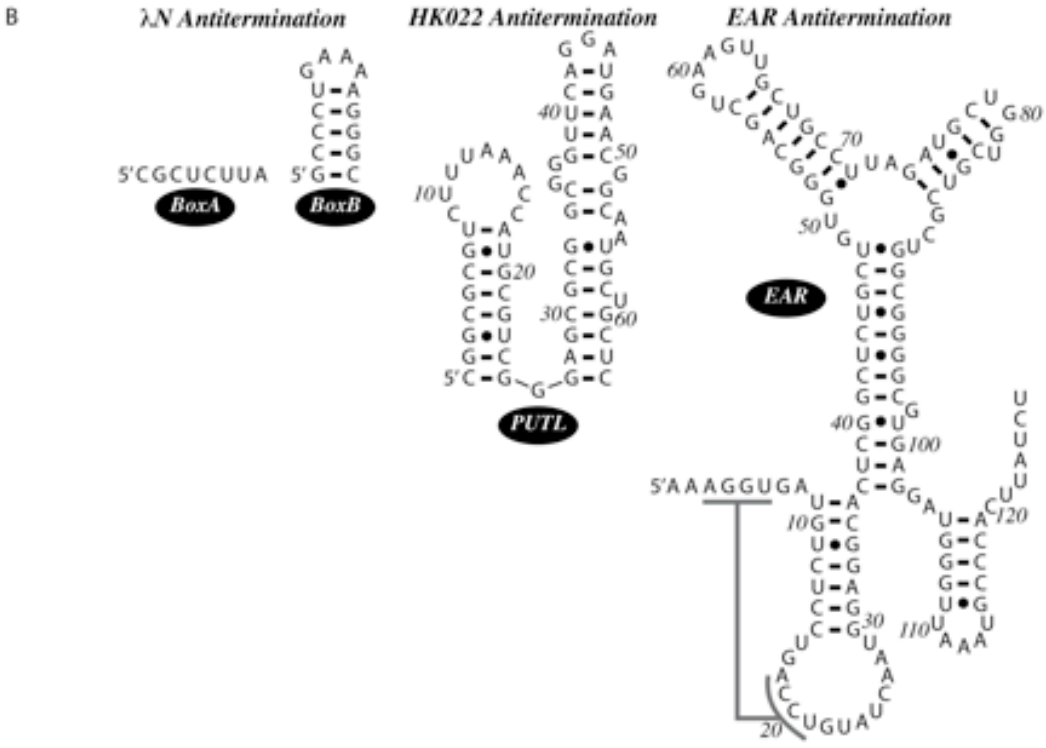
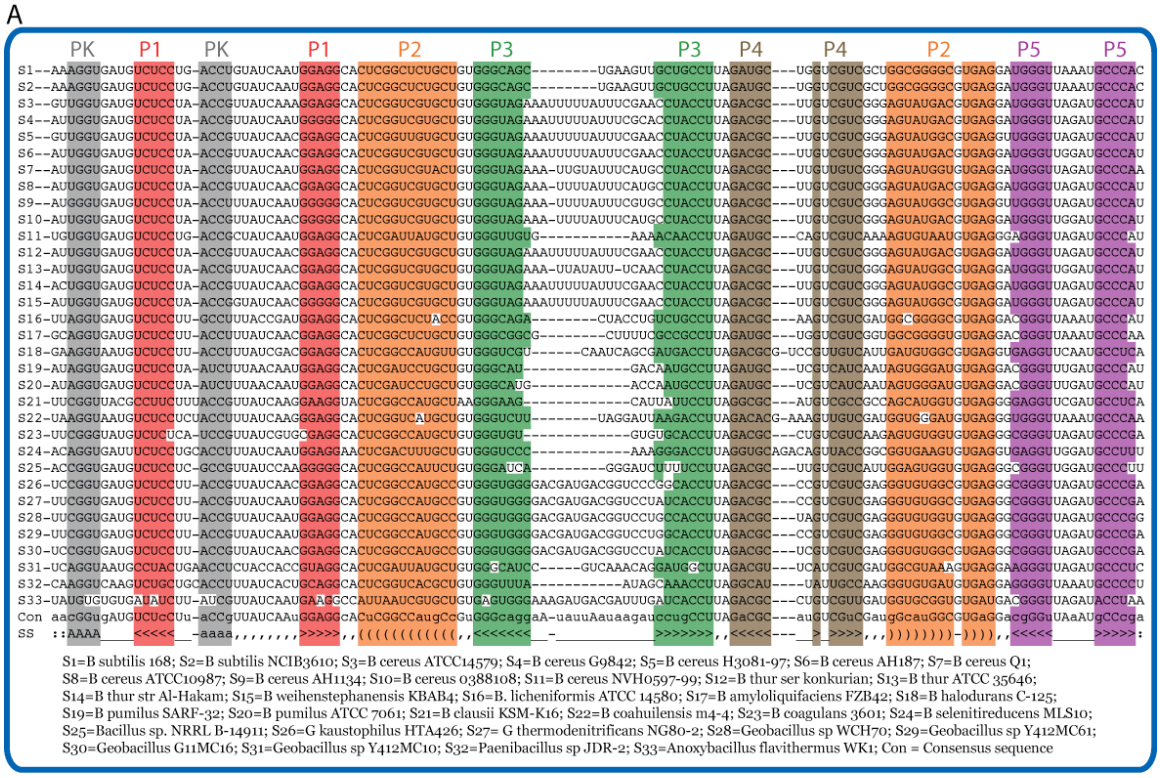

λN Antitermination
HK022 Antitermination
EAR Antitermination
5'CGCUCUUA
5'G
BoxA
BoxB
PUTL
EAR
5'AAAGGUGA

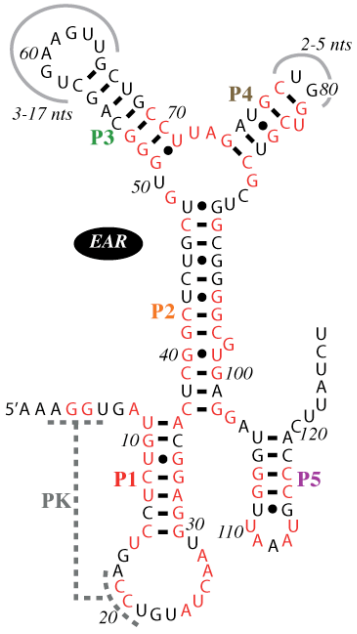
Figure 3-1. Comparison of the EAR element with RNA determinants from other processive antitermination systems. (A) Schematic depiction of the major classes of previously discovered processive antitermination systems (Roberts *et al.*, 2008; Weisberg and Gottesman, 1999). In this depiction, they are grouped based upon the nature of their effector interactions. For example, processive antitermination systems that require RNA-binding proteins are shown in the left panel. Those requiring DNA-binding proteins are depicted in the central panels. Finally, the *put* RNA element, which alone can instigate processive antitermination, is depicted in the panel on the right. The EAR element appears to exhibit a combination of these characteristics. For example, the EAR antitermination element is required for expression of bacterial biofilm and capsular polysaccharides, similar to RfaH antitermination. However, unlike RfaH antitermination and akin to *put* antitermination, EAR-mediated antitermination is promoted by an RNA element. Gold-colored lines = DNA. Black lines = nascent RNA. Blue oval = RNA polymerase. Circles = elongation and antitermination factors that participate in antitermination. (B) RNA determinants required for antitermination. This panel shows the difference in sequence and structural complexity between the RNA-based antitermination determinants for λN antitermination, *HK022 put* antitermination, and EAR antitermination, respectively.

A

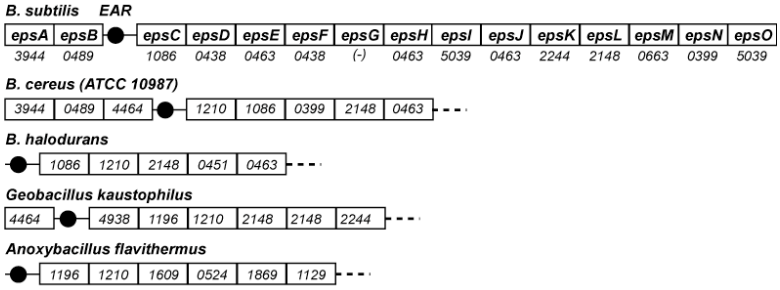


S1=B subtilis 168; S2=B subtilis NCIB3610; S3=B cereus ATCC14579; S4=B cereus G9842; S5=B cereus H3081-97; S6=B cereus AH187; S7=B cereus Q1;
S8=B cereus ATCC10987; S9=B cereus AH1134; S10=B cereus 0388108; S11=B cereus NVH0597-99; S12=B thur ser konkurian; S13=B thur ATCC 35646;
S14=B thur str Al-Hakam; S15=B weihenstephanensis KBAB4; S16=B. licheniformis ATCC 14580; S17=B amyloliquifaciens FZB42; S18=B halodurans C-125;
S19=B pumilus SARF-32; S20=B pumilus ATCC 7061; S21=B clausii KSM-K16; S22=B coahuilensis m4-4; S23=B coagulans 3601; S24=B selenitireducens MLS10;
S25=Bacillus sp. NRRL B-14911; S26=G kaustophilus HTA426; S27= G thermodenitrificans NG80-2; S28=Geobacillus sp WCH70; S29=Geobacillus sp Y412MC61;
S30=Geobacillus G11MC16; S31=Geobacillus sp Y412MC10; S32=Paenibacillus sp JDR-2; S33=Anoxybacillus flavithermus WK1; Con = Consensus sequence

B ***eps*-Associated RNA (EAR)**



C ***Representative arrangements of the EAR element***



Clusters of Orthologous Groups (COG) Assignments

0399	*NDP-hexose 3, 4-dehydratase*	*1210*	*UDP-glucose pyrophosphorylase*
0438	*Glycosyltransferase*	*1609*	*Transcriptional regulator*
0451	*Nucleoside diphosphate sugar epimerase*	*1869*	*Ribose transport*
0463	*Glycosyltransferase*	*2148*	*Lipopolysaccharide sugar transferase*
0489	*ATPase for chromosome partitioning*	*2244*	*Membrane protein - export of teichoic acids*
0524	*Sugar kinase*	*3944*	*Capsular polysaccharide biosynthesis*
0663	*Carbonic anhydrases/acetyltransferases*	*4464*	*Capsular polysaccharide biosynthesis*
1086	*Nucleoside diphosphate sugar epimerase*	*4938*	*Uncharacterized*
1129	*ABC transport*	*5039*	*Exopolysaccharide biosynthesis protein*
1196	*Chromosome segregation ATPase*		

Figure 3-2. Comparative sequence analysis of EAR elements**.** (A) Alignment of EAR elements from Bacillales. Regions exhibiting base-pairing potential are highlighted by different colors (P1-P5). "PK" indicates the presence of a potential pseudoknot. The consensus sequence and secondary structure arrangement are depicted within the last two rows. (B) Sequence and secondary structure of the *Bacillus subtilis* EAR element as predicted by sequence analysis and structural probing (see Fig. 3-3), located within the *epsB-C* intergenic region. Red positions are conserved in at least 95% of sequences. Variable loop regions are marked by grey semicircular lines. (C) Representative arrangements of EAR elements and their nearest genes from several species, including: *B. subtilis*, *B. cereus*, *B. halodurans*, *G. kaustophilus*, and *A. flavithermus*. The EAR element is denoted by black circle (●). All *eps* genes are shown as white rectangles and are labeled by the appropriate COG category (Cluster of Orthologous Group).

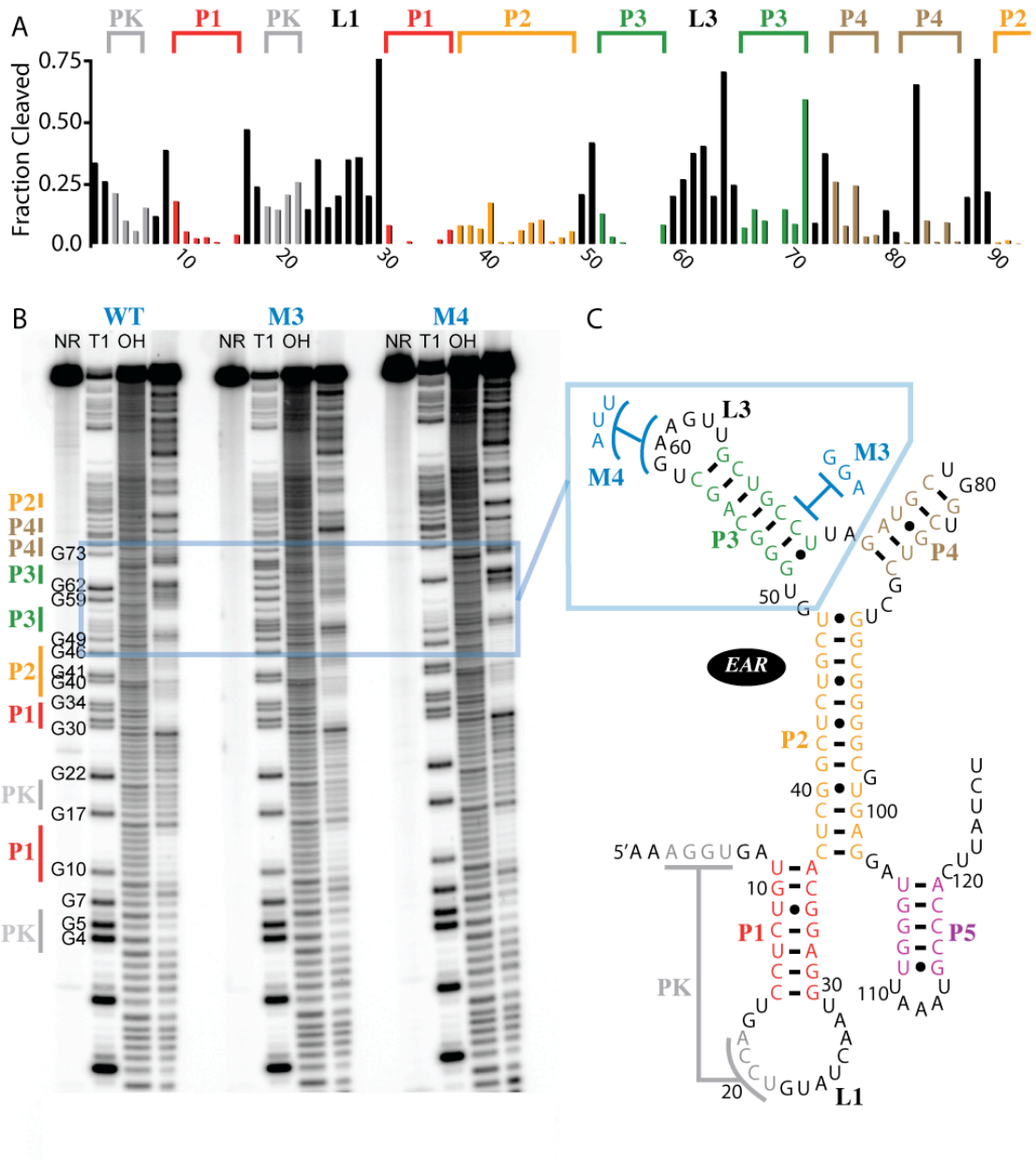


Figure 3-3. In-line probing of wild-type and mutant EAR molecules. In this assay, bands qualitatively correlate with unpaired residues. These data suggest that the overall global configurations do not differ between wild-type, M3, and M4 molecules. (A) Quantification of the in-line probing analysis. Nucleotide position is shown on the x-axis as depicted in the secondary structure, while y-axis represents the degree of cleavage at a particular position as calculated using SAFA (Das *et al.*, 2005). (B-C) The band patterns overall are nearly identical for wild-type, M3, and M4, with the exception of the highlighted region between G49 and G73 (B). This region of the molecule was mapped to the P3 helix, the location of the M3 and M4 mutations. Specifically, the M3 construct exhibited a significant disruption of the P3 helix compared to wild-type (notice the appearance of bands between G49 and G73), while the M4 mutation, which is located within the P3 terminal loop, showed no significant changes of the P3 helix (C).

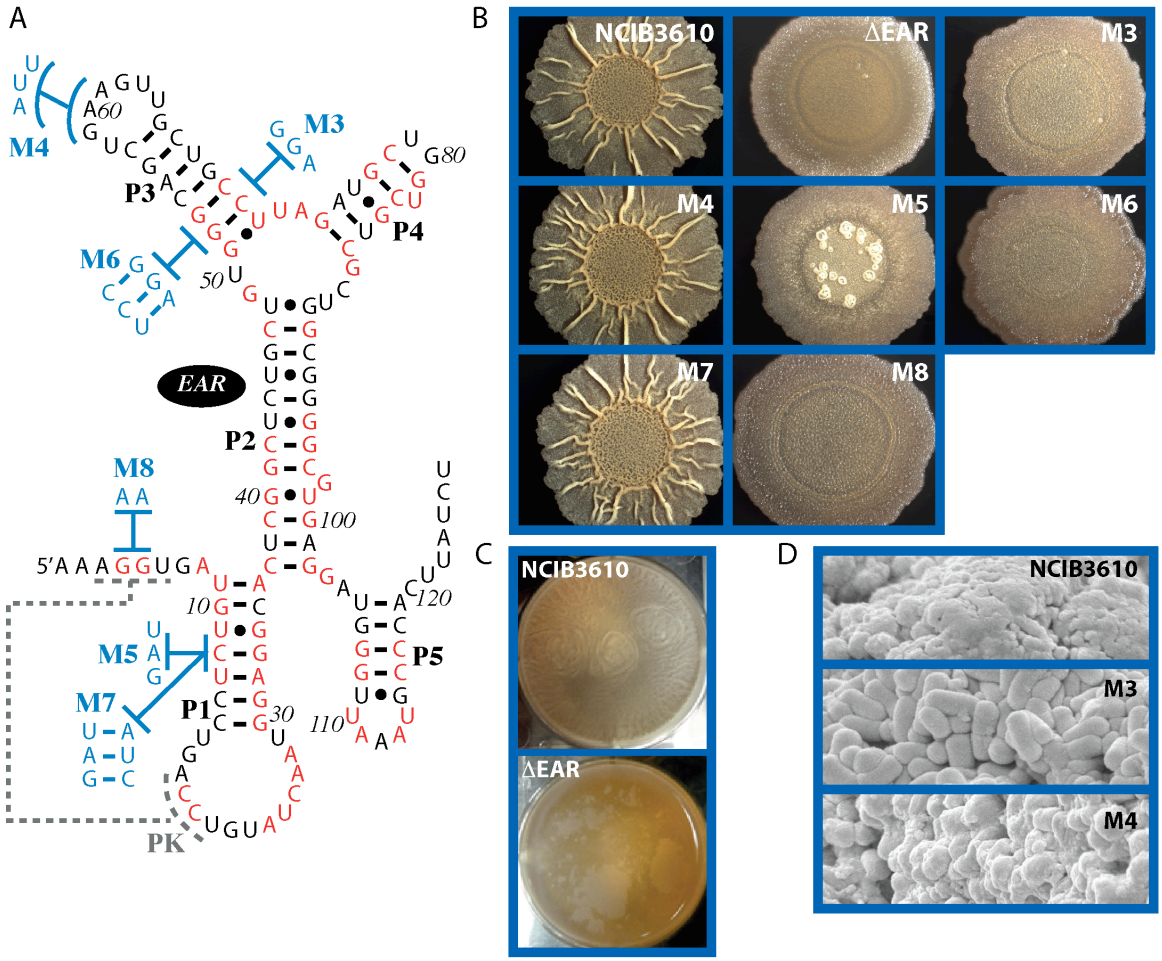


Figure 3-4. The EAR element is required for *B. subtilis* biofilm formation and exopolysaccharide production. (A) Secondary structure arrangement for the *B. subtilis* EAR element. Colored in red are positions that are conserved in at least 95% of sequences. Regions that can form base pairing interactions are labeled "P1" through "P5". "PK" denotes a potential pseudoknot. Positions of site-directed mutagenesis (labeled as "M3" through "M8") are denoted in blue. (B) Complex colony morphology of bacterial strains containing mutations of the EAR element. Complex colony architectures are characteristic of biofilm communities (Branda *et al.,* 2001; Aguilar *et al.,* 2007). (C) Top-down view of pellicle formation for wild-type *B. subtilis* NCIB 3610 and ΔEAR. (D) Scanning electron micrographs of the surface of biofilms for the wild-type strain and cells containing mutations in the EAR element.

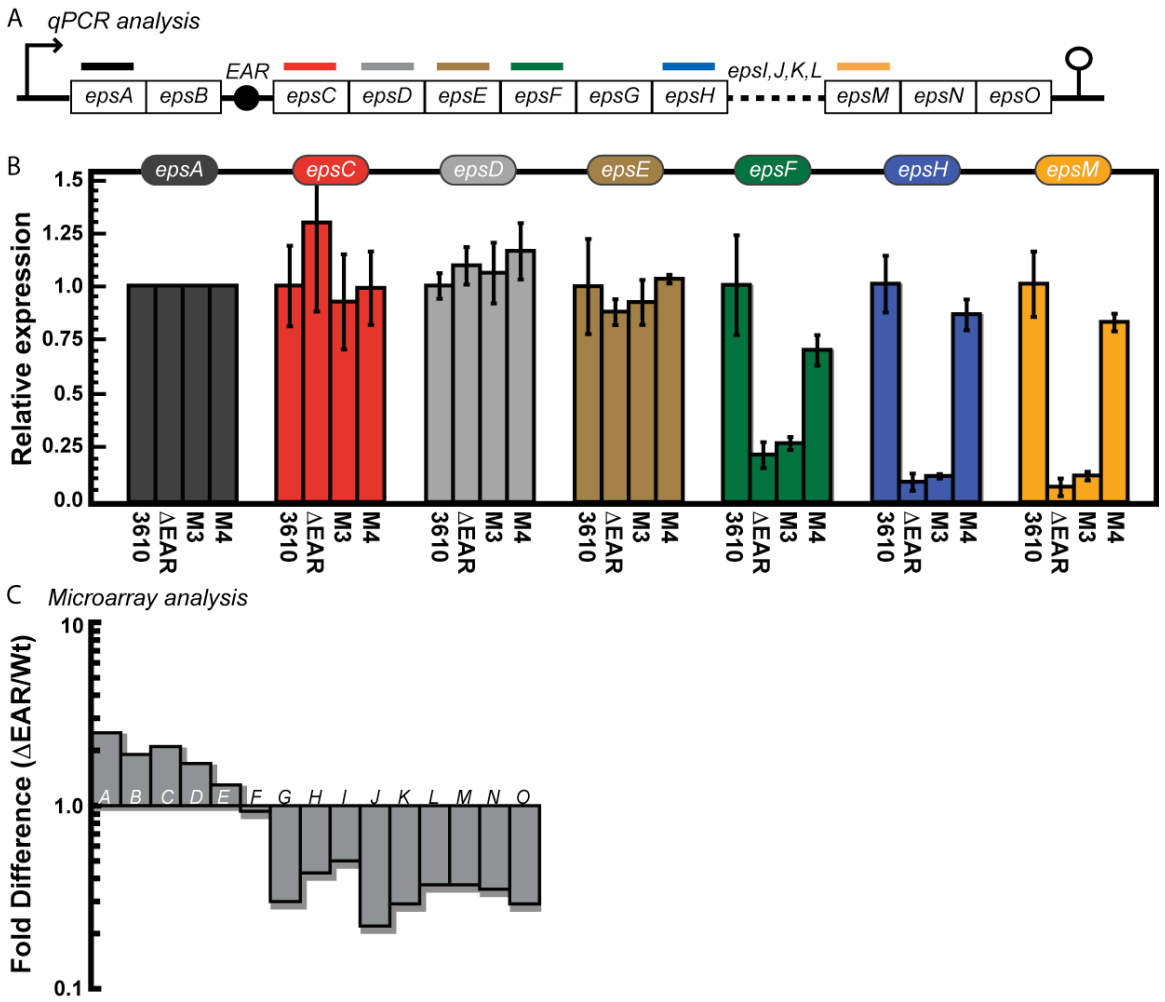


Figure 3-5 The EAR element promotes expression of distally located genes. (A) Schematic of the *eps* operon with colored lines representing the general locations that were chosen for qPCR analyses. (B) Relative expression level of *eps* genes in EAR mutants as compared to NCIB 3610. The relative expression of *epsA* was used to normalize the data from each sample (see Supplementary Methods for more details). (C) Microarray analysis of ΔEAR as compared to NCIB3610. Total RNA was extracted from a total of three 48-hour colonies cultured on MSgg solid medium (see representative colony pictures in Fig. 3-4) and used for microarray analysis. Differences in expression of *eps* genes between ΔEAR and NCIB3610 are shown above. MSgg is a medium condition that has been previously observed to promote biofilm formation in *B. subtilis*.

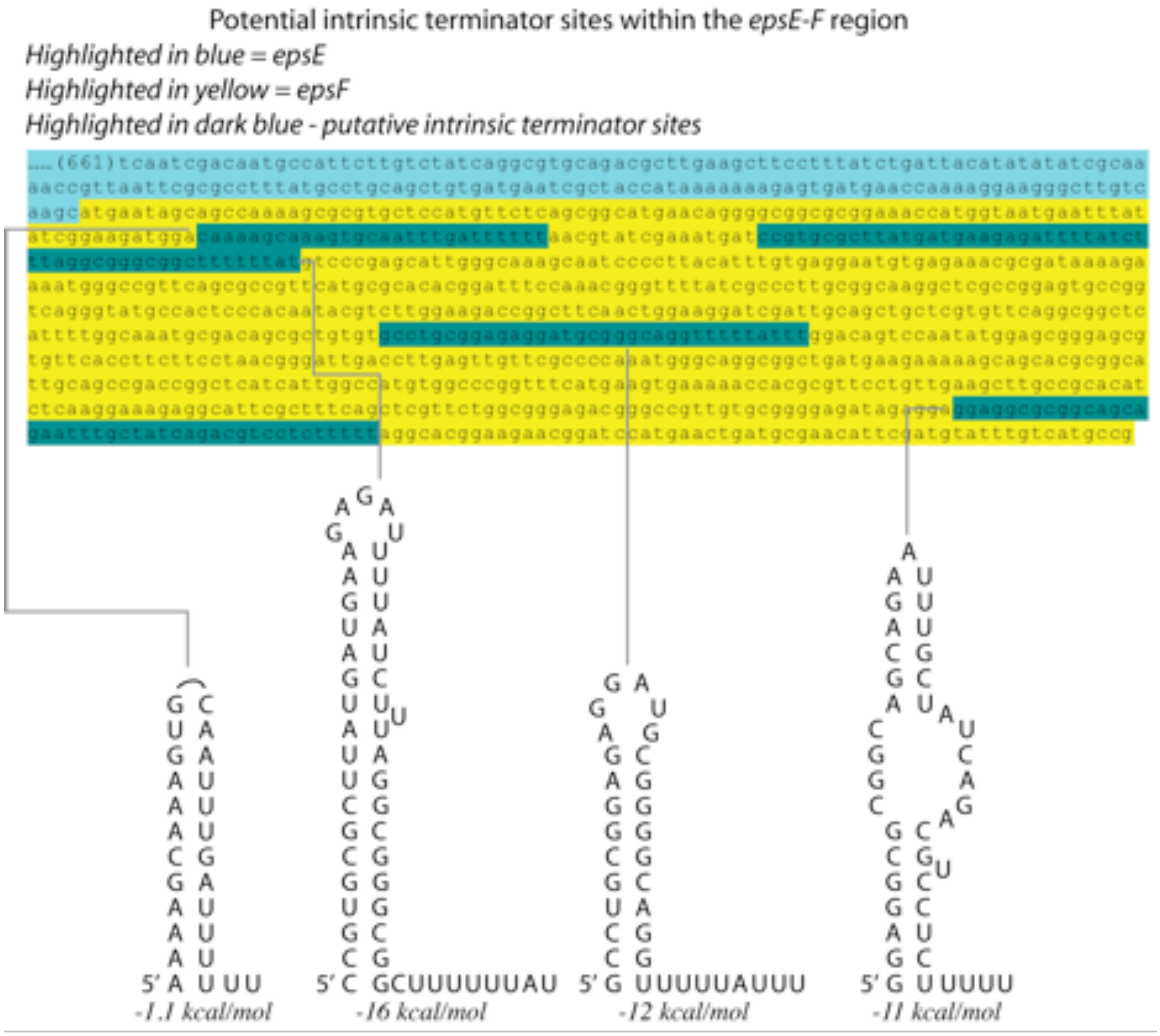


Figure 3-6. Candidate intrinsic terminator sequences located within the *epsF* region. The nucleotide sequence from the *epsE-F* region is shown above. Only the N-terminal half of the *epsF* coding sequence is shown above. Several candidate intrinsic terminator hairpins within this region are indicated above by dark blue shaded portions. For each terminator candidate, the predicted secondary structure and their calculated free energy were shown below the genomic sequence. It is worth noting that these terminator candidates occur within the least conserved portion of the *epsF* gene, perhaps suggesting that this portion of the gene is evolutionarily malleable enough to allow evolution of hairpin motifs while maintaining in-frame capacity for *epsF*.

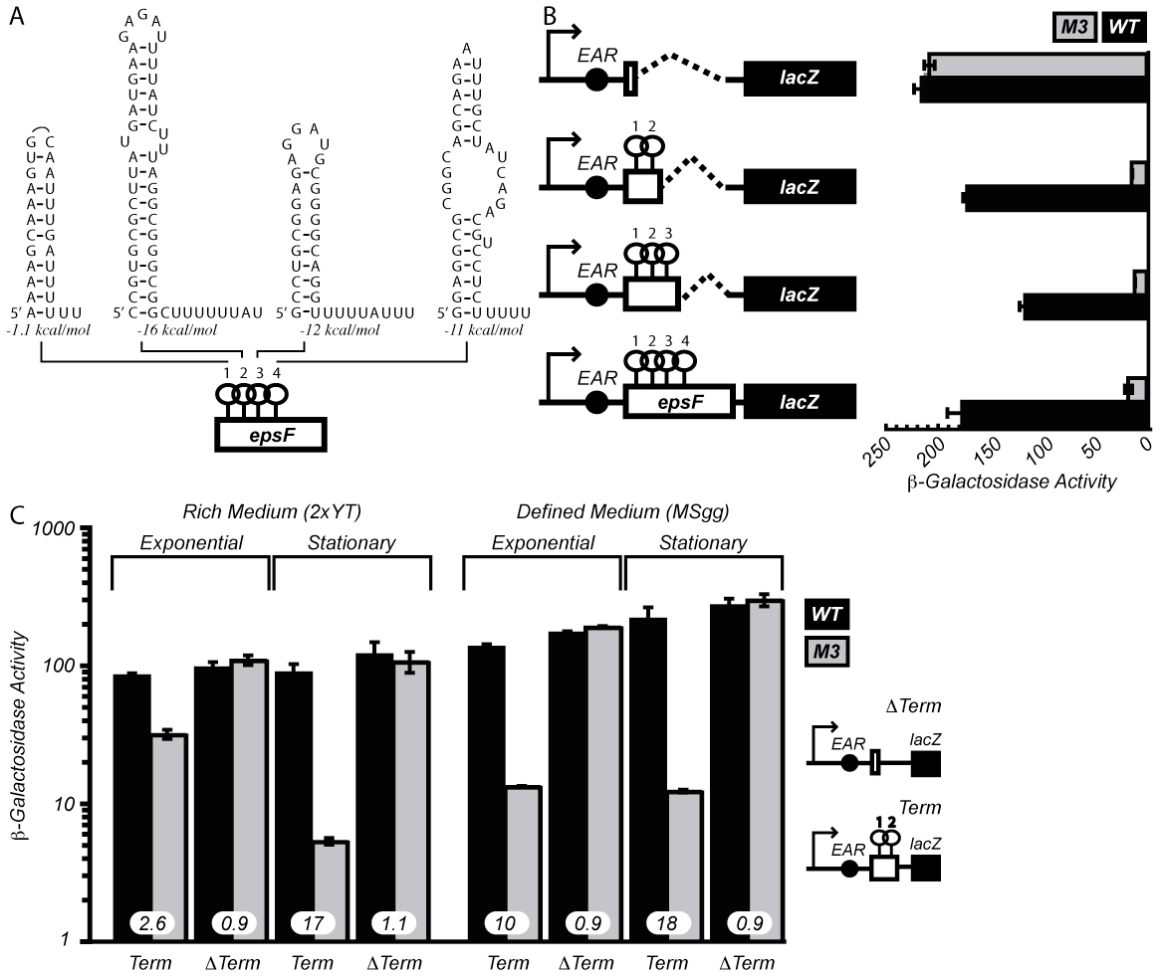


Figure 3-7. EAR-assisted read-through of intrinsic termination within the *epsF* region. (A) Putative intrinsic termination sites located within the *epsF* gene (shown within the coding sequence in Fig. 3-6). (B) Constructs containing different truncations of the *epsF* region, located downstream of the EAR element and upstream of a *lacZ* gene, are shown schematically in the left panel. Mutation of the EAR element results in lowered *lacZ* expression for each of these constructs that contain terminator sites, suggesting that the EAR element indeed promotes read-through of intrinsic terminators within *epsF*. These experiments were conducted in strain 168. Expression of the EAR-*epsF*-*lacZ* reporter fusion is identical between NCIB3610 and 168 (data not shown). All other data presented in the manuscript results from experimentation using NCIB3610.

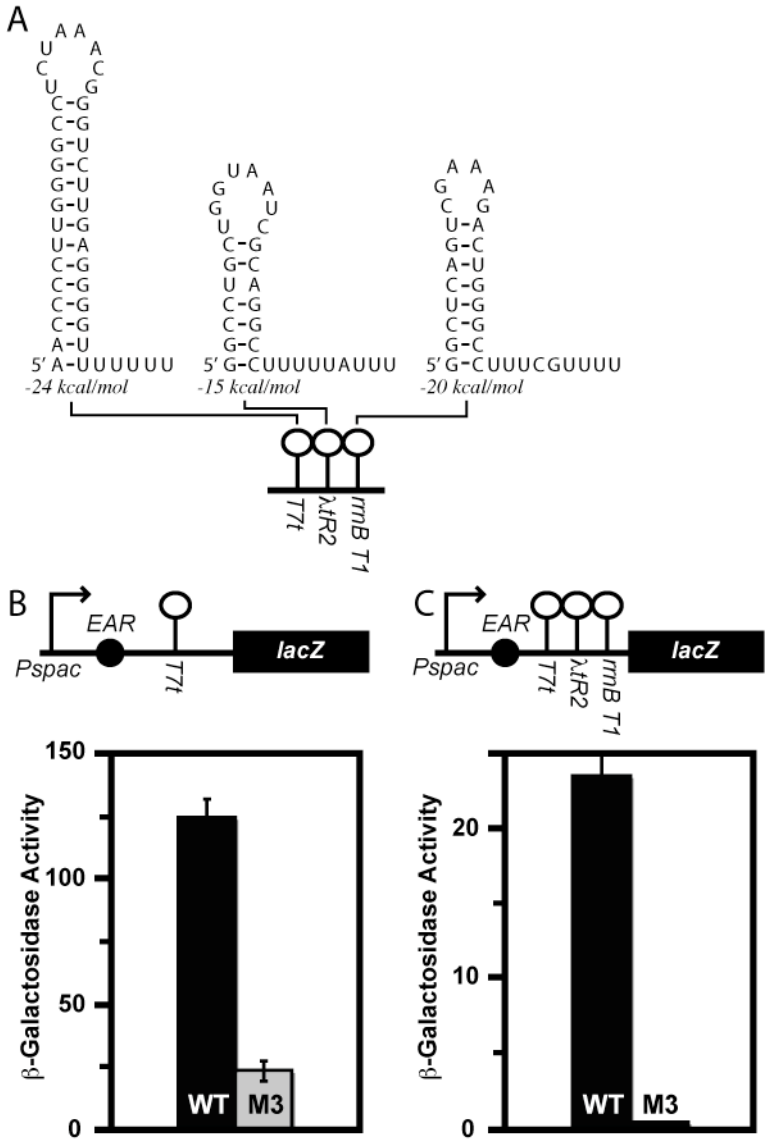


Figure 3-8. The EAR element promotes read-through of heterologous intrinsic terminators. (A) The heterologous intrinsic terminators used in this study are shown schematically. (B) A schematic representation of the antitermination assay construct is shown. The EAR element was placed upstream of a well-characterized, strong intrinsic transcription terminator from the T7 bacteriophage (T7t) (Reynolds and Chamberlin, 1992), which itself was upstream of a *lacZ* reporter. Expression was measured for constructs with a wild-type or M3 EAR element. (C) Similarly, a wild-type or M3 EAR element was placed upstream of a tandem arrangement of three different intrinsic transcription terminators (T7t-λtR2-rrnB T1; Artsimovitch *et al.*, 2000), also upstream of *lacZ*.

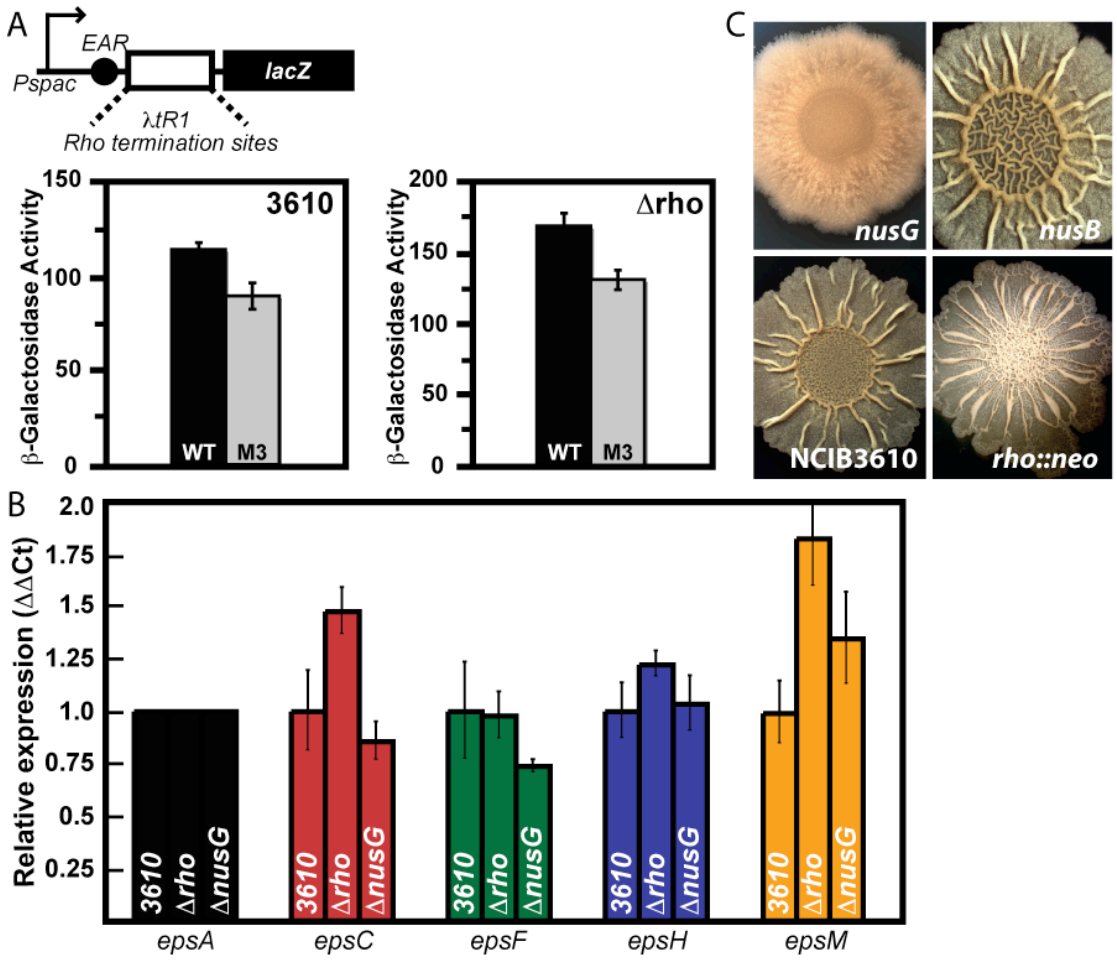


Figure 3-9. The EAR element does not promote read-through of Rho termination sites or require NusG and NusB. (*A*) A wild-type or mutant EAR element was placed upstream of a well-characterized Rho termination site ('λtR1'; Artsimovitch *et al.*, 2000), followed by *lacZ*. *lacZ* expression of these constructs was measured for a wild-type bacterium and in a *rho* mutant strain (DH1450; gift from D. Kearns – Indiana University). (*B*) Relative expression level of *eps* genes in *rho* and *nusG* mutants as compared to NCIB 3610. (*C*) Complex colony morphology of wild-type NCIB3610, *rho*, *nusG,* and *nusB* mutants. Although *rho* and *nusG* mutants show different colony morphology as compared to NCIB 3610, the deletion of both elongation factors do not significantly affect the expression of *eps* operon as measured by qPCR analysis in (*B*). In contrast, deletion of *nusB* does not affect colony morphology. These data suggest Rho, NusG, and NusB proteins might not play an important role in EAR antitermination mechanism.

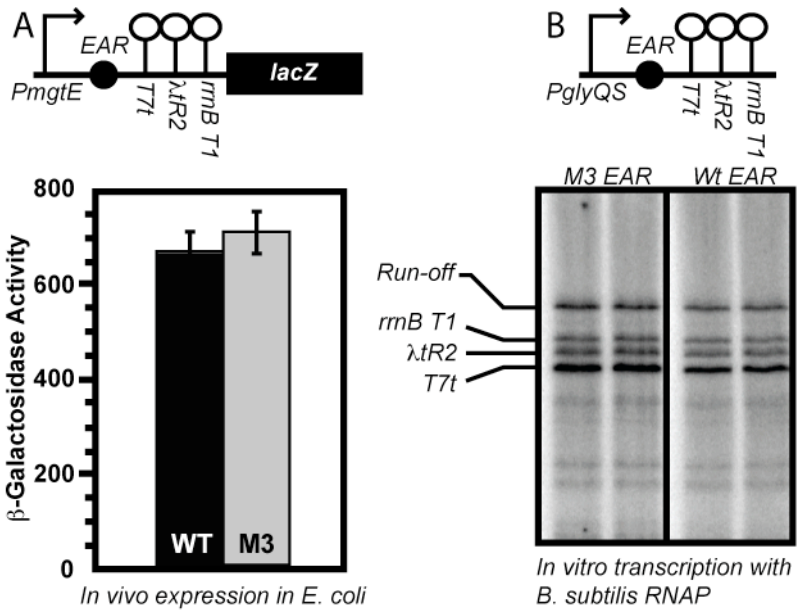


Figure 3-10. EAR antitermination may require additional cellular co-factors. (A) A schematic representation of the antitermination assay construct is shown. A wild-type or M3 EAR element was placed upstream of a tandem arrangement of three different intrinsic transcription terminators (T7t-λtR2-rrnB T1; Reynolds and Chamberlin, 1992; Artsimovitch *et al.*, 2000), also upstream of *lacZ*. Expression of these constructs was then assessed within an *E. coli* heterologous host. (B) We also assessed the ability of the EAR element to promote antitermination *in vitro*, in the absence of additional cellular components. A schematic representation of the PCR-generated DNA template for the *in vitro* antitermination assay is shown. Representative data from a single-round transcription assay using *B. subtilis* RNA polymerase is shown in replicates. Termination at all three heterologous termination sites is marked accordingly, as are transcripts resulting from run-off transcription. Quantification of wild-type and M3 EAR constructs revealed a similar overall level of run-off transcription for both, suggesting that antitermination activity is impaired *in vitro*. Similar results were obtained using a multi-round transcription assay format.

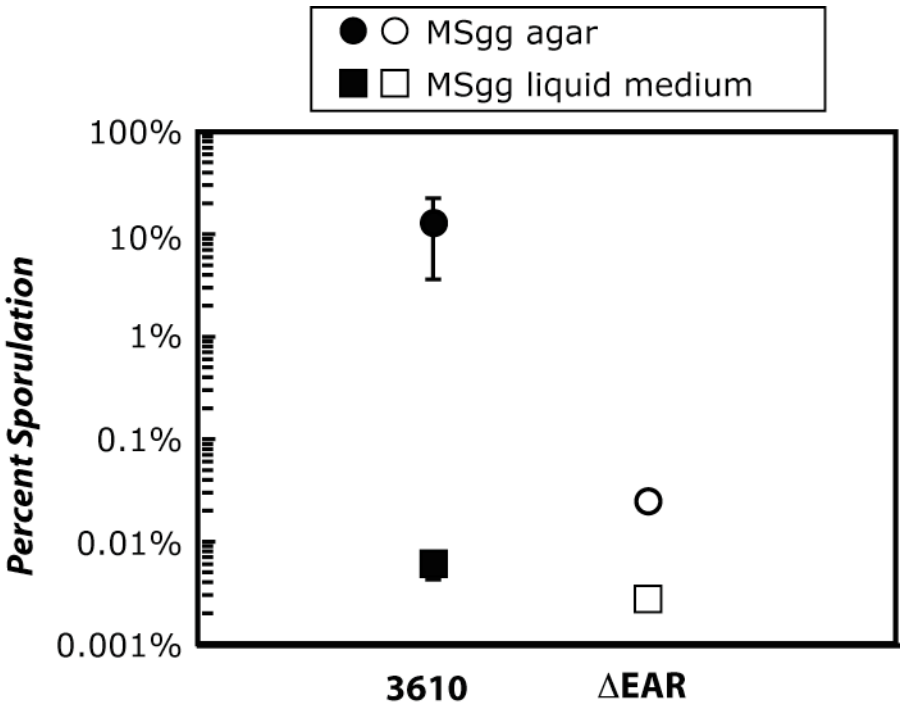
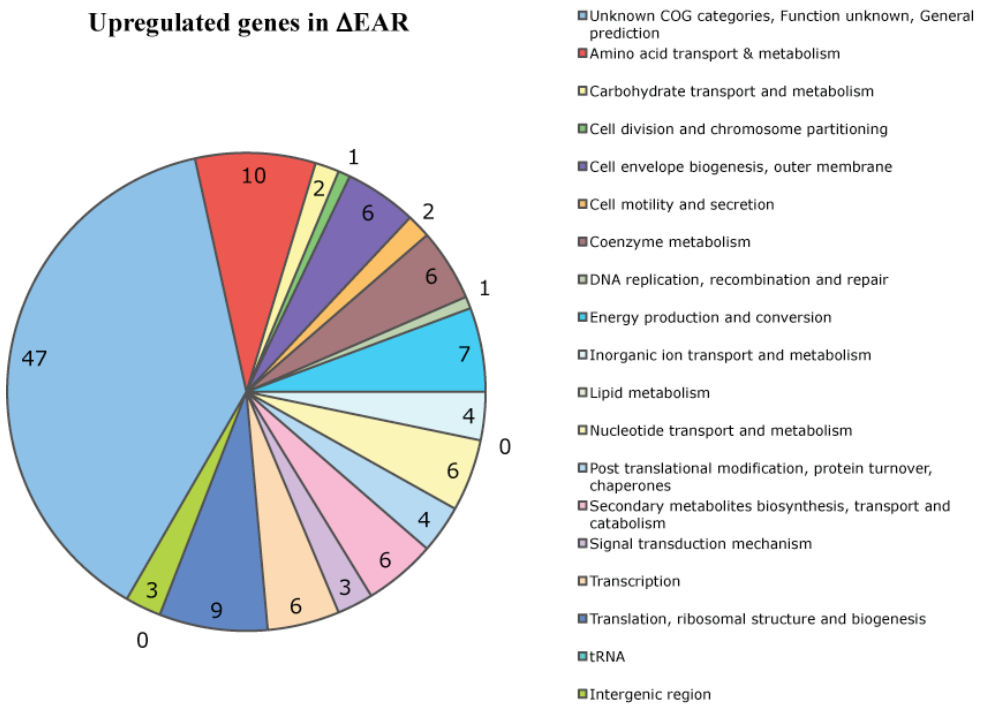


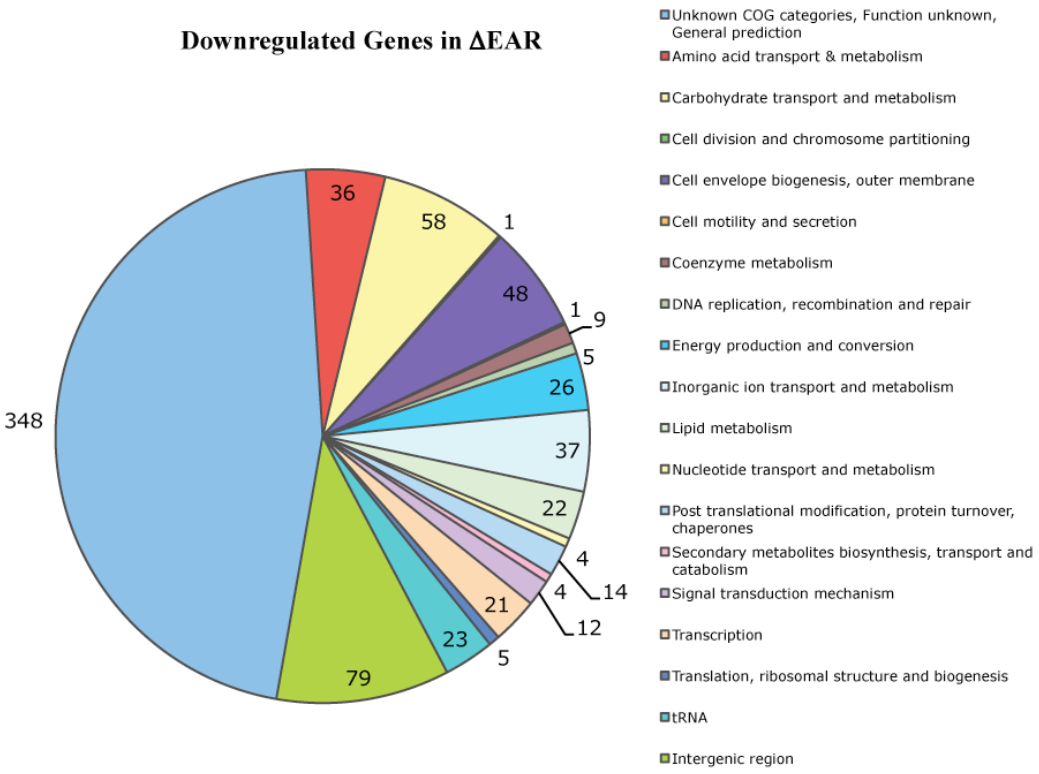
Figure 3-11. Deletion of the EAR element negatively affects sporulation. Sporulation assays were done using NCIB3610 (closed) and ΔEAR (open) cells harvested from either complex colonies grown on solid MSgg agar (circle) or from cultures grown in MSgg liquid medium (square). y-axis represents the percentage of spores from the chloroform-treated cells compared to the non-treated control. Only the cells grown on solid MSgg agar were affected by EAR deletion.



Increase

COG Functional Category	# of genes
Unknown COG categories, Function unknown, General prediction	47
Amino acid transport & metabolism	10
Carbohydrate transport and metabolism	2
Cell division and chromosome partitioning	1
Cell envelope biogenesis, outer membrane	6
Cell motility and secretion	2
Coenzyme metabolism	6
DNA replication, recombination and repair	1
Energy production and conversion	7
Inorganic ion transport and metabolism	4
Lipid metabolism	0
Nucleotide transport and metabolism	6
Post translational modification, protein turnover, chaperones	4
Secondary metabolites biosynthesis, transport and catabolism	6
Signal transduction mechanism	3
Transcription	6
Translation, ribosomal structure and biogenesis	9
tRNA	0
Intergenic region	3
Total	123

Figure 3-12. Microarray analysis of ΔEAR as compared to NCIB3610: graphical representation of transcripts that are increased in the ΔEAR mutant. The expression level of approximately 123 different genes and intergenic regions were significantly increased in the ΔEAR strain as compared to wild-type bacteria when cultured on solid medium under biofilm-promoting condition. Genes and intergenic regions that were upregulated were grouped based on their COG functional categories and plotted as pie charts. The categories and the number of genes or intergenic regions in it are also listed in the table format.



Decrease

COG Functional Category	# of genes
Unknown COG categories, Function unknown, General prediction	348
Amino acid transport & metabolism	36
Carbohydrate transport and metabolism	58
Cell division and chromosome partitioning	1
Cell envelope biogenesis, outer membrane	48
Cell motility and secretion	1
Coenzyme metabolism	9
DNA replication, recombination and repair	5
Energy production and conversion	26
Inorganic ion transport and metabolism	37
Lipid metabolism	22
Nucleotide transport and metabolism	4
Post translational modification, protein turnover, chaperones	14
Secondary metabolites biosynthesis, transport and catabolism	4
Signal transduction mechanism	12
Transcription	21
Translation, ribosomal structure and biogenesis	5
tRNA	23
Intergenic region	79
Total	753

Figure 3-13. Microarray analysis of ΔEAR as compared to NCIB3610: graphical representation of transcripts that are decreased in the ΔEAR mutant. The expression level of approximately 753 different genes and intergenic regions were significantly decreased in the ΔEAR strain as compared to wild-type bacteria when cultured on solid medium under biofilm-promoting condition. Genes and intergenic regions that were downregulated were grouped based on their COG functional categories and plotted as pie charts. The categories and the number of genes or intergenic regions in it are also listed in the table format.

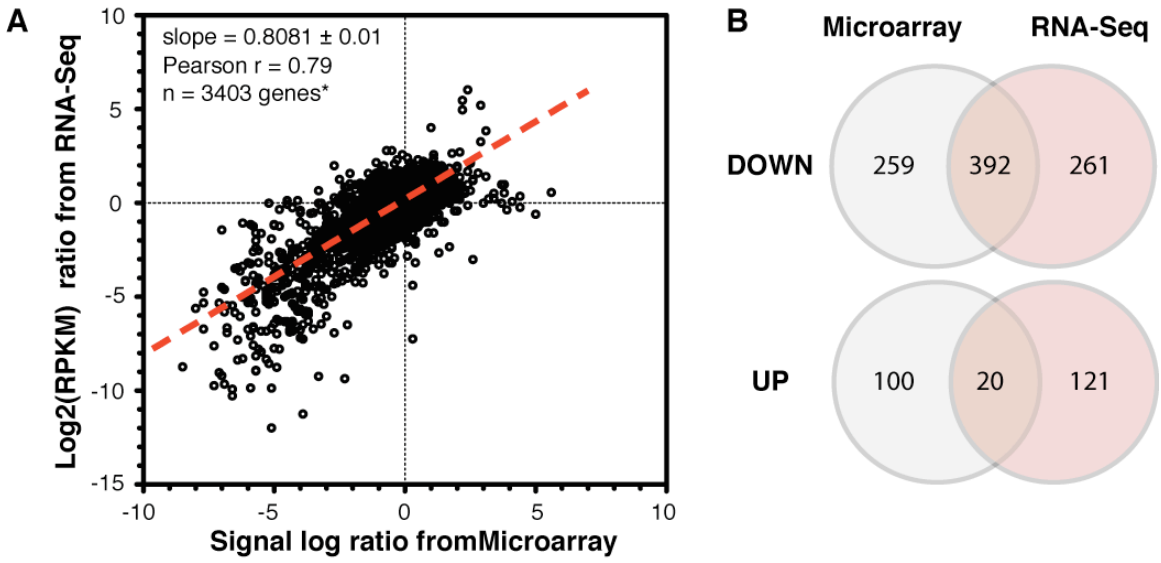
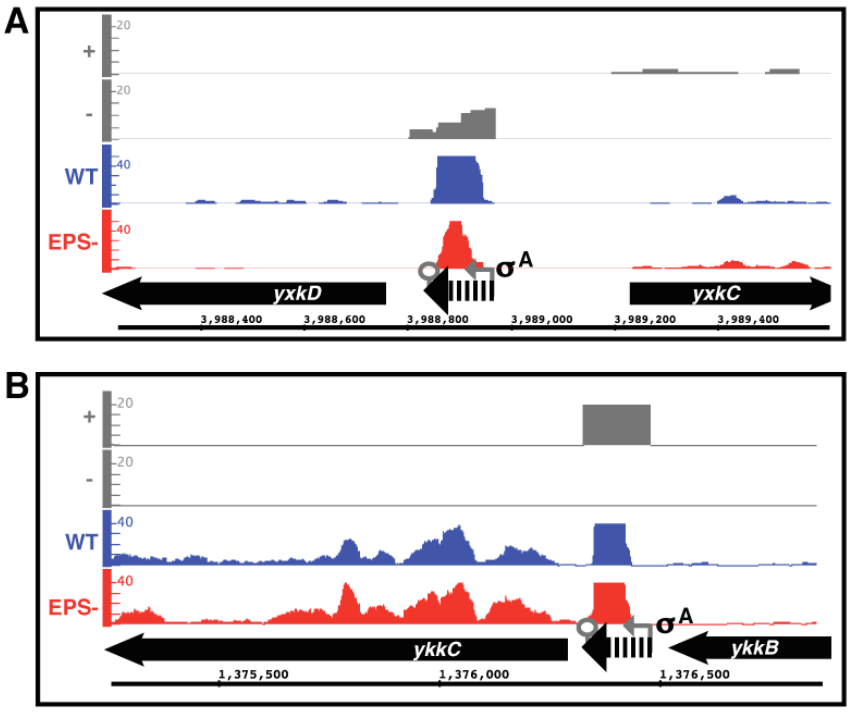


Figure 3-14. A positive correlation between the microarray and the RNA-Seq transcriptomic analyses. (A) Comparison of the gene expression changes observed by microarray (*x*-axis) and RNA-Seq (*y*-axis). Fold-changes observed by microarray were shown as signal log ratio calculated using Affymetrix Gene Chip Operating System (GCOS). Fold-changes observed by RNA-Seq were shown as $\log_2$ of the RPKM ratio calculated according to Mortazavi *et al.* (2008). Due to differences in genomic information used in the two analyses (*e.g.,* gene names; see text), only 3403 genes present in both datasets (denoted by O) were used to measure the correlation. (B) Venn diagram showing the overlaps between the microarray (grey circle) and the RNA-Seq (red circle) datasets. The numbers represent genes with at least 3 fold lower or higher expressions compared to wild-type.



C

		Fold change (EPS- vs WT)
yxkD	RNA	-4.7
yxkD	coding	-4.4
ykkC	RNA	1.0
ykkC	coding	-0.7

Figure 3-15. The expression and genomic context of *B. subtilis yxkD* and *ykkC* riboswitches. (A-B) The genomic locus of each RNA element 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. The 5`-transcription start site mapping data for both the positive and negative genomic strands are shown in grey (described in Chapter 4). Dashed arrow indicates the transcriptional unit of the RNA of interest. If available, the transcriptional start site is denoted by grey arrow and the potential transcription terminator is denoted by grey circle. The predicted sigma factor is also shown in bold near the transcription start site (DBTBS; Sierro *et al.*, 2008). (C) Quantification of the expression changes for each riboswitch and their corresponding downstream gene within the wild-type and EPS deficient biofilm. Fold-change is calculated based on RPKM measurement (Mortazavi *et al.*, 2008). Negative value corresponds to a lower expression in EPS- strain, while positive value corresponds to a higher expression in EPS- strain.

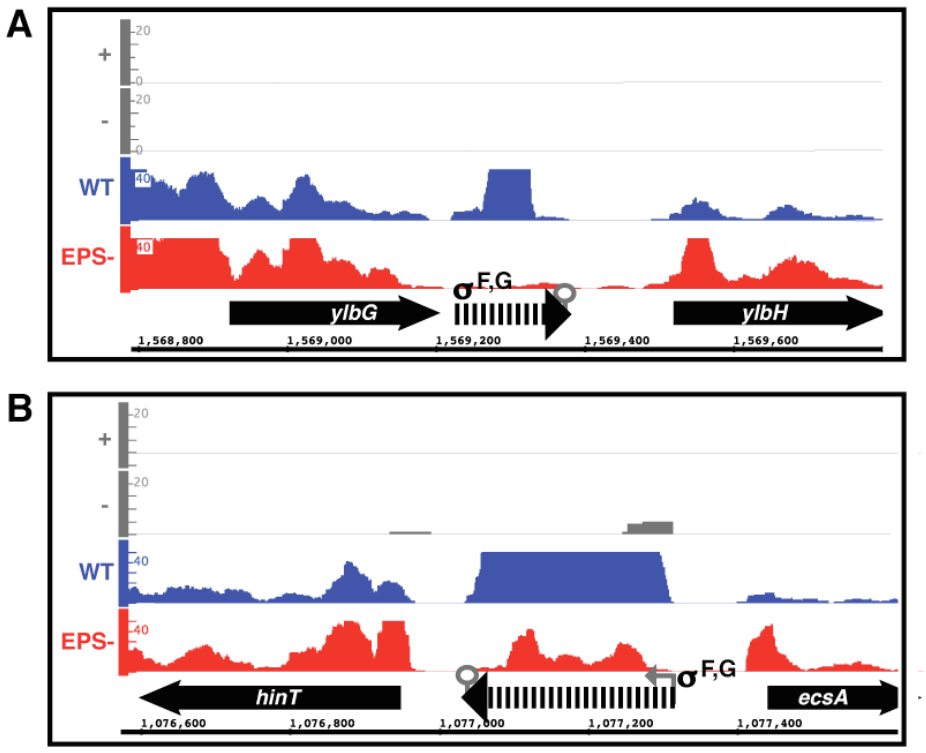


Figure 3-16. The expression and genomic context of the *ylbG-ylbH* and *hinT-ecsA* RNAs. (A-B) The genomic locus of each RNA element is shown at the bottom with the two flanking genes. The distribution of cDNA reads obtained from the RNA-Seq analysis of the wild-type and EPS deficient biofilm colonies are represented by the blue and red graphs, respectively. The 5`-transcription start site mapping data for both the positive and negative genomic strands are shown in grey (described in Chapter 4). Dashed arrow indicates the transcriptional unit of the RNA of interest. If available, the transcriptional start site is denoted by grey arrow and the potential transcription terminator is denoted by grey circle. The predicted sigma factor is also shown in bold near the transcription start site (DBTBS; Sierro *et al.*, 2008).

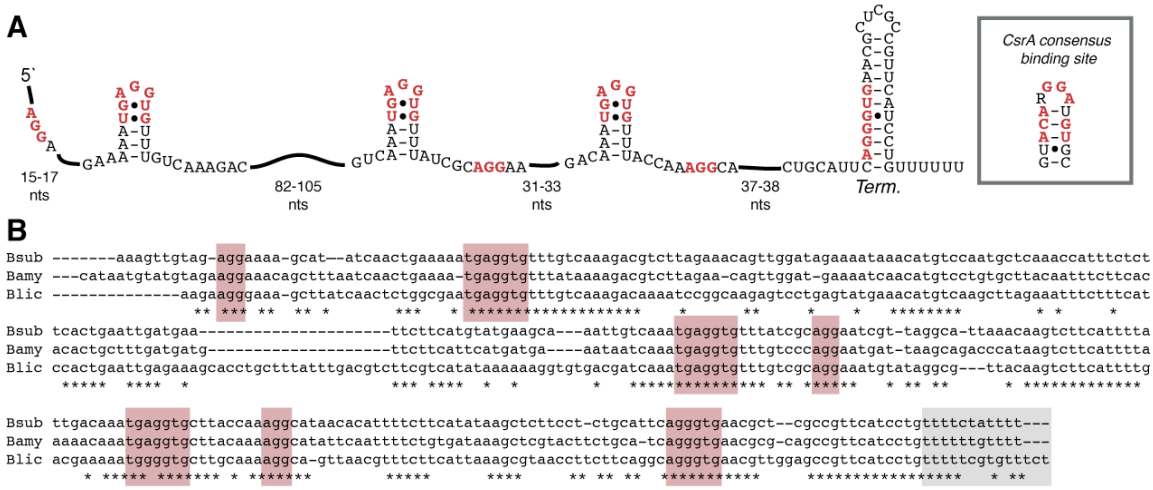


Figure 3-17. The candidate CsrA-sequestering RNA in *Bacillus subtilis*. (A) The predicted secondary structure of the *hinT-ecsA* RNA showing the potential CsrA-binding stem-loop structures. The conserved "5`-AGG-3`" motif is colored in red. Black lines represent the variable regions. The consensus binding site for *E. coli* CsrA (Dubey *et al.*, 2005) is shown in the inset for comparison. (B) Sequence alignment of *hinT-ecsA* from several *Bacillus* species. Potential CsrA binding sites and the "5-AGG-3" motif are shaded in red. The end of the terminator element is shaded in grey.

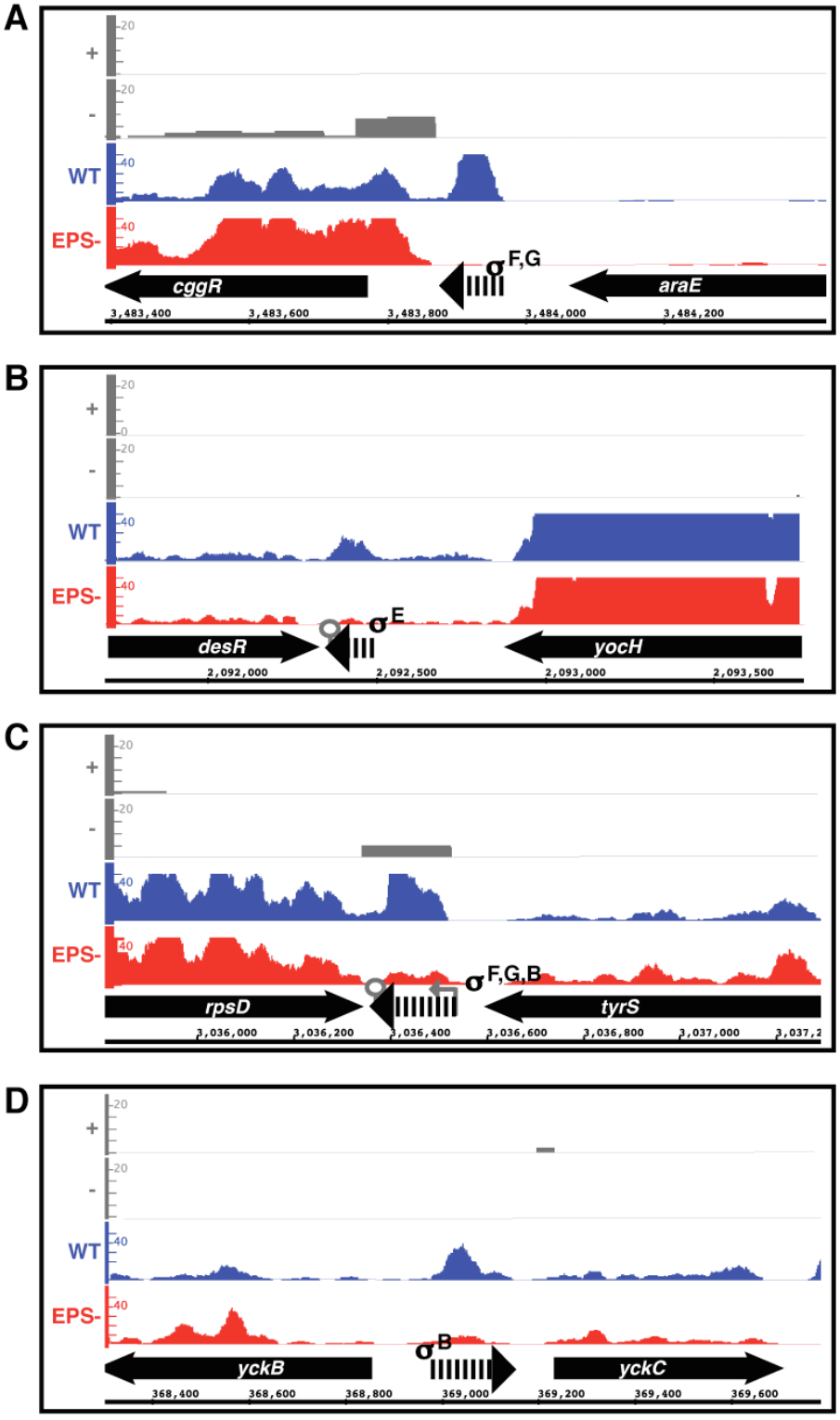

A
WT
EPS-
σF,G
cggR
araE
3,483,400
3,483,600
3,483,800
3,484,000
3,484,200
B
WT
EPS-
σE
desR
yocH
2,092,000
2,092,500
2,093,000
2,093,500
C
WT
EPS-
σF,G,B
rpsD
tyrS
3,036,000
3,036,200
3,036,400
3,036,600
3,036,800
3,037,000
D
WT
EPS-
σB
yckB
yckC
368,400
368,600
368,800
369,000
369,200
369,400
369,600

Figure 3-18. The expression and genomic context of the *cggR-araE*, *desR-yocH*, *rpsD-tyrS*, and *yckB-yckC* RNAs. (A-D) The genomic locus of each RNA element 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. The 5`-transcription start site mapping data for both the positive and negative genomic strands are shown in grey (described in Chapter 4). Dashed arrow indicates the transcriptional unit of the RNA of interest. If available, the transcriptional start site is denoted by grey arrow and the potential transcription terminator is denoted by grey circle. The predicted sigma factor is also shown in bold near the transcription start site (DBTBS; Sierro *et al.*, 2008).

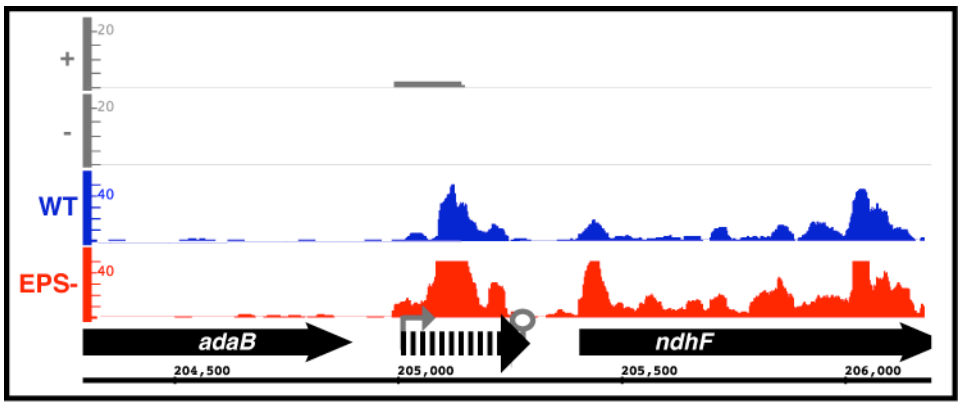


Figure 3-19. The expression and genomic context of the *adaB-ndhF* RNA. The genomic locus of this RNA element is shown at the bottom with the two flanking genes.The distribution of cDNA reads obtained from the RNA-Seq analysis of the wild-type and EPS deficient biofilm colonies are represented by the blue and red graphs, respectively. The 5`-transcription start site mapping data for both the positive and negative genomic strands are shown in grey (described in Chapter 4). Dashed arrow indicates the transcriptional unit of the RNA of interest. If available, the transcriptional start site is denoted by grey arrow and the potential transcription terminator is denoted by grey circle. The predicted sigma factor is also shown in bold near the transcription start site (DBTBS; Sierro *et al.*, 2008).

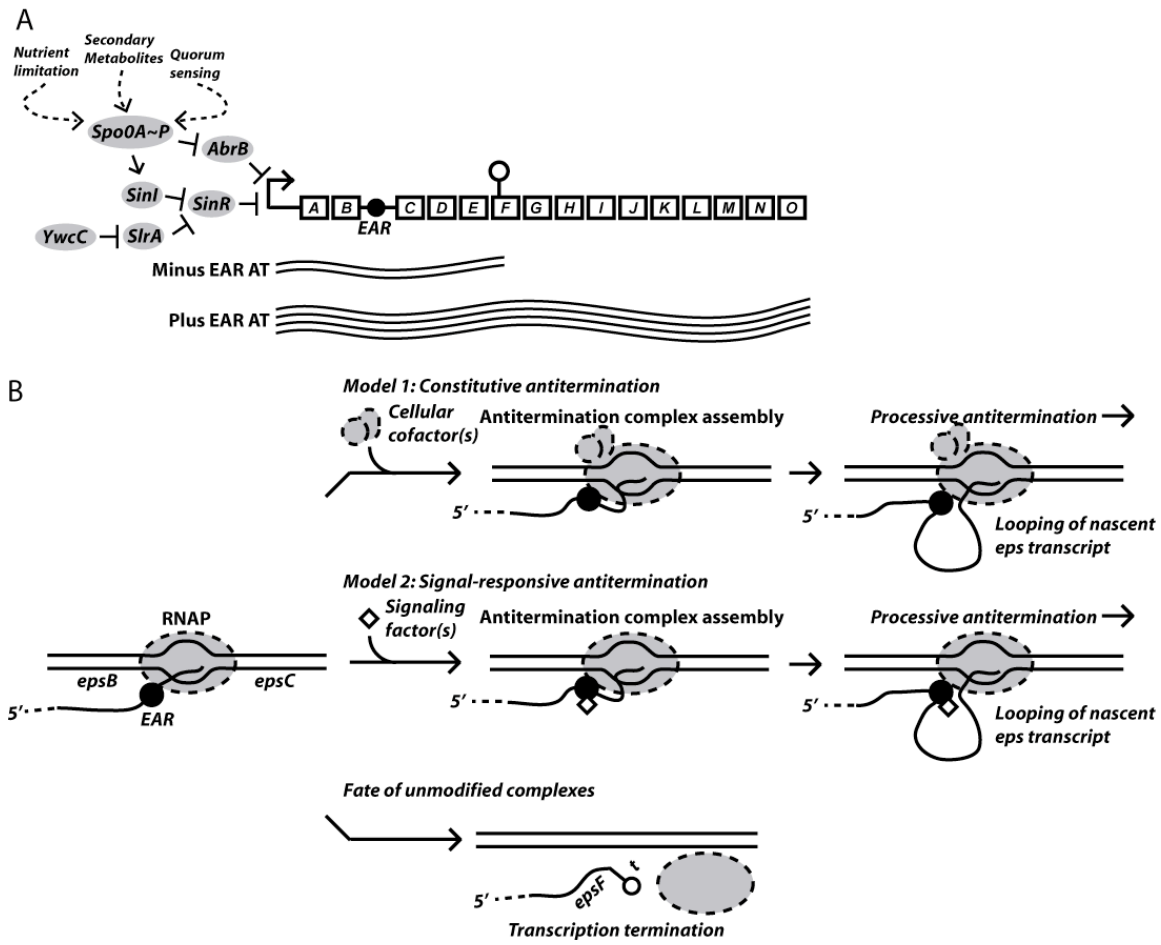


Figure 3-20. The model for EAR-mediated antitermination of the *eps* operon. (A) The *eps* operon is directly repressed by the master regulator for biofilm formation, SinR. Under biofilm-promoting conditions, inactivation of SinR (by SinI and SlrA) and repression of AbrB leads to increased transcription of the *eps* operon. Activation of EAR antitermination allows for complete synthesis of the long *eps* transcript. In the absence of EAR antitermination, most of the transcripts will be terminated at the *epsF* locus. (B) From the data presented herein, we hypothesize that the EAR element promotes *eps* synthesis via two possible models, which are not mutually exclusive. Constitutive antitermination could result from cellular co-factors that assemble with EAR and the transcription elongation complex to promote processive antitermination. Alternatively, processive antitermination may require an unknown signal, in addition to the EAR element.

Table 3-1. Expression of representative cell type-specific markers in EPS deficient biofilm

Gene	Function	Fold-change (EPS- vs WT)[a]	
		Microarray[b]	RNA-Seq[b]
Spores			
sigF	forespore-specific sporulation sigma factor	-4.6	-2.2
sigG	mother cell-specific sporulation sigma factor	-24.3	-13.3
sigE	forespore-specific sporulation sigma factor	-18.4	-6.0
spoIIIC	late mother cell-specific sporulation sigma factor (C-terminal portion of SigK)	-12.1	-6.3
spoIVCB	late mother cell-specific sporulation sigma factor (N-terminal portion of SigK)	-8.0	-6.2
Competent cells			
comK	late competence regulator	-3.0	-3.6
*comGA**	late competence gene required for exogenous DNA binding	-1.3	-7.7
Miners			
aprE	Extracellular alkaline serine protease, subtilisin	-5.3	-4.4
bpr	bacillopeptidase	-5.7	-2.9
Matrix producers			
tasA (cotN)*	matrix protein	7.0	9.5
epsA (yveK)*	exopolysaccharide production	2.5	3.2
Cannibals			
sdpA(yvaW)*	cannibalism toxin, Sdp	3.0	3.5
skfA (ybcO)*	cannibalism toxin, Skf	-4.3	-2.0
Motile cells			
hag	flagellin protein	-1.1	1.3
Surfactin-producers			
*srfAA**	surfactin synthethase	-1.3[c]	-1.1

*Only the first gene in the operon is shown. The expression levels for other genes in the operon can be found in Appendix I. The relevant operons are: *comGA-GG*, *yqxM-sipW-tasA*, *epsA-O (yveK-T, yvfA-F)*, *sdpABC* (*yvaWXY*), *skfA-H* (*ybcOPST, ybdABDE*), *srfAA-AD*. Previous gene names are listed in parentheses.

[a]Fold change is calculated based on microarray signal ratio and RNA-Seq RPKM measurement (Mortazavi *et al.*, 2008). Negative value corresponds to a lower expression in EPS- strain, while positive value corresponds to a higher expression in EPS- strain.

[b]For EPS- strains, ΔEAR was used in microarray experiment while M3 was used in RNA-Seq (see Irnov and Winkler, 2010).

[c]*srfAA* was represented by multiple probes in the microarray chips. This value is based on the average probes intensity for *srfAA*.

Table 3-2. Expression of various sigma and transcription factors in EPS deficient biofilm

Gene	Function	Fold-change (EPS- *vs* WT)[a]	
		Microarray[b]	RNA-Seq[b]
DOWN			
sigF	forespore-specific sporulation sigma factor	-4.6	-2.1
sigE	mother cell-specific sporulation sigma factor	-18.4	-6.3
sigG	forespore-specific sporulation sigma factor	-24.3	-12.5
spoIIIC	late mother cell-specific sporulation sigma factor (C-terminal portion of SigK)	-8.0	-6.3
spoIVCB	late mother cell-specific sporulation sigma factor (N-terminal portion of SigK)	-12.1	-6.3
sigV	extracytoplasmic sigma factor	-4.3	-2.0
sigY	extracytoplasmic sigma factor	-2.6	-4.6
spoVT	regulates SigG target genes	-2.6	-2.6
spoIIID	regulates SigE and SigK target genes	-500.0	-500.0
gerE	regulates SigK target genes	-59.7	-142.9
rsfA (ywfV)	anti-anti SigE	-26.0	-50.0
gerR (ylbO)	regulates SigE and SigK target genes	-18.4	-5.6
acoR	acetoin degradation	-4.3	-3.0
fadR (ysiA)	fatty acid degradation	-3.7	-4.4
tnrA	regulates gene expression under nitrogen-limiting growth condition	-3.0	-2.7
comK	late competence transcription factor	-3.0	-3.6
gntR	gluconate utilization	-4.0	-3.6
pucR (yunI)	purine catabolism	-3.5	-5.6
LITTLE/NO CHANGE			
sigH	post-exponential phase gene expression	2.1	1.5
sigD	regulates cell surface proteins, motility, autolysis	2.0	-1.3
sigB	general stress-response sigma factor	-1.1	-2.6
sigA	major vegetative sigma factor	1.6	1.2
sigI	heat-shock induced sigma factor	1.7	1.3
sigL	regulates carbon and nitrogen metabolism genes	-1.5	-1.1
sigX	extracytoplasmic sigma factor	2.1	1.5
sigM (yhdM)	extracytoplasmic sigma factor	1.9	1.1
sigW	extracytoplasmic sigma factor	1.4	1.6
sigZ	extracytoplasmic sigma factor	1.2	1.6
ylaC	extracytoplasmic sigma factor	1.6	1.2
sinR	regulates genes involved in biofilm formation	-1.4	1.2
spo0A	regulates the initiation of cellular differentiation	-1.3	-1.1
abrB	transition state regulator	1.1	-1.1
abh	SigX-activated AbrB homolog	-1.2	-1.1
codY	GTP and branched amino acid-responsive global transcription regulator	1.5	1.2
degU	regulates genes involved in exoprotease production, motility, biofilm	1.5	1.4
comA	quorum-sensing dependent competence regulator	1.7	1.4

[a]Fold change is calculated based on microarray signal ratio and RNA-Seq RPKM measurement (Mortazavi *et al.*, 2008). Negative value corresponds to a lower expression in EPS- strain, while positive value corresponds to a higher expression in EPS- strain.

[b]For EPS- strains, ΔEAR was used in microarray experiment while M3 was used in RNA-Seq (see Irnov and Winkler, 2010).

Table 3-3. Candidate biofilm-specific regulatory RNAs

	SIZE[a] (nts)	PREV GENE	NEXT GENE	ORIENTATION[b]	#cDNA WT	#cDNA EPS-	FOLD CHANGE[c] (EPS- *vs* WT)	DIST.[e]	NAME
DOWN									
Peak3857	85	*yneK*	*cotM*	← → ←	203	0	-98.6[d]	1	ncr992[f]
Peak7015	80	*cggR*	*araE*	← ← ←	94	1	-94.8	1,2	-
Peak3022	113	*ylbG*	*ylbH*	→ ← →	276	8	-34.8	5,6,7,8, 9	*ylbH* orphan riboswitch[g,i]
Peak2043	275	*hinT*	*ecsA*	← ← →	4167	108	-38.9	1,2	ncr1670[f]
Peak5380	137	*yrpD*	*yrpE*	→ → ←	1307	51	-25.8	1,3,4	-[h]
Peak4213	133	*desR*	*yocH*	→ ← ←	60	12	-5.0	(-)	*yocG-yocH*[h]
Peak7939	101	*yxkD*	*yxkC*	← ← →	352	76	-4.7	Diverse	*ykkC/yxkD* orphan riboswitch[g]
Peak6013	180	*rpsD*	*tyrS*	→ ← ←	162	50	-3.3	1	ncr2360[f]
Peak726	107	*yckB*	*yckC*	← → →	64	19	-3.4	1	-
UP									
Peak827	156	*adaB*	*ndhF*	→ ← →	95	314	3.3	(-)	ncr181[f]

[a]Size is calculated as the length of continuous nucleotide positions with 5 or more cDNA reads.

[b]The orientation of candidate RNA is determined based on our previous transcription start site analysis (Irnov *et al.*, 2010) and/or the presence of transcription terminator elements.

[c]Fold change is calculated based on RPKM measurement (Mortazavi *et al.*, 2008). Negative value indicates candidate RNAs with lower expression in EPS- strain, while positive value indicates candidate RNAs with higher expression in EPS- strain.

[d]# of cDNA in EPS- strain is counted as 1 for calculating fold change.

[e](-): *B. subtilis only*, (1) *B. amyloliquefaciens*, (2) *B. licheniformis*, (3) *B. pumilus*, (4) *B. megaterium*, (5) *B. anthracis*, (6) *B. cereus*, (7) *B. halodurans*, (8) *O. iheyensis*, (9) *T. tencongensis*, (diverse) also present outside of the Firmicutes.

[f]see Irnov *et al.,* 2010.

[g]see Barrick *et al.*, 2004.

[h]see Schmalish *et al.*, 2010.

[i]Although it was previously annotated as a riboswitch, this element is actually a *trans*-acting small RNA. See text for discussion.

Table 3-4. Previously identified small regulatory RNAs not affected by EPS deficiency

NAME	PREV GENE	NEXT GENE	#cDNA WT	#cDNA EPS-	FOLD CHANGE[a] (EPS- *vs* WT)	REF.[b]
ncr2768	*pbpG*	*ywhD*	21	21	1	1
ncr1575/ncr10	*vmlR*	*ydgF*	241	414	-1.7	1,2
ncr952	*mutL*	*ymzD*	180	303	1.7	1
ncr1015	*pps*	*xynA*	61	51	-1.2	1
ncr738	*ykwD*	*pbpH*	24	31	1.3	1
ncr560/ncr18	*yhaZ*	*yhaX*	2248	2906	1.3	1,2
ncr1957	*ypbR*	*ypbQ*	26	73	2.78	1
ncr1566	*cspC*	*ydeB*	39	50	1.3	1
ncr1755/35	*zosA*	*ykvY*	56	86	1.5	1,2
ncr1733/ncr26	*ykcC*	*htrA*	537	758	1.4	1,2
ncr977	*cotU*	*thyA*	18	7	-2.6	1
ncr2752/ncr75	*rho*	*ywjI*	35	86	2.4	1,2
ncr6	*yceI*	*yceJ*	18	18	1	2
ncr14	*bdhA*	*ydjM*	22	54	2.4	2
ncr42	*cwlS*	*yojK*	27	39	1.4	2
ncr47	*ypjP*	*ypiP*	35	38	1.1	2
ncr59	*yrkN*	*yrkL*	16	27	1.7	2
ncr63	*accD*	*ytsJ*	26	59	2.3	2
FsrA	*ykuI*	*ykuJ*	112	95	-1.2	3
RsaE	*yizD*	*yjbH*	747	1619	2.2	1,2
SurA	*yndK*	*yndL*	115	89	-1.3	4

[a]Fold change is calculated based on RPKM measurement (Mortazavi *et al.*, 2008). Negative value indicates candidate RNAs with lower expression in EPS- strain, while positive value indicates candidate RNAs with higher expression in EPS- strain. We consider < 3-fold change as not significant.

[b]References: (1) Irnov *et al.*, 2010; (2) Rasmussen *et al.*, 2009; (3) Gaballa *et al.*, 2008; (4) Silvaggi *et al.*, 2005.

CHAPTER FOUR

Deep sequencing-based analysis of the *Bacillus subtilis* transcriptome: Discovering novel regulatory RNAs

Introduction

Many RNA-mediated genetic control mechanisms have already been demonstrated to be significantly important for *Bacillus subtilis*. For example, at least 4% of the genome is believed to be subject to control by *cis*-acting regulatory RNAs alone (Irnov *et al.*, 2006) and is still certain to underestimate the full degree of RNA-mediated regulation. However, little is known about the importance or mechanisms of *trans*-acting sRNAs. The majority of small RNAs studies have occurred with *E. coli* and other proteobacterial species, and it is unclear how meaningful they are for other bacteria. To that end, recent efforts have begun to uncover candidate sRNAs in non-proteobacterial species. Recently, sRNAs have been identified in Gram-positive bacteria, including *Bacillus anthracis, Listeria monocytogenes*, *Staphylococcus aureus*, *Streptococcus* species, and *Streptomyces coelicolor* (Christiansen *et al.*, 2006; Halfmann *et al.*, 2007; Mandin *et al.*, 2007; Boisset *et al.*, 2007; Swiercz *et al.*, 2008; Geissmann *et al.*, 2009; Nielsen *et al.*, 2010; Passalacqua *et al.*, 2009; Perez *et al.*, 2009; Tsui *et al.*, 2010; Romby and Charpentier, 2010). However, the common 'rules' for sRNA regulation, as well as their involvement with RNA-binding proteins, are not yet apparent from these analyses (Romby and Charpentier, 2010; Jousselin *et al.*, 2009). Historically, *Bacillus subtilis* has been used as the benchmark model microorganism for Gram-positive bacteria. Yet study of sRNA regulation in this organism is still in an early stage. Until recently, only about a

dozen of 'non-housekeeping' sRNAs have been identified in this organism. Therefore, rigorous analysis of the importance and molecular mechanisms for sRNA regulation in *B. subtilis* is important for elucidation of posttranscriptional regulation in Firmicutes and other Gram-positive bacteria.

Historically, identification of regulatory RNAs has been more accidental than methodical. Most were discovered from genetic and physiology-based study of specific genes or pathways of interest. More recently, researchers have been utilizing more systematic approaches that use various computational analyses with experimental verification of candidates. For instance, thermodynamic stability of intergenic sequences has been used to predict novel regulatory RNAs in several microorganisms (Workman and Krogh, 1999; Rivas and Eddy, 2000; Clote *et al.*, 2005; Washietl *et al.*, 2005; Uzilov *et al.*, 2006). Many candidate regulatory RNAs have also been identified by locating genetic elements that are common for RNA-based regulation. Specifically, using algorithms for identification of promoter and terminator elements in noncoding regions, investigators have identified many *trans*-acting regulatory RNAs in *E. coli* (Argaman *et al.*, 2001). Additionally, the prediction of terminator and antiterminator pairings within *B. subtilis* intergenic regions proved to be a relatively successful approach for identification of established regulatory RNAs, as well as the prediction of numerous additional candidates (Merino and Yanofsky, 2005). So far, the most successful approach for finding novel regulatory RNAs has been via comparative sequence analyses, which take into account primary sequence and secondary structure conservation (Barrick *et al.*, 2004; Corbino *et al.*, 2005; Weinberg *et al.*, 2007; Gardner *et al.*, 2009; Voss *et al.*, 2009; Xu *et al.*, 2009; Weinberg *et al.*, 2010). However, poorly conserved or newly evolved

regulatory RNAs might not be detected by these analyses. In the past several years, many investigators have started to use more global transcriptomic approaches, such as genomic tiling array and high-throughput deep sequencing, for discovering novel RNAs (*e.g.,* Livny *et al.*, 2008; Sittka *et al.*, 2008; Rasmussen *et al.*, 2009; Toledo-Arana *et al.*, 2009; Sharma *et al.*, 2010; Tsui *et al.*, 2010). In addition to being unbiased, these approaches also provide gene expression information that is often useful for identifying the biological roles of various RNA elements.

A recent publication describing identification of *B. subtilis* regulatory RNAs via mapping of transcriptionally active regions by high-density oligonucleotide tiling arrays successfully identified most known sRNAs, as well other new candidates, from exponential growth condition (Rasmussen *et al.*, 2009). Herein, we describe the discovery of sRNAs and other putative regulatory RNA elements using a sister technique to the latter study. Specifically, we utilized a differential RNA-sequencing (dRNA-seq) approach that is selective for transcriptional start sites (TSS) (Sharma *et al.*, 2010). Our approach successfully recovers most known regulatory RNAs as well as a portion of the microarray-predicted sRNAs, while also identifying many unique candidates. This catalog of candidate regulatory RNA elements will serve as important reference point for comprehensive analyses of sRNA regulation in *B. subtilis* and other *Bacillus* species.

Results and Discussion

Identification of regulatory RNAs using differential RNA-sequencing (dRNA-Seq)

Bacillus subtilis strain 168 cells were cultured in glucose minimal medium until stationary phase, whereupon total RNA was extracted using standard methods. Most

bacterial RNAs, including both mRNAs and sRNAs, contain a triphosphate moiety at their 5' terminus, whereas processed transcripts, such as rRNAs and tRNAs, contain a 5' monophosphate. To specifically enrich our samples for primary transcripts, half of the sample was treated with terminator exonuclease that preferentially degrades 5'-monophosphorylated RNAs. cDNA libraries were then prepared from these two pools (‘unenriched’ and ‘enriched’) and analyzed by 454 pyrosequencing. After 5`-linker and poly-A tail removal, a total of 406,531 cDNA reads (>15 nucleotides in length) could be successfully mapped to the *B. subtilis* genome. Most of these sequences corresponded to ribosomal RNAs and tRNAs, which were excluded from further analysis. At the end, 25,675 and 44,098 cDNA reads were obtained for the unenriched and enriched libraries, respectively, and visualized with the Integrated Genome Browser (Appendix II) (IGB, Affymetrix). The majority mapped to intergenic regions, due to the enrichment for cDNA reads located proximal to the transcription start site (Sharma *et al.*, 2010). Based on these data we were able to identify approximately 600 potential transcription start sites in the *B. subtilis* genome, which appeared to be modestly increased nearer to the origin of replication (data not shown). From this analysis, classes of transcriptional start sites could be identified for both sense and antisense RNAs (Fig. 4-1). Also, signals that were likely to correspond to long 5' mRNA leader regions could be identified for some genes, whereas other cDNA reads appeared likely to correspond to small regulatory RNAs (sRNAs) (Fig. 4-1). We concentrated our efforts on the latter classes, which are most likely to include *cis*- and *trans*-acting regulatory RNAs.

Identification of 'long' 5' leader regions

B. subtilis employs a wide variety of *cis*-acting regulatory RNA elements (Irnov *et al.*, 2006). In contrast to the average length of approximately 360 nts (+/-150 nts) for 5' leader regions of transcripts including *cis*-acting regulatory RNAs, the average overall leader length based on the TSS map peaks around 35 nts (Fig. 4-2) (34). This observation suggests that 'long' 5' leader regions are likely to occur only when specialized functions are encrypted within them. One possible function of a long 5' leader region is to incorporate structural elements that affect the stability of the overall transcript. Alternatively, some long 5' leader regions include within them sequence and structural components that help guide intracellular mRNA localization. However, the most likely explanation for a long 5' mRNA leader region is due to inclusion of a *cis*-acting, signal-responsive regulatory RNA. Therefore, unbiased experimental methods capable of identifying long 5' leader regions, such as high-throughput sequencing of transcriptional start sites, offer a potentially powerful approach for discovery of new regulatory RNA elements. Current bioinformatics-based approaches are likely to include bias for phylogenetically widespread and highly conserved regulatory RNAs. In contrast, unbiased mapping of transcriptional start sites is expected to uncover 5' mRNA leader regions without regard for phylogenetic distribution, even when they include poorly conserved, recently evolved, or highly degenerate *cis*-acting regulatory sequences.

To that end, we investigated whether 454 pyrosequencing of *B. subtilis* stationary phase RNAs was capable of identifying previously established long 5' leader regions. Previous data have established a minimum of 24 protein-responsive *cis*-acting regulatory RNAs, 19 tRNA-responsive *cis*-acting regulatory RNAs, 32 metabolite-responsive *cis*-

acting regulatory RNAs and one metal-sensing regulatory RNA (Irnov *et al.*, 2006; Dann *et al.*, 2007). Long 5' leader regions were correctly identified in our data set for 68%, 68%, and 67% of tRNA-, metabolite-, and protein-responsive regulatory RNAs, respectively (data not shown). Moreover, a qualitative assessment of putative start sites determined in our dataset matched on average within one nucleotide of the previously established start sites (as cataloged by DBTBS; Sierro *et al.*, 2008). Of the ~600 putative transcription start sites identified herein, 93 were located at least 100 nts away from the downstream gene and did not already correspond to a known long leader region (data not shown). There are bound to be false positives within this data set, *i.e.,* start sites that do not correspond to synthesis of a long UTR. For example, it is possible that transcription could initiate upstream of a gene for synthesis of a separate, unique sRNA gene, having nothing to do with expression of the downstream gene. Therefore, as a conservative strategy for specifically identifying long leader regions, we assessed most closely those cDNA reads that start within an intergenic region but that cumulatively overlap with the downstream coding sequence (or that end within 10 nucleotides of the downstream gene). We assumed that this arrangement would result in the highest confidence for assigning leader regions. A total of 40 examples fit this description (Table 4-1). The fact that the cDNA signals for the remaining 53 stop upstream of the downstream coding region does not automatically eliminate them as corresponding to 5' leader regions. Indeed, many previously established *cis*-acting regulatory RNAs resemble the latter pattern, presumably due to the presence of an intrinsic terminator element before the coding region. Also, several transcripts that have been demonstrated to contain long 5' leader regions of unknown function appear in our data (*e.g., srfAA*, *yxbB*), although their cDNA reads do

not fully continue into the downstream gene.

Interestingly, several of the transcripts with newly identified long leader regions can be grouped relative to their expected functional roles. For example, five such transcripts contained moderately 'long' 5' leader regions (104, 104, 119, 121, and 143 nts) upstream of genes encoding ribosomal protein homologues. Post-initiation regulation of ribosomal protein genes is common in bacteria. Oftentimes, ribosomal proteins (r-proteins) bind to structural motifs located within the leader region to control expression of the downstream genes, in order to coordinate their overall stoichiometry with other r-proteins (Nomura *et al.*, 1980; Grundy and Henkin, 1991; Choonee *et al.*, 2007; Li *et al.*, 1997). It is possible that the moderately long 5' leader regions identified herein are required for similar regulatory mechanisms. Certain RNase enzymes have also been demonstrated to posttranscriptionally autoregulate their expression by interacting within their 5' leader region (Schuck *et al.*, 2009). Therefore, the fact that the recently identified RNase Y (*ymdA*) gene appears to be preceded by a 164 nt leader region could be suggestive of a similar mechanism. Also, transcripts encoding for certain core transcription elongation subunits (*rpoB*, *greA, nusA – located in an operon with ylxS*,) also appear from our data to contain a long 5' leader region, suggesting they also may be subjected to post-initiation control.

Another functionally related group of transcripts within this list encode for central metabolism genes. For example, the 5' leader regions for certain tricarboxylic acid cycle transcripts, including pyruvate dehyrogenase (*pdhA*), citrate synthase (*citZ*), and succinate dehydrogenase (*odhA*), appear to be 212, 199, and 100 nts in length. Similarly, the 5' leader region for an oxidative phosphorylation gene, menaquinol oxidase (*qoxA*), is

248 nts in length. A single glycolysis-related transcript, which encodes for fructose 1,6 bisphosphate aldolase (*fbaA*), also contains a long 5' leader region (111 nts). It remains to be determined whether there are any common sequence or structural features between the 5' leader portions of these central metabolism transcripts. Further experimentation will be required to assess whether the new 5' leader regions identified in this study contain within them elements that are important for posttranscriptional regulation of their associated genes.

Identification and detection of small RNAs

One of our primary motivations for performing the experimentation described herein was to validate previously established sRNAs and, more preferably, to discover new examples. There have been 14 non-housekeeping sRNAs identified previously in *B. subtilis*, although only a few have been studied in detail. Two have been identified as 6S RNAs (6S-1 and 6S-2; Barrick *et al.*, 2005; Trotochaud and Wassarman, 2005). Other studies have predicted a small suite of candidates, some of which may be under control of sporulation-specific sigma factors (Rasmussen *et al.*, 2009; Silvaggi *et al.*, 2006; Saito *et al.*, 2009). Of these candidates, mRNA targets have been experimentally identified for only a few. For example, two antisense RNAs have been demonstrated to regulate a toxin gene (*txpA*) and an unknown gene (*yabE*), respectively (Silvaggi *et al.*, 2005; Eiamphungporn and Helmann, 2009). One sRNA, SR1, controls expression of a transcriptional activator of arginine catabolism, AhrC (Heidrich *et al.*, 2006), while another, FsrA, controls iron-responsive genes (*sdhCAB, citB, yvfW, leuCD*) (Gaballa *et al.*, 2008). Recent discoveries in *S. aureus* revealed an sRNA (RsaE) that is widely

conserved amongst Gram-positive species, including *B. subtilis*. It appears to target central metabolism genes and *cstA*, which encodes for a 'carbon starvation' gene (Geissmann *et al.*, 2009). A more recent investigation, which examined the global transcriptional profile of *B. subtilis* by high-density oligonucleotide tiling arrays, resulted in identification of 54 new sRNA candidates (Rasmussen *et al.*, 2009). This raised the total number of sRNAs proposed for *B. subtilis* to approximately 70. To find novel sRNA candidates using 454 pyrosequencing of stationary phase RNAs, we searched for cDNA peaks that occurred specifically and entirely within intergenic regions, and which oftentimes included an identifiable intrinsic transcription terminator at the 3' terminus. Of the 14 previously identified sRNAs, our analysis recovered seven: 6S-1, 6S-2, *fsrA*, *bsrE*, *bsrF* (SR2), *bsrG*, and *bsrH* (Table 4-2, Fig. 4-3; Barrick *et al.*, 2005; Trotochaud and Wassarman, 2005; Saito *et al.*, 2009; Gaballa *et al.*, 2008). Additionally, pRNA, a small RNA oligonucleotide that is synthesized by RNA polymerase using 6S as a template, could be detected for 6S-1 but not 6S-2 (Fig. 4-3). The complicated relationship between the 6S and pRNA expression profiles will be addressed more fully in a different, future publication (personal communication, R. Hartmann). SurA, another previously identified sRNA, appeared from our analysis to be an antisense RNA since it appeared to overlap with the adjacent *yndL* gene. A putative sRNA was also previously identified within the *polC-ylxS* locus. This particular candidate was not specifically found within our analysis; instead, our data exhibited cDNA reads at the same locus but that appeared to correspond to a long 5' leader region for the *ylxS* gene. The remaining previously identified sRNA candidates (*bsrC, bsrD, bsrI,* SR1*,* SurC) couldn't be detected in our dataset, potentially due to limited expression during our growth conditions. SurC, for example, is exclusively

expressed during sporulation (Silvaggi *et al.*, 2006). Also, from the 54 putative sRNAs identified by Rasmussen *et al.* (2009), we detected 11 (20%) total. Finally, our analysis detected 50 new unique sRNA candidates (Table 4-2, Table 4-3). We did not specifically investigate whether these RNAs exhibited putative open reading frames; therefore, we cannot exclude that a subset might encode for small peptides.

An interesting feature of sRNAs in Gram-negative bacteria is their phylogenetic distribution. For example, it is not uncommon to find sRNAs that are well conserved among the gamma-proteobacterial species. It is not yet clear why these sRNAs have not evolved more rapidly among these organisms but is generally assumed that the primary sequence and secondary structure conservation for certain sRNAs has been retained to maintain intermolecular interactions with a common mRNA target. However, it is also possible that certain sRNAs exhibit phylogenetic conservation because they are constructed from exceptionally successful structural scaffolds, which are optimized for both interactions with target mRNAs and protection against RNases. Of the sRNA candidates identified in this study, most can be identified only in *B. subtilis* or the most closely related *Bacillus* species that have been sequenced. However, a few *B. subtilis* sRNA candidates also appeared to be present in genome sequences of other Bacillales species (Table 4-2, Table 4-3). Overall, this suggests that the *B. subtilis* sRNAs are likely to be more limited in their phylogenetic distribution than their proteobacterial counterparts.

Most striking in its phylogenetic distribution is RsaE, which has been identified in two prior studies (Geissmann *et al.*, 2009; Rasmussen *et al.*, 2009) and can be found in diverse Gram-positive bacteria, including *Staphylococcus*, *Lysinibacillus, Geobacillus,*

Listeria and *Bacillus* species. In *B. subtilis*, the top mRNA candidate for interaction with RsaE is *cstA*, which encodes for an uncharacterized carbon homeostasis protein. However, this gene does not appear to be a target in *Staphylococcus* species. Therefore, it is still unclear why this particular sRNA exhibits such high, albeit lineage-sporadic, distribution.

Several sRNA candidates that were identified herein but that were also discovered by prior studies (*bsrE*, *bsrH*, ncr39, ncr10, and ncr60) can also be found within the genomes of other Firmicutes, most often for *Bacillus* species (Table 4-2, Table 4-3). Comparative sequence alignments of these sRNA genes reveals several instances of covarying residues within putative helices, which together predict the occurrence of secondary structure features common for each sRNA class (data not shown). Additionally, a few novel sRNAs discovered by our current analysis appear to be conserved amongst genomes of a few other *Bacillus* species. For example, ncr1015 can be identified in the genomes of *Bacillus subtilis, Bacillus amyloliquefaciens, Bacillus licheniformis, Brevibacillus* species, and *Paenibacillus* species. Similarly, ncr2637 can be found in *Anoxybacillus flavithermus*, *Baillus subtilis*, *Bacillus amyloliquefaciens, Bacillus licheniformis*, and *Bacillus pumilus.* It is not yet obvious why these particular sRNA candidates are conserved in these other organisms, although a common mRNA target would be the primary assumption.

These data together help create an inventory of sRNA candidates in *B. subtilis*. However, demonstrating they are functionally required for genetic regulation is a challenging endeavor. Three experimental methods are traditionally used to add more confidence in individual sRNA candidates: (1) independent detection by alternative

experimentation (*e.g.,* by Northern blot analysis), (2) demonstration of a reliance upon Hfq for stability, and (3) prediction and validation of mRNA targets. The role(s) of Hfq in Gram-positive bacteria is still poorly defined; therefore, this was not taken into account in the current study. Instead, we chose 11 of the longest and most highly expressed sRNA candidates for validation by Northern blot analyses. All of the sRNA candidates that were chosen for Northern analysis could be successfully detected (Fig. 4-4, 4-5, 4-7). Also, several appeared to be subjected to intracellular processing, given that they corresponded to lengths shorter than their predicted size (Fig. 4-4, 4-5). Other sRNA candidates were not assessed thusly as they appeared to exhibit lowered expression levels that are likely to be within the range of detection by deep sequencing methodology but not by Northern blot analyses. Therefore, although much more experimentation is still yet required, preliminary experimentation on a subset of the candidate sRNAs appeared to validate their intracellular presence.

In order to begin assessing putative mRNA targets for these sRNA motifs we subjected the sRNA candidates to analysis by TargetRNA, a program designed to search for interrupted base pairing interactions within intergenic regions (Tjaden *et al.*, 2006). As it is difficult to differentiate false positives from actual mRNA targets using this software alone, we interpret these predictions with caution. Only a subset of the target predictions, which exhibited particularly low estimated p-values, is highlighted in Fig. 4-4 and 4-5. One possible explanation for the lack of mRNA targets for certain sRNA candidates is that the latter may target portions of protein coding sequences to affect mRNA stability or translation (Pfeiffer *et al.*, 2009), an interaction that is not addressed by current prediction software. Additional experimentation will be required in order to

determine whether these and other mRNAs represent actual targets for the sRNA candidates newly identified herein.

Prophage regions contain sRNAs and multiple RNA-based toxin-antitoxin systems

The genome of *B. subtilis* contains several prophages (*SPβ, skin, PBSX*) and prophage-like (*pro1-pro7*) regions, which are typically characterized by higher-than-background A+T nucleotide composition (Zahler *et al.*, 1977; Wood *et al.*, 1990; Takemaru *et al.*, 1995; Nicolas *et al.*, 2002). Prophages, like plasmids, conjugative transposons, and introns, are mobile elements that can be transferred horizontally, occasionally causing genomic rearrangements in bacteria. These elements often carry beneficial traits, such as antibiotic resistance cassettes or virulence factors, that could help the host adapt to their environment.

From our analysis, we detected 16 putative sRNAs originating mainly from the *SPβ*, *skin*, *pro6*, and *pro7* loci (Fig. 4-6). Some of these sRNA candidates, in fact, were the highest expressed sRNAs in our dataset (data not shown). None of the putative sRNAs described herein (Table 4-2, Table 4-3) were identified within the *PBSX* or *pro1-pro5* regions. It is generally assumed that a subset of phage genes expressed during lysogenic phase either confer a particular selective advantage for the host or are important for maintaining the phage-host equilibrium. We predict that some of the sRNAs identified herein are likely to perform similar functions. However, it remains to be determined whether these sRNAs target genes within the phage loci or specific host genes, although there is precedence for both scenarios in other prokaryotes as well as eukaryotes (Pfeffer *et al.*, 2004; Sullivan *et al.*, 2005; Castillo-Keller *et al.*, 2006;

Gottwein *et al.*, 2007; Grey *et al.*, 2007).

Interestingly, six of the sRNA candidates within phage-like regions are predicted to interact in pairs through antisense interactions. These three sRNA pairs are co-organized within intergenic regions in distinct tail-to-tail arrangements; their 3' terminal ~100 nucleotides overlap and therefore are predicted to interact via antisense pairings (Fig. 4-7). Several of the sRNAs that we predict to be organized in this manner have been identified previously, although their corresponding antisense partners were not (Saito *et al.*, 2009). Specifically, our data suggest that sRNA candidates *bsrE*, *bsrG*, and *bsrH*, which were identified previously, pair through intermolecular antisense interactions with newly identified *ncr1019*, *ncr1058*, and *ncr1155,* respectively. For convenience, we refer to these various sRNAs as *bsrE, bsrG, bsrH, as-bsrE*, *as-bsrG*, and *as-bsrH* in order to denote their antisense pairings. The three pairs of these RNA molecules are located within different prophage or prophage-like regions: *bsrE*/*as-bsrE* in *pro6*, *bsrG*/*as-bsrG* in *SPβ*, and *bsrH*/*as-bsrH* in *skin*. In addition, we also noticed that one of the pairs (*bsrH*/*as-bsrH*) is situated adjacent to a previously established toxin-antitoxin (TA) system, *txpA*/*ratA* (Fig. 4-7; Silvaggi *et al.*, 2005). Of particular note is that the *txpA/ratA* TA system shares a similar overall arrangement with the newly identified antisense-based RNA pairs (Fig. 4-7). Most TA modules consist of two components: a stable toxin and a labile antitoxin. The *txpA/ratA* system represents a typical type I TA system that includes an mRNA encoding for a short, toxic peptide (TxpA) and an antitoxin that is comprised of an antisense RNA (RatA). In contrast, type II TA systems rely on a protein factor as the antitoxin (Gerdes *et al.*, 2005; Gerdes and Wagner, 2007; Fozo *et al.*, 2008).

Based on these observations we examined each of the 6 sRNA candidates for

small open reading frames. Interestingly, only one sRNA from each pairings exhibited the potential to encode for a peptide of approximately 30 amino acids and included an appropriately spaced ribosome binding site (*as-bsrE, bsrG, bsrH*; Fig. 4-7). All three peptides are predicted to contain a single α-helical transmembrane domain of approximately 20 amino acids with several additional charged residues at the C-terminus. This arrangement is consistent with type I toxins (Fig. 4-7; Fozo *et al.*, 2008). TxpA encodes a similar peptide but with a modestly longer C-terminus. The remaining sRNAs (*bsrE*, *as-bsrG*, *as-bsrH*) did not exhibit any similar peptide-encoding potential, consistent with a potential role as an antitoxin. However these latter RNAs shared some primary sequence features and a common overall secondary structure arrangement, which consists of four stem-loop regions (Fig. 4-7). Approximately 100 nucleotides, located between two of the helical elements (P2 and P4), exhibited base-pairing potential to the 3`-end of the respective peptide-encoding mRNAs. RatA appears to have a similar secondary structure arrangement but with a longer 5`-portion (data not shown). The molecular mechanisms for how these antisense RNAs control toxin expression are still unclear. However, it has been proposed previously (Silvaggi *et al.*, 2005) that extensive 3`-end base pairing could promote simultaneous degradation of both RNAs.

Type I TA systems were originally discovered as an important component of plasmid maintenance mechanisms in *E. coli* (Gerdes *et al.*, 1985). More recently they have also been discovered in many bacterial chromosomes (Fozo *et al.*, 2008; Pedersen and Gerdes, 1999; Fozo *et al.*, 2010; Kwong *et al.*, 2010). It has been theorized that coupling TA systems to control of plasmid replication would ensure that plasmid-free cells are killed by toxin accumulation, a phenomenon termed ‘postsegregational killing’

(Hayes, 2003). Similarly, the *txpA/ratA* antisense module was proposed to be important for ensuring propagation of an accompanying phage genome in host cells (Silvaggi *et al.*, 2005). Our analysis therefore uncovers three more potential TA systems that are distributed in different prophage regions, suggesting that RNA-based TA mechanisms could be more common than previously recognized. Interestingly, we also note that two of the toxin-encoding mRNAs (*as-bsrE* and *bsrG*) are predicted to contain a ResD binding site within their putative promoter regions, suggesting a potential linkage between toxin expression and oxygen limitation (data not shown; Hartig *et al.*, 2004).

Identification of antisense RNAs

In addition to the newly identified toxin-antitoxin systems described above, our analysis also revealed other putative antisense RNAs (asRNAs). These different asRNAs could be assigned to several different categories based upon the arrangement with their target mRNA. A subset of the asRNAs exhibited the potential to fully pair with the entire target mRNA, while others pair with only a portion of the target mRNA through head-to-head (5`-overlap) or tail-to-tail (3`-overlap) interactions. In total, 29 candidate asRNAs were identified in our analysis (Table 4-4). Two of them (*ratA* and *as-yabE*) were previously shown to regulate expression of target mRNAs (the toxin-encoding *txpA* mRNA and *yabE*, an mRNA encoding for a cell-wall binding protein, respectively) (Appendix II; Silvaggi *et al.*, 2005; Eiamphungporn and Helmann, 2009). SurA, which overlaps a σ^K-regulated transcript, *yndL*, has been shown to accumulate during sporulation but its regulatory capabilities remain to be demonstrated (Silvaggi *et al.*, 2006). Ten of the asRNA candidates discovered herein were also identified via the high-

density tiling array analysis (Rasmussen *et al.*, 2009). These include asRNAs for the major vegetative growth phase sigma factor *sigA*, the teichoic acid biosynthesic enzyme, *ggaA*, a leucine biosynthetic enzyme, *leuA*, a choline transporter, *opuBD*, and a cryptic SpoIIIJ-associated protein, *jag*. The remaining asRNAs identified herein are novel and are predicted to pair with a variety of characterized and uncharacterized genes (Table 4-4).

One potentially interesting asRNA candidate is *ncr1430*, which overlaps with the 5`leader region of *bglP* (Table 4-4; Fig. 4-8). *bglP* encodes for a sugar phosphotransferase system (PTS) component that is involved in the utilization of β-glucosides, such as arbutin and salicin. It is transcribed as an operon with *bglH*, which encodes for the enzyme to metabolize the sugars (Le Coq *et al.*, 1995). The full synthesis of the *bglPH* operon had been shown to be under regulatory control by CcpA-mediated catabolite repression and a transcription attenuation mechanism mediated by the RNA antiterminator (RAT) element located within the 5` leader region (Aymerich and Steinmetz, 1992). Under inducing conditions and in the absence of glucose, an antiterminator protein, LicT, binds to the RAT element to stabilize formation of an antiterminator structure, thereby allowing transcription to proceed to the downstream coding region (Kruger and Hecker, 1995). Based upon its location, *ncr1430* is predicted to base pair with the *bglPH* mRNA within the region between the RAT element and the downstream *bglP* open reading frame, perhaps to repress translation. We hypothesize that this arrangement could provide yet a second layer of posttranscriptional regulation to allow finer control over the level of proteins needed to import and process the appropriate

sugar. If true, these data highlight a unique example of regulation of a single gene by a transcriptional mechanism as well as by both *cis*- and *trans*-acting regulatory RNAs.

Conclusion and Future Directions

Bacillus bacterial species, which are typically motile, aerobic endospore-forming microorganisms, have been isolated from diverse locales including soil, water sources, and plant root systems (Earl *et al.*, 2008). Contributing to its adaptive abilities, *B. subtilis* is capable of differentiating from dividing cells to metabolically inactive spores, which are resistant to many chemicals, irradiation and desiccation. As mutually exclusive alternative lifestyles, *B. subtilis* can also initiate developmental pathways that culminate with cells that are competent (capable of DNA uptake and primed for homologous recombination), or function as dedicated producers of biofilm extracellular matrix constituents (Aguilar *et al.*, 2007; Lopez *et al.*, 2009). Indeed, *B. subtilis* can form a multicellular community, consisting of spatially and temporally located cellular subtypes (Aguilar *et al.*, 2007; Lopez *et al.*, 2009; Lopez and Kolter, 2010). Responsible in part for these biological properties are many transcription factors and a suite of alternative sigma factors that regulate transcription of the developmental pathway genes. However, given their general importance in Gram-negative bacteria, it stands to reason that RNA-based regulatory strategies are also likely to be important for coordination of multicellular behaviors and developmental pathways.

Collectively, our results argued for broader roles for small RNA regulators in *B. subtilis*. In combination with other data, we increase the number of potential small RNAs in *B. subtilis* to upwards of 100 candidates. The next step will be to identify the biological

functions of these RNAs. We hypothesize that due to the differences in several key proteins involved in RNA metabolism between Gram-positive and Gram-negative bacteria (*e.g.*, ribonucleases, Hfq, and Rho), future studies of these noncoding RNAs will reveal novel mechanisms and add to the repertoire of bacterial RNA-based genetic regulatory strategies.

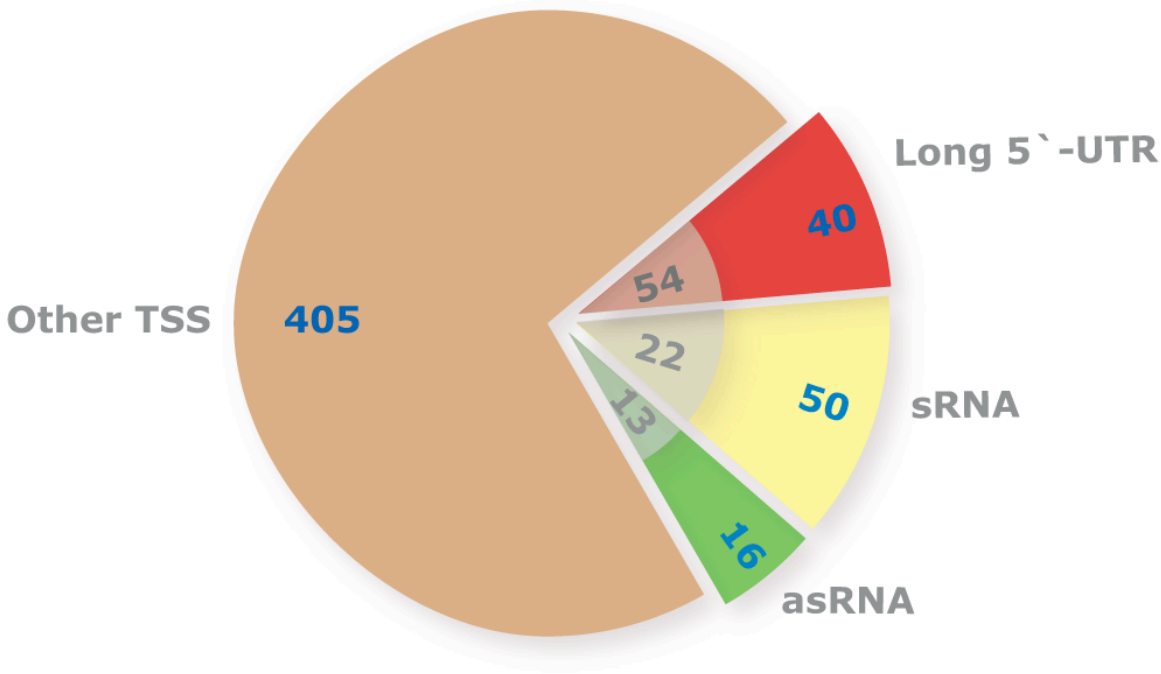


Figure 4-1. Chart depicting the total number of transcriptional start sites (TSS) detected in this study and their relationship with different classes of regulatory RNAs. The total number of TSS that was assigned to correspond to novel leader regions (40), sRNA candidates (50), and antisense RNAs (16) (asRNA) are shown above. The light-shaded numbers (*e.g.,* 3, 22, and 54, respectively) correspond to the number of previously identified RNA elements detected in our analysis within the same categories.

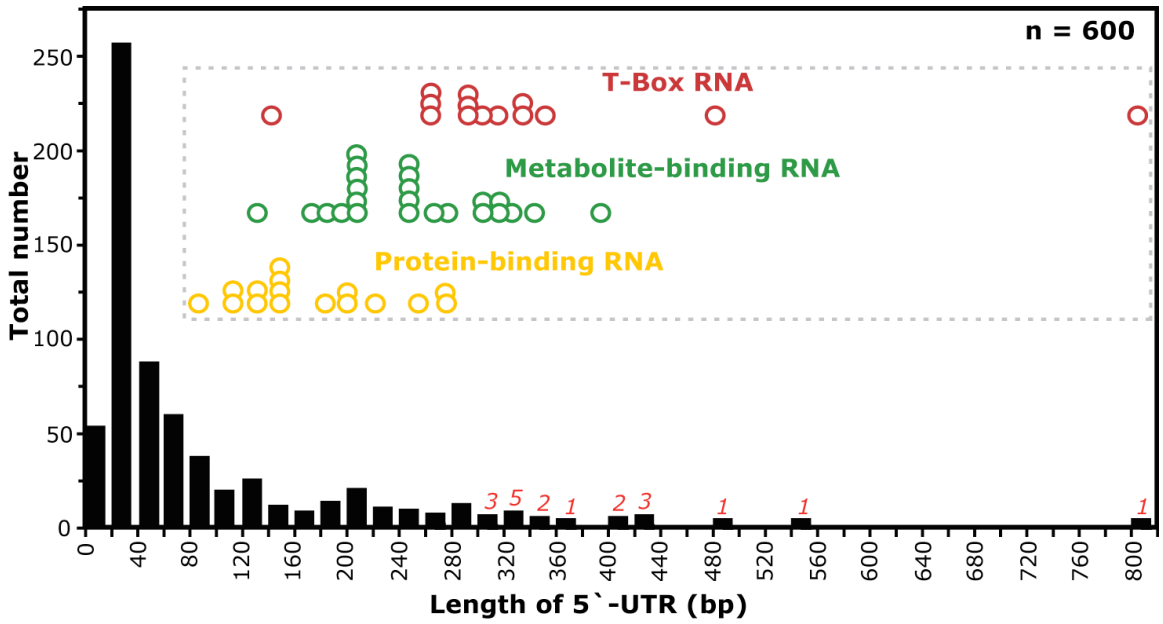


Figure 4-2. Length distribution of *B. subtilis* 5`-leader regions**.** In total, 600 transcription start sites (TSS) were identified based on the dRNA-seq analysis. The length of putative leader regions was calculated as the distance from the start of the cDNA reads to the start of the downstream coding region. The data is presented as a histogram with bin width = 20. The length of various UTRs containing known *cis*-regulatory RNAs are shown in the inset with each circle representing a specific UTR.

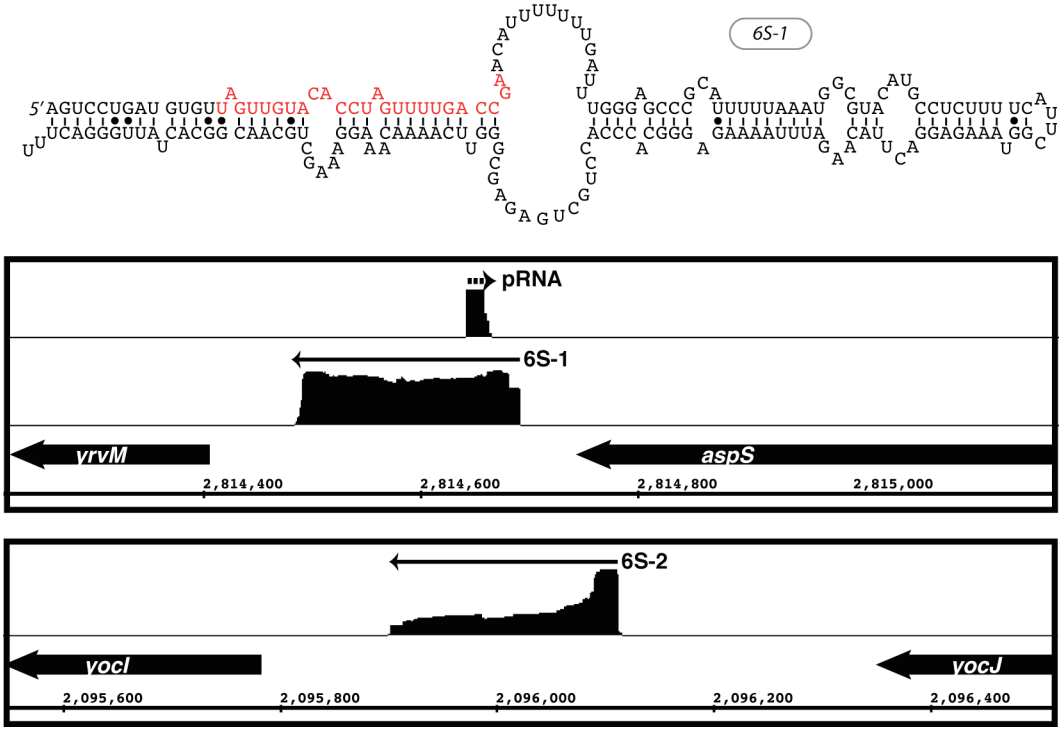


Figure 4-3. Visualization of cDNA reads for *B. subtilis* 6S RNAs. The upper panel shows the sequence and predicted secondary structure for 6S-1. Denoted in red is the sequence which is transcribed by RNAP from 6S as an RNA template into the short product RNA (pRNA). The bottom panel shows the genomic context and the distribution of cDNA reads mapped to both 6S-1 and 6S-2 loci. Arrows denote the direction of transcription. The cDNA reads for 6S-1 and 6S-2 are shown in the same relative scale. In contrast, the cDNA reads corresponding to the pRNA are approximately ten-fold less abundant compared to the 6S and, thus, are shown as a close-up for visualization purposes.

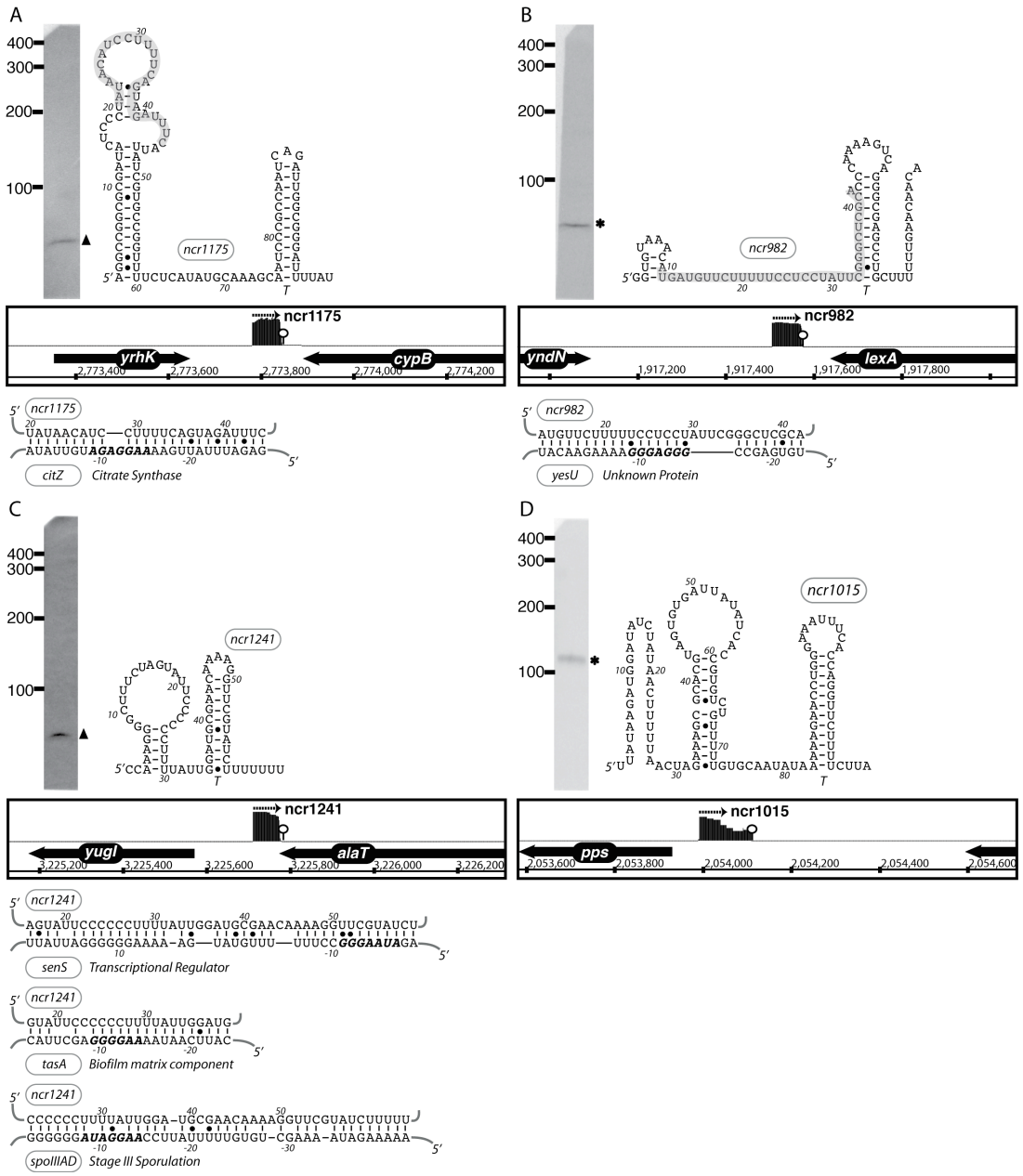


Figure 4-4. The expression, predicted secondary structure, and genomic context of *B. subtilis* sRNA candidates: *ncr1175* (A), *ncr982* (B), *ncr1241* (C), *ncr1015* (D). For each of these RNAs, the expression level was assessed by Northern blotting using total samples obtained from stationary phase cells cultured in minimal media. '∗' indicates that the size of sRNA detected by Northern blotting is in agreement with the size of the putative sRNA as predicted by sequencing data. '▲' denotes an sRNA with a different predicted size (*e.g.*, due to processing or termination events). The genomic locus of each sRNA is shown with its cDNA hits and the two flanking genes. The transcriptional unit is indicated by an arrow. An open circle denotes a potential intrinsic terminator. Candidate mRNA targets, as predicted by TargetRNA software, for are also included in the figure. The region of the sRNA predicted to associate with the target mRNA is highlighted in grey. Secondary structures were predicted using RNAfold and RNAz software.

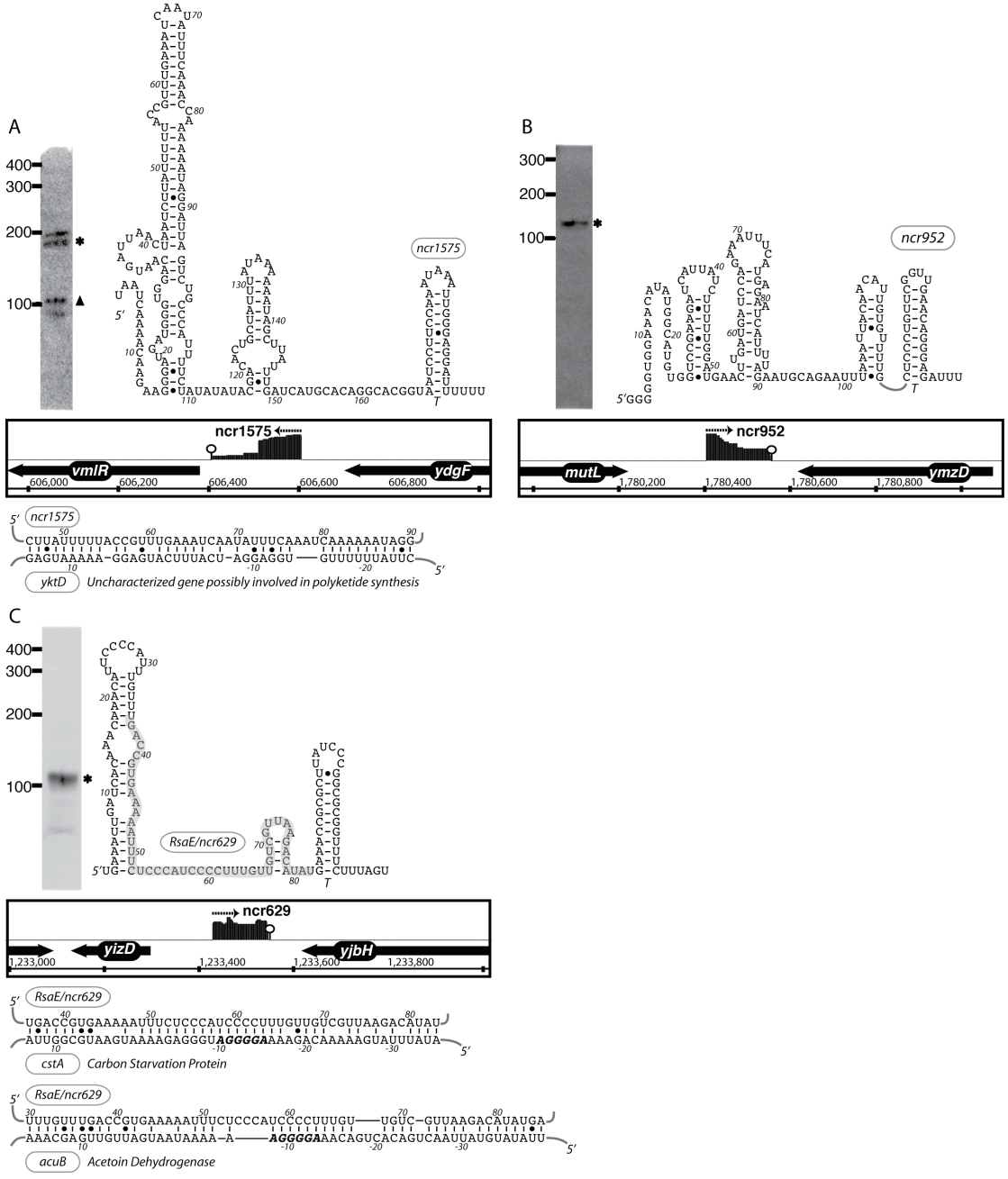


Figure 4-5. The expression, predicted secondary structure, and genomic context of *B. subtilis* sRNA candidates: *ncr1575* (A), *ncr952* (B), *RsaE/ncr629* (C). For each of these RNAs, the expression level was assessed by Northern blotting using total samples obtained from stationary phase cells cultured in minimal media. '∗' indicates that the size of sRNA detected by Northern blotting is in agreement with the size of the putative sRNA as predicted by the sequencing data. '▲' denotes an sRNA with a different predicted size (*e.g.*, due to processing or termination events). The genomic locus of each sRNA is shown with its cDNA hits and the two flanking genes. The transcriptional unit is indicated by an arrow. An open circle denotes a potential intrinsic terminator. Candidate mRNA targets, as predicted by TargetRNA software, for are also included in the figure. The region of the sRNA predicted to associate with the target mRNA is highlighted in grey.

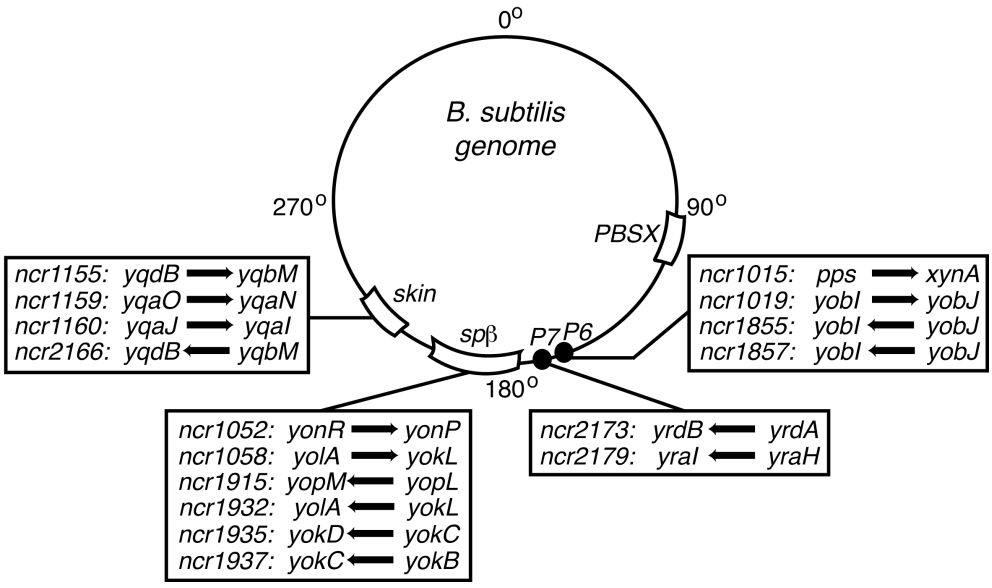


Figure 4-6. Putative sRNAs encoded within prophage regions. 16 sRNA candidates (denoted by arrow) originated from prophage or prophage-like regions (*SPβ*, *skin*, *P6*, *P7*) and are shown relative to their genomic location. Genes immediately upstream and downstream of the sRNA are also listed.

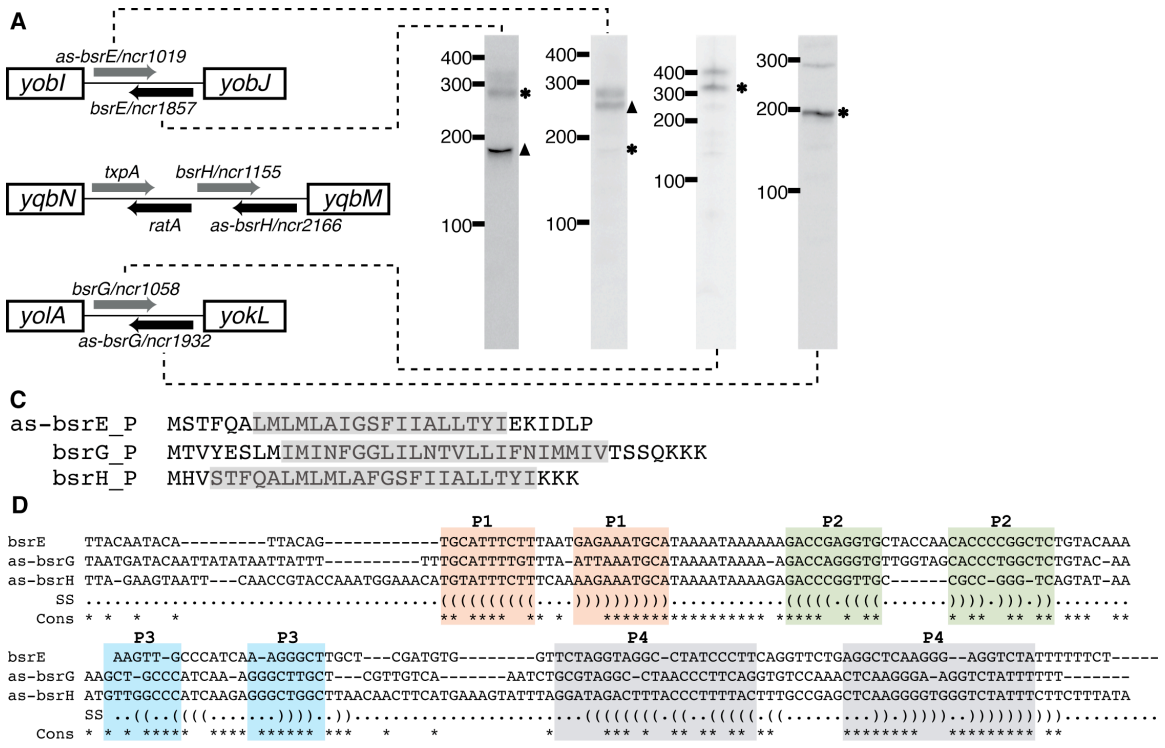


Figure 4-7. Novel toxin-antitoxin systems predicted by deep sequencing analysis. (A) Genomic locus of three new toxin-antitoxin systems in *B. subtilis*. The toxic protein (grey arrow) and the RNA antitoxin (black arrow) are all arranged in tail-to-tail configuration. Note that the *txpA* and *ratA* system had been previously characterized (46). (B) Northern blotting of the toxin and antitoxin RNAs. '*' indicates that the size of the sRNA as predicted by Northern blot is in agreement with sequencing data. '▲' denotes sRNA with different predicted size (*eg*. due to processing or termination events). The expression level for *bsrH* and *as-bsrH* were too low to be detected by Northern blotting in our analysis. (C) Putative sequences for *as-bsrE*, *bsrG*, and *bsrH* toxins. Predicted membrane spanning regions are highlighted in grey. (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.

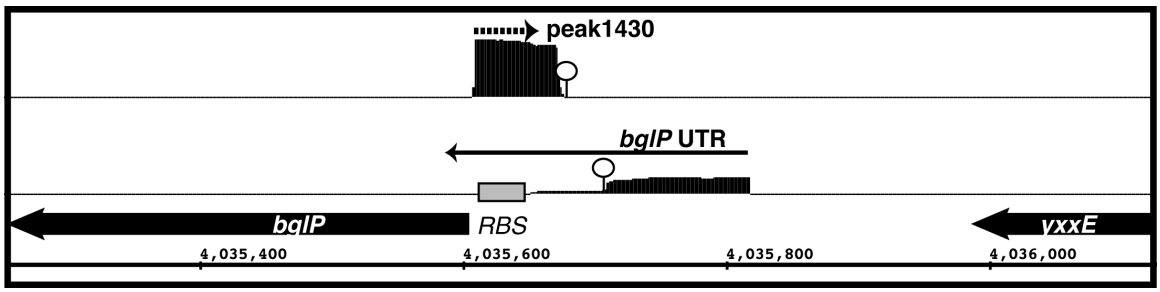


Figure 4-8. Novel arrangement of an antisense RNA predicted to base-pair with the *bglP* 5' leader region, which includes a cis-acting regulatory RNA. IGB representation of cDNA reads corresponding to *ncr1430* (top) and *bglP* UTR (bottom). Each transcriptional unit is represented by an arrow and the potential intrinsic terminator region is shown by a circle. *ncr1430* RNA is predicted to base-paired with the ribosomal binding site of *bglP* (denoted by grey box).

Table 4-1. Candidates for long 5' leader regions

	UTR size[a] (NTS)	cDNA #[b]	DOWNSTREAM GENE	REFERENCES	DISTRIBUTION[i]
ncr2261	264	30	*clpX*	Gerth et al. 1996[c]	2,7,8
ncr102	214	51	*rpoB*	Boor et al. 1995	2,7,8
ncr776	212	30	*pdhA*	Gao et al. 2002[d]	7
ncr2328	199	91	*citZ*	Jin and Sonenshein 1994	2,7
ncr1471	189	19	*yybN*	*this work*	(-)
ncr2639	187	8	*cwlO*	Bisicchia et al. 2007	2
ncr1422	183	8	*yxjJ*	*this work*	(-)
ncr921	180	748	*ylxS*	Shazand et al. 1993	2,7,8
ncr1521	180	8	*ycdA*	*this work*	(-)
ncr2264	179	19	*tig*	*this work*	2,7,8
ncr1554	165	48	*ydbN*	*this work*	2,7
ncr942	164	13	*ymdA*	*this work*	2,3,4,7,8,9
ncr551	145	8	*yhdT*	*this work*	2,7,8
ncr2017	143	8	*ypfD*	*this work*	2,7,8
ncr2103	128	9	*sodA*	*this work*	2,7
ncr948	124	21	*spoVS*	Resnekov et al. 1995	2,7,8
ncr812	123	12	*ylbK*	*this work*	2,12
ncr2692	122	17	*tagD*	Mauel et al. 1995	(-)
ncr2879	121	65	*yybS*	*this work*	2,7
ncr95	121	21	*rplK*	*this work*	Diverse
ncr2755	111	34	*fbaA*	*this work*	2,7,8
ncr2498	110	11	*dhbA*	Rowland and Taber 1996	(-)
ncr1278	106	8	*yuxN*	*this work*	2
ncr2243	104	156	*rplU*	Barrick et al. 2004[e], *this work*	Diverse
ncr2896	104	14	*rpmH*	Ogasawara et al. 1985[f]	1,2,3,4,7,8,9,12
ncr665	258	16	*rex*	*this work*	2
ncr1323	96	14	*pelC*	*this work*	(-)
ncr1443	420	617	*yxbB*	*this work*	(-)
ncr2793	248	4	*qoxA*	*this work*	2,7,8
ncr2190	238	5	*greA*	*this work*	2,7
ncr1167	201	2	*yraL*	*this work*	3,4,7,9,10,11,12,14
ncr2367	153	5	*aroA*	*this work*	2,8
ncr2579	131	3	*nhaK*	*this work*	(-)
ncr1576	129	4	*ydhB*	*this work*	(-)
ncr627	128	3	*mecA*	Kong et al. 1993[g]	2,7,8
ncr421	127	6	*yebE*	*this work*	(-)
ncr2789	126	4	*ywcI*	*this work*	2
ncr1261	120	6	*yuiD*	*this work*	(-)
ncr925	119	2	*rpsO*	Yao and Bechhofer 2009	Diverse
ncr1889	100	6	*odhA*	Resnekov et al. 1992[h]	2,7,8

[a]Candidates for long 5' leader regions ("5' UTR") are selected based on cDNA signals from intergenic regions (of the enriched sample), which include reads that overlap with or end within 10 nucleotides of the downstream gene. 5' UTR size is calculated as the distance from the start of cDNA signals up to the start of the coding region.

[b]cDNA # is calculated as the average number of cDNAs corresponding to the first 15 nucleotides from the 5`-end of the overall peak in the enriched sample. Only potential UTRs represented by 7

or more cDNA hits are shown in this table. The remaining candidates are listed in Supplementary Table 1.

[c]There are 2 transcription start site (TSS) detected for *clpX*. The TSS described here is 180 nts upstream of the previously characterized start site (Gerth et al. 1996), which was also observed in our dataset.

[d]The TSS for *pdhA* detected herein is 139 nts upstream of the one previously characterized by Gao et al. (2002)

[e]A long 5' UTR upstream of *rplU* was predicted previously by Barrick et al. (2004) based on sequence conservation.

[f]Two TSS upstream of *rpmH* are observed, in agreement with Ogasawara et al. (1985)

[g]We detected another potential Transcription Start Site (TSS) for *mecA* 74 nts upstream of the one described in Kong et al. 1993. The latter TSS is also observed in our dataset.

[h]The TSS for *odhA* detected here is 13 nts upstream of the one previously characterized by Resnekov et al. (1992).

[i]Organisms to which blast hits could be detected are denoted as: (-) *Bacillus subtilis* only (1) *Anoxybacillus flavithermus*, (2) *Bacillus amyloliquefaciens*, (3) *Bacillus anthracis*, (4) *Bacillus cereus*, (5) *Bacillus clausii*, (6) *Bacillus intermedius*, (7) *Bacillus licheniformis*, (8) *Bacillus pumilus*, (9) *Bacillus thuringiensis*, (10) *Bacillus weihenstephanensis*, (11) *Brevibacillus brevis*, (12) *Geobacillus species*, (13) *Lysinibacillus sphaericus*, (14) *Paenibacillus* species, (Diverse) all of the above Bacillales families with the addition of *Staphylococcus*, *Listeria*, *Streptococcus*, and *Lactobacillus* species.

Table 4-2. Predicted sRNAs

	PEAK[a] (NTS)	START	END	cDNA #	PREV GENE	NEXT GENE	GENE DIRECTION	NAME	DISTRIBUTION[b]
ncr1160	70	2697037	2697106	920	*yqaJ*	*yqaI*	/-/+/-/	-	(-)
ncr1159	50	2692882	2692931	352	*yqaO*	*yqaN*	/-/+/-/	-	(-)
ncr982	80	1917501	1917580	324	*yndN*	*lexA*	/+/+/-/	-	8
ncr1058	297	2273533	2273829	260	*yolA*	*yokL*	/-/+/-/	ncr46/bsrG	2,7
ncr1562	60	532583	532642	197	*ydcO*	*ydcP*	/+/-/+/	-	8
ncr1932	184	2273701	2273884	112	*yolA*	*yokL*	/-/-/-/	-	2
ncr1175	107	2773780	2773886	90	*yrhK*	*cypB*	/+/+/-/	-	2
ncr1937	69	2283685	2283753	89	*yokC*	*yokB*	/-/-/-/	-	(-)
ncr2768	56	3852061	3852116	54	*pbpG*	*ywhD*	/+/-/-/	-	2,6,8
ncr724	56	1451260	1451315	46	*stoA*	*zosA*	/+/+/+/	-	(-)
ncr1857	260	2069869	2070128	44	*yobI*	*yobJ*	/-/-/-/	bsrE	2,3,4,7,8
ncr1019	171	2069821	2069991	35	*yobI*	*yobJ*	/-/+/-/	ncr39	2,7,8
ncr1575	199	606407	606605	34	*vmlR*	*ydgF*	/-/-/-/	ncr10	2,3,4,9,10,13
ncr2184	232	2779137	2779368	32	*yrzI*	*yrhG*	/-/-/-/	ncr60	2,7,8
ncr471	173	820666	820838	24	*yfmI*	*yfmG*	/-/+/+/	-	(-)
ncr264	248	376678	376925	23	*hxlR*	*srfAA*	/+/+/+/	-	(-)
ncr952	151	1780404	1780554	20	*mutL*	*ymzD*	/+/+/-/	-	2
ncr2424	58	3146126	3146183	20	*mntA*	*menC*	/-/-/-/	-	7
ncr1915	59	2208755	2208813	18	*yopM*	*yopL*	/-/-/-/	-	(-)
ncr629	117	1233429	1233545	16	*yizD*	*yjbH*	/-/+/-/	ncr22/rsaE	Diverse
ncr1015	120	2053989	2054108	14	*pps*	*xynA*	/-/+/-/	-	2,7,11,14
ncr1155	233	2678729	2678961	13	*yqdB*	*yqbM*	/+/+/-/	ncr58/bsrH	2,5,7,8,11
ncr1241	128	3225697	3225824	13	*yugI*	*alaT*	/-/+/-/	-	2,7
ncr2299	99	2913485	2913583	13	*trxA*	*xsa*	/-/-/-/	-	(-)
ncr738	58	1467704	1467761	13	*ykwD*	*pbpH*	/-/+/+/	-	2
ncr969	58	1868404	1868461	12	*ymzA*	*ymaA*	/+/+/+/	-	2
ncr2637	74	3573045	3573118	11	*yvcI*	*trxB*	/-/-/-/	-	1,2,7,8
ncr178	124	199857	199980	11	*glmM*	*glmS*	/+/+/+/	-	2
ncr560	230	1056390	1056619	11	*yhaZ*	*yhaX*	/-/+/+/	ncr18	2
ncr992	72	1925548	1925619	7	*yneK*	*cotM*	/-/+/-/	-	2
ncr620	100	1219702	1219801	7	*trpS*	*oppA*	/-/+/+/	-	2,7
ncr585	201	1150478	1150678	7	*gerPA*	*yisI*	/-/+/-/	ncr20	(-)
ncr1957	60	2316348	2316407	6	*ypbR*	*ypbQ*	/-/-/-/	-	2

[a]Candidates for small RNA are selected based on cDNA signals from the intergenic regions (of the enriched sample), which do not correspond to a known gene, with at least 25 nucleotides distance from both the upstream and downstream genes. The length is measured from the start to the end of the cDNA signals.

[b]Organisms to which blast hits could be detected are denoted as: (-) *Bacillus subtilis* only (1) *Anoxybacillus flavithermus*, (2) *Bacillus amyloliquefaciens*, (3) *Bacillus anthracis*, (4) *Bacillus cereus*, (5) *Bacillus clausii*, (6) *Bacillus intermedius*, (7) *Bacillus licheniformis*, (8) *Bacillus pumilus*, (9) *Bacillus thuringiensis*, (10) *Bacillus weihenstephanensis*, (11) *Brevibacillus brevis*, (12) *Geobacillus species*, (13) *Lysinibacillus sphaericus*, (14) *Paenibacillus* species, (Diverse) all of the above Bacillales families with the addition of *Staphylococcus* species.

Table 4-3. Predicted sRNAs exhibiting lowered abundance

	PEAK[a] (NTS)	START	END	cDNA #	PREV GENE	NEXT GENE	GENE DIRECTION	NAME	DISTRIBUTION[b]
ncr1855	94	2069075	2069168	5	*yobI*	*yobJ*	/-/-/-/	-	2,7,8
ncr2507	85	3302792	3302876	5	*yuzG*	*guaC*	/-/-/+/	-	(-)
ncr1670	68	1077246	1077313	5	*hinT*	*ecsA*	/-/-/+/	-	2,7
ncr214	108	275609	275716	5	*garD*	*ycbJ*	/+/+/+/	-	(-)
ncr2360	188	3036340	3036527	5	*rpsD*	*tyrS*	/+/-/-/	-	2
ncr2173	97	2734262	2734358	5	*yrdB*	*yrdA*	/-/-/-/	-	(-)
ncr2185	159	2780319	2780477	4	*yrhG*	*yrhF*	/-/-/-/	-	2,7
ncr2166	83	2678994	2679076	4	*yqdB*	*yqbM*	/+/-/-/	-	2,7
ncr1876	140	2099817	2099956	4	*yozO*	*yozC*	/-/-/-/	-	2
ncr1566	79	559532	559610	4	*cspC*	*ydeB*	/+/-/-/	-	2,7,8
ncr976	52	1900528	1900579	4	*yncF*	*cotU*	/+/+/-/	-	(-)
ncr2160	259	2647405	2647663	3	*yqeG*	*yqeF*	/-/-/-/	-	2,7,12
ncr1421	223	3996388	3996610	3	*pepT*	*yxjJ*	/+/+/+/	-	(-)
ncr1755	187	1453368	1453554	3	*zosA*	*ykvY*	/+/-/+/	ncr35	(-)
ncr1733	107	1357727	1357833	3	*ykcC*	*htrA*	/+/-/-/	ncr26	(-)
ncr1118	86	2540930	2541015	3	*yqhR*	*yqhQ*	/-/+/+/	-	2,7
ncr2339	57	2991183	2991239	3	*ytsJ*	*dnaE*	/-/-/-/	-	(-)
ncr2665	77	3631679	3631755	3	*yvyD*	*yvzG*	/-/-/-/	-	(-)
ncr1935	104	2282621	2282724	2	*yokD*	*yokC*	/+/-/-/	-	(-)
ncr1052	183	2221800	2221982	2	*yonR*	*yonP*	/-/+/+/	ncr44	(-)
ncr721	112	1446806	1446917	2	*ykzR*	*ykvR*	/+/+/+/	ncr34	(-)
ncr181	158	204991	205148	2	*adaB*	*ndhF*	/+/+/+/	ncr4	(-)
ncr465	51	796025	796075	2	*yetO*	*ltaSA*	/+/+/+/	-	(-)
ncr2897	118	157	274	2	start[c]	*dnaA*	/+/+/+	-	(-)
ncr2857	103	4122960	4123062	2	*yyzE*	*yydK*	/-/-/+/	-	(-)
ncr977	106	1901991	1902096	2	*cotU*	*thyA*	/-/+/+/	-	(-)
ncr2752	256	3804713	3804968	2	*rho*	*ywjI*	/-/-/-/	ncr75	2,7,8
ncr1565	61	554497	554557	2	*yddR*	*yddS*	/+/-/+/	-	(-)
ncr2658	60	3625573	3625632	2	*ftsE*	*cccB*	/-/-/-/	-	2
ncr826	72	1596338	1596409	2	*sbp*	*ftsA*	/+/+/+/	-	2,8
ncr1221	73	3072289	3072361	2	*ythP*	*ytzE*	/-/+/+/	-	2
ncr2179	111	2752023	2752133	2	*yraI*	*yraH*	/-/-/-/	-	2,7

[a]Candidates for small RNA are selected based on cDNA signals from the intergenic regions (of the enriched sample), which do not correspond to a known gene, with at least 25 nucleotides distance from both the upstream and downstream genes. The length is measured from the start to the end of cDNA signals.

[b]Organisms to which blast hits could be detected are denoted as: (-) *Bacillus subtilis* only (1) *Anoxybacillus flavithermus*, (2) *Bacillus amyloliquefaciens*, (3) *Bacillus anthracis*, (4) *Bacillus cereus*, (5) *Bacillus clausii*, (6) *Bacillus intermedius*, (7) *Bacillus licheniformis*, (8) *Bacillus pumilus*, (9) *Bacillus thuringiensis*, (10) *Bacillus weihenstephanensis*, (11) *Brevibacillus brevis*, (12) *Geobacillus species*, (13) *Lysinibacillus sphaericus*, (14) *Paenibacillus* species.

[c]Start indicates the start of the genomic replication (0^o)

Table 4-4. Predicted novel antisense RNA (asRNA) candidates

	PEAK[a] (NTS)	START	END	cDNA #	OPPOSITE GENE	OVERLAP[b]	asRNA DIRECTION	NAME
ncr2706	47	3738263	3738309	114	*ywqA*	N	-	
ncr1430	70	4035606	4035675	58	*bglP*	5`	+	
ncr1687	24	1154737	1154760	26	*wprA*	G	-	
ncr1296	236	3460215	3460450	26	*opuBD*	G	+	shd102
ncr1207	231	2997369	2997599	15	*ytoI*	G	+	shd84
ncr1334	103	3669968	3670070	12	*ggaA*	3`	+	shd112
ncr1812	239	1915034	1915272	11	*yndL*	5`	-	SurA
ncr1265	218	3307602	3307819	10	*yutK*	G	+	
ncr1193	130	2892567	2892696	10	*leuA*	G	+	shd83
ncr2153	101	2641931	2642031	9	*comER*	G	-	
ncr1383	204	3862340	3862543	8	*ywfM*	3`	+	shd115
ncr1135	160	2600156	2600315	7	*sigA*	3`	+	shd77
ncr1186	17	2849374	2849390	7	*nadB*	G	+	
ncr1006	219	2002176	2002394	6	*yoeA*	5`	+	
ncr1799	25	1775732	1775756	6	*mutS*	5`	-	
ncr1479	232	4213213	4213444	6	*jag*	G	+	shd127
ncr1046	71	2190676	2190746	5	*yoqZ*	G	+	shd60
ncr1557	159	519216	519374	5	*ndoA*	3`	-	shd23
ncr2058	110	2483671	2483780	5	*yqzJ*	G	-	
ncr394	76	646225	646300	5	*ydiF*	G	+	shd26
ncr2160	259	2647405	2647663	3	*sda*	G	-	
ncr1351	227	3747160	3747386	2	*mbl*	3`	+	
ncr1565	61	554497	554557	2	*yddR*	3`	-	
ncr2885	106	4186388	4186493	2	*yyaQ*	3`	-	
ncr1546	50	452734	452783	2	*mtlD*	3`	-	
ncr507	30	924181	924210	2	*yfhD*	3`	+	
ncr2410	249	3123766	3124014	2	*ytoA*	3`	-	

[a]Candidates for antisense RNA are selected based on cDNA signals (of the enriched sample) that either start within genes with the opposite orientation or end within 50 nucleotides away from such genes. The length is measured from the start to the end of cDNA signals.

[b]Overlapping region is classified as follow:

G = asRNA is fully complementary to the opposite gene

5` = asRNA 5` end is complementary to the 5` region of the opposite gene

3` = asRNA 3` end is complementary to the 3` region of the opposite gene

N = no overlap; the distance between 3`-end of the asRNA and the opposite gene is less than 50 nts and there is no predicted intrinsic terminator.

CHAPTER FIVE

MATERIALS AND METHODS

Chemicals and oligonucleotides

All chemicals and enzymes, unless otherwise noted, were obtained from Sigma-Aldrich and New England Biolabs, respectively. DNA oligonucleotides were purchased either from Integrated DNA Technologies, Inc. or Sigma-Aldrich. Exact nucleotide sequences and brief description of DNA oligonucleotides used in these studies can be found in Appendix IV.

Strains and growth conditions

The *B. subtilis* strains used in *glmS* studies (Chapter 2) were all derivatives from strain BR151. The *B. subtilis* strains used in EAR (Chapter 4) and deep sequencing studies (Chapter 5) were all derivatives from either strain NCIB3610 or 168 as noted in the text and figure legends. All *B. subtilis* strains were obtained from Bacillus Genetic Stock Center (BGSC, Ohio).

All *B. subtilis* strains were typically cultured at 37^0C 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), 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 37^0C in LB (10 g tryptone, 5 g yeast extract, 5 g NaCl per liter), except for the RNase J1 experiment (Chapter 2) in which RM medium

(2% casamino acids, 0.2% glucose, 1 mM $MgCl_2$, 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl per liter; adjusted to pH 7.4 with NaOH) was used instead.

To analyze colony architecture, 5 μl of an overnight culture in 2xYT were 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.

When appropriate, antibiotics were included at the following concentrations: 50 μg/ml ampicillin (for *E. coli*), 5 μg/ml chloramphenicol (30 μg/ml for *E. coli*), 100 μg/ml spectinomycin, 1 μg/ml erythromycin plus 25 μg/ml lincomycin, 5 μg/ml neomycin, and 5 μg/ml tetracycline. Tryptophan, lysine, methionine were added to a final concentration of 50 μg/ml for all experiments with strain BR151 and its derivatives. Only tryptophan was added for all experiments with strain 168 and its derivatives. Arabinose, xylose, and IPTG were used to induce specific gene expression with final concentrations of 0.2% (w/v), 1.5% (w/v), and 80 μM, respectively.

Strains construction

Construction of lacZ fusion strains

All *lacZ* fusions, unless otherwise noted, was made by cloning the region of interest into pDG1661 plasmid (BGSC, Ohio). Plasmid pDG1661 contains regions of the *amyE* gene that flank the *lacZ* fusions and was thereby used to integrate these sequences

within the *amyE* gene via double homologous recombination. Oligonucleotides used to make all *lacZ* fusions are listed in Appendix IV.

Construction of inducible glmS strain

The 'left' and 'right' regions of the *glmS* gene as well as an *erm* cassette were first amplified and then cloned into plasmid pBGSC6 (BGSC, Ohio) via restriction sites added during PCR amplification. This plasmid was used for allelic replacement of the *glmS* gene with the *erm* cassette via double-crossover homologous recombination. Next, the *glmS* coding sequence was cloned into plasmid pDG148 and transformed into *B. subtilis* cells (strain BR151, BGSC, Ohio) using methods described elsewhere (Dann *et al.*, 2007; Jarmer *et al.*, 2002). The resulting cells were then transformed with the pBGSC6-based, *erm*-containing, *glmS*-deletion plasmid and plated on glucose minimal medium under conditions that included 1 mM IPTG. The resulting transformants were screened for IPTG-dependence. The region encompassing the original *glmS* locus was then PCR-amplified for a subset of IPTG-dependent isolates and subjected to DNA sequencing analysis in order to verify replacement of the *glmS* gene with *erm*.

Construction of B. subtilis strain with inducible RNase J1 and RNase J2 deletion

Strain GP41, containing a xylose-inducible copy of *rnjA* (RNase J1) (generous gift from Jörg Stülke, University of Göttingen, Germany) was cultured in rich media in the presence or absence of 1.5 % xylose. The original GP41 strain maintains the inducible copy of *rnjA* via a chloramphenicol resistance cassette. However, for our analyses, plasmid pCm::Nm (BGSC, Ohio) was used to swap *cat* for *neo*, encoding for neomycin

resistance, for construction of strain GP41-neo. This then allowed for transformation of GP41 with pDG1661 plasmids that contained *glmS-lacZ* fusions with selection for chloramphenicol. To construct a strain containing disruption of *rnjB* (RNase J2) as well as the xylose-inducible control of *rnjA* (RNase J1), GP41-neo cells were transformed with chromosomal DNA extracted from strain SSB348 (gift from H. Putzer, Institut de Biologie Physico-Chimique, Paris; Even *et al.*, 2005) and the resulting transformants were selected for resistance to neomycin and tetracycline. These cells were also assayed for xylose dependency.

Construction of E. coli strain with inducible RNase J1 and hammerhead-rpsT

rnjA gene (RNase J1) was amplified from *B. subtilis* chromosomal DNA and cloned into plasmid pBAD/*Myc*-His A (Invitrogen) under an arabinose-inducible promoter using NcoI and HindIII restriction sites added during PCR amplification. A stop codon was included in the insert such that the resulting protein would not contain *myc*- and *his*-epitopes. The *rpsT* gene was amplified from the chromosomal DNA of *E. coli* strain DH5α. The *B. subtilis glmS* promoter and hammerhead ribozyme were then fused to the 5`-end of *rpsT* by multi-round PCR. A 19-bp sequence tag (5`-AGACCCCCACCCCGATTCTT-3`; Celesnik *et al.*, 2007) was also fused to the 3`-untranslated region of *rpsT* gene via PCR amplification. The resulting fragment was cloned into plasmid pACYC184 (New England Biolabs). Both plasmids, containing inducible RNaseJ1 and hammerhead-*rpsT* constructs, were then transformed into *E. coli* strain DH5α.

Construction of EAR mutant and reporter strains

Unless otherwise noted, all mutant strains were generated by directly transforming competent *B. subtilis* NCIB3610 cells with plasmid DNA (~5-10 μg) prepared from methylation-deficient *E. coli* strains (INV110, Invitrogen). DNA was transformed into *B. subtilis* using a modified version of a previously published protocol (Anagnostopolous and Spizizen, 1961).

In-frame deletions: ΔEAR

To generate the marker-less ΔEAR deletion construct, the region upstream and downstream of EAR was PCR amplified from NCIB3610 chromosomal DNA. The two fragments, containing ~20 nucleotides overlap, were then fused together via a 2nd round PCR reaction. The resulting PCR was digested with *EcoRI* and *BamHI* and sub-cloned into pMAD (Arnaud *et al.*, 2004), which carries a temperature sensitive origin of replication, erythromycin resistance cassette and constitutively active *lacZ* gene. Correct clones were confirmed by DNA sequencing. The plasmid was used to transform NCIB3610 strains at the permissive temperature for plasmid replication (30^{o}C) with selection for resistance to erythromycin plus lincomycin. Isolates resulting from Campbell integration were obtained after culturing the cells overnight in 2xYT broth at the restrictive temperature (37^{o}C) and plating serial dilutions at 37^{o}C with selection for resistance to erythromycin plus lincomycin. These isolates were blue when grown on solid medium containing bromo-chloro-indolyl-galactopyranoside (X-gal; 40-80 μg/ml). To cure the plasmid, the strain was incubated overnight in 2xYT broth without shaking and without antibiotic at 30^{o}C, followed by shaking incubation for 5 hours at 30^{o}C and 3

hours at 37°C. The cells were then serially diluted and plated on TBAB at 37°C in the absence of antibiotic. Individual colonies were patched onto TBAB plates and TBAB plates containing 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 from these isolates was isolated and used as templates for diagnostic PCR reactions and subsequent DNA sequencing reactions to confirm mutagenesis of the targeted genomic locus. ΔNusG and ΔNusB strains were constructed in a similar manner.

EAR mutant strains: M3, M4, M5, M6, M7, M8

All pMAD plasmids (Arnaud *et al.*, 2005) containing EAR mutations were transformed into ΔEAR as a recipient strain, instead of NCIB3610, using the same protocol that was employed for making in-frame gene deletion strains (described above). The EAR mutations were created via QuikChange mutagenesis (Stratagene).

EAR- T7t-lacZ transcriptional fusion

The *B. subtilis* EAR region and *lacZ* gene were amplified from NCIB3610 chromosomal DNA andpDG1661, respectively. The reverse primer for amplifying EAR and the forward primer for amplifying *lacZ* gene each contain a portion of the previously characterized T7 terminator sequence, 'T7t', which could be reassembled after a subsequent PCR-sewing step (Reynolds *et al.*, 1992). The two fragments were subjected to a 2nd round PCR-sewing reaction in order to generate an EAR sequence that is followed by T7t and the *lacZ* gene. The resulting PCR was digested with *HindIII* and

SphI and subcloned into pHyperSpank (gift from Gurol Suel, Dallas, TX). These oligonucleotide primers also incorporated *SalI* and *NheI* restriction sites into the region between the EAR-T7 terminator sequence and the T7 terminator-*lacZ* region, respectively, which were then used for subsequent plasmid constructions. NCIB3610 cells were then transformed with the resulting plasmid with selection for resistance to spectinomycin. Individual colonies were screened by diagnostic PCR reactions for integration of the plasmid into the *amyE* locus via double homologous recombination. The remaining EAR-terminator-*lacZ* fusion strains were created by cloning various sequences into the EAR-T7t-*lacZ* plasmid via SalI/NheI restriction sites (see Irnov *et al.*, 2010 for further descriptions and sequences for each construct).

β-galactosidase activity assays

Assays for *lacZ* expression were conducted as described elsewhere (Dann et al. 2007). Briefly, cell pellets were resuspended in 1.0 ml Z buffer (Miller, 1972). 10 µl of toluene was added and the cell suspension was vortexed for 30 seconds whereupon the toluene was evaporated under a hood for 15 minutes. 0.15 ml (0.6 mg) orthonitrophenyl-β-*D*-galactoside (ONPG) was added to 0.75 ml permeabilized cell suspension to initiate the enzymatic assay (time, T_i). Reactions were terminated (time, T_f) by addition of 0.375 ml 1M Na_2CO_3, centrifuged briefly, and analyzed for absorbance at A_{420}. Data was analyzed relative to control reactions containing Z buffer alone. Miller Units were calculated using the equation $[(1000 \times A_{420})/((T_f - T_i) \times 0.75 \text{ ml} \times A_{600})]$. All A_{420} readings were taken within 20 minutes of reaction termination.

Sporulation assays

To look at the role of the *glmS* ribozyme in sporulation, wild-type BR151 and M9 strains were induced into sporulation by the resuspension method as described elsewhere (Sterlini and Mandelstam, 1969; Nicholson and Setlow, 1990). For microscopy analysis, 5 µl of the sporulating cells were spotted onto a 1.5% agarose pad (in 1x PBS), dried for 1 hour at 37^{o}C, and viewed at 60X magnification via phase contrast microscopy, using an Olympus IX81 motorized inverted microscope.

To look at the role of the EAR element in sporulation, NCIB3610 and ΔEAR cells were harvested either from 48 hours colonies grown on MSgg agar plates at 30^{o}C or 48 hours liquid cultures grown shaking in MSgg media at 37^{o}C. For colonies, cells were first transferred onto saline-EDTA and passed through 22G syringe multiple times to break up the extracellular matrix. The cell suspensions were treated with 10% chloroform (v/v) or saline-EDTA (control), was 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.

RNA extraction

Total RNA was harvested from *B. subtilis* strains cultured at 37^{o}C as described in the text. These cells were pelleted, resuspended in LETS buffer (0.1M LiCl, 10mM EDTA, 10mM TrisHCl, 1% SDS), vortexed with acid washed glass beads (Sigma-Aldrich) for 4 min, and incubated at 55°C for 5 min. The resulting solution was subjected to extraction by TRIzol reagent (Invitrogen) according to the manufacturer instructions.

The RNA was then concentrated by ethanol precipitation and quantified via absorbance spectroscopy. To extract total RNA from biofilm, the cells were lifted off of the agar plates, transferred directly onto LETS buffer, and passed through 22G syringe multiple times to break up the extracellular matrix. The resulting cell suspensions were then vortexed with glass beads and subjected to TRIzol extraction as described above.

Northern blot analyses

Total RNA samples (10-20 μg) were heated at 65°C for 10 min in 1x 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 either 6% denaturing (8 M urea) polyacrylamide electrophoresis or formaldehyde denaturing 1-2% agarose gel 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, hybridized overnight in UltraHyb buffer (Ambion) with either 5′-radiolabeled (^{32}P) DNA oligonucleotide or internally-radiolabeled (^{32}P) probe RNA. DNA oligonucleotide probe was hybridized overnight to the blots at 42°C and washed 2x for 15 minutes using low stringency wash buffer followed by 30 minutes wash using high stringency buffer. Antisense RNA probes was hybridized overnight at 68^{o}C and washed 2x for 5 minutes with low stringency wash buffer (1x SCC, 0.1% SDS, 1 mM EDTA) at room temperature followed by a high stringency wash (0.2x SCC, 0.1% SDS, 1 mM EDTA) at 68°C for 15 minutes. Radioactive bands were visualized using ImageQuant software and a Typhoon PhosphorImager (Molecular Dynamics).

Primer extension analyses

15-30 μg total RNA was treated with 10 units of DNase I (Roche) in the presence of 0.25 mM $MgCl_2$ at 37°C for 30 min. Following phenol extraction and ethanol precipitation steps, the RNA was subjected to reverse transcription reactions using the appropriate DNA primer and Transcriptor reverse transcriptase (Roche), according to the manufacturer instructions. Dideoxy DNA sequencing ladders were generated using the same DNA oligonucleotide and the Thermo Sequenase Cycle Sequencing Kit (USB). As a control for total RNA quantities, 5S RNA was reverse transcribed using a separate 5S-specific 5′-radiolabeled DNA primer.

PABLO

PABLO (Phosphorylation Assay By Ligation of Oligonucleotides) was performed as described in Celesnik *et al.* (2007) with some minor modifications. RNA was mixed with 2 μM of the 5` DNA oligonucleotide and a "splint" DNA oligonucleotide in a final volume of 25 μl, heated to 75 ^{o}C for 5 minutes and slowly cooled to 30 ^{o}C. Reaction mixtures containing 2x T4 DNA ligase buffer (Roche), 2 mM ATP, and 10 units of T4 DNA ligase (Roche) was added and incubated at 37 ^{o}C for 4 hours. The reactions were then phenol extracted, ethanol precipitated, resolved by 6 % denaturing PAGE and analyzed by Northern blotting using an antisense RNA probe as described above. These assays were performed using either synthetic RNA (0.02 pmol) or total RNA (10 μg). The latter RNA was harvested from a strain (GP41) containing a wild-type *glmS* ribozyme-*lacZ* fusion and that was depleted for RNase J1 (*i.e.,* cultured in the absence of

xylose). The synthetic RNA was transcribed *in vitro* in the presence of 2.5 mM GlcN6P (thereby generating the 3`-cleaved product with a 5`hydroxyl group) from PCR templates. Details for *in vitro* transcription are described below. The RNA samples were either treated with T4 polynucleotide kinase (New England Biolabs) under standard conditions or left untreated.

Quantitative real-time RT-PCR

Total RNA from strains was isolated as described above. Approximately 4 μg RNA was then incubated in 20 μl reactions with 2 units DNase (Roche) and 0.5 mM $MgCl_2$. These reactions were incubated at 37°C for 30 minutes followed by 75°C for 10 minutes. The resulting RNA samples were employed for synthesis of cDNA libraries, which were generated using the iScript Select cDNA synthesis kit (BioRad) as per manufacturer instructions. A total of 0.5 – 1.0 μg RNA and random hexamer primers were included in these cDNA synthesis reactions. Control reactions lacking reverse transcriptase were also prepared for each different RNA sample. All reactions were prepared in triplicate in MicroAmp optical 384-well reaction plates (ABI). Each reaction contained 150 nM primers, 5 μl of 2x SYBR Green master mix and 25 ng cDNA templates in a total volume of 10 μl. The cycling parameters used for these experiments were as follows: 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min, using an ABI 7900HT Fast Real-Time PCR instrument. The number of cycles required to reach threshold fluorescence (C_t) was determined through use of Sequence Detection Systems version 2.2.2 software (Applied Biosystems) using automatic baseline and threshold determination and averaged from triplicate samples. For *glmS* experiment,

C_t values for 5S control RNAs were subtracted from experimental values to give ΔC_t values, which were averaged from data performed in triplicate. These values were used to establish the ratio of transcript abundances for ribozyme-containing transcripts during conditions of high and low GlcN6P. The formula used to calculate this ratio was:

$2^{(\Delta C_t low - \Delta C_t high)}$.

The expression of different *eps* genes was normalized to the amount of transcription of the *eps* operon (as measured by *epsA* expression level) and the fold-change relative to wild-type ($\Delta\Delta C_t$) was calculated using by the following formula:

Fold change = $2^{-(\Delta C_t mutant - \Delta C_t 3610)}$; with $\Delta C_t = \text{avg}C_{t(\text{gene of interest})} - \text{avg}C_{t(epsA)}$.

Standard deviation values were calculated as described elsewhere (Bookout *et al.,* 2006).

Scanning Electron Microscopy (SEM)

B. subtilis colonies were prepared on solid MSgg medium as described above and fixed and prepared for SEM by the UT Southwestern Molecular and Cellular Imaging Facility using standard protocols. Representative images were captured using a FEI XL30 ESEM microscope at 10,000X magnification.

RNA structural probing

DNA templates for *in vitro* studies of different RNAs were created by PCR using appropriate oligonucleotide primers. PCR products were then prepared using the QIAGEN PCR clean-up kit. RNAs were synthesized at 37ºC in vitro from 25 – 50 μl reaction mixtures that included ~10 – 30 pmol templates, 30 mM Tris-HCl (pH 8.0), 10 mM DTT, 0.1% Triton X-100, 0.1 mM spermidine-HCl, 2.5-5.0 mM each NTP (Roche),

40 mM $MgCl_2$, and ~50 μg ml^{-1} T7 RNA polymerase. Reactions were terminated after 2.5 hours with 2x volume 8 M urea. Products were resolved by denaturing 6% PAGE and excised, cut into ~ 1 mm cubes, and equilibrated in 200 mM NaCl, 10 mM Tris-HCl (pH 7.5), and 10 mM EDTA (pH 8.0) for < 2 hours at 23 ºC. Passively eluted RNAs were then ethanol precipitated and quantified by A_{260}. An extra G was added at the 5'-terminus to facilitate T7 transcription.

RNA substrates for in-line probing were dephosphorylated using calf intestinal alkaline phosphatase (New England Biolabs) and 5'-radiolabeled using T4 polynucleotide kinase (New England Biolabs) and γ-^{32}P ATP (Amersham). Reactions contained ~1 nM RNA, 50 mM Tris-HCl (pH 8.3) and 100 mM KCl. RNAs were incubated at 25 ºC for ~40 hours and products were resolved by 10% PAGE adjacent to control lanes containing: partial digestion by RNase T1 (cleavage after G), a hydroxyl cleavage ladder (cleavage at every position), and an aliquot of unreacted RNA.

All polyacrylamide (National Diagnostics) gels were visualized using a PhosphorImager (Amersham) and quantified with SAFA (Das *et al.*, 2005) and ImageQuant software (Molecular Dynamics).

Single-round transcription assay

DNA template for *in vitro* transcription assay were created by amplifying the appropriate constructs from the EAR-terminator-*lacZ* fusion plasmids described above. A promoter sequence from *B. subtilis glyQ* gene was added through the forward DNA primer. Elongation complexes were formed with a 40 nM concentration of DNA template, 44 nM *B. subtilis* RNAP, and 100 nM SigA in 100 ul transcription buffer (200

mM Tris-HCl pH 8.0, 20 mM NaCl, 14 mM $MgCl_2$, 0.1 mM EDTA, 14 mM β-mercaptoethanol). In this construct, elongation complexes can be halted at nucleotide position 17 when the transcription is initiated in the absence of CTP (with 100 μM ApU, 2.5 μM ATP, 2.5 μM GTP, 1 μM UTP; in the presence of [α-^{32}P] UTP). Halted complexes were formed with 10 minutes incubation at 37 ºC. Transcription was restarted by the addition of all nucleotides at 200 μM concentration and heparin to 100 μg/ml. All reactions were quenched after 15 minutes incubation at 37 ºC with equal volume of 2x Urea 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).

Microarray and RNA-Seq sample preparation

For the EAR experiments, total RNA was harvested from wild-type *B. subtilis* NCIB 3610 and ΔEAR colonies as described above. For microarray, samples were processed and hybridized to GeneChip *B. subtilis* Genome Arrays (Affymetrix) by the UT Southwestern Microarray Core Facility following the manufacturer's recommendations. For RNA-Seq, total RNA was first treated with RQ1 DNase (Promega) at 37^{o}C for 30 minutes. Following phenol extraction and ethanol precipitation steps, the RNA was then subjected to rRNA depletion using the MicrobExpress kit (Ambion) following the recommended protocol. Approximately 1 μg of the rRNA-depleted total RNA was used as a starting material to generate cDNA libraries using Illumina mRNA-Seq kit without the polyA-enrichment step. Additionally, the cDNA was purified using Qiagen PCR Cleanup kit (~100 nts cut off) after adaptor ligation, instead of following the recommended gel extraction step. The processed samples were then submitted to the UT Southwestern Microarray Core Facility for sequencing on Illumina

GAXII genome analyzer. The resulting cDNA sequences were mapped onto *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). More details on the analysis will be provided in Appendix I.

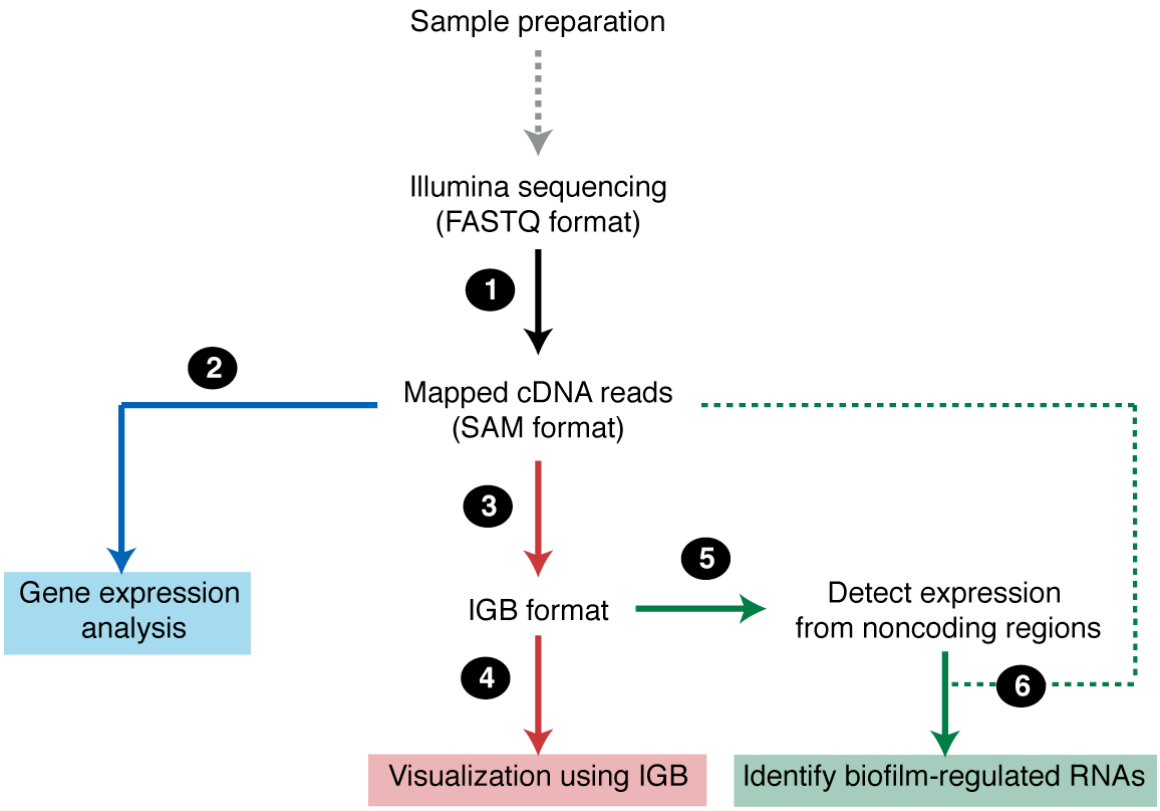
454 pyrosequencing analysis

cDNA cloning and pyrosequencing was performed as described (Berezikov *et al.,* 2006) except without size fractionation of RNA. However, the total RNA was split into two samples and for one sample primary transcripts containing 5`-PPP were enriched by treatment of the total RNA samples with Terminator 5'-phosphate-dependent exonuclease (Epicentre). Upon treatment of both samples with tobacco acid pyrophosphatase to generate 5'-monophosphates for linker ligation, RNA samples were poly(A)-tailed using poly(A) polymerase followed by ligation of a RNA adapter to the 5´-phosphate of the small RNAs. First-strand cDNA synthesis was then performed using an oligo(dT)-adapter primer and M-MLV H- reverse transcriptase. A detailed protocol for the enrichment, cDNA library generation and subsequent sequencing steps are described (Sharma *et al.*, 2010). cDNA libraries were sequenced on a Roche FLX sequencer at the M. D. Anderson Cancer Center DNA Analysis Core Facility. 5`-linker and poly-A tail removal were performed using custom-made Python scripts. The resulting cDNA sequences were then mapped onto *B. subtilis* genome (NC_000964.3) using 'segemehl' software (39). Mapped reads were visualized using Integrated Genome Browser (IGB). More details on the analysis will be provided in Appendix II.

APPENDIX I

RNA-Seq analysis of the wild-type *vs* EPS deficient colonies

Data Analysis

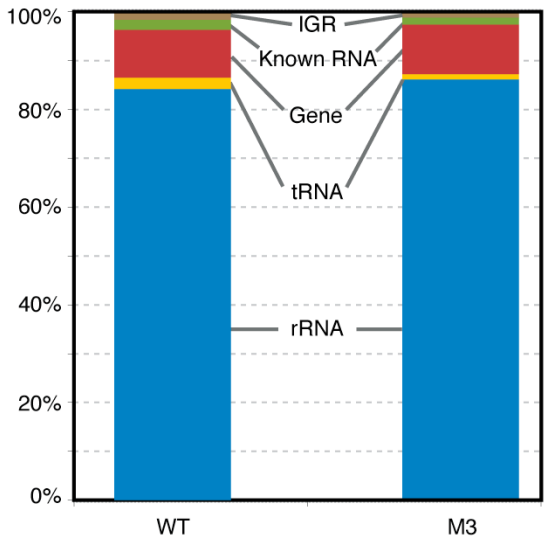


1. Mapping cDNA reads onto the *B. subtilis* genome

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 - released in 2010 and the M3 version) using the "Burrows-Wheeler Aligner" software (version 0.5.8a; Li and Durbin, 2009; freely available at http://bio-bwa.sourceforge.net/) with the following parameters:

```
Genome indexing:
   >bwa index -a is [genome_file]

Alignment:
   >bwa aln -t 2 -l 20 -k 2 -n 4 [genome_file] [input] > [output_1]

These parameters will allow up to 4 mismatches with a maximum of 2
mismatches allowed for the first 20 nucleotides (seed sequence).

Conversion to SAM format:
   >bwa samse -n 100 [genome_file] [output_1] [input] > [output_2]

These parameters will allow the inclusion of cDNA reads with up to
100 hits in the genome. We chose an arbitrary cut off of 100 to make
sure that all the cDNA reads that can be mapped onto B. subtilis
genome will be included in the output. For cDNA reads with multiple
hits in the genome, only the first mapped coordinate (as determined
by BWA) will be used for further analysis.


[genome_file] = genomic sequence in FASTA format
[input]       = cDNA reads from Illumina sequencing in the default
                FASTQ format
[output_1]    = aligned cDNA reads generated by BWA
                (with ".sai" extension)
[output_2]    = aligned cDNA reads in SAM format
                (with ".sam" extension)

Further details for FASTQ and SAM formatting are available at:
   http://maq.sourceforge.net/fastq.shtml
   http://samtools.sourceforge.net/SAM1.pdf
```

In total, 24,210,581 and 24,408,252 cDNA reads were successfully mapped onto the genome for the wild-type and M3 samples, respectively. Despite of the rRNA depletion steps (see Materials and Methods), ~80% of cDNA reads still originated from various rRNA loci. A better rRNA (and tRNA) depletion methods might be needed in the future to increase the proportion of cDNA reads from other coding and noncoding regions.



	WT	M3
rRNA	20,382,328	21,034,695
tRNA	552,515	251,963
Gene	2,370,022	2,466,639
Known RNA	504,683	338,408
Intergenic Region (IGR)	401,033	316,547
TOTAL	24,210,581	24,408,252

2. Gene expression analysis

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" (Li *et al.*, 2009; freely available at http://samtools.sourceforge.net/) with the following parameters:

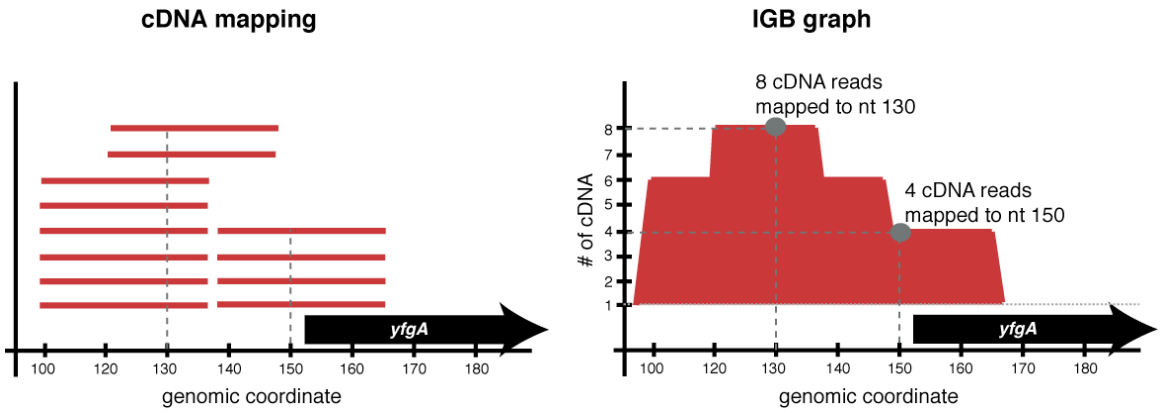
```
>samtools view -S -F 4 [input_SAM_format] > [output_mapped_SAM]
```

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)
```

3-4. Conversion to the IGB format for data visualization

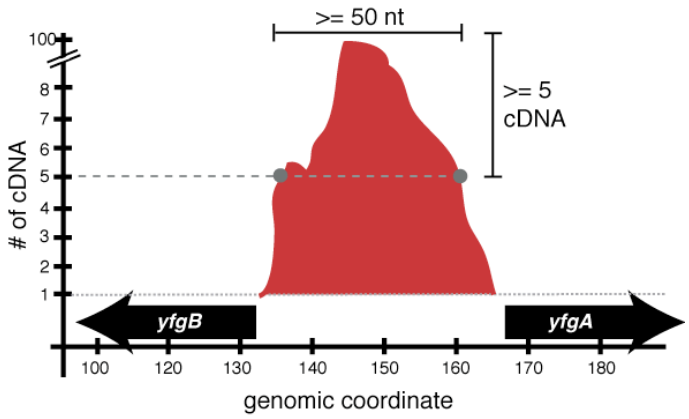
Integrated Genome Browser (IGB; freely available at http://www.bioviz.org/igb/) was used to visualize all of our deep sequencing analyses (see examples in Chapter 3 and 4). A typical IGB graph shows the number of cDNA reads that are mapped onto a particular nucleotide position in the genome.



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. The length of the input file will be equal to the size of the genome (*i.e.*, 4,215,606 nucleotides for *B. subtilis*). The data will be visualized using "NC_000964_3.fas" and "NC_000964_3.gff" as the source of genomic sequence and gene information, respectively (available from NCBI).

5-6. Identification of biofilm-dependent regulatory RNAs

To identify biofilm-specific RNAs, we utilized the ".gr" file generated in step #3 to detect expressions from the noncoding portions of the *B. subtilis* genome. Specifically, we looked at potential transcription units in the wild-type sample represented by 50 or longer continuous nucleotide positions with at least 5 cDNA hits (also referred to as "peaks"). The cutoff values that we used here were arbitrarily determined based on the known sizes of regulatory RNAs and the background observed in our dataset.



From our dataset, we successfully extracted 249 peaks that fit the abovementioned criteria. As expected, many of them correspond to tRNAs and previously characterized regulatory RNAs (*e.g.*, riboswitches, toxin-antitoxin systems, and small RNAs). We reasoned that the *bonafide* biofilm-regulated RNAs should show a differential expression between the two datasets. To that end, we quantified the expression of each "peak" by RPKM (as described in step 2) in both the wild-type and M3 datasets. In total, we identified 9 potential biofilm-regulated regulatory RNAs with decreased expressions and 1 RNA with increased expression in the EPS deficient biofilm (shown in Table 3-3). The characteristics and potential role(s) of these RNAs are discussed in Chapter 3.

Expressions of additional genes mentioned in Chapter 3

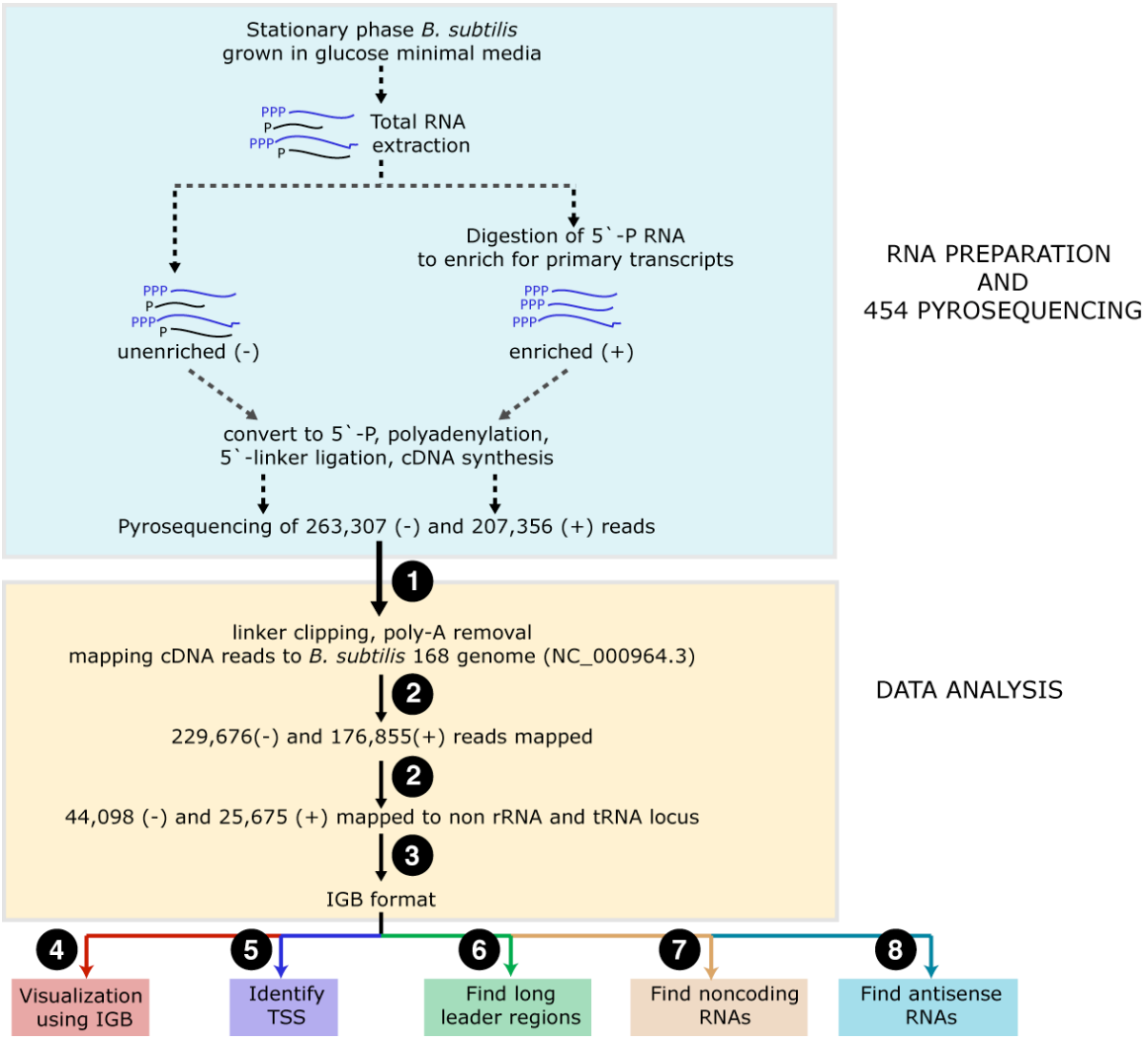
Gene	Fold-change (EPS- vs WT)	
	Microarray	RNA-Seq
ComK-regulated genes		
comC	-4.0	-2.8
comEA	-2.5	-2.2
comEB	1.2	1.1
comEC	-6.5	1.2
comER	-68.6	-17.6
comFA	3.2	1.1
comFB	1.9	1.4
comFC	1.9	1.1
comGA	-1.3	-7.7
comGB	1.3	-9.1
comGE	-1.5	-3.0
comGF	-1.5	2.0
comGG	-2.8	-1.0
comK	-3.6	-3.6
yyaF	48.5	1.5
yneA	-1.1	-1.3
yneB	1.1	1.0
recA	-1.1	1.1
nucA	1.3	1.1
nin	-1.1	-2.2
ssbA	2.6	1.3
ssbB	2.6	-1.9
ywfM	-2.6	1.1
maf	1.1	-1.5
yvyF	1.1	1.0
yvrP	2.1	2.4
yvzC	-4.0	-5.0
ybdK	1.4	1.8
sacX	-1.6	-4.2
rapH	1.5	1.4
recA	-1.1	1.1
***srf* operon**		
srfAA	-1.9	-1.0
srfAB	-1.3	-1.1
srfAC	1.0	1.0
srfAD	1.0	-1.33
***dhb* operon**		
dhbA	2.3	1.4
dhbB	4.0	1.5
dhbC	2.0	1.2
dhbE	2.0	1.6
dhbF	2.1	1.3

Gene	Fold-change (EPS- vs WT)	
	Microarray	RNA-Seq
pps operon		
ppsA	3.0	2.1
ppsB	2.3	2.2
ppsC	2.5	2.0
ppsD	3.0	1.9
ppsE	2.3	1.8
***yqxM* operon**		
yqxM	8.6	14.3
sipW	7.5	9.6
tasA	7.0	9.5
***sdp* operon**		
sdpA (yvaW)	3.0	3.5
sdpB (yvaX)	4.3	2.3
sdpC (yvaY)	1.1	1.5
***skf* operon**		
skfA (ybcO)	-4.3	-2.0
skfB (ybcP)	-1.6	-1.6
skfC (ybcS)	-1.3	1.0
skfD (ybcT)	1.2	-
skfE (ybdA)	1.1	1.3
skfF (ybdB)	-1.4	1.4
skfG (ybdD)	1.0	1.3
skfH (ybdE)	1.2	1.5
***eps* operon**		
epsA (yveK)	2.5	3.2
epsB (yveL)	1.9	1.2
epsC (yveM)	2.1	5.2
epsD (yveN)	1.7	4.1
epsE (yveO)	1.1	5.3
epsF (yveP)	-1.1	1.7
epsG (yveQ)	-3.7	-1.8
epsH (yveR)	-2.3	-1.8
epsI (yveS)	-2.0	-2.3
epsJ (yveT)	-4.6	-4.0
epsK (yvfA,B)	-3.2, -3.5	-1.3
epsL (yvfC)	-2.6	-1.0
epsM (yvfD)	-2.6	-1.2
epsN (yvfE)	-2.8	-1.6
epsO (yvfF)	-3.5	1.1

APPENDIX II

Differential RNA Sequencing with 5` transcription start site (TSS) enrichment

Data Analysis



1. Processing the raw cDNA reads

For this experiment, we obtained 263,307 and 205,356 total cDNA reads from unenriched and 5`-enriched samples, respectively. Due to the sequencing

methodology (see above; further details are available in Materials and Methods), the resulting cDNA reads will contain 5`-linker and 3`-polyA sequences.

We used an in-house python script to automate the removal of those sequences (shown in the schematic above). 5`-linker sequences were identified and removed using the "pairwise2" biopython module with the following paramaters:

```
pairwise2.align.localms([linker_sequence],[input_seq], 2,-3,-2,
                        -0.2,one_alignment_only=1)

The scoring system used in this case is similar to the one used
by BLAST for match, mismatch, gap opening, gap elongation,
respectively.
```

In this case, only cDNA reads with a maximum of 4 mismatches (including insertion and deletion) were tolerated and used for subsequent analysis. 3`-polyA sequences were detected and removed using python's "regular expression"-based approach. Since the oligonucleotide used for cDNA synthesis contained 9 'A' residues at the 3`-end, we considered 8 or more continuous 'A' residues as "polyA" tail and was deleted. We chose 8 'A' as an optimum compromised based on a preliminary analysis looking at ~5000 cDNA reads. In some instances, we found cDNA reads containing non-'A' residues following the polyA tail. These trailing nucleotides will also be removed by our current approach. cDNA reads with less than 8 'A' residues at the 3`-end will trimmed up to the closest non-'A' nucleotide. It is interesting to note that we also identified cDNA reads with correct 5`-linker sequences but did not contain

polyA sequences (<10%). These cDNAs might be derived from incomplete second-strand synthesis of long cDNAs or atypical reverse transcription reactions. Regardless, these sequences were included in the downstream analyses since they contained the correct 5`-linker sequences. 262,677 and 203,232 cDNA reads were retained from the unenriched and enriched samples, respectively, after the linker and polyA removal steps.

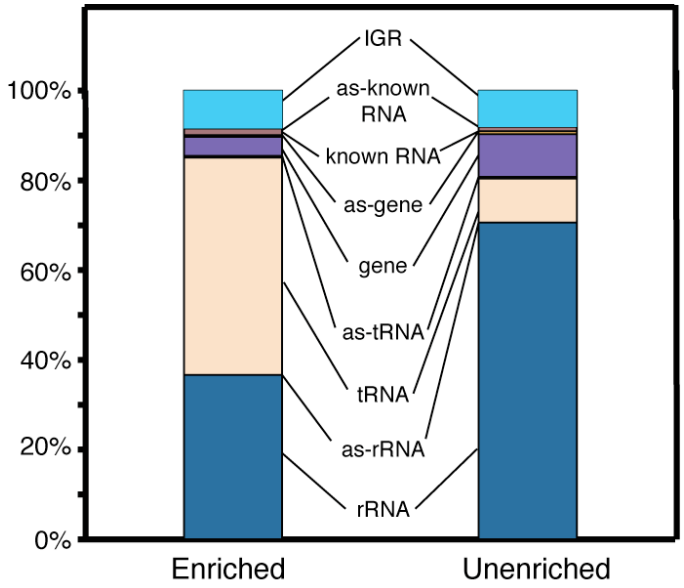
2. Mapping cDNA reads onto the *B. subtilis* genome

We utilized the "segemehl" software (version 0.0.5; Hoffman *et al.*, 2009; freely available at http://www.bioinf.uni-leipzig.de/Software/segemehl/) to map the processed cDNA reads onto the latest *B. subtilis* genome (NC_000964.3; released in 2010). To reduced false positives, we only mapped cDNA reads longer than 15 nucleotides (the same cutoff was also used in several other publications). The following parameters were used for the mapping:

```
Genome indexing:
      > segemehl.x -x [index_file] -d [genome_file]

Alignment:
      >segemehl.x -i [index_file] -d [genome_file] -q
       [input_file]  -m 15 -E 0.1 -A 80 -H 2

Using these parameters, only cDNA reads that could be mapped with
at least 80% accuracy and e-value < 0.1 were reported.
```

In total, 229,676 (87.4%) and 176,855 (87%) cDNA reads were successfully mapped onto *B. subtilis* genome for the unenriched and enriched datasets, respectively. The average length of the cDNA reads was ~100 nucleotides. It is important to note that, in contrast to the Illumina RNA-Seq described in Chapter 3, the methodology used to

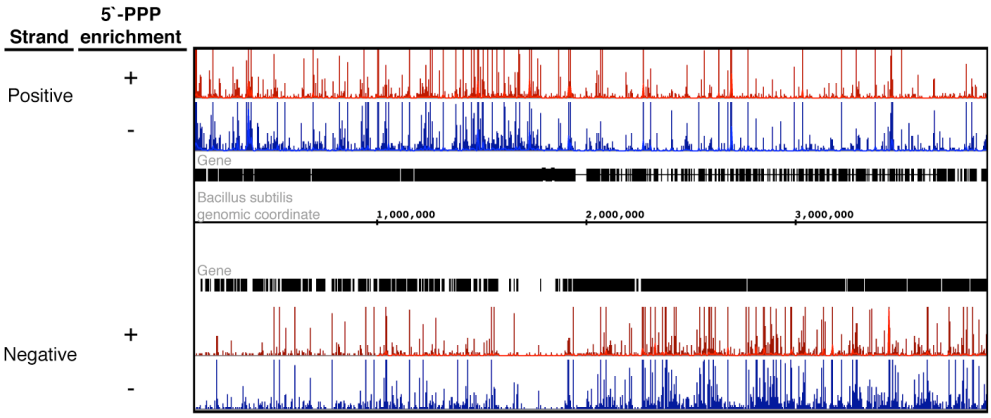
enrich for the transcriptional start site (TSS) was strand-specific, thus we could definitively mapped the cDNA reads onto the positive and negative genomic strands. All cDNA reads mapped to various rRNA and tRNA loci were excluded from further analyses.



	Enriched	Unenriched
rRNA	64,788	162,124
as-rRNA	7	10
tRNA	86,381	23,422
as-tRNA	4	22
Gene	7,651	21,817
as-gene	512	1,475
Known RNA	2,620	2,252
as-known RNA	13	25
IGR	14,879	18,529
TOTAL	176,855	229,676

3-4. Conversion to the IGB format for data visualization

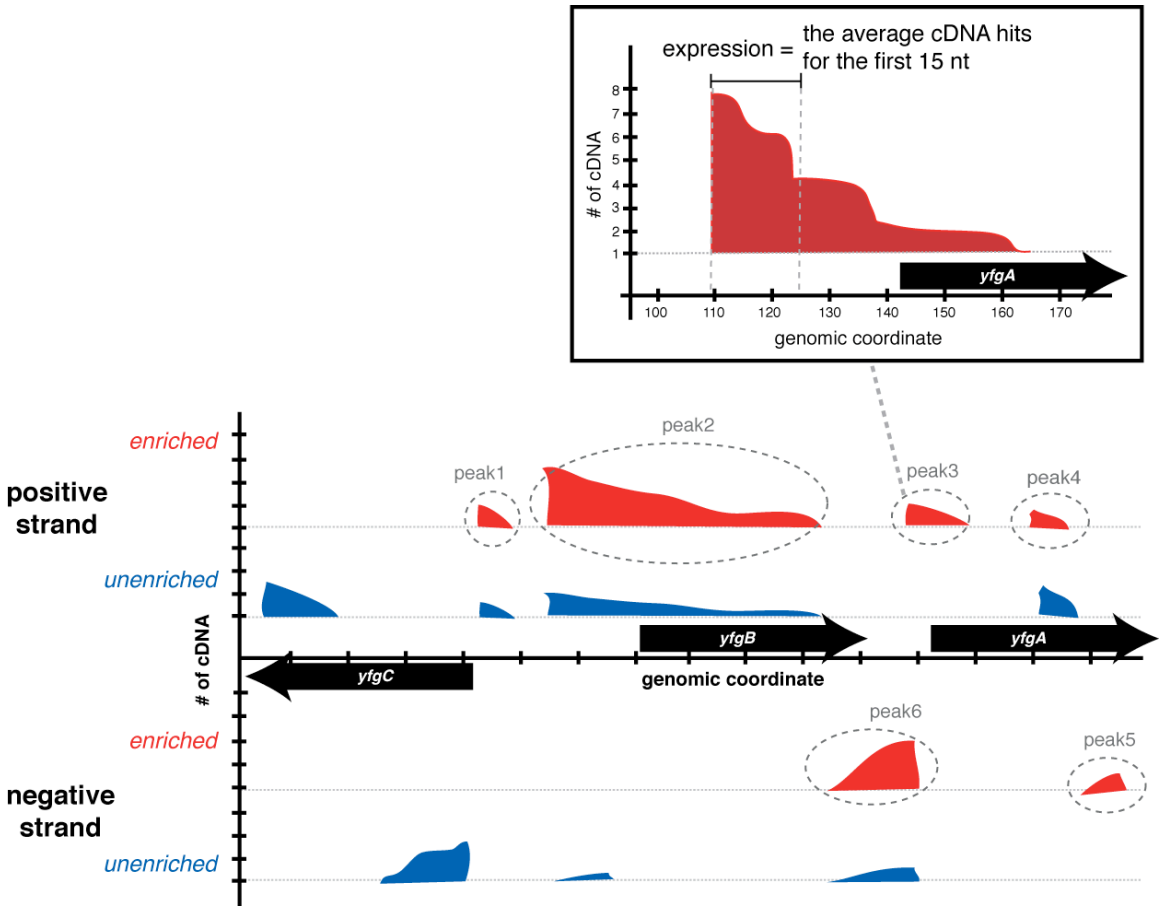
The output from "segemehl" was converted into the IGB-appropriate format using an in-house python script. More details on the IGB format can be found in Appendix I. The genomic overview is shown below using IGB. The upper panel represents the cDNA reads mapped to the positive genomic strand. The lower panel represents the cDNA reads mapped to the negative genomic strand. The Red and blue graphs denote data from the enriched and unenriched datasets, respectively. As expected, the distribution of cDNA reads correlates with the genes distribution.



5. TSS identification

We utilized the ".gr" file generated in step #3 to first identify all genomic regions (both the positive and negative strands) covered by at least one cDNA reads from the enriched dataset (herein referred to as "peaks"). We reasoned that due to the specific shape of the enrichment peaks (higher near the 5`-end), the expression of these peaks could be approximated by the average number of cDNA hits at the nucleotide positions proximal to their 5`-end. In this case, we chose to use the first 15 nucleotides since it was also used as a cutoff for mapping cDNA reads in step #2. The fold-enrichment was calculated as the ratio of the peak expression in the enriched and unenriched datasets. Thus, a higher ratio would indicate a higher probability of it being the real TSS.

```
Fold-enrichment = enriched_peak_expression/unenriched_peak_expression
```



It is important to note that our analysis ignored peaks that are present exclusively in the unenriched dataset (*e.g.*, the peak on the negative strand for the *yfgC* gene). It is possible that these peaks could correspond to RNA processing sites, which mostly generate transcripts with 5`-monophosphate. However, due to the lack of well-characterized examples (as positive controls), we decided not to further examine these peaks. In the future, it might be interesting to analyze these peaks and potentially use them as "markers" for identifying processing sites. In addition, we also excluded

potential TSSs that were found within coding regions. We plan to assess this class of TSSs in the near future.

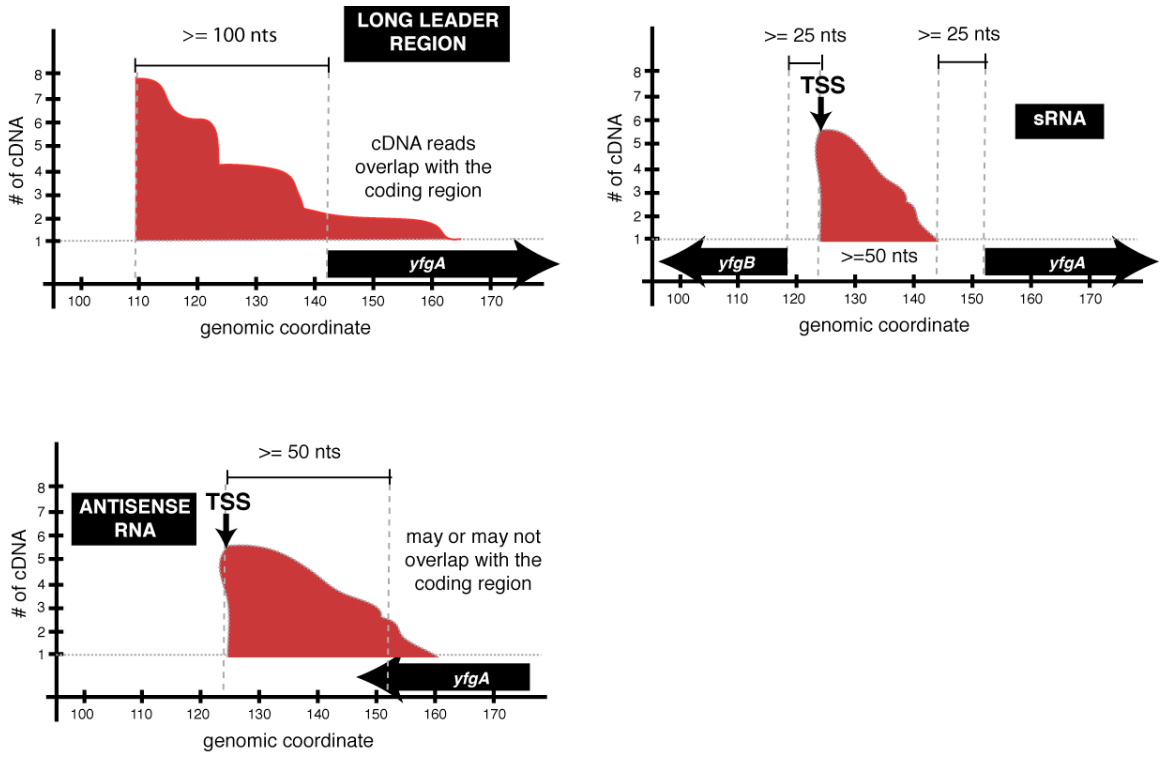
Overall, we successfully cataloged ~600 peaks as potential TSSs using the following criteria: located upstream of genes transcribed in the same direction, started within the noncoding regions, and represented by at least 2 cDNA reads. To assess the accuracy of our TSS dataset, we compared a subset of these peaks to the experimentally characterized TSS cataloged by DBTBS (Sierro *et al.*, 2008; http://dbtbs.hgc.jp/). On average, our predicted TSSs matched to within one nucleotide of the previously characterized TSSs.

6. Identification of long 5`-leader regions

In most cases, a long 5`-leader region implies regulatory function. Most of the known *cis*-acting regulatory RNAs (*i.e.*, protein-sensing RNAs, tRNA-sensing RNAs, and metabolite-sensing RNAs) were located within leader regions of 100 nucleotides or longer. Hence, to identify more examples of long leader regions, we specifically focused on enriched TSSs located at least 100 nucleotides upstream open reading frames. In addition, we exlusively looked at the TSSs with cDNA reads overlapping the downstream coding regions (*e.g.*, peak2 in the schematic shown above). Following this approach, we successfully identified 40 long leader regions shown in Table 4-1.

It is important to note that we tried to be more conservative in our prediction by not including peaks with cDNA reads ended upstream of the downstream coding regions. We note that some of the known *cis*-acting regulatory RNAs also showed this profile presumably due to the transcriptional attenuation mechanisms. However, it is

also possible that these peaks represent independent transcripts. Thus, in the absence of other data, we opted to exclude these regions from consideration.



7. Identification of noncoding RNAs

Using the peaks identified in step #5, we cataloged potential new noncoding RNA based on the following criteria: located exclusively within the noncoding regions, at least 50 nucleotides in length, and at least 25 nucleotides away from the neighboring genes. We arbitrarily determined the 25 nucleotides cutoff based on looking at the overall peaks distribution. In total, we obtained 54 candidate noncoding RNAs which are listed in Table 4-2 and Table 4-3. We note that some of these candidate RNAs might encode for small peptides. However, this does not eliminate the possibility of the RNA playing a regulatory role in addition to being a protein-encoding mRNA (for instance, SgrS in *E. coli*).

8. Identification of antisense RNAs

Antisense RNAs (asRNAs) are not well-characterized in *B. subtilis*. In fact, there are only two known examples of antisense RNAs in this organism (*i.e.*, *ratA* and as-*yabE*). Both of these RNAs are arranged in a tail-to-tail conformation relative to their targets (overlapping at the 3`-ends as shown in the schematic above). However, antisense RNAs can also be found in a head-to-head conformation (overlapping at the 5`-ends). As such, we searched for potential antisense RNAs with both conformations in our datasets. We specifically focused on the TSSs predicted to be located within 50 nucleotides of coding regions transcribed in the opposite directions. TSSs located further than 50 nucleotides away would probably indicate potential noncoding RNAs. In this case, we decided to include all the peaks whether or not their cDNA reads overlap with the downstream coding regions. The 29 candidate asRNAs identified through this analysis were listed in Table. 4-4.

APPENDIX III

List of *Bacillus subtilis* small RNAs

sRNA	Location (°)	Start	End	Orientation	Size	Left Gene	Right Gene	Transcription/ Sigma Factors	Function	Conditions	References
Abundant/Housekeeping RNAs											
scRNA/4.5S RNA	1	26434	26740	/+/+/+/	328	*tadA*	*dnaX*		component of signal recognition particle (with Ffh)	ALL	Nakamura et al. 1992
6S-2	179	2095899	2096126	/-/-/-/	228	*yocI*	*yocJ*		Transcription regulator	ALL	Barrick et al. 2005; Trotochau and Wassarman 2005
rnpB	194	2273701	2273884	/-/-/-/	408	*ypsC*	*ypsB*		RNA component of RNAse P enzyme	ALL	-
6S-1	240	2814483	2814693	/-/-/-/	211	*yrvM*	*aspS*		Transcription regulator	ALL	Barrick et al. 2005; Trotochau and Wassarman 2005
ssrA/tmRNA	295	3450709	3451232	/-/-/-/	524	*yvaG*	*smpB*		degradation of mRNA without stop codon	ALL	Ito et al. 2002
Small RNAs with known functions											
as-yabE	4	49990	50013	/+/-/+/	24	*yabE*	*rnmV*	ECF sigma factors	antisense RNA to yabE		Eiamphungporn and Helmann 2009
ncr1575/ncr10 /vmlR	52	606407	606605	/-/-/-/	199	*vmlR*	*ydgF*		antibiotic resistance	exp, stat	Ohki et al. 2005
fsrA/ncr746/ncr36	127	1483553	1483811	/+/+/+/	259	*ykuI*	*ykuJ*	Fur	Iron-regulation	exp, stat	Gaballa et al. 2008
SR1	131	1534084	1534289	/-/+/-/	205	*slp*	*speA*	CCpN, CcpA	arginine catabolism	Gluconeogenic conditions	Licht et al. 2005
ratA	229	2678343	2678586	/-/+/-/	244	*txpA*	*yqbM*		RNA antitoxin to txpA(yqdB)		Silvaggi et al. 2005
Small RNAs with predicted functions											

ncr1857 /bsrE	177	2069869	2070128	/-/-/-/	260	*yobI*	*yobJ*		*RNA antitoxin*	stat	Saito et al. 2009
ncr1019/ncr39 /as-bsrE	177	2069821	2069991	/-/+/-/	171	*yobI*	*yobJ*		*toxic peptide*	exp_LB, stat	Irnov et al. 2010; Rasmussen et al. 2009
ncr1155/ncr58 /bsrH	229	2678729	2678961	/+/+/-/	233	*yqdB*	*yqbM*		*toxic peptide*	exp_LB, stat	Saito et al. 2009
ncr2166/as-bsrH	229	2678994	2679076	/+/-/-/	83	*yqdB*	*yqbM*		*RNA antitoxin*	stat	Irnov et al. 2010
ncr1058/ncr46 /bsrG	194	2273533	2273829	/-/+/-/	297	*yolA*	*yokL*		*toxic peptide*	exp	Saito et al. 2009
ncr1932/as-bsrG	194	2273701	2273884	/-/-/-/	184	*yolA*	*yokL*		*RNA antitoxin*	stat	Irnov et al. 2010
Previously identified small RNAs with unknown functions											
bsrC	40	474329	474597	/+/-/+/	280	*ydaG*	*ydaH*			exp	Saito et al. 2009
ncr629/ncr22/ rsaE	105	1233429	1233545	/-/+/-/	117	*yizD*	*yjbH*			exp, stat	Irnov et al. 2010; Rasmussen et al. 2009
SurA	164	1914993	1915255	/+/-/+/	280	*yndK*	*yndL*	AbrB, SigA		sporulation, biofilm	Silvaggi et al. 2005
ncr1021/ncr40 /bsrF/SR2	178	2079095	2079225	/+/+/-/	117	*yobO*	*csaA*	CodY		exp, stat	Preis et al. 2009
SurC	224	2625953	2626043	/-/+/-/	90	*dnaJ*	*dnaK*	SigK		sporulation	Silvaggi et al. 2005
bsrI	287	3360971	3361189	/-/-/-/	280	*yurI*	*yurZ*			exp	Saito et al. 2009
bsrD	47	-	-	/+/?/-/	70	*yddM*	*yddN*			exp	Saito et al. 2009
Small RNAs with potential for coding regions (other than the toxin-antitoxin systems above)											
yckB-yckC	32	368981	369087	/-/+/+/	107	*yckB*	*yckC*	*SigB*	*small peptide*	EPS+ biofilm	Irnov_unpublished
ncr1733/ncr26	116	1357727	1357833	/+/-/-/	107	*ykcC*	*htrA*		*small peptide*	exp	Irnov et al. 2010; Rasmussen et al. 2009
yrpD-yrpE	234	2740745	2740881	/+/+/-/	137	*yrpD*	*yrpE*		*small peptide*	EPS+ biofilm, sporulation	Irnov_unpublished; Schmalish et al. 2010
ncr2184/ncr60	237	2779137	2779368	/-/-/-/	232	*yrzI*	*yrhG*		*small peptide*	exp_M9, stat	Irnov et al. 2010; Rasmussen et al. 2009

ncr73	314	3672158	3672728	/-/-/+/	571	*yvzI*	*yvzE*		*small peptide*	exp	Rasmussen et al. 2009
ncr79	341	3996389	3996607	/+/+/+/	219	*pepT*	*yxjJ*		*small peptide*	exp	Rasmussen et al. 2009
ncr82	350	4097016	4097344	/-/+/+/	329	*yxbC*	*yxbB*		*small peptide*	exp_M9	Rasmussen et al. 2009
Other uncharacterized small RNAs											
ncr2897	0	157	274	/+/+/+/	118	*start*	*dnaA*			stat	Irnov et al. 2010
ncr2	5	56150	56302	/+/-/+/	153	*spoVG*	*gcaD*			exp_M9	Rasmussen et al. 2009
ncr3	13	154753	155103	/+/-/+/	351	*rpsI*	*ybaJ*			exp_M9	Rasmussen et al. 2009
ncr178	17	199857	199980	/+/+/+/	124	*glmM*	*glmS*			stat	Irnov et al. 2010
ncr181/ncr4/adaB-ndhF	18	204991	205148	/+/+/+/	156	*adaB*	*ndhF*			exp_M9, EPS-biofilm	Irnov et al. 2010; Rasmussen et al. 2009; Irnov_unpublished
ncr214	24	275609	275716	/+/+/+/	108	*garD*	*ycbJ*			stat	Irnov et al. 2010
ncr6	27	318974	319170	/+/+/-/	197	*yceI*	*yceJ*			exp	Rasmussen et al. 2009
ncr264	32	376678	376925	/+/+/+/	248	*hxlR*	*srfAA*			stat	Irnov et al. 2010
ncr1562	45	532583	532642	/+/-/+/	60	*ydcO*	*ydcP*			stat	Irnov et al. 2010
ncr1565	47	554497	554557	/+/-/+/	61	*yddR*	*yddS*			stat	Irnov et al. 2010
ncr1566	48	559532	559610	/+/-/-/	79	*cspC*	*ydeB*			stat	Irnov et al. 2010
ncr8	48	560674	561156	/-/+/-/	483	*ydeB*	*ydzE*			exp_M9	Rasmussen et al. 2009
ncr9	51	591936	592220	/+/+/-/	285	*ydfJ*	*nap*			exp_M9	Rasmussen et al. 2009
ncr12	56	659158	659332	/-/-/-/	175	*ydzW*	*ydiR*			exp_LB	Rasmussen et al. 2009
ncr13	57	663344	663562	/+/+/+/	219	*ydjA*	*ydjB*			exp	Rasmussen et al. 2009
ncr14	58	679090	679220	/-/-/+/	131	*bdhA*	*ydjM*			exp	Rasmussen et al. 2009
ncr465	68	796025	796075	/+/+/+/	51	*yetO*	*ltaSA*			stat	Irnov et al. 2010
ncr16	69	809409	809561	/+/-/+/	153	*yfmS*	*ymfR*			exp_M9	Rasmussen et al. 2009
ncr471	70	820666	820838	/-/+/+/	173	*yfmI*	*yfmG*			stat	Irnov et al. 2010
ncr17	87	1018648	1018800	/-/+/+/	153	*phoA*	*lytE*			exp_LB	Rasmussen et al. 2009
ncr560/ncr18	90	1056390	1056619	/-/+/+/	230	*yhaZ*	*yhaX*			exp, stat	Irnov et al. 2010; Rasmussen et al. 2009
ncr1670/hinT-ecsA	92	1077246	1077313	/-/-/+/	275	*hinT*	*ecsA*	*SigF, SigG*		stat, EPS+ biofilm	Irnov et al. 2010
ncr19	95	1117866	1117930	/+/+/+/	65	*yxhD*	*yxjA*			exp	Rasmussen et al. 2009
ncr585/ncr20	98	1150478	1150678	/-/+/-/	201	*gerPA*	*yisI*			exp_M9	Irnov et al. 2010; Rasmussen et al. 2009

ncr620	104	1219702	1219801	/-/+/+/	100	*trpS*	*oppA*		stat	Irnov et al. 2010
ncr25	109	1279031	1279381	/-/+/-/	351	*yjdJ*	*ctaO*		exp_M9	Rasmussen et al. 2009
ncr33	123	1435278	1435430	/-/+/-/	153	*motA*	*clpE*		exp_M9	Rasmussen et al. 2009
ncr721/ncr34	124	1446806	1446917	/+/+/+/	112	*ykzR*	*ykvR*		exp_M9, stat	Irnov et al. 2010; Rasmussen et al. 2009
ncr724	124	1451260	1451315	/+/+/+/	56	*stoA*	*zosA*		stat	Irnov et al. 2010
ncr1755/ncr35	124	1453368	1453554	/+/-/+/	187	*zosA*	*ykvY*		exp	Irnov et al. 2010; Rasmussen et al. 2009
ncr738	125	1467704	1467761	/-/+/+/	58	*ykwD*	*pbpH*		stat	Irnov et al. 2010
ncr37	130	1527937	1528111	/+/-/+/	175	*ykyA*	*pdhA*		exp_LB	Rasmussen et al. 2009
Rc-ylbH	134	1569223	1569335	/+/-/+/	113	*ylbG*	*ylbH*	*SigF, SigG*	EPS+ biofilm	Irnov_unpublished
ncr826	136	1596338	1596409	/+/+/+/	72	*sbp*	*ftsA*		stat	Irnov et al. 2010
ncr952	152	1780404	1780554	/+/+/-/	151	*mutL*	*ymzD*		stat	Irnov et al. 2010
ncr969	160	1868404	1868461	/+/+/+/	58	*ymzA*	*ymaA*		stat	Irnov et al. 2010
ncr976	162	1900528	1900579	/+/+/-/	52	*yncF*	*cotU*		stat	Irnov et al. 2010
ncr977	162	1901991	1902096	/-/+/+/	106	*cotU*	*thyA*		stat	Irnov et al. 2010
ncr982	164	1917501	1917580	/+/+/-/	80	*yndN*	*lexA*		stat	Irnov et al. 2010
ncr992/yneK-cotM	164	1925548	1925619	/-/+/-/	85	*yneK*	*cotM*		stat, EPS+ biofilm	Irnov et al. 2010
ncr1015	175	2053989	2054108	/-/+/-/	120	*pps*	*xynA*		stat	Irnov et al. 2010
ncr1855	177	2069075	2069168	/-/-/-/	94	*yobI*	*yobJ*		stat	Irnov et al. 2010
desR-yocH/yocG-yocH	179	2092364	2092496	/+/-/-/	133	*desR*	*yocH*	SigE	EPS+ biofilm, sporulation	Irnov_unpublished; Schmalish et al. 2010
ncr1876	179	2099817	2099956	/-/-/-/	140	*yozO*	*yozC*		stat	Irnov et al. 2010
ncr42	181	2116746	2117030	/-/+/-/	285	*cwlS*	*yojK*		exp_M9	Rasmussen et al. 2009
ncr1915	189	2208755	2208813	/-/-/-/	59	*yopM*	*yopL*		stat	Irnov et al. 2010
ncr43	190	2219992	2220276	/-/-/-/	285	*yonT*	*yonS*		exp_M9	Rasmussen et al. 2009
ncr1052/ncr44	190	2221800	2221982	/-/+/+/	183	*yonR*	*yonP*		exp_LB, stat	Irnov et al. 2010; Rasmussen et al. 2009
ncr45	190	2226070	2226618	/+/+/+/	549	*yonN*	*yonK*		exp	Rasmussen et al. 2009
ncr1935	195	2282621	2282724	/+/-/-/	104	*yokD*	*yokC*		stat	Irnov et al. 2010
ncr1937	195	2283685	2283753	/-/-/-/	69	*yokC*	*yokB*		stat	Irnov et al. 2010

ncr47	196	2299176	2299416	/-/+/-/	241	*ypjP*	*ypiP*		exp_M9	Rasmussen et al. 2009
ncr48	197	2307661	2307813	/+/-/+/	153	*ugtP*	*cspD*		exp_M9	Rasmussen et al. 2009
ncr1957	198	2316348	2316407	/-/-/-/	60	*ypbR*	*ypbQ*		stat	Irnov et al. 2010
ncr50	202	2367684	2367946	/-/+/-/	263	*ypiA*	*aroE*		exp_M9	Rasmussen et al. 2009
ncr52	212	2477751	2477925	/+/-/-/	175	*yqjL*	*rnz*		exp_M9	Rasmussen et al. 2009
ncr53	213	2488824	2488976	/-/+/-/	153	*yqjC*	*yqjB*		exp_M9	Rasmussen et al. 2009
ncr1118	217	2540930	2541015	/-/+/+/	86	*yqhR*	*yqhQ*		stat	Irnov et al. 2010
ncr56	226	2647051	2647291	/-/-/+/	241	*yqeG*	*sda*		exp_LB	Rasmussen et al. 2009
ncr2160	226	2647405	2647663	/-/-/-/	259	*yqeG*	*yqeF*		stat	Irnov et al. 2010
ncr57	226	2647695	2647869	/+/+/-/	175	*sda*	*yqeF*		exp_M9, stat	Rasmussen et al. 2009
ncr1159	230	2692882	2692931	/-/+/-/	50	*yqaO*	*yqaN*		stat	Irnov et al. 2010
ncr1160	230	2697037	2697106	/-/+/-/	70	*yqaJ*	*yqaI*		stat	Irnov et al. 2010
ncr59	231	2707768	2708052	/+/-/-/	285	*yrkN*	*yrkL*		exp_M9, stat	Rasmussen et al. 2009
ncr2173	234	2734262	2734358	/-/-/-/	97	*yrdB*	*yrdA*		stat	Irnov et al. 2010
ncr2179	235	2752023	2752133	/-/-/-/	111	*yraI*	*yraH*		stat	Irnov et al. 2010
ncr1175/yrhK-yrhJ	237	2773780	2773886	/+/+/-/	107	*yrhK*	*cypB*	SigD	stat, sporulation	Irnov,Irnov_unpublished, Schmalish et al. 2010
ncr2185	237	2780319	2780477	/-/-/-/	159	*yrhG*	*yrhF*		stat	Irnov et al. 2010
ncr2299	249	2913485	2913583	/-/-/-/	99	*trxA*	*xsa*		stat	Irnov et al. 2010
ncr63	255	2989649	2989801	/-/+/-/	153	*accD*	*ytsJ*		exp_LB	Rasmussen et al. 2009
ncr2339	255	2991183	2991239	/-/-/-/	57	*ytsJ*	*dnaE*		stat	Irnov et al. 2010
ncr2360/rpsD-tyrS	259	3036340	3036527	/+/-/-/	180	*rpsD*	*tyrS*	*SigF, SigG, SigB*	stat, EPS+ biofilm	Irnov et al. 2010, Irnov_unpublished
ncr64	259	3036396	3036570	/+/+/-/	175	*rpsD*	*tyrS*		exp_M9	Rasmussen et al. 2009
ncr1221	262	3072289	3072361	/-/+/+/	73	*ythP*	*ytzE*		stat	Irnov et al. 2010
ncr66	265	3107025	3107177	/+/-/-/	153	*ytvA*	*yttB*		exp	Rasmussen et al. 2009
ncr2424	269	3146126	3146183	/-/-/-/	58	*mntA*	*menC*		stat	Irnov et al. 2010
ncr1241	275	3225697	3225824	/-/+/-/	128	*yugI*	*alaT*		stat	Irnov et al. 2010
ncr2507	282	3302792	3302876	/-/-/+/	85	*yuzG*	*guaC*		stat	Irnov et al. 2010
ncr70	292	3419450	3419646	/+/+/-/	197	*fhuD*	*lysP*		exp_M9	Rasmussen et al. 2009
cggR-araE	298	3483888	3483967	/-/-/-/	80	*cggR*	*araE*	*SigF, SigG*	EPS+ biofilm	Irnov_unpublished
ncr2637	305	3573045	3573118	/-/-/-/	74	*yvcI*	*trxB*		stat	Irnov et al. 2010
ncr2658	310	3625573	3625632	/-/-/-/	60	*ftsE*	*cccB*		stat	Irnov et al. 2010

ncr2665	310	3631679	3631755	/-/-/-/	77	*yvyD*	*yvzG*	stat	Irnov et al. 2010
ncr74	324	3788243	3788439	/-/+/-/	197	*atpI*	*upp*	exp	Rasmussen et al. 2009
ncr2752/ncr75	325	3804713	3804968	/-/-/-/	256	*rho*	*ywjI*	exp_LB, stat	Irnov et al. 2010; Rasmussen et al. 2009
ncr2768	329	3852061	3852116	/+/-/-/	56	*pbpG*	*ywhD*	stat	Irnov et al. 2010
ncr77	337	3941821	3942215	/+/-/+/	395	*epr*	*sacX*	exp_M9	Rasmussen et al. 2009
ncr1421	341	3996388	3996610	/+/+/+/	223	*pepT*	*yxjJ*	stat	Irnov et al. 2010
ncr2857	352	4122960	4123062	/-/-/+/	103	*yyzE*	*yydK*	stat	Irnov et al. 2010
ncr83	356	4172042	4172096	/-/+/-/	55	*yyzJ*	*yyzK*	exp_M9	Rasmussen et al. 2009
ncr84	358	4187411	4187496	/+/+/-/	86	*yyaP*	*tetB*	exp	Rasmussen et al. 2009

APPENDIX IV

LIST OF OLIGONUCLEOTIDES

	Sequence (5`→3`)	Notes
irv5	AAAGAATTCTAAAAAACAATTGACCGTTTATGCCACATGTTGTAAAATCAAGCTTGCGCCCGAACTAAGCG	Forward primer for moving the transcription start site of *glmS* to the site of ribozyme self-cleavage. Through use of this oligo, the endogenous *glmS* promoter region was grafted to the G+1 at the site of self-cleavage.
irv13	GTTCGGTATGGGAACGGG	Reverse primer for *B. subtilis* 5S RNA.
irv18	CCCTACTCTCGCATGGGGAG	Reverse primer for *E. coli* 5S rRNA (rrfB); about 100 nt from the 5' end; to be used for Northern blot
irv27	GATCGCACCATGGCAAAATTTGTAAAAAATGATCAG	Forward primer for amplification of RNase J1 (ykqC); put into pBADMycHis-A with irv28
irv28	TGCGATCAAGCTTTTAAACCTCCATAATGATCGG	Reverse primer amplification of RNase J1; put into pBADMycHis-A with irv27
irv43	TAATACGACTCACTATAGGGTCCCCTCCTACATG	Forward primer with T7 promoter to make anti-sense probe to glmS UTR. Corresponds to nucleotides +70 to +81 relative to self-cleavage site. Used in combination with irv44.
irv44	CTTGTTCTTATTTTCTCAATAGG	Reverse primer to make anti-sense probe to glmS UTR. Corresponds to nucleotides -60 to -38 relative to self-cleavage site. Use with irv43.
irv104	TTTGGTGGCGATAGCGAAGAG	Reverse primer to synthesize anti-sense probe to B. subtilis 5S RNA. Corresponds to nucleotides +1 to +21 of rrnB. Used in combination with irv105
irv105	TAATACGACTCACTATAGGGCTTGGCGGCGTCCTACTCTC	Forward primer with T7 promoter to make anti-sense probe to B. subtilis 5S RNA. Corresponds to nucleotides +96 to +116 of rrnB. Used in combination with irv104.
irv120	CGACTACCATCGGCGCTGAA	Primer used for synthesis of anti-sense probe to *B. subtilis* 5S RNA. Corresponds to nucleotides +61 to +80 relative to *rrnB* translation start site.
irv134	AAGATCGGGGTGGGGGTCTAAGATCGGGGTGGGGGTCTCAGCAAATTGGCGATTAAGCCAGTTTGTTGATC	Reverse primer for E. coli hammerhead-rpsT experiment, to be used with irv133; goes down to 14 nts after rpsT stop codon, add rpstT "tag" (18 nts, to separate this from endogenous copy of rpsT)
irv135	GATCTCTAGAATAAAAAAACCCGCTTGCGCGGGCTTTTTCACAAAGCTAAGAATCGGGGTGGGGGTCT	Reverse primer for E. coli hammerhead-rpsT, to be used with irv132, will add terminator; overlap with rpsT tag (18 nts), basically ends at the natural terminator of rpsT
irv136	AAGATCGGGGTGGGGGTCT	Reverse primer for *rpsT* "tag", to be used for Northern blot of *rpsT* so it can be differentiated from the chromosomal copy of *rpsT*
irv144	GTAAAATCAAGCTTGATTATAATAAGATCCCGCTCGAGCGGGATCAGTGATCTGAAGAGCCGAAAG	Forward primer for E. coli hammerhead-rpsT experiment, includes 10 last nt from glmS promoter, 50 nts into the hammerhead construct (see Celesnik et al 2007); 25 nts overlap with irv146

irv145	AAAAGCATGCACAATTGACCGTTTATGCCACATGTTGTAAAATCAAGCTTGATTATAATAAGATCCC	Forward primer for E. coli hammerhead-rpsT experiment, include glmS promoter + 15 nts overlap with hammerhead; use after irv144. This primer comes from dividing irv132 into two parts to avoid primer dimer
irv146	GATCAGTGATCTGAAGAGCCGAAAGGCGAAACACGCGTAAGCGTGTCATCACTACGTAACGAGTGCC	Forward primer for E. coli hammerhead-rpsT experiment, start at nt 30 of hammerhead construct (see Celesnik et al 2007); 30 nts overlap with irv145; 20 nts overlap to rpsT transcript
JC10	GGTTTGAATTCGATAATTGTGAGACATACGG	Forward primer for amplification of *B. subtilis glmS* intergenic region. Adds EcoR1 site upstream of promoter.
JC11	TAAAATTGGATCCGCATCAAGCTGACCG	Reverse primer for amplification of *glmS* UTR. Corresponds to nucleotides +24 to +39 relative to *glmS* translation start site. Adds BamHI site.
JC29	GACTTTCTTCGTCGACTTTTACAATCTTAGGAGG	Forward primer for amplification of *B. subtilis glmS* coding region. Adds restriction site for cloning into plasmid pDG148 (BGSC, Ohio)
JC30	CACCCCTTTAGATAAGCATGCCCAAAGGGGTTAAAC	Reverse primer for amplification of *B. subtilis glmS* coding region. Adds restriction site for cloning into plasmid pDG148 (BGSC, Ohio)
JC43	TTACGGCTGTGATCTGCACACT	Reverse primer for *glmS* UTR, used for primer extension analyses. Corresponds to nucleotides +82 to +103 relative to ribozyme self-cleavage site.
WCW257	AGCGGTTGGATCCTTCGGTCCTCCGATC	Forward primer for amplification of *erm* from plasmid pMUTIN4 (BGSC, Ohio).
WCW258	AGTCATTTCTAGATCTCACTGCAGAGATCCC	Reverse primer for amplification of *erm* from plasmid pMUTIN4 (BGSC, Ohio).
WCW259	CACCAGTCTAGACACTGGCAACTCAAGAGC	Forward primer for amplification of the dowstream portion of *B. subtilis glmS* gene.
WCW260	TATTACTCTGCAGTAACACTCTTCGCAAGG	Reverse primer for amplification of the dowstream portion of *B. subtilis glmS* gene.
WCW261	AACTGAAGAATTCGTCCGTGTCATGG	Forward primer for amplification of the upstream portion of *B. subtilis glmS* UTR.
WCW407	TAATACGACTCACTATAGGGCGGCGTCCTACTCTCAC	Reverse oligo for amplification of *B. subtilis* 5S RNA. Used for synthesis of an antisense RNA probe Adds a T7 RNA polymerase promoter.
WCW408	GCGGAATTCAAGCTTTTTGGTGGCGATAGCGAAG	Forward oligo for amplification of *B. subtilis* 5S RNA.
WCW409	GCGGAATTCAAGCTTGCGGAATTCAAGCTTGTTCTTATTTTCTC	Forward oligo for amplification of *glmS* ribozyme. Used for synthesis of an antisense RNA probe.
WCW410	TAATACGACTCACTATAGGGCTTCGCATCAAGCTGACCG	Reverse oligo for amplification of *glmS* ribozyme. Used for synthesis of an antisense RNA probe. Adds a T7 RNA polymerase promoter.
WCW417	CATAAGAGAATTCTGAGACATACGGCAAAG	Forward primer for amplification of *glmS* promoter region.
WCW418	ATTGAGAGGATCCGAACAAGACAAGCTTG	Reverse primer for amplification of *glmS* promoter region.
irv126	TCCTCGGATCCTGGTGAAGGTAAATCG	Forward primer for amplification of the upstream portion of the EAR element; used to delete EAR
irv129	GATCGAATTCCACCGCAGGCATGATACTTTC	Reverse primer for amplification of the downstream portion of the EAR element; used to

		delete EAR
irv293	CCGCAGCTAAACCGATAAAAG	Forward primer for amplification of the downstream portion of the EAR element; used to delete EAR
irv308	CTTTTATCGGTTTAGCTGCGGGGTCAGGAGACATCACCTTTAC	Reverse primer for amplification of the upstream portion of the EAR element; used to delete EAR
irv321	GTGGGCAGCTGAAGTTGCTGGGATAGATGCTGGTCGTCGCTGG	Forward oligo to create the M3 EAR mutant using Pfu mutagenesis
irv322	CCAGCGACGACCAGCATCTATCCCAGCAACTTCAGCTGCCCAC	Reverse oligo to create the M3 EAR mutant using Pfu mutagenesis
irv323	CTGCTGTGGGCAGCTATTGTTGCTGCCTTAGATG	Forward oligo to create the M4 EAR mutant using Pfu mutagenesis
irv324	CATCTAAGGCAGCAACAATAGCTGCCCACAGCAG	Reverse oligo to create the M4 EAR mutant using Pfu mutagenesis
irv335	GTTTTTGTAAAGGTGATGTAGCCTGACCTGTATCAATG	Forward oligo to create the M5 EAR mutant using Pfu mutagenesis
irv336	CATTGATACAGGTCAGGCTACATCACCTTTACAAAAAC	Reverse oligo to create the M5 EAR mutant using Pfu mutagenesis
irv339	CTCGGCTCTGCTGTTCCCAGCTGAAGTTGC	Forward oligo to create the M6 EAR mutant using Pfu mutagenesis
irv340	GCAACTTCAGCTGGGAACAGCAGAGCCGAG	Reverse oligo to create the M6 EAR mutant using Pfu mutagenesis
irv341	CCTGTATCAATGGCTACACTCGGCTCTGC	Forward oligo to create the M7 EAR mutant using Pfu mutagenesis
irv342	GCAGAGCCGAGTGTAGCCATTGATACAGG	Reverse oligo to create the M7 EAR mutant using Pfu mutagenesis
irv343	CTACTAGTTTTTGTAAAAATGATGTCTCCTGACCTG	Forward oligo to create the M8 EAR mutant using Pfu mutagenesis
irv344	CAGGTCAGGAGACATCATTTTTACAAAAACTAGTAG	Reverse oligo to create the M8 EAR mutant using Pfu mutagenesis
irv352	GATCGAATTCCCTTCAGCGTGTGAAAC	Forward oligo to amplify the upstream portion of *nusG*; used to delete NusG
irv353	CTCATAACCAGAGTACGTGTGAACAAC	Reverse oligo to amplify the upstream portion of *nusG*; used to delete NusG
irv354	CACACGTACTCTGGTTATGAGTTCGTTAATATGTTTGGCCGTGAA	Forward oligo to amplify the downstream portion of *nusG*; used to delete NusG
irv355	AGCTGGATCCGAGCGTTAAACTCC	Reverse oligo to amplify the downstream portion of *nusG*; used to delete NusG
irv356	TCGGCGTCACCCTTATTATG	Forward oligo for qPCR measurement of *epsA*
irv357	TTTCGCTGGATGTCACTGAG	Reverse oligo for qPCR measurement of *epsA*
irv358	TTGGGAAAGTCAGCAGAACC	Forward oligo for qPCR measurement of *epsC*
irv359	CATTGCCGGATGTGTTACTG	Reverse oligo for qPCR measurement of *epsC*
irv360	ACTCTGACATTGCCCAAACC	Forward oligo for qPCR measurement of *epsH*
irv361	TCCATGAAAACCCTCTGACC	Reverse oligo for qPCR measurement of *epsH*
irv364	CAGTAAGCTTCGTCATCTTTGATTCC	Forward oligo for amplification of the EAR element used in making EAR-terminator-*lacZ* constructs
irv365	CGTTTAGAGGCCCCAAGGGGTTATGCTAAAGCGTCGACATGATAGCTGAGGAG	Reverse oligo for amplification of the EAR element used in making EAR-terminator-*lacZ* constructs; incorporate half of the T7 terminator sequence

irv366	GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGCTAGCCTAGGATCCCCAGCTTGTTG	Forward oligo for amplification of *lacZ* used in making EAR-terminator-*lacZ* constructs; incorporate half of the T7 terminator sequence
irv367	GATCGCATGCCTGATCGATAGTACATAATG	Reverse oligo for amplification of *lacZ* used in making EAR-terminator-*lacZ* constructs
irv368	CTCAAGGAAAGAGGCATTCG	Forward oligo for qPCR measurement of *epsF*
irv369	GTTCATGGATCCGTTCTTCC	Reverse oligo for qPCR measurement of *epsF*
irv370	ACATTGATTCACCCGTCAGC	Forward oligo for qPCR measurement of *epsM*
irv371	CCGTATTGATGATGCAGTGG	Reverse oligo for qPCR measurement of *epsM*
irv378	TAATACGACTCACTATAGGGAACACTATTCCTACTAGTT	Forward oligo for EAR amplification used for the inline probing experiment; incorporate T7 promoter sequence
irv379	CGGTTTTCGTCTCTCCAAAAAATC	Reverse oligo for EAR amplification used for the inline probing experiment; incorporate T7 promoter sequence
irv398	CTGCTTACCGCTGTCAGCTTG	Forward oligo corresponding t0 nts 100 of *epsC*
irv399	AATAAAAAGGCCTGCGATTACCAGCAGGCCTGTTATTAGCTCAGTACAAAAAACCCCTCAAGACCCG	Reverse oligo used to add lambda tR2 terminator sequence to the EAR-terminator-*lacZ* constructs
irv400	CTCATGCTAGCCAGATTAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGAAATAAAAAGGCCTGCGATTACCAG	Reverse oligo used to add *rrnB* T1 terminator sequence
irv415	CAGTGAATTCATGGCAAAGGTGTC	Forward oligo to amplify the upstream portion of *nusB*; used to delete NusB
irv416	GCTTATTCTCTTGCTGTTCTTCTTTTCAT	Reverse oligo to amplify the upstream portion of *nusB*; used to delete NusB
irv417	ATGAAAAGAAGAACAGCAAGAGAATAAGCGATTGCCAATGTTGACCGTGC	Forward oligo to amplify the downstream portion of *nusB*; used to delete NusB
irv418	ATGCGGATCCTGAACGAGAATGC	Reverse oligo to amplify the downstream portion of *nusB*; used to delete NusB
irv423	GGGGAAAAGGTTGAGTTACCGTATGCGGAAGAGGTAAAGCC	Forward oligo used to add lambda tRa terminator sequence to the EAR-terminator-lacZ constructs
irv424	TACTGCTAGCTCAGTTCCAAGC	Reverse oligo used to add lambda tRa terminator sequence to the EAR-terminator-lacZ constructs
irv430	TGACCCGCTATTTCATGACG	Forward oligo for qPCR measurement of *epsC*
irv431	TTTACGGGCTCTCCCATATC	Reverse oligo for qPCR measurement of *epsC*
irv432	AAGGCCTCGGTATGAATGTG	Forward oligo for qPCR measurement of *epsD*
irv433	CAGCACTGTCACCGATTTTG	Reverse oligo for qPCR measurement of *epsD*
irv434	CGGCGGTCATTTACGTATTC	Forward oligo for qPCR measurement of *epsE*
irv435	TCACAGCTGCAGGCATAAAG	Reverse oligo for qPCR measurement of *epsE*
irv436	CATTCGCTTTCAGCTCGTTC	Forward oligo for qPCR measurement of *epsF*; 2nd independent probe
irv437	GTTCATGGATCCGTTCTTCC	Reverse oligo for qPCR measurement of *epsF*; 2nd independent probe
irv445	caggggatcccggtcTacccgtaatcttacgtc	Reverse oligo to remove the SalI site within the *lacZ* coding region
irv456	GCATGCTAGCTTCATGGATCC	Reverse oligo to amplify all 4 *epsF* terminators
irv457	TCCCGCTAGCGAAGAAGGTG	Reverse oligo to amplify 3 *epsF* terminators
irv458	CTCAGCTAGCTCACAAATGTAAGG	Reverse oligo to amplify 2 *epsF* terminators
irv459	CACTGCTAGCAACATGGAGCAC	Reverse oligo to amplify region upstream of the *epsF* terminators
irv467	GTCAGTCGACAATGCCATTCTTG	Forward oligo to amplify the *epsF* terminators

irv468	GCTTGTTTTTTTCTGTAATCTCAGTGTGcgtcatctttgattcc	Forward oligo to amplify the EAR-3 terminators-*lacZ* construct; incorporate the second half of the *mgtE* promoter sequence; for use in *E. coli*
irv469	GATCGAATTCACTTTTTTGAGTTGACATAGCTTGTTTTTTTCTGTAATCTCAGTG	Forward oligo to amplify the EAR-3 terminators-*lacZ* construct; incorporate the first half of the *mgtE* promoter sequence; for use in *E. coli*
irv509	CCTGACTTTTGGGTGCGAGCCCGAATAGGAGG	Oligonucleotide for detecting ncr982 by Northern blot
irv510	CGTAAACAGTCATTTTCCCACCCCTTTCATAG	Oligonucleotide for detecting ncr1058 (bsrG) by Northern blot
irv511	GCATATGAGAAAACCGGCACGATAATGAAATC	Oligonucleotide for detecting ncr1175 by Northern blot
irv513	ACCTAGAACCACATCGAGCAAGCCCTTTGATG	Oligonucleotide for detecting ncr1857 (bsrE) by Northern blot
irv514	GCCTGGAATGTTGACATAGCATCACCCCTTTC	Oligonucleotide for detecting ncr1019 by Northern blot
irv515	TAGTTAATCATTGTCCACCCATCATCCC	Oligonucleotide for detecting ncr1575 by Northern blot
irv516	CTCACATTTGGATCATCAAGTTCATCC	Oligonucleotide for detecting ncr952 by Northern blot
irv517	ACAACAAAGGGGATGGGAGAAATTTTTCACGG	Oligonucleotide for detecting ncr629 (rsaE) by Northern blot
irv518	GACACGGTGATATAATCACACTACGTGCGCTT	Oligonucleotide for detecting ncr1015 by Northern blot
irv519	GAAATGTTGACACGTGCATCACCCCCTTTCAT	Oligonucleotide for detecting ncr1155 (bsrH) by Northern blot
irv524	TCTTGATCTGATAGAGGGGTTGGGGAACAGA	Oligonucleotide for detecting ncr746 (fsrA) by Northern blot
irv527	TGACAACGAGCAAGCCCTTTGATGGGCAGCT	Oligonucleotide for detecting ncr1932 by Northern blot
irv530	ACGAACCTTTTGTTCGCATCCAATAAAAGG	Oligonucleotide for detecting ncr1241 by Northern blot

BIBLIOGRAPHY

Aguilar, C., Vlamakis, H., Losick, R. and Kolter, R. 2007. Thinking about Bacillus subtilis as a multicellular organism. *Curr Opin Microbiol* 10(6): 638-43.

Altuvia, S. 2007. Identification of bacterial small non-coding RNAs: experimental approaches. *Curr Opin Microbiol* 10(3): 257-61.

Amster-Choder, O. and Wright, A. 1990. Regulation of activity of a transcriptional anti-terminator in E. coli by phosphorylation in vivo. *Science* 249(4968): 540-2.

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G., Margalit, H. and Altuvia, S. 2001. Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. *Curr Biol* 11(12): 941-50.

Arnold, T. E., Yu, J. and Belasco, J. G. 1998. mRNA stabilization by the ompA 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA* 4(3): 319-30.

Artsimovitch, I. and Landick, R. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* 97(13): 7090-5.

Artsimovitch, I. and Landick, R. 2002. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* 109(2): 193-203.

Aymerich, S. and Steinmetz, M. 1992. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proc Natl Acad Sci U S A* 89(21): 10410-4.

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

Babitzke, P. and Romeo, T. 2007. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* 10(2): 156-63.

Bai, U., Mandic-Mulec, I. and Smith, I. 1993. SinI modulates the activity of SinR, a developmental switch protein of Bacillus subtilis, by protein-protein interaction. *Genes Dev* 7(1): 139-48.

Barrick, J. E., Corbino, K. A., Winkler, W. C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N.*, et al.* 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A* 101(17): 6421-6.

Barrick, J. E., Sudarsan, N., Weinberg, Z., Ruzzo, W. L. and Breaker, R. R. 2005. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* 11(5): 774-84.

Beier, L., Nygaard, P., Jarmer, H. and Saxild, H. H. 2002. Transcription analysis of the Bacillus subtilis PucR regulon and identification of a cis-acting sequence required for PucR-regulated expression of genes involved in purine catabolism. *J Bacteriol* 184(12): 3232-41.

Bejerano-Sagie, M. and Xavier, K. B. 2007. The role of small RNAs in quorum sensing. *Curr Opin Microbiol* 10(2): 189-98.

Belasco, J. G., Beatty, J. T., Adams, C. W., von Gabain, A. and Cohen, S. N. 1985. Differential expression of photosynthesis genes in R. capsulata results from segmental differences in stability within the polycistronic rxcA transcript. *Cell* 40(1): 171-81.

Blair, K. M., Turner, L., Winkelman, J. T., Berg, H. C. and Kearns, D. B. 2008. A molecular clutch disables flagella in the Bacillus subtilis biofilm. *Science* 320(5883): 1636-8.

Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A. C., Benito, Y*., et al.* 2007. Staphylococcus aureus RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev* 21(11): 1353-66.

Borukhov, S., Lee, J. and Laptenko, O. 2005. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol* 55(5): 1315-24.

Bouvier, M., Sharma, C. M., Mika, F., Nierhaus, K. H. and Vogel, J. 2008. Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol Cell* 32(6): 827-37.

Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R. and Kolter, R. 2001. Fruiting body formation by Bacillus subtilis. *Proc Natl Acad Sci U S A* 98(20): 11621-6.

Branda, S. S., Vik, S., Friedman, L. and Kolter, R. 2005. Biofilms: the matrix revisited. *Trends Microbiol* 13(1): 20-6.

Branda, S. S., Chu, F., Kearns, D. B., Losick, R. and Kolter, R. 2006. A major protein component of the Bacillus subtilis biofilm matrix. *Mol Microbiol* 59(4): 1229-38.

Braun, F., Le Derout, J. and Regnier, P. 1998. Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of Escherichia coli. *EMBO J* 17(16): 4790-7.

Breaker, R. R. 2004. Natural and engineered nucleic acids as tools to explore biology. *Nature* 432(7019): 838-45.

Britton, R. A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W. C., Mathy, N., Tobin, C., Daou, R., Szyk, J.*, et al.* 2007. Maturation of the 5' end of Bacillus subtilis 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol Microbiol* 63(1): 127-38.

Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E. and Rosch, P. 2010. A NusE:NusG complex links transcription and translation. *Science* 328(5977): 501-4.

Burns, C. M., Richardson, L. V. and Richardson, J. P. 1998. Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. *J Mol Biol* 278(2): 307-16.

Burns, C. M., Nowatzke, W. L. and Richardson, J. P. 1999. Activation of Rho-dependent transcription termination by NusG. Dependence on terminator location and acceleration of RNA release. *J Biol Chem* 274(8): 5245-51.

Burova, E., Hung, S. C., Sagitov, V., Stitt, B. L. and Gottesman, M. E. 1995. Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. *J Bacteriol* 177(5): 1388-92.

Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. and Nudler, E. 2008. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* 320(5878): 935-8.

Carpousis, A. J. 2002. The Escherichia coli RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. *Biochem Soc Trans* 30(2): 150-5.

Castillo-Keller, M., Vuong, P. and Misra, R. 2006. Novel mechanism of Escherichia coli porin regulation. *J Bacteriol* 188(2): 576-86.

Celesnik, H., Deana, A. and Belasco, J. G. 2007. Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal. *Mol Cell* 27(1): 79-90.

Chai, Y., Kolter, R. and Losick, R. 2009. Paralogous antirepressors acting on the master regulator for biofilm formation in Bacillus subtilis. *Mol Microbiol* 74(4): 876-87.

Chen, C. Z., Li, L., Lodish, H. F. and Bartel, D. P. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303(5654): 83-6.

Choonee, N., Even, S., Zig, L. and Putzer, H. 2007. Ribosomal protein L20 controls expression of the Bacillus subtilis infC operon via a transcription attenuation mechanism. *Nucleic Acids Res* 35(5): 1578-88.

Christiansen, J. K., Nielsen, J. S., Ebersbach, T., Valentin-Hansen, P., Sogaard-Andersen, L. and Kallipolitis, B. H. 2006. Identification of small Hfq-binding RNAs in Listeria monocytogenes. *RNA* 12(7): 1383-96.

Chu, F., Kearns, D. B., McLoon, A., Chai, Y., Kolter, R. and Losick, R. 2008. A novel regulatory protein governing biofilm formation in Bacillus subtilis. *Mol Microbiol* 68(5): 1117-27.

Ciampi, M. S. 2006. Rho-dependent terminators and transcription termination. *Microbiology* 152(Pt 9): 2515-28.

Clote, P., Ferre, F., Kranakis, E. and Krizanc, D. 2005. Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA* 11(5): 578-91.

Coburn, G. A. and Mackie, G. A. 1999. Degradation of mRNA in Escherichia coli: an old problem with some new twists. *Prog Nucleic Acid Res Mol Biol* 62: 55-108.

Cochrane, J. C., Lipchock, S. V. and Strobel, S. A. 2007. Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. *Chem Biol* 14(1): 97-105.

Collins, J. A., Irnov, I., Baker, S. and Winkler, W. C. 2007. Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev* 21(24): 3356-68.

Commichau, F. M., Rothe, F. M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer, E., Volker, U. and Stulke, J. 2009. Novel activities of glycolytic enzymes in Bacillus subtilis: interactions with essential proteins involved in mRNA processing. *Mol Cell Proteomics* 8(6): 1350-60.

Condon, C., Squires, C. and Squires, C. L. 1995. Control of rRNA transcription in Escherichia coli. *Microbiol Rev* 59(4): 623-45.

Condon, C. 2003. RNA processing and degradation in Bacillus subtilis. *Microbiol Mol Biol Rev* 67(2): 157-74, table of contents.

Condon, C. 2007. Maturation and degradation of RNA in bacteria. *Curr Opin Microbiol* 10(3): 271-8.

Condon, C. and Putzer, H. 2002. The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res* 30(24): 5339-46.

Corbino, K. A., Barrick, J. E., Lim, J., Welz, R., Tucker, B. J., Puskarz, I., Mandal, M., Rudnick, N. D. and Breaker, R. R. 2005. Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol* 6(8): R70.

Cromie, M. J., Shi, Y., Latifi, T. and Groisman, E. A. 2006. An RNA sensor for intracellular Mg(2+). *Cell* 125(1): 71-84.

Dam, P., Olman, V., Harris, K., Su, Z. and Xu, Y. 2007. Operon prediction using both genome-specific and general genomic information. *Nucleic Acids Res* 35(1): 288-98.

Dambach, M. D. and Winkler, W. C. 2009. Expanding roles for metabolite-sensing regulatory RNAs. *Curr Opin Microbiol* 12(2): 161-9.

Dann, C. E., 3rd, Wakeman, C. A., Sieling, C. L., Baker, S. C., Irnov, I. and Winkler, W. C. 2007. Structure and mechanism of a metal-sensing regulatory RNA. *Cell* 130(5): 878-92.

Darfeuille, F., Unoson, C., Vogel, J. and Wagner, E. G. 2007. An antisense RNA inhibits translation by competing with standby ribosomes. *Mol Cell* 26(3): 381-92.

Deana, A. and Belasco, J. G. 2005. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev* 19(21): 2526-33.

de Hoon, M. J., Makita, Y., Nakai, K. and Miyano, S. 2005. Prediction of transcriptional terminators in Bacillus subtilis and related species. *PLoS Comput Biol* 1(3): e25.

Deutscher, M. P. 2006. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res* 34(2): 659-66.

Diwa, A., Bricker, A. L., Jain, C. and Belasco, J. G. 2000. An evolutionarily conserved RNA stem-loop functions as a sensor that directs feedback regulation of RNase E gene expression. *Genes Dev* 14(10): 1249-60.

Doherty, G., Fogg, M., Wilkinson, A. and Lewis, P. 2010. Small subunits of RNA polymerase: localisation, levels and implications for core enzyme composition. *Microbiology*.

Doudna, J. A. and Cech, T. R. 2002. The chemical repertoire of natural ribozymes. *Nature* 418(6894): 222-8.

Dreyfus, M. and Regnier, P. 2002. The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell* 111(5): 611-3.

Dubey, A. K., Baker, C. S., Romeo, T. and Babitzke, P. 2005. RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* 11(10): 1579-87.

Earl, A. M., Losick, R. and Kolter, R. 2008. Ecology and genomics of Bacillus subtilis. *Trends Microbiol* 16(6): 269-75.

Eiamphungporn, W. and Helmann, J. D. 2009. Extracytoplasmic function sigma factors regulate expression of the Bacillus subtilis yabE gene via a cis-acting antisense RNA. *J Bacteriol* 191(3): 1101-5.

Emory, S. A., Bouvet, P. and Belasco, J. G. 1992. A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli. *Genes Dev* 6(1): 135-48.

Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Brechemmier-Baey, D. and Putzer, H. 2005. Ribonucleases J1 and J2: two novel endoribonucleases in B.subtilis with functional homology to E.coli RNase E. *Nucleic Acids Res* 33(7): 2141-52.

Fedor, M. J. and Williamson, J. R. 2005. The catalytic diversity of RNAs. *Nat Rev Mol Cell Biol* 6(5): 399-412.

Fisher, S. H. 1999. Regulation of nitrogen metabolism in Bacillus subtilis: vive la difference! *Mol Microbiol* 32(2): 223-32.

Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Regnier, P. and Hajnsdorf, E. 2003. The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res* 31(24): 7302-10.

Fozo, E. M., Hemm, M. R. and Storz, G. 2008. Small toxic proteins and the antisense RNAs that repress them. *Microbiol Mol Biol Rev* 72(4): 579-89, Table of Contents.

Fozo, E. M., Makarova, K. S., Shabalina, S. A., Yutin, N., Koonin, E. V. and Storz, G. 2010. Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic Acids Res* 38(11): 3743-59.

Fujita, M. and Losick, R. 2005. Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev* 19(18): 2236-44.

Fujita, Y. and Fujita, T. 1987. The gluconate operon gnt of Bacillus subtilis encodes its own transcriptional negative regulator. *Proc Natl Acad Sci U S A* 84(13): 4524-8.

Gaballa, A., Antelmann, H., Aguilar, C., Khakh, S. K., Song, K. B., Smaldone, G. T. and Helmann, J. D. 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(33): 11927-32.

Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., Wilkinson, A. C., Finn, R. D., Griffiths-Jones, S.*, et al.* 2009. Rfam: updates to the RNA families database. *Nucleic Acids Res* 37(Database issue): D136-40.

Gartner, D., Degenkolb, J., Ripperger, J. A., Allmansberger, R. and Hillen, W. 1992. Regulation of the Bacillus subtilis W23 xylose utilization operon: interaction of the Xyl repressor with the xyl operator and the inducer xylose. *Mol Gen Genet* 232(3): 415-22.

Geissmann, T., Chevalier, C., Cros, M. J., Boisset, S., Fechter, P., Noirot, C., Schrenzel, J., Francois, P., Vandenesch, F.*, et al.* 2009. A search for small noncoding RNAs in Staphylococcus aureus reveals a conserved sequence motif for regulation. *Nucleic Acids Res* 37(21): 7239-57.

Gerdes, K., Larsen, J. E. and Molin, S. 1985. Stable inheritance of plasmid R1 requires two different loci. *J Bacteriol* 161(1): 292-8.

Gerdes, K., Christensen, S. K. and Lobner-Olesen, A. 2005. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* 3(5): 371-82.

Gerdes, K. and Wagner, E. G. 2007. RNA antitoxins. *Curr Opin Microbiol* 10(2): 117-24.

Gilbert, S. D. and Batey, R. T. 2006. Riboswitches: fold and function. *Chem Biol* 13(8): 805-7.

Glatz, E., Nilsson, R. P., Rutberg, L. and Rutberg, B. 1996. A dual role for the Bacillus subtilis glpD leader and the GlpP protein in the regulated expression of glpD: antitermination and control of mRNA stability. *Mol Microbiol* 19(2): 319-28.

Gollnick, P. and Babitzke, P. 2002. Transcription attenuation. *Biochim Biophys Acta* 1577(2): 240-50.

Gonzalez-Pastor, J. E., Hobbs, E. C. and Losick, R. 2003. Cannibalism by sporulating bacteria. *Science* 301(5632): 510-3.

Gorke, B. and Vogel, J. 2008. Noncoding RNA control of the making and breaking of sugars. *Genes Dev* 22(21): 2914-25.

Gottesman, S. 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet* 21(7): 399-404.

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

Gottwein, E., Mukherjee, N., Sachse, C., Frenzel, C., Majoros, W. H., Chi, J. T., Braich, R., Manoharan, M., Soutschek, J.*, et al.* 2007. A viral microRNA functions as an orthologue of cellular miR-155. *Nature* 450(7172): 1096-9.

Greive, S. J. and von Hippel, P. H. 2005. Thinking quantitatively about transcriptional regulation. *Nat Rev Mol Cell Biol* 6(3): 221-32.

Grey, F., Meyers, H., White, E. A., Spector, D. H. and Nelson, J. 2007. A human cytomegalovirus-encoded microRNA regulates expression of multiple viral genes involved in replication. *PLoS Pathog* 3(11): e163.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R. and Bateman, A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* 33(Database issue): D121-4.

Grunberg-Manago, M. 1999. Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu Rev Genet* 33: 193-227.

Grundy, F. J. and Henkin, T. M. 1991. The rpsD gene, encoding ribosomal protein S4, is autogenously regulated in Bacillus subtilis. *J Bacteriol* 173(15): 4595-602.

Gutierrez-Preciado, A., Henkin, T. M., Grundy, F. J., Yanofsky, C. and Merino, E. 2009. Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiol Mol Biol Rev* 73(1): 36-61.

Halfmann, A., Kovacs, M., Hakenbeck, R. and Bruckner, R. 2007. Identification of the genes directly controlled by the response regulator CiaR in Streptococcus pneumoniae: five out of 15 promoters drive expression of small non-coding RNAs. *Mol Microbiol* 66(1): 110-26.

Hambraeus, G., Persson, M. and Rutberg, B. 2000. The aprE leader is a determinant of extreme mRNA stability in Bacillus subtilis. *Microbiology* 146 Pt 12: 3051-9.

Hamon, M. A., Stanley, N. R., Britton, R. A., Grossman, A. D. and Lazazzera, B. A. 2004. Identification of AbrB-regulated genes involved in biofilm formation by Bacillus subtilis. *Mol Microbiol* 52(3): 847-60.

Hampel, K. J. and Tinsley, M. M. 2006. Evidence for preorganization of the glmS ribozyme ligand binding pocket. *Biochemistry* 45(25): 7861-71.

Hartig, E., Geng, H., Hartmann, A., Hubacek, A., Munch, R., Ye, R. W., Jahn, D. and Nakano, M. M. 2004. Bacillus subtilis ResD induces expression of the potential regulatory genes yclJK upon oxygen limitation. *J Bacteriol* 186(19): 6477-84.

Hayes, F. 2003. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* 301(5639): 1496-9.

Heidrich, N., Chinali, A., Gerth, U. and Brantl, S. 2006. The small untranslated RNA SR1 from the Bacillus subtilis genome is involved in the regulation of arginine catabolism. *Mol Microbiol* 62(2): 520-36.

Henkin, T. M. 2008. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* 22(24): 3383-90.

Hershberg, R., Altuvia, S. and Margalit, H. 2003. A survey of small RNA-encoding genes in Escherichia coli. *Nucleic Acids Res* 31(7): 1813-20.

Hoch, J. A. 1993. Regulation of the phosphorelay and the initiation of sporulation in Bacillus subtilis. *Annu Rev Microbiol* 47: 441-65.

Horsburgh, M. J., Thackray, P. D. and Moir, A. 2001. Transcriptional responses during outgrowth of Bacillus subtilis endospores. *Microbiology* 147(Pt 11): 2933-41.

Houman, F., Diaz-Torres, M. R. and Wright, A. 1990. Transcriptional antitermination in the bgl operon of E. coli is modulated by a specific RNA binding protein. *Cell* 62(6): 1153-63.

Ingham, C. J., Dennis, J. and Furneaux, P. A. 1999. Autogenous regulation of transcription termination factor Rho and the requirement for Nus factors in Bacillus subtilis. *Mol Microbiol* 31(2): 651-63.

Ingham, C. J. and Furneaux, P. A. 2000. Mutations in the ss subunit of the Bacillus subtilis RNA polymerase that confer both rifampicin resistance and hypersensitivity to NusG. *Microbiology* 146 Pt 12: 3041-9.

Irnov, Kertsburg, A. and Winkler, W. C. 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., Sharma, C. M., Vogel, J. and Winkler, W. C. 2010. Identification of regulatory RNAs in Bacillus subtilis. *Nucleic Acids Res* doi: 10.1093/nar/gkq454.

Irnov, I. and Winkler, W. C. 2010. A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. *Mol Microbiol* 76(3): 559-75.

Jiang, X. and Belasco, J. G. 2004. Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. *Proc Natl Acad Sci U S A* 101(25): 9211-6.

Jiang, X., Diwa, A. and Belasco, J. G. 2000. Regions of RNase E important for 5'-end-dependent RNA cleavage and autoregulated synthesis. *J Bacteriol* 182(9): 2468-75.

Joanny, G., Le Derout, J., Brechemier-Baey, D., Labas, V., Vinh, J., Regnier, P. and Hajnsdorf, E. 2007. Polyadenylation of a functional mRNA controls gene expression in Escherichia coli. *Nucleic Acids Res* 35(8): 2494-502.

Johnnidis, J. B., Harris, M. H., Wheeler, R. T., Stehling-Sun, S., Lam, M. H., Kirak, O., Brummelkamp, T. R., Fleming, M. D. and Camargo, F. D. 2008. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature* 451(7182): 1125-9.

Jousselin, A., Metzinger, L. and Felden, B. 2009. On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol* 17(9): 399-405.

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

Kawamoto, H., Koide, Y., Morita, T. and Aiba, H. 2006. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol* 61(4): 1013-22.

Kearns, D. B., Chu, F., Branda, S. S., Kolter, R. and Losick, R. 2005. A master regulator for biofilm formation by Bacillus subtilis. *Mol Microbiol* 55(3): 739-49.

King, R. A., Banik-Maiti, S., Jin, D. J. and Weisberg, R. A. 1996. Transcripts that increase the processivity and elongation rate of RNA polymerase. *Cell* 87(5): 893-903.

King, R. A., Markov, D., Sen, R., Severinov, K. and Weisberg, R. A. 2004. A conserved zinc binding domain in the largest subunit of DNA-dependent RNA polymerase modulates intrinsic transcription termination and antitermination but does not stabilize the elongation complex. *J Mol Biol* 342(4): 1143-54.

Klein, D. J. and Ferre-D'Amare, A. R. 2006. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* 313(5794): 1752-6.

Klein, D. J., Wilkinson, S. R., Been, M. D. and Ferre-D'Amare, A. R. 2007. Requirement of helix P2.2 and nucleotide G1 for positioning the cleavage site and cofactor of the glmS ribozyme. *J Mol Biol* 373(1): 178-89.

Klinkert, B. and Narberhaus, F. 2009. Microbial thermosensors. *Cell Mol Life Sci* 66(16): 2661-76.

Klumpp, S. and Hwa, T. 2008. Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. *Proc Natl Acad Sci U S A* 105(47): 18159-64.

Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S.*, et al.* 2003. Essential Bacillus subtilis genes. *Proc Natl Acad Sci U S A* 100(8): 4678-83.

Kobayashi, K. 2008. SlrR/SlrA controls the initiation of biofilm formation in Bacillus subtilis. *Mol Microbiol* 69(6): 1399-410.

Komissarova, N., Velikodvorskaya, T., Sen, R., King, R. A., Banik-Maiti, S. and Weisberg, R. A. 2008. Inhibition of a transcriptional pause by RNA anchoring to RNA polymerase. *Mol Cell* 31(5): 683-94.

Kruger, S. and Hecker, M. 1995. Regulation of the putative bglPH operon for aryl-beta-glucoside utilization in Bacillus subtilis. *J Bacteriol* 177(19): 5590-7.

Kunst, F., Msadek, T., Bignon, J. and Rapoport, G. 1994. The DegS/DegU and ComP/ComA two-component systems are part of a network controlling degradative enzyme synthesis and competence in Bacillus subtilis. *Res Microbiol* 145(5-6): 393-402.

Kushner, S. R. 2004. mRNA decay in prokaryotes and eukaryotes: different approaches to a similar problem. *IUBMB Life* 56(10): 585-94.

Kwong, S. M., Jensen, S. O. and Firth, N. 2010. Prevalence of Fst-like toxin-antitoxin systems. *Microbiology* 156(Pt 4): 975-7; discussion 77.

Landick, R. and Yanofsky, C. (1987). Transcription attenuation. In *Eschericia coli and Salmonella typhimurium: Cellular and molecular biology* (eds. F. C. Neidhart), pp. 1276-301. American Society for Microbiology, Washington, D.C.

Landt, S. G., Abeliuk, E., McGrath, P. T., Lesley, J. A., McAdams, H. H. and Shapiro, L. 2008. Small non-coding RNAs in Caulobacter crescentus. *Mol Microbiol* 68(3): 600-14.

Lau, L. F., Roberts, J. W. and Wu, R. 1982. Transcription terminates at lambda tR1 in three clusters. *Proc Natl Acad Sci U S A* 79(20): 6171-5.

Le Coq, D., Lindner, C., Kruger, S., Steinmetz, M. and Stulke, J. 1995. New beta-glucoside (bgl) genes in Bacillus subtilis: the bglP gene product has both transport and regulatory functions similar to those of BglF, its Escherichia coli homolog. *J Bacteriol* 177(6): 1527-35.

Li, J., Horwitz, R., McCracken, S. and Greenblatt, J. 1992. NusG, a new Escherichia coli elongation factor involved in transcriptional antitermination by the N protein of phage lambda. *J Biol Chem* 267(9): 6012-9.

Li, J., Mason, S. W. and Greenblatt, J. 1993. Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. *Genes Dev* 7(1): 161-72.

Li, X., Lindahl, L., Sha, Y. and Zengel, J. M. 1997. Analysis of the Bacillus subtilis S10 ribosomal protein gene cluster identifies two promoters that may be responsible for transcription of the entire 15-kilobase S10-spc-alpha cluster. *J Bacteriol* 179(22): 7046-54.

Li, Z., Pandit, S. and Deutscher, M. P. 1998. Polyadenylation of stable RNA precursors in vivo. *Proc Natl Acad Sci U S A* 95(21): 12158-62.

Liu, J. M., Livny, J., Lawrence, M. S., Kimball, M. D., Waldor, M. K. and Camilli, A. 2009. Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res* 37(6): e46.

Liu, M. Y., Yang, H. and Romeo, T. 1995. The product of the pleiotropic Escherichia coli gene csrA modulates glycogen biosynthesis via effects on mRNA stability. *J Bacteriol* 177(10): 2663-72.

Liu, N., Bezprozvannaya, S., Williams, A. H., Qi, X., Richardson, J. A., Bassel-Duby, R. and Olson, E. N. 2008. microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart. *Genes Dev* 22(23): 3242-54.

Livny, J., Teonadi, H., Livny, M. and Waldor, M. K. 2008. High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS One* 3(9): e3197.

Livny, J. and Waldor, M. K. 2007. Identification of small RNAs in diverse bacterial species. *Curr Opin Microbiol* 10(2): 96-101.

Lopez, D., Fischbach, M. A., Chu, F., Losick, R. and Kolter, R. 2009a. Structurally diverse natural products that cause potassium leakage trigger multicellularity in Bacillus subtilis. *Proc Natl Acad Sci U S A* 106(1): 280-5.

Lopez, D., Vlamakis, H. and Kolter, R. 2009b. Generation of multiple cell types in Bacillus subtilis. *FEMS Microbiol Rev* 33(1): 152-63.

Lopez, D., Vlamakis, H., Losick, R. and Kolter, R. 2009c. Cannibalism enhances biofilm development in Bacillus subtilis. *Mol Microbiol* 74(3): 609-18.

Lopez, D., Vlamakis, H., Losick, R. and Kolter, R. 2009d. Paracrine signaling in a bacterium. *Genes Dev* 23(14): 1631-8.

Lopez, D. and Kolter, R. 2010. Extracellular signals that define distinct and coexisting cell fates in Bacillus subtilis. *FEMS Microbiol Rev* 34(2): 134-49.

Lopez de Saro, F. J., Yoshikawa, N. and Helmann, J. D. 1999. Expression, abundance, and RNA polymerase binding properties of the delta factor of Bacillus subtilis. *J Biol Chem* 274(22): 15953-8.

Luo, X., Hsiao, H. H., Bubunenko, M., Weber, G., Court, D. L., Gottesman, M. E., Urlaub, H. and Wahl, M. C. 2008. Structural and functional analysis of the E. coli NusB-S10 transcription antitermination complex. *Mol Cell* 32(6): 791-802.

Maamar, H., Raj, A. and Dubnau, D. 2007. Noise in gene expression determines cell fate in Bacillus subtilis. *Science* 317(5837): 526-9.

Mackie, G. A. 1998. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* 395(6703): 720-3.

Mandal, M. and Breaker, R. R. 2004. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* 11(1): 29-35.

Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L. and Breaker, R. R. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306(5694): 275-9.

Mandin, P., Repoila, F., Vergassola, M., Geissmann, T. and Cossart, P. 2007. Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets. *Nucleic Acids Res* 35(3): 962-74.

Martin-Verstraete, I., Charrier, V., Stulke, J., Galinier, A., Erni, B., Rapoport, G. and Deutscher, J. 1998. Antagonistic effects of dual PTS-catalysed phosphorylation on the Bacillus subtilis transcriptional activator LevR. *Mol Microbiol* 28(2): 293-303.

Mascher, T., Hachmann, A. B. and Helmann, J. D. 2007. Regulatory overlap and functional redundancy among Bacillus subtilis extracytoplasmic function sigma factors. *J Bacteriol* 189(19): 6919-27.

Mason, S. W., Li, J. and Greenblatt, J. 1992. Direct interaction between two Escherichia coli transcription antitermination factors, NusB and ribosomal protein S10. *J Mol Biol* 223(1): 55-66.

Mathy, N., Benard, L., Pellegrini, O., Daou, R., Wen, T. and Condon, C. 2007. 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* 129(4): 681-92.

Mathy, N., Hebert, A., Mervelet, P., Benard, L., Dorleans, A., de la Sierra-Gallay, I. L., Noirot, P., Putzer, H. and Condon, C. 2010. Bacillus subtilis ribonucleases J1 and J2 form a complex with altered enzyme behaviour. *Mol Microbiol* 75(2): 489-98.

McCarthy, T. J., Plog, M. A., Floy, S. A., Jansen, J. A., Soukup, J. K. and Soukup, G. A. 2005. Ligand requirements for glmS ribozyme self-cleavage. *Chem Biol* 12(11): 1221-6.

McDowall, K. J., Kaberdin, V. R., Wu, S. W., Cohen, S. N. and Lin-Chao, S. 1995. Site-specific RNase E cleavage of oligonucleotides and inhibition by stem-loops. *Nature* 374(6519): 287-90.

McLaren, R. S., Newbury, S. F., Dance, G. S., Causton, H. C. and Higgins, C. F. 1991. mRNA degradation by processive 3'-5' exoribonucleases in vitro and the implications for prokaryotic mRNA decay in vivo. *J Mol Biol* 221(1): 81-95.

Merino, E. and Yanofsky, C. 2005. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet* 21(5): 260-4.

Moller, T., Franch, T., Udesen, C., Gerdes, K. and Valentin-Hansen, P. 2002. Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. *Genes Dev* 16(13): 1696-706.

Mooney, R. A., Schweimer, K., Rosch, P., Gottesman, M. and Landick, R. 2009. Two structurally independent domains of E. coli NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J Mol Biol* 391(2): 341-58.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* (7): 621-28.

Mota, L. J., Tavares, P. and Sa-Nogueira, I. 1999. Mode of action of AraR, the key regulator of L-arabinose metabolism in Bacillus subtilis. *Mol Microbiol* 33(3): 476-89.

Murray, E. J., Strauch, M. A. and Stanley-Wall, N. R. 2009. SigmaX is involved in controlling Bacillus subtilis biofilm architecture through the AbrB homologue Abh. *J Bacteriol* 191(22): 6822-32.

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

Nehrke, K. W. and Platt, T. 1994. A quaternary transcription termination complex. Reciprocal stabilization by Rho factor and NusG protein. *J Mol Biol* 243(5): 830-9.

Nicolas, P., Bize, L., Muri, F., Hoebeke, M., Rodolphe, F., Ehrlich, S. D., Prum, B. and Bessieres, P. 2002. Mining Bacillus subtilis chromosome heterogeneities using hidden Markov models. *Nucleic Acids Res* 30(6): 1418-26.

Nielsen, J. S., Lei, L. K., Ebersbach, T., Olsen, A. S., Klitgaard, J. K., Valentin-Hansen, P. and Kallipolitis, B. H. 2010. Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in Listeria monocytogenes. *Nucleic Acids Res* 38(3): 907-19.

Nodwell, J. R. and Greenblatt, J. 1993. Recognition of boxA antiterminator RNA by the E. coli antitermination factors NusB and ribosomal protein S10. *Cell* 72(2): 261-8.

Nudler, E. and Gottesman, M. E. 2002. Transcription termination and anti-termination in E. coli. *Genes Cells* 7(8): 755-68.

Nomura, M., Yates, J. L., Dean, D. and Post, L. E. 1980. Feedback regulation of ribosomal protein gene expression in Escherichia coli: structural homology of ribosomal RNA and ribosomal protein MRNA. *Proc Natl Acad Sci U S A* 77(12): 7084-8.

Papenfort, K. and Vogel, J. 2009. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res Microbiol* 160(4): 278-87.

Pasman, Z. and von Hippel, P. H. 2000. Regulation of rho-dependent transcription termination by NusG is specific to the Escherichia coli elongation complex. *Biochemistry* 39(18): 5573-85.

Passalacqua, K. D., Varadarajan, A., Ondov, B. D., Okou, D. T., Zwick, M. E. and Bergman, N. H. 2009. Structure and complexity of a bacterial transcriptome. *J Bacteriol* 191(10): 3203-11.

Pedersen, K. and Gerdes, K. 1999. Multiple hok genes on the chromosome of Escherichia coli. *Mol Microbiol* 32(5): 1090-102.

Perego, M. and Hoch, J. A. 1996. Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in Bacillus subtilis. *Proc Natl Acad Sci U S A* 93(4): 1549-53.

Perez, N., Trevino, J., Liu, Z., Ho, S. C., Babitzke, P. and Sumby, P. 2009. A genome-wide analysis of small regulatory RNAs in the human pathogen group A Streptococcus. *PLoS One* 4(11): e7668.

Pfeffer, S., Zavolan, M., Grasser, F. A., Chien, M., Russo, J. J., Ju, J., John, B., Enright, A. J., Marks, D.*, et al.* 2004. Identification of virus-encoded microRNAs. *Science* 304(5671): 734-6.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C. and Vogel, J. 2009. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol* 16(8): 840-6.

Piggot, P. J. and Hilbert, D. W. 2004. Sporulation of Bacillus subtilis. *Curr Opin Microbiol* 7(6): 579-86.

Ramesh, A. and Winkler, W. C. 2010. Magnesium-sensing riboswitches in bacteria. *RNA Biol* 7(1): 77-83.

Rasmussen, S., Nielsen, H. B. and Jarmer, H. 2009. The transcriptionally active regions in the genome of Bacillus subtilis. *Mol Microbiol* 73(6): 1043-57.

Ratnayake-Lecamwasam, M., Serror, P., Wong, K. W. and Sonenshein, A. L. 2001. Bacillus subtilis CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev* 15(9): 1093-103.

Reizer, J., Bachem, S., Reizer, A., Arnaud, M., Saier, M. H., Jr. and Stulke, J. 1999. Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of Bacillus subtilis. *Microbiology* 145 (Pt 12): 3419-29.

Reynolds, R., Bermudez-Cruz, R. M. and Chamberlin, M. J. 1992. Parameters affecting transcription termination by Escherichia coli RNA polymerase. I. Analysis of 13 rho-independent terminators. *J Mol Biol* 224(1): 31-51.

Richardson, J. P. 1990. Rho-dependent transcription termination. *Biochim Biophys Acta* 1048(2-3): 127-38.

Rivas, E. and Eddy, S. R. 2000. Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics* 16(7): 583-605.

Romby, P. and Charpentier, E. 2010. An overview of RNAs with regulatory functions in gram-positive bacteria. *Cell Mol Life Sci* 67(2): 217-37.

Roberts, J. W. 1969. Termination factor for RNA synthesis. *Nature* 224(5225): 1168-74.

Roberts, J. W., Shankar, S. and Filter, J. J. 2008. RNA polymerase elongation factors. *Annu Rev Microbiol* 62: 211-33.

Roth, A., Nahvi, A., Lee, M., Jona, I. and Breaker, R. R. 2006. Characteristics of the glmS ribozyme suggest only structural roles for divalent metal ions. *RNA* 12(4): 607-19.

Roth, A., Winkler, W. C., Regulski, E. E., Lee, B. W., Lim, J., Jona, I., Barrick, J. E., Ritwik, A., Kim, J. N., *et al.* 2007. A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat Struct Mol Biol* 14(4): 308-17.

Roth, A. and Breaker, R. R. 2009. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* 78: 305-34.

Schmalisch, M., Maiques, E., Nikolov, L., Camp, A. H., Chevreux, B., Muffler, A., Rodriguez, S., Perkins, J. and Losick, R. 2010. Small genes under sporulation control in the Bacillus subtilis genome. *J Bacteriol* 192(20): 5402-12.

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

Schuck, A., Diwa, A. and Belasco, J. G. 2009. RNase E autoregulates its synthesis in Escherichia coli by binding directly to a stem-loop in the rne 5' untranslated region. *Mol Microbiol* 72(2): 470-8.

Schwalbe, H., Buck, J., Furtig, B., Noeske, J. and Wohnert, J. 2007. Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angew Chem Int Ed Engl* 46(8): 1212-9.

Sen, R., King, R. A., Mzhavia, N., Madsen, P. L. and Weisberg, R. A. 2002. Sequence-specific interaction of nascent antiterminator RNA with the zinc-finger motif of Escherichia coli RNA polymerase. *Mol Microbiol* 46(1): 215-22.

Shahbabian, K., Jamalli, A., Zig, L. and Putzer, H. 2009. RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. *EMBO J* 28(22): 3523-33.

Sharma, C. M., Darfeuille, F., Plantinga, T. H. and Vogel, J. 2007. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* 21(21): 2804-17.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J*., et al.* 2010. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* 464(7286): 250-5.

Sharma, C. M. and Vogel, J. 2009. Experimental approaches for the discovery and characterization of regulatory small RNA. *Curr Opin Microbiol* 12(5): 536-46.

Sharp, J. S. and Bechhofer, D. H. 2005. Effect of 5'-proximal elements on decay of a model mRNA in Bacillus subtilis. *Mol Microbiol* 57(2): 484-95.

Sharrock, R. A., Gourse, R. L. and Nomura, M. 1985. Defective antitermination of rRNA transcription and derepression of rRNA and tRNA synthesis in the nusB5 mutant of Escherichia coli. *Proc Natl Acad Sci U S A* 82(16): 5275-9.

Sierro, N., Makita, Y., de Hoon, M. and Nakai, K. 2008. DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Res* 36(Database issue): D93-6.

Silvaggi, J. M., Perkins, J. B. and Losick, R. 2005. Small untranslated RNA antitoxin in Bacillus subtilis. *J Bacteriol* 187(19): 6641-50.

Silvaggi, J. M., Perkins, J. B. and Losick, R. 2006. Genes for small, noncoding RNAs under sporulation control in Bacillus subtilis. *J Bacteriol* 188(2): 532-41.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C. M., Rolle, K., Binnewies, T. T., Hinton, J. C. and Vogel, J. 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4(8): e1000163.

Skordalakes, E. and Berger, J. M. 2006. Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell* 127(3): 553-64.

Sonenshein, A. L. 2000. Control of sporulation initiation in Bacillus subtilis. *Curr Opin Microbiol* 3(6): 561-6.

Soukup, G. A. and Breaker, R. R. 1999. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* 5(10): 1308-25.

Squires, C. L., Greenblatt, J., Li, J. and Condon, C. 1993. Ribosomal RNA antitermination in vitro: requirement for Nus factors and one or more unidentified cellular components. *Proc Natl Acad Sci U S A* 90(3): 970-4.

Storz, G., Altuvia, S. and Wassarman, K. M. 2005. An abundance of RNA regulators. *Annu Rev Biochem* 74: 199-217.

Steinmetz, M., Le Coq, D. and Aymerich, S. 1989. Induction of saccharolytic enzymes by sucrose in Bacillus subtilis: evidence for two partially interchangeable regulatory pathways. *J Bacteriol* 171(3): 1519-23.

Steinmetz, M. (1993). Carbohydrate catabolism: pathways, enzymes, genetic regulation, and evolution. . In *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics* (eds. A. L. Sonenshein, Hoch, J.A., Losick, R.), pp. 157-70. Am. Soc. Microbiol., Washington, DC.

Stulke, J. and Hillen, W. 2000. Regulation of carbon catabolism in Bacillus species. *Annu Rev Microbiol* 54: 849-80.

Stulke, J. 2002. Control of transcription termination in bacteria by RNA-binding proteins that modulate RNA structures. *Arch Microbiol* 177(6): 433-40.

Sudarsan, N., Hammond, M. C., Block, K. F., Welz, R., Barrick, J. E., Roth, A. and Breaker, R. R. 2006. Tandem riboswitch architectures exhibit complex gene control functions. *Science* 314(5797): 300-4.

Suel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. and Elowitz, M. B. 2007. Tunability and noise dependence in differentiation dynamics. *Science* 315(5819): 1716-9.

Sullivan, C. S., Grundhoff, A. T., Tevethia, S., Pipas, J. M. and Ganem, D. 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* 435(7042): 682-6.

Sullivan, S. L. and Gottesman, M. E. 1992. Requirement for E. coli NusG protein in factor-dependent transcription termination. *Cell* 68(5): 989-94.

Sutherland, I. W. 2001. The biofilm matrix--an immobilized but dynamic microbial environment. *Trends Microbiol* 9(5): 222-7.

Swiercz, J. P., Hindra, Bobek, J., Haiser, H. J., Di Berardo, C., Tjaden, B. and Elliot, M. A. 2008. Small non-coding RNAs in Streptomyces coelicolor. *Nucleic Acids Res* 36(22): 7240-51.

Takemaru, K., Mizuno, M., Sato, T., Takeuchi, M. and Kobayashi, Y. 1995. Complete nucleotide sequence of a skin element excised by DNA rearrangement during sporulation in Bacillus subtilis. *Microbiology* 141 (Pt 2): 323-7.

Tjaden, B., Goodwin, S. S., Opdyke, J. A., Guillier, M., Fu, D. X., Gottesman, S. and Storz, G. 2006. Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res* 34(9): 2791-802.

Timmermans, J. and Van Melderen, L. 2010. Post-transcriptional global regulation by CsrA in bacteria. *Cell Mol Life Sci* 67(17): 2897-908.

Tinsley, R. A., Furchak, J. R. and Walter, N. G. 2007. Trans-acting glmS catalytic riboswitch: locked and loaded. *RNA* 13(4): 468-77.

Tock, M. R., Walsh, A. P., Carroll, G. and McDowall, K. J. 2000. The CafA protein required for the 5'-maturation of 16 S rRNA is a 5'-end-dependent ribonuclease that has context-dependent broad sequence specificity. *J Biol Chem* 275(12): 8726-32.

Tojo, S., Matsunaga, M., Matsumoto, T., Kang, C. M., Yamaguchi, H., Asai, K., Sadaie, Y., Yoshida, K. and Fujita, Y. 2003. Organization and expression of the Bacillus subtilis sigY operon. *J Biochem* 134(6): 935-46.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T.*, et al.* 2009. The Listeria transcriptional landscape from saprophytism to virulence. *Nature* 459(7249): 950-6.

Trotochaud, A. E. and Wassarman, K. M. 2005. A highly conserved 6S RNA structure is required for regulation of transcription. *Nat Struct Mol Biol* 12(4): 313-9.

Tsui, H. C., Mukherjee, D., Ray, V. A., Sham, L. T., Feig, A. L. and Winkler, M. E. 2010. Identification and characterization of noncoding small RNAs in Streptococcus pneumoniae serotype 2 strain D39. *J Bacteriol* 192(1): 264-79.

Urban, J. H. and Vogel, J. 2008. Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. *PLoS Biol* 6(3): e64.

Uzilov, A. V., Keegan, J. M. and Mathews, D. H. 2006. Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. *BMC Bioinformatics* 7: 173.

Valentin-Hansen, P., Johansen, J. and Rasmussen, A. A. 2007. Small RNAs controlling outer membrane porins. *Curr Opin Microbiol* 10(2): 152-5.

Vanderpool, C. K. and Gottesman, S. 2004. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol* 54(4): 1076-89.

Vecerek, B., Moll, I. and Blasi, U. 2007. Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. *EMBO J* 26(4): 965-75.

Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R. and Blasi, U. 2008. The C-terminal domain of Escherichia coli Hfq is required for regulation. *Nucleic Acids Res* 36(1): 133-43.

Verhamme, D. T., Kiley, T. B. and Stanley-Wall, N. R. 2007. DegU co-ordinates multicellular behaviour exhibited by Bacillus subtilis. *Mol Microbiol* 65(2): 554-68.

Vlamakis, H., Aguilar, C., Losick, R. and Kolter, R. 2008. Control of cell fate by the formation of an architecturally complex bacterial community. *Genes Dev* 22(7): 945-53.

Vogel, J. and Papenfort, K. 2006. Small non-coding RNAs and the bacterial outer membrane. *Curr Opin Microbiol* 9(6): 605-11.

von Hippel, P. H. and Yager, T. D. 1991. Transcript elongation and termination are competitive kinetic processes. *Proc Natl Acad Sci U S A* 88(6): 2307-11.

von Hippel, P. H. and Yager, T. D. 1992. The elongation-termination decision in transcription. *Science* 255(5046): 809-12.

Voss, B., Georg, J., Schon, V., Ude, S. and Hess, W. R. 2009. Biocomputational prediction of non-coding RNAs in model cyanobacteria. *BMC Genomics* 10: 123.

Wadler, C. S. and Vanderpool, C. K. 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc Natl Acad Sci U S A* 104(51): 20454-9.

Washietl, S., Hofacker, I. L., Lukasser, M., Huttenhofer, A. and Stadler, P. F. 2005. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* 23(11): 1383-90.

Wakeman, C. A. and Winkler, W. C. 2009. Structural probing techniques on natural aptamers. *Methods Mol Biol* 535: 115-33.

Wang, Z., Gerstein, M. and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1): 57-63.

Wassarman, K. M. 2007. 6S RNA: a regulator of transcription. *Mol Microbiol* 65(6): 1425-31.

Wassarman, K. M., Zhang, A. and Storz, G. 1999. Small RNAs in Escherichia coli. *Trends Microbiol* 7(1): 37-45.

Waters, L. S. and Storz, G. 2009. Regulatory RNAs in bacteria. *Cell* 136(4): 615-28.

Weinberg, Z., Barrick, J. E., Yao, Z., Roth, A., Kim, J. N., Gore, J., Wang, J. X., Lee, E. R., Block, K. F., *et al.* 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* 35(14): 4809-19.

Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H. and Breaker, R. R. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* 11(3): R31.

Weisberg, R. A. and Gottesman, M. E. 1999. Processive antitermination. *J Bacteriol* 181(2): 359-67.

Wickiser, J. K., Winkler, W. C., Breaker, R. R. and Crothers, D. M. 2005. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* 18(1): 49-60.

Winkelman, J. T., Blair, K. M. and Kearns, D. B. 2009. RemA (YlzA) and RemB (YaaB) regulate extracellular matrix operon expression and biofilm formation in Bacillus subtilis. *J Bacteriol* 191(12): 3981-91.

Winkler, W. C. 2005. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr Opin Chem Biol* 9(6): 594-602.

Winkler, W. C. and Breaker, R. R. 2005. Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* 59: 487-517.

Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. and Breaker, R. R. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428(6980): 281-6.

Wood, H. E., Dawson, M. T., Devine, K. M. and McConnell, D. J. 1990. Characterization of PBSX, a defective prophage of Bacillus subtilis. *J Bacteriol* 172(5): 2667-74.

Workman, C. and Krogh, A. 1999. No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. *Nucleic Acids Res* 27(24): 4816-22.

Xiao, C., Calado, D. P., Galler, G., Thai, T. H., Patterson, H. C., Wang, J., Rajewsky, N., Bender, T. P. and Rajewsky, K. 2007. MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. *Cell* 131(1): 146-59.

Xu, F. and Cohen, S. N. 1995. RNA degradation in Escherichia coli regulated by 3' adenylation and 5' phosphorylation. *Nature* 374(6518): 180-3.

Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. A. and Kosik, K. S. 2009a. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell* 137(4): 647-58.

Xu, X., Ji, Y. and Stormo, G. D. 2009b. Discovering cis-regulatory RNAs in Shewanella genomes by Support Vector Machines. *PLoS Comput Biol* 5(4): e1000338.

Yao, S., Blaustein, J. B. and Bechhofer, D. H. 2007. Processing of Bacillus subtilis small cytoplasmic RNA: evidence for an additional endonuclease cleavage site. *Nucleic Acids Res* 35(13): 4464-73.

Yakhnin, H., Pandit, P., Petty, T. J., Baker, C. S., Romeo, T. and Babitzke, P. 2007. CsrA of Bacillus subtilis regulates translation initiation of the gene encoding the flagellin protein (hag) by blocking ribosome binding. *Mol Microbiol* 64(6): 1605-20.

Ye, R., Rehemtulla, S. N. and Wong, S. L. 1994. Glucitol induction in Bacillus subtilis is mediated by a regulatory factor, GutR. *J Bacteriol* 176(11): 3321-7.

Zahler, S. A., Korman, R. Z., Rosenthal, R. and Hemphill, H. E. 1977. Bacillus subtilis bacteriophage SPbeta: localization of the prophage attachment site, and specialized transduction. *J Bacteriol* 129(1): 556-8.

Zellmeier, S., Hofmann, C., Thomas, S., Wiegert, T. and Schumann, W. 2005. Identification of sigma(V)-dependent genes of Bacillus subtilis. *FEMS Microbiol Lett* 253(2): 221-9.