Post-Transcriptional Control of Multicellularity in Bacillus subtilis

Post-transcriptional Control of Multicellularity in *Bacillus subtilis*

A dissertation presented by

Aaron DeLoughery

to

The Department of Molecular and Cellular biology

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In the subject of

Biochemistry

Harvard University

Cambridge, Massachusetts

February 2016

Dissertation Advisor: Professor Richard Losick Aaron DeLoughery

Word did not find any entries for your table of contents.

Post-transcriptional Control of Multicellularity in *Bacillus subtilis*

Abstract

Bacteria are capable of sensing changes in their environment and responding to those changes by altering patterns of gene expression. There are many mechanisms by which bacteria can sense internal and external signals. The most well studied systems for this, two-component systems, are composed of a histidine kinase that can act as a detector and a response regulator that when activated by the histidine kinase can carry out a function. In the case of regulating gene expression, once a signal is detected, the response regulator can directly act as a transcription factor or control the activity of other transcription factors to determine the genes that are transcribed and the genes that are not. Historically, much of gene regulation has been thought to be regulated at the level of transcription. There are, however, several processes that take place after a gene is transcribed that ultimately determine the level of a gene product These include mRNA stability, translation initiation, translation efficiency, and protein degradation, all of which the cell could potentially target to fine-tune protein levels. Many such post-transcriptional mechanisms of regulation have been identified, such as controlled proteolysis and the regulation of translation by riboswitches. Here I report on mechanisms operating at the level of translational efficiency and mRNA stability that control entry into a multicellular state in the bacterium *Bacillus subtilis*.

Like many bacteria, *B. subtilis* can form architecturally complex communities known as biofilms. Biofilm formation by *B. subtilis* is largely governed by a dedicated repressor for matrix production, SinR. At the transcriptional level, biofilm formation is regulated by the response regulator Spo0A, which turns on the gene for the anti-repressor SinI. SinI, in turn, binds to and inactivates SinR. The work that I have been a part of and that is presented here has revealed two additional mechanisms in which SinR is regulated post-transcriptionally. The first is a simple mechanism in which *B. subtilis* can sense changes in nutrient availability without a dedicated protein or RNA. Instead certain synonymous serine codons cause

pausing when serine levels are reduced, and enrichment for these codons in *sinR* decreases its rate of translation under biofilm-inducing conditions, contributing to derepression of matrix genes.

The second, which is the main focus of this thesis, acts at the level of the stability of the mRNA for *sinR*. This mechanism was unexpectedly uncovered while determining the function of three mysterious proteins, YlbF, YmcA, and YaaT, which are required for biofilm formation. I have found that YlbF, YmcA, and YaaT form a three-protein complex that interacts with and is required for the full activity of the ribonuclease RNase Y. This complex is required for biofilm formation because it destabilizes the mRNA for *sinR*. In addition RNA-sequencing experiments reveal a global role for the complex in RNA turnover. Turnover of many known targets of RNase Y, including mRNA and non-coding RNAs, are dependent on the complex.

In light of these findings, the decision to form a multicellular community can be seen as the convergence of three regulatory pathways operating at three different levels of the expression of the gene for the master regulatory protein for biofilm formation SinR.

Table of contents

Introduction and Background 1

Chapter 1: YlbF and YmcA are required for biofilm formation and are in a complex with YaaT 11
- **1.1** Mutants lacking YlbF or YmcA are defective for biofilm formation and do not express matrix production operons 12
- **1.2** What are YlbF and YmcA? 14
- **1.3** Mutants lacking YlbF or YmcA show additional stationary phase defects 17
- **1.4** YlbF interacts with YaaT 17
- **1.5** What is YaaT? 22
- **1.6** Deletion of *yaaT* reduces biofilm formation and impairs growth in LB 22
- **1.7** Other proteins co-purify with YlbF 24
- **Discussion** 24

Chapter 2: The complex of YlbF, YmcA, and YaaT plays a role in biofilm formation in a manner independent of the response regulator Spo0A 27
- **2.1** Spo0A activity is not dependent on the complex of YlbF, YmcA, and YaaT under biofilm-inducing conditions 28
- **2.2** The absence of YlbF, YmcA, or YaaT, but not Spo0A, impairs growth in Luria-Bertani (LB) medium 31
- **2.3** The genes for *ylbF*, *ymcA*, and *yaaT* are widely conserved amongst species of Gram-positive bacteria that lack *spo0A* and all other components of the phosphorelay 33
- **2.4** Spo0B and Spo0F, as well as all other components of the phosphorelay, do not co-purify with YlbF in *B. subtilis* 33
- **Discussion** 33

Chapter 3: Complexes of YlbF, YmcA, and YaaT interact with the endoribonuclease RNase Y to control the mRNA level of the matrix gene repressor SinR 35
- **3.1** The level of SinR protein is higher in *ylbF* and *ymcA* mutants 36
- **3.2** The increase of SinR protein is sufficient to block biofilm formation 36
- **3.3** A small reduction of SinR levels restores biofilm formation to mutants of *ylbF* 40
- **3.4** The level of *sinR* mRNA is higher in *ylbF* and *ymcA* mutants 40
- **3.5** How does the absence of YlbF and YmcA cause increased *sinR* mRNA levels? 44
- **3.6** The gene for RNase Y is near the gene for YmcA. The genes for RNase Y and YaaT show a a striking pattern of co-conservation 44
- **3.7** Mutants lacking RNase Y are defective for biofilm formation 45
- **3.8** YlbF and YmcA interact directly and independently with RNase Y 49
- **3.9** Cleavage of an additional well-studied target of RNase Y depends on YlbF and YmcA 51
- **Discussion** 53

Chapter 4: RNA-sequencing reveals a wider role for YlbF and YmcA in RNA turnover and processing 56
- **4.1** *B. subtilis* RNase Y and its role in RNA turnover 57
- **4.2** Experimental design, a novel technique for RNA sequencing 57
- **4.3** Mutation of components of the YlbF-YmcA-YaaT-Rny complex lead to increased mRNA levels of other components. A new *yaaT* transcript 58

4.4 Some known targets of RNase Y are affected by YlbF and YmcA, while others are not 60
4.5 The turnover of many riboswitches depends on *YlbF* 63
Discussion 67

Summary 70

Bibliography 75

Appendix A Materials and Methods 80

Appendix B: A Serine Sensor for Multicellularity in a Bacterium 93

List of figures

Figure 1. Biofilm formation by *Bacillus subtilis* 5
Figure 2. Mutants of *ylbF* and *ymcA* are defective for biofilm formation 13
Figure 3. What is YlbF? 15
Figure 4. What is YmcA? 16
Figure 5. Pull-down with YlbF reveals potential interacting partners 19
Figure 6. What is YaaT? 21
Figure 7. YlbF, YmcA, and YaaT function independently of Spo0A 29
Continued 30
Figure 8. Mutants of *ylbF*, *ymcA*, and *yaaT* do not phenocopy mutants of *spo0A*, *spo0B*, or *spo0F* 32
Figure 9. Mutants of the complex have heightened levels of SinR protein 38
Continued 39
Figure 10. Mutants of the complex have heightened levels of *sinR* mRNA 42
Continued 43
Figure 11. The gene for RNase Y is co-conserved with *yaaT* 46
Figure 12. An RNase Y mutant phenocopies mutants of the complex 47
Continued 48
Figure 13. Bacterial two-hybrid assays show that YlbF and YmcA interact directly with RNase Y 50
Figure 14. Cleavage of the *cggR-gapA* mRNA by RNase Y is dependent on YlbF and YmcA 52
Figure 15. A model 55
Figure 16. A new *yaaT-yabA* transcript in mutants of *ymcA* 59
Figure 17. *cggR-gapA* processing in mutants of *ymcA* 61
Figure 18. Riboswitch abundance in mutants of *ymcA* 65

List of Tables

Table 1. List of proteins that co-purify with MBP-YlbF from *B. subtilis* grown under biofilm inducing conditions 23
Continued 24

Table 2 Change in abundance of mRNA in mutants lacking *ymcA* under biofilm conditions 62

Table 3. Change in abundance of riboswitches in mutants lacking *ymcA* 66

Table A1. Strains used in this study 88

Acknowledgements

First I would like to acknowledge my advisor Richard Losick. He has provided endless support and encouragement throughout the past 6 years. I admire his enthusiam for science, his creativity in designing experiments, and his dedication to teaching and mentoring students and young scientists. I feel more than grateful to have been a member of his lab. There are many current and former Losick lab members to acknowledge: Sara Leiman, Alicia DeFrancesco, Bjorn Traag, Emma Nagy, Niels Bradshaw, Lucy Foulston, Tom Norman, Anna Wang Erickson, Sarah Wacker, Nadya Masloboeva, Vanina Dengler, Matt Cabeen, Jon Russell, Alex Elsholz, Yunrong Chai, Leah Silverstein, and Yun Chen. In particular, I want to acknowledge Yunrong Chai who was my mentor while I was a rotation student in the lab and for the years following. We have had a lot of valuable conversations, and have continued to be helpful to each other. Also, members of the Burton lab were just as much a part of our community, Marina Besprozvannaya, Bijou Bose, Scott Chilton, Valerie Pivorunas, and Hannah Foster. They have all been helpful when thinking about my project and troubleshooting experiments. Moreover, they have been great colleagues, and friends, to work with and get to know over the past years. I have really enjoyed all of the Friday beer hours, fantasy football leagues, coffee breaks, and lab parties that we have had. I have also mentored three rotation students, Emma Nagy, Beste Mutlu, and Mohammed Mostajo-Radji, and an undergraduate student Eliza Llewellen, which has been great.

My path towards research started at Indiana University in the lab of professor Yves Brun. As my undergraduate advisor he provided an immense amount of mentorship and inspiration for me and he has certainly become a role model for me. In his lab I worked with Rey Allen, Melanie Lawler, Mike Trimble, and Patrick Curtis. I would also like to acknowledge great educators, my high school biology and chemistry teachers, Robert Hamilton and John Thompson.

In the department of Molecular and Cellular Biology there have been many people who have contributed to my training and my research. Teaching with Briana Burton, Karine Gibbs, Mary Ellen Wiltrout, Tom Torello, and Martha Zepeda have been rewarding experiences. Their energy and passions for science education have sparked my own. The faculty members who have been a part of my

Dissertation Advisor Committee, and who have contributed significantly to the progress of my work and have provided advice about my future. This includes Vlad Denic, Karine Gibbs, Briana Burton, Ethan Garner, and the chair of the committee Susan Mango.

Collaborations have been an important part of my graduate work. I would like to thank C. Lee, Alan Grossman, and Ole Skoovgard for their efforts to test an early hypthesis. I would like to thank Arvind Subramaniam, Yun Chen, and Yunrong Chai for there work on the serine story. Finally, I would also like to thank Gene-Wei Li, Ariel Schieler, Darren Parker, and Jean-Benoit Lalanne for carrying out and teaching me to do RNA-sequencing, which has informed experiments and future directions.

My classmates; Sarah Douglas, Kyle Mcelroy, Laila Akhmetova, Hannah Chen, Jiang He, Bumjin Nakoon, Jamie Wester, Ania Puszynska, Ghazaleh Ashrafi, Alix Lacoste, Nikhil Sharma, Alan Rodrigues, and Melis Yilmaz Balban, who have made the past years much more fun and much more pleasant.

The support of friends and family has made this possible, especially through some difficult times. The friends I have made here have been a big influence and support to me, Arndt Luemers, Jonathan Schneiderman, Michael Dickstein, and Justin Thaler. I can't thank enough my mother Stephanie DeLoughery, father Jay DeLoughery, sister Mandy DeLoughery, and twin brother Matt DeLoughery. The love and unconditional support they have provided, not just throughout this process, but throughout my life have made this possible. This has been written in the loving memory of a friend J.G. and a nephew C.J.C.

Introduction and Background

Discovering the function of uncharacterized genes can be a difficult task, but also a rewarding one. Even in the well-studied bacterium, *Bacillus subtilis*, the function of many genes remains unknown. A genetics approach to biology usually reveals genes that are required for a certain phenotype or a certain biological function. Often, the genes are identified based on what happens to the organism when they are disrupted, which indicates a connection between the gene's product and the phenotype, but once identified, the task becomes to determine the molecular function of that gene, i.e. what the gene's product is doing when it is present and how that function explains the original phenotype.

In bacteria, gene expression is tightly regulated to meet the needs of the cell. Optimized growth and survival require the production of sufficient levels of needed gene products, mostly proteins but also RNAs, without the wasteful or toxic production of excess or inappropriate ones. Moreover, bacteria are able to change the levels of gene products as their growth or environmental conditions change. Thus it is important that bacteria are able to sense and respond to their environment or changes in internal cell state by altering levels of gene products. One way in which the model organism, *B. subtilis*, responds to changing conditions is by turning on a program of gene expression that leads to biofilm formation.

The work presented here focuses on the genes *ylbF*, *ymcA*, and *yaaT* of *B. subtilis* that were originally identified due to their role in biofilm formation in 2004. The functions of these genes have

since remained a mystery. Previous members of the lab as well as another group have worked towards understanding their function. There have been several ideas about how they might be working, most of which have proposed a role in the known and well-studied regulatory network that controls biofilm formation. One such idea was published by the laboratory of professor David Dubnau in 2013. They propose a role for YlbF, YmcA, and YaaT in accelerating the phosphorylation of the response regulator Spo0A, a core regulatory component of biofilm formation. However, I did not find an effect on Spo0A in mutants lacking YlbF or YmcA under biofilm inducing conditions, so I continued to look for their function.

Somewhat unexpectedly, I found that YlbF, YmcA, and YaaT play a broader role in bacterial RNA turnover. The topic of RNA turnover and regulation, although somewhat outside of what our laboratory historically studies, is vitally important towards understanding how bacteria regulate gene products and respond to changing conditions. The uncovering of a role for YlbF, YmcA, and YaaT in RNA turnover provides novel insight into RNA turnover in *B. subtilis*. It also identifies an unrecognized layer in which biofilm formation is regulated independently of the Spo0A pathway, a type of regulation that is also post-transcriptional.

The structure of this thesis is as follows: In this section background information about two topics, biofilm formation by *B. subtilis* and RNA turnover in bacteria, will be reviewed. Original work will then be presented in four chapters. Chapter 1 describes the phenotypes of *ylbF* and *ymcA* mutants, as well as the identification of a related third gene of unknown function *yaaT*. Chapter 2 revisits the findings of Carabetta et al 2013, which proposes a function for these genes. Chapter 3 provides evidence for a function in which YlbF, YmcA, and YaaT act with and affect the activity of the ribonuclease RNase Y, and that the biofilm defect is specifically caused by an effect on the mRNA of *sinR*, the master repressor of biofilm formation. Finally chapter 4 provides evidence that these genes play a more global role RNA turnover and processing.

In addition, a study is presented in the appendix of this thesis describing the finding of yet another mechanism in which biofilm formation is regulated post-transcriptionally. In this case the regulation is

occurring at the level of translational efficiency. It was found that certain serine codons cause ribosomal pausing while the others do not under biofilm conditions, and that in the case for the master repressor of biofilm formation this acts as a simple sensor to trigger matrix production.

Biofilm Formation by *B. subtilis*

Like many bacteria, the soil-dwelling bacterium, *B. subtilis* is capable of forming architecturally complex, multicellular communities known as biofilms. Biofilms are defined by the production of an extracellular matrix, which acts as a scaffold to hold cells together (Aguilar, Vlamakis, Losick, & Kolter, 2007; Branda, Gonzalez-Pastor, Ben-Yehuda, Losick, & Kolter, 2001; Kolter & Greenberg, 2006). A genetic screen in the wild isolate of *B. subtilis* 3610 revealed many genes that are required for biofilm formation. In this way, the genes for matrix production and many of their regulators were identified.

The matrix produced by *B. subtilis* is composed of two materials, a polysaccharide and a hydrophobic protein fiber. The extracellular polysaccharide (EPS) is produced and exported by an operon called *epsA-O*. The amyloid-like fiber, a polymer of TasA, is produced and secreted by the *tapA-sipW-tasA operon* (Branda et al., 2004; Chai, Chu, Kolter, & Losick, 2008; Chu, Kearns, Branda, Kolter, & Losick, 2006). In it's natural environment, *B. subtilis* produces biofilms on a variety of surfaces including plant roots (Beauregard, Chai, Vlamakis, Losick, & Kolter, 2013; Y. Chen et al., 2012). In the lab, *B. subtilis* biofilms are grown on a specific medium as either a colony on solid agar or as a floating pellicle in liquid. In both cases the production of matrix leads to the development of architecturally complex structures, such as wrinkles (Figure 1A) (Branda et al., 2001).

The production of a biofilm requires a specific program of gene expression, which is tightly regulated. Matrix production in *B. subtilis* is in part activated under conditions of nutrient deprivation. In nutrient rich conditions, the genes for matrix production are repressed by the dedicated repressor of biofilm formation, SinR (figure 1B). SinR is a DNA binding protein that binds to the promoters of matrix genes, preventing their transcription (Chu et al., 2006; Kearns, Chu, Branda, Kolter, & Losick, 2005). Matrix genes are de-repressed by the production of SinI, an antagonist of SinR (Bai, Mandic-Mulec, &

Smith, 1993). SinI interacts directly with SinR to prevent it from binding to DNA. Thus in the presence of SinI, matrix genes are turned on.

SinI is regulated transcriptionally by the phosphorylated form of the response regulator Spo0A (Molle et al., 2003). Spo0A has been fairly well studied in the context of its role as the master regulator of sporulation in *B. subtilis*. Spo0A is phosphorylated by a phosphorelay that is composed of two phosphotransfer proteins, Spo0B and Spo0F, and five histidine kinases, KinA-E (Jiang, Shao, Perego, & Hoch, 2000; McLoon, Kolodkin-Gal, Rubinstein, Kolter, & Losick, 2011). When activated, the histidine kinases first phosphorylate Spo0F. The phosphoryl group is then transferred to Spo0B and then to Spo0A (Tzeng, Zhou, & Hoch, 1998). The phosphorylation of Spo0A affects its DNA binding activity allowing it to act as a transcription factor. Spo0A~P binds near the promoter for *sinI* activating its transcription (Chai et al., 2008). The genes for *sinI* and *sinR* are next to each other on the chromosome (figure 1C).

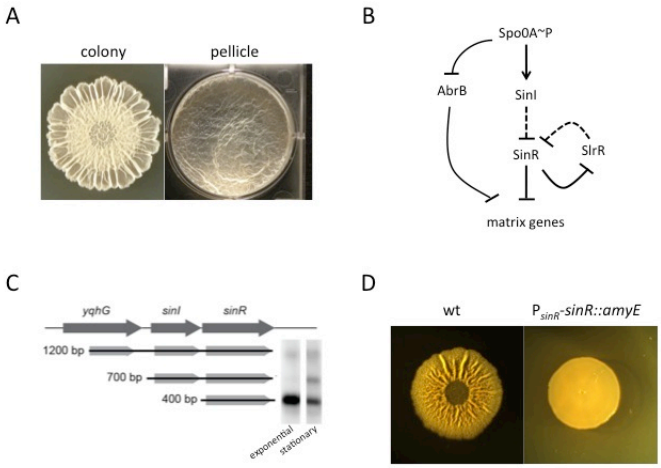

Figure 1: Biofilm formation by *Bacillus subtilis*

A. Photographs of a colony biofilm and pellicle biofilm produced by wild type 3610 in the biofilm inducing medium MSgg. Biofilms were grown for 72 hours at 37°C.

B. Schematic depicting the important regulators of biofilm formation and how they interact to regulate matrix production. Dashed line indicates protein-protein interaction and solid lines indicate transcription regulation.

C. Representation of the *sinI/R* operon and the three transcripts produced by its transcription. A northern blot with a probe against *sinR* shows the presence of these transcripts.

D. Biofilm formation is ultrasensitive to levels of SinR. Shown are colony biofilms of wild type 3610 and a mutant harboring two copies of the gene for SinR.

There are two other proteins that contribute to the regulation of matrix production that are of interest here. The first, AbrB, is a global regulator of many stationary phase genes including the *eps* and *tapA* operons (Banse, Chastanet, Rahn-Lee, Hobbs, & Losick, 2008; Chu et al., 2008; Hamon & Lazazzera, 2001; Hamon, Stanley, Britton, Grossman, & Lazazzera, 2004). In this case, AbrB acts as a repressor of matrix genes. This repression is also alleviated by Spo0A~P, which represses the gene for AbrB (Molle et al., 2003). The second is a protein, SlrR, that acts as an additional antagonist of SinR (Chu et al., 2008). SlrR and SinR act in a double negative feedback look because the gene for SlrR is itself repressed by SinR, and when produced SlrR binds to titrates SinR (Chai et al., 2008; Norman, Lord, Paulsson, & Losick, 2013). This creates a bistable switch between a non-biofilm producing state and a biofilm producing state. Under biofilm inducing conditions, the production of SinI causes a switch from a high repression by SinR to low repression by SinR resulting in high levels of SlrR (Chai, Norman, Kolter, & Losick, 2010). The expression of matrix operons is ultrasensitive to the levels of SinR. Thus small perturbations in SinR abundance can have a profound effect on biofilm formation (figure 1D) (Chai, Norman, Kolter, & Losick, 2011; Norman et al., 2013). This ultra sensitivity is created by the molecular titration of SinR by SinI and the cooperative nature of SinR DNA binding. SinR activity is regulated transcriptionally through the expression of its antagonists SinI and SlrR.

RNA turnover in bacteria

In bacteria most RNAs have a very short half-life, in particular mRNAs. This provides an advantage to bacteria by allowing them to respond very quickly to changing conditions. At a basic level, the transcriptional control of gene expression requires mRNA turnover. If mRNAs were very stable, changes in gene transcription would not quickly result in changes of protein production, especially for genes that are being turned off. At an additional level, the regulated degradation of specific mRNAs can very quickly change what proteins are being translated by removing its mRNA. In this way the cell can regulate protein levels very quickly because it does not require the transcription and production of new protein. While microbiologists have appreciated the significance of RNA turnover, a lot of the details

remain unclear. Much of our understanding about RNA turnover come from *E. coli*, which as a model for RNA degradation has some weaknesses, especially for Gram-positive bacteria. This is because a majority of RNA degradation in *E. coli* is carried out by one protein, RNase E, which is essential, and its gene is not conserved in most Gram-positive species (Babitzke & Kushner, 1991; Melefors & von Gabain, 1991; Mudd, Krisch, & Higgins, 1990; Ono & Kuwano, 1979; Taraseviciene, Miczak, & Apirion, 1991). The interest in RNA turnover in Gram-positive bacteria and the efforts being put forth towards its understanding have increased significantly in the past few years.

The main players in RNA turnover are a class of enzymes called ribonucleases. There are many types of ribonucleases found in the bacterial world. The best studied is RNase E from *Escherichia coli*. RNase E has a well-conserved catalytic domain at its N-terminus and a less well-conserved C-terminus that contains a membrane-binding helix (Khemici, Poljak, Toesca, & Carpousis, 2005; McDowall & Cohen, 1996; Vanzo et al., 1998). RNase E acts as an endoribonuclease that has been shown to cleave predominantly single stranded regions of RNA that are A/U rich, but a distinct nucleotide recognition sequence has not been identified, and may not exist (McDowall, Lin-Chao, & Cohen, 1994). In addition to its endonuclease activity, it has also been shown to act as an exonuclease on the 5' end of RNA, with a preference for mono-phosphorylated nucleotides (Mackie, 1998). The significance of the phosphorylation state of the 5' nucleotide is that, upon transcription, the 5' nucleotide is tri-phosphorylated, which stabilizes the RNA. The phosphorylation state of the 5' nucleotide is changed by a class of enzymes that can convert nucleotide tri-phosphates to nucleotide mono-phosphates, destabilizing the RNA (Celesnik, Deana, & Belasco, 2007; Deana, Celesnik, & Belasco, 2008; Richards et al., 2011). RNase E is essential, but work with temperature sensitive mutants of RNase E have revealed that most RNA turnover in *E. coli* is dependent on this protein. *E. coli* does have additional non-essential ribonucleases that play a minor role in RNA turnover, but they alone are not sufficient for viability.

In *B. subtilis* a majority of RNA turnover and its initiation are carried out by the combined activity of two ribonucleases, RNase Y and RNase J1 (Deikus & Bechhofer, 2011; Durand, Gilet, Bessieres, Nicolas, & Condon, 2012; Lehnik-Habrink, Newman, et al., 2011; Yao, Richards, Belasco, &

Bechhofer, 2011). RNase Y carries out the endoribonuclease activity and RNaseJ1 carries out the 5' activity (Even et al., 2005; Mathy et al., 2007). While there are some similarities in the catalytic domains of these proteins to RNase E, neither closely resembles RNase E. Historically, both RNase Y and RNase J1 were thought to be essential genes (Durand et al., 2012). Their role in bulk RNA turnover was determined in cells that were depleted of these nucleases. An important study in 2011 found that RNase Y plays both a role in general mRNA turnover as well as a role in targeting specific RNAs for degradation (Lehnik-Habrink, Schaffer, et al., 2011). Many targets of RNase Y were identified in this study using microarray analysis, but a recognition sequence for RNase Y was not. Similarly to RNase E in *E. coli*, RNase Y has been shown to localize to the membrane in *B. subtilis*, through a N-terminal transmembrane domain and by interacting with glycolytic enzymes (Newman et al., 2012). It has recently been appreciated that RNase Y is not actually essential in *B. subtilis*, although mutants are impaired for growth and defective in a variety of ways (Figaro et al., 2013). It has been shown in *Streptococcus pyogenes* and *Staphylococcus aureus* that the homologs of RNase Y regulate some virulence genes (L. H. Chen, Emory, Bricker, Bouvet, & Belasco, 1991; Marincola et al., 2012). Rinse J1 is essential in the presence of an endogenous pseudophage that *B. subtilis* carries, but may not be in its absence (Figaro et al., 2013). Although mutants lacking RNase J1 are extremely impaired for growth. It should be noted that these ribonucleases do not just degrade RNAs, they also play a variety of roles in the maturation of non-coding RNAs (Gilet, DiChiara, Figaro, Bechhofer, & Condon, 2015).

There are three main pathways by which a RNA molecule is degraded. For a majority of RNAs, turnover is initiated by the direct-access pathway, which starts with an endonucleolytic cleavage by RNase E in *E. coli* or RNase Y in *B. subtilis*. The cleaved product is no longer protected from degradation, and is quickly hydrolyzed by nucleases. In the second pathway, initiation happens at the 5' end. In *E. coli* this is also carried out by RNase E, but in *B. subtilis* is carried out by RNase J1 (Babitzke & Kushner, 1991; Durand et al., 2012; Laalami et al., 2013; Lehnik-Habrink, Schaffer, et al., 2011; Melefors & von Gabain, 1991). In this case, both enzymes have a strong preference for mono-phosphorylated 5' nucleotides (Deana et al., 2008; Richards et al., 2011). For this pathway, the initiating

event is usually the dephosphorylation of the 5' nucleotide to a monophosphate. These two pathways together are essential in both *E. coli* and *B. subtilis*. There is also a third pathway in which the ribonuclease PNPase acts on the 3' end of a RNA as a 3' to 5' exonuclease (Richards, Mehta, & Karzai, 2006). Both *E. coli* and *B. subtilis* have PNPase, which plays a role in RNA turnover under normal conditions, but is not sufficient for carrying out enough RNA turnover to produce viable cells.

In both *E. coli* and *B. subtilis*, the half-lives of RNAs vary over a large range. This implies that RNAs are not all being indiscriminately degraded at the same rate, but that some feature of the RNA determines its stability. At the same time no clear features, either primary sequence or secondary structure, have been identified for either RNase E or RNase Y (Hui, Foley, & Belasco, 2014). How RNA degradation of different RNAs varies and how this is regulated is not well understood. It's thought that RNase E and RNase Y might target a fairly common feature, and that it is the protection of these features from the nuclease that prevents cleavage. The protection could be provided by secondary structure, but could also be created by the variety of cellular machines and proteins that bind to RNA, such as ribosomes or RNA chaperones (Arnold, Yu, & Belasco, 1998; Braun, Le Derout, & Regnier, 1998; Yarchuk, Jacques, Guillerez, & Dreyfus, 1992). The regulation of the stability of specific RNAs and the mechanisms by which they can be altered is of great interest to scientists today.

The role of RNA turnover in biofilm formation by *B. subtilis*

The connection between biofilm formation in *B. subtilis* and RNA turnover is the three protein complex of YlbF, YmcA, and YaaT. Presented here is the finding that this complex, which is required for biofilm formation, interacts with RNase Y to regulate the level of the mRNA for *sinR*, the dedicated repressor of biofilm formation. In addition, YlbF, YmcA, and YaaT are found to play a role in global RNA turnover with RNase Y. Results from RNA-sequencing experiments reveal some information about the extent of YlbF, YmcA, and YaaT activity. The interaction of these proteins with RNase Y and their effect on RNase Y activity provides novel insight into RNA turnover in *B. subtilis*. Moreover the

regulation of biofilm formation based on mRNA stability provides a new level of regulation for this behavior, which could provide an additional mechanism for *B. subtilis* to recognize specific signals.

Chapter 1

YlbF and YmcA are required for biofilm formation and are in a complex with YaaT

An ongoing challenge is to identify all of the genes involved in biofilm formation and to understand how they function. Here I report on two genes, *ylbF* and *ymcA*, whose role in biofilm formation has remained mysterious since their identification in 2004. This chapter provides background information about these proteins, reports the effect of the loss of these proteins, and also describes the identification of a third interacting partner, YaaT. It has recently been published that these proteins play a role in accelerating the phosphorylation of Spo0A, a finding that will be revisited in Chapter 2.

Contributions: Part of this work was published as: (DeLoughery, Dengler, Chai, & Losick, 2016)

DeLoughery, A., Dengler, V., Chai, Y., & Losick, R. (2016). Biofilm formation by Bacillus subtilis requires an endoribonuclease-containing multisubunit complex that controls mRNA levels for the matrix gene repressor SinR. *Mol Microbiol, 99*(2), 425-437. doi: 10.1111/mmi.13240

All of the experiments presented in this section were completed by me. A few strains were previously constructed by Yunrong Chai. Significant intellectual contributions were by Richard Losick and myself. The motivation for this work is based off previous work by Yunrong Chai, Frances Chu, and Daniel Kearns.

1.1. Mutants lacking YlbF or YmcA are defective for biofilm formation and do not express matrix production operons.

YlbF and YmcA are required for biofilm formation in *B. subtilis*. These proteins were identified through a transposon mutagenesis screen (Branda et al., 2004). In the screen, mutants lacking these proteins were selected because they failed to produce architecturally complex colonies. Insertional deletions of these genes, *ylbF* and *ymcA,* are blocked for biofilm formation. Reassessing the mutant's ability to form structured colonies under biofilm conditions reconfirmed these results (Figure 2A, 2B). The mutants failed to produce biofilms as colonies on solid agar or as floating pellicles in liquid media. The resulting colonies were flat, shiny, and completely lacked wrinkles, indicating a block in biofilm formation. Deletion of *sinR*, the master repressor of biofilm formation, restores matrix production to mutants of *ylbF* and *ymcA* (Kearns et al., 2005). This finding provides evidence that the block in biofilm formation was occurring at the level of regulation and that the mutants were still capable of physically forming biofilms when relieved of repression. To quantitatively assess the ability of *ylbF* and *ymcA* mutants to form biofilms, the expression of the *eps* operon was examined in the mutants under biofilm conditions. Using a *lacZ* reporter it was found that there is no expression of the *eps* operon in cells lacking either YlbF or YmcA under biofilm conditions (Figure 2C).

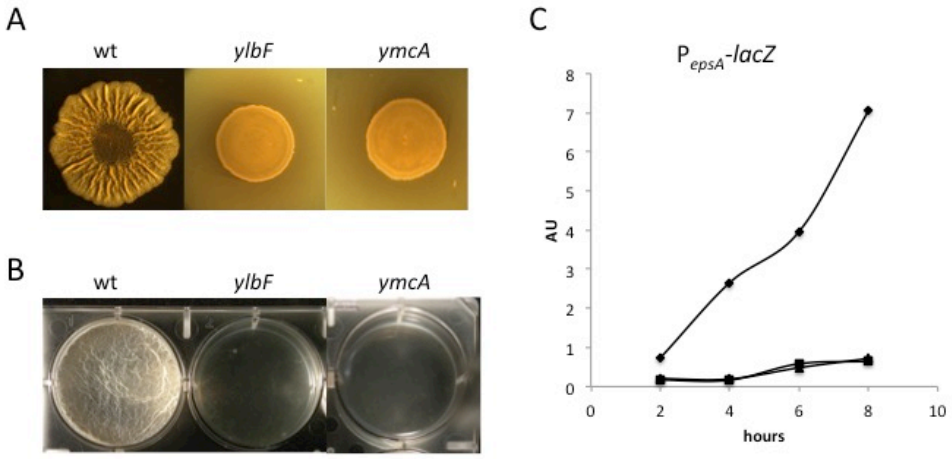

Figure 2: Mutants of *ylbF* and *ymcA* are defective for biofilm formation.

A. Colony morphology phenotypes of the wild strain (3610), a mutant of *ylbF* (RL4168), and a mutant of *ymcA* (RL4169) grown for 72 h on MSgg 1.5% agar (all images are at the same scale).

B. Pellicle biofilm phenotypes of the wild strain (3610), a mutant of *ylbF* (RL4168), and a mutant of *ymcA* (RL4169) grown for 72 h on MSgg 1.5% agar (all images are at the same scale).

C. Synthesis of β-galactosidase under the control of the *epsA* promoter (P_{epsA}–*lacZ*) in liquid shaking MSgg. Shown is a comparison of synthesis by the wild type (◆) with that of *ylbF* (■), *ymcA* (●)

1.2. What are YlbF and YmcA?

YlbF is a 149 amino acid protein, encoded by the first gene of a two-gene operon with *ylbG* (Figure 3A). The cause of the biofilm defect is due to the loss of YlbF and not to a polar effect on *ylbG* based on complementation. Sequence analysis of YlbF does not reveal the presence of any domains of known function, with the exception of a putative metallothionein-like domain at the C-terminus. This domain has four highly conserved cysteine residues that are required for YlbF function. The gene for YlbF is conserved amongst Gram-positive bacteria, including the pathogens *Listeria monocytogenes*, *Staphylococcus aureus*, and *Enterococcus faecalis* (Figure 3B). Mutants of these homologous genes have not been reported to have phenotypes in these other organisms, nor has a function been described. Tiling array experiments have shown that the expression of *ylbF* does not change substantially under many growth and stress conditions, with the exception of a slight increase after treatment with the oxidizing agents hydrogen peroxide or paraquat as well as a slight increase after exposure to cold.

YmcA is a 143 amino acid protein that is expressed from the second gene in a two-gene operon with *ymcB*. The expression of *ymcA* at an ectopic locus restores biofilm formation to the mutant. The amino acid sequence of YmcA shows a high degree of similarity to YlbF (19% identity and 51% similarity) (figure 4A). The gene for YmcA is also similarly conserved amongst Gram-positive bacteria (Figure 4B). YmcA and YlbF are known to directly interact with each other based on a yeast 2-hybrid experiment as well as pull-down experiments with purified protein. Sequence analysis, like with YlbF, reveals little, but a structure of YmcA has been solved for residues 2 through 124. In the structure, YmcA forms a dimer. Each monomer contains 5 alpha helices that conform to approximately the shape of a capital L, with each monomer folding back on itself such that the N-terminus and C-terminus are near to each other (Figure 4C). While a structure of YlbF has not been solved, secondary structure predictions reveal the likely presence of the same 5 alpha helices (figure 4A).

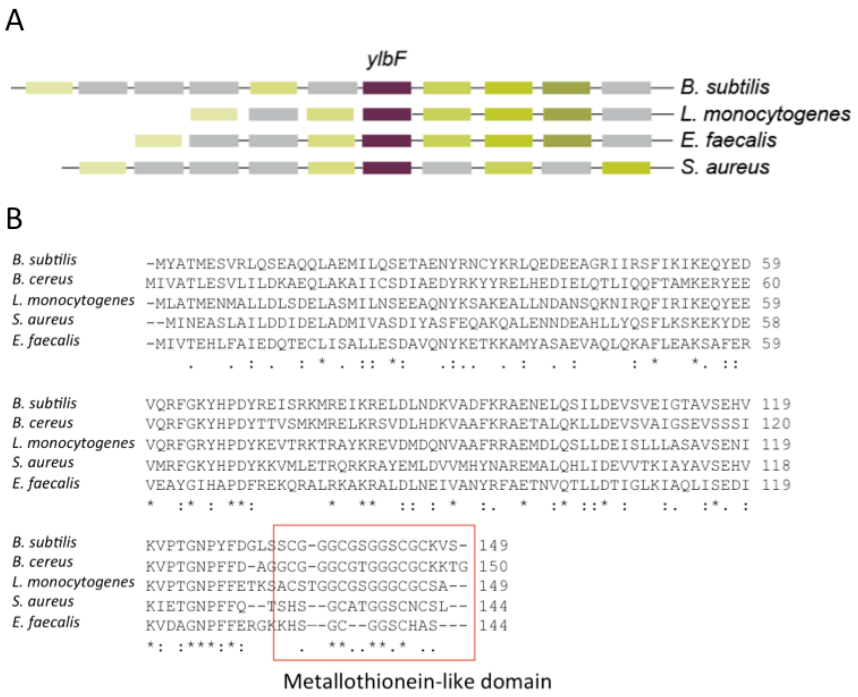

Figure 3. What is YlbF?

A. Schematic showing *ylbF* (magenta) and its homologs in *Listeria monocytogenes, Staphylococcus aureus*, *and Enterococcus faecalis* with its neighboring genes on the chromosome represented in a contrasting color. The neighboring genes are colored such that matching colors represent genes that are homologs to the respective gene in *B. subtilis*. The percentage identity for the proteins are as follows: YlbF (*L. monocytogenes* 54%, *S. aureus* 40%, *E. faecalis* 33%).

B. Sequence alignment of YlbF and it homologs in *Bacillus cereus, Staphylococcus aureus*, *Enterococcus faecalis*, and *Listeria monocytogenes* from ClustalW. Asterisks demark highly conserved residues. The red box indicates a predicted putative metallothionien-like domain.

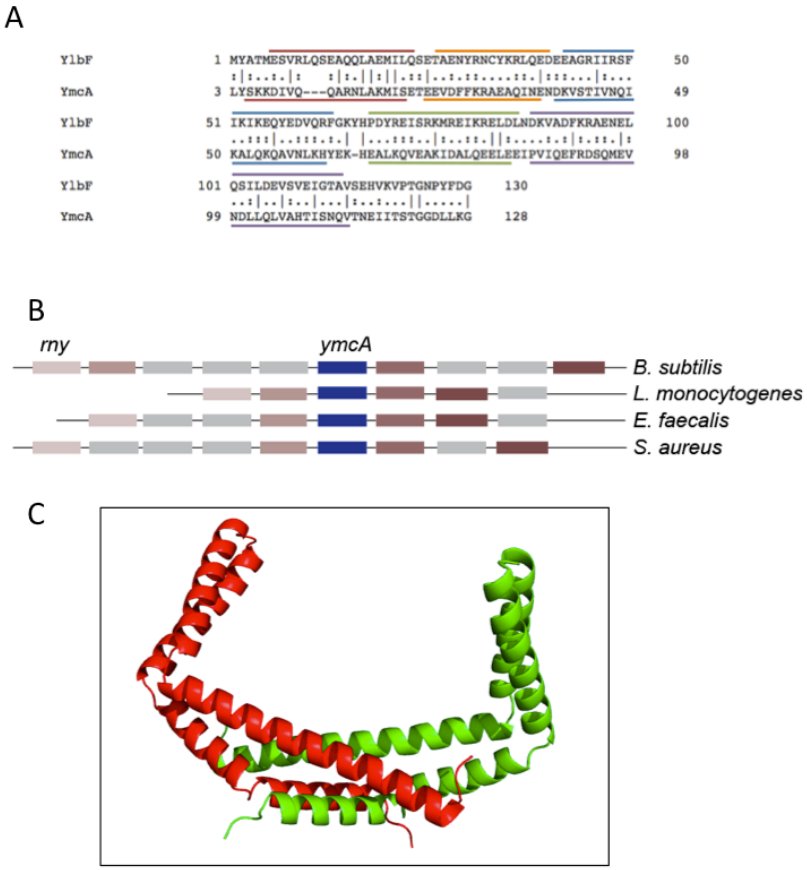

Figure 4. What is YmcA?

A. Alignment of ylbf with ymcA, bars represent predicted alpha helices

B. Schematic showing *ylbF ymcA* (blue) and its homologs in *Listeria monocytogenes, Staphylococcus aureus*, *and Enterococcus faecalis* with their neighboring genes on the chromosome represented in a contrasting color. The neighboring genes are colored such that matching colors represent genes that are homologs to the respective gene in *B. subtilis*. The percentage identity for the proteins is as follows: YmcA (*L. monocytogenes* 50%, *S. aureus* 36%, *E. faecalis* 31%).

C. The structure of a dimer of YmcA. Pdb 2PIH.

1.3. Mutants lacking YlbF or YmcA show additional stationary phase defects:

In addition to a biofilm defect, mutants of *ylbF* and *ymcA* are reported to be defective for other stationary phase behaviors. YlbF is known to be needed for competence in *B. subtilis*. Mutants of YlbF have been reported to have a 10,000-fold decrease in sporulation efficiency in the lab strain 168. While I have seen a small effect on sporulation, I do not see as strong of a defect as had been reported in the laboratory strain PY79. In the wild strain 3610 I see only a very small effect on sporulation (not shown).

In addition to the reported phenotypes, I have found that mutants lacking YlbF or YmcA are impaired for growth in LB medium (see figure 7). Specifically, cultures of the mutants show identical growth to wild type during exponential phase, but at the onset of early stationary phase show a delay in growth that results in a dip in the culture density. The mutants continue growth after the delay and ultimately reach a density slightly lower than wild type. Growth curves carried out in a defined medium with varying concentrations of glucose provide evidence that this delay corresponds to depletion of glucose from the growth medium (not shown). Mutants of *ylbF* and *ymcA* do not show growth impairment in the biofilm inducing medium MSgg, which does not contain glucose.

1.4 YlbF interacts with YaaT

In attempts to discover the molecular function of YlbF in biofilm formation and in general, I carried out a pull-down experiment to find potential interacting partners. The identification of an interaction with a protein of known function could have provided a clue towards YlbF's function, but any novel interacting proteins would be interesting. The experiment was carried out by expressing a Maltose Binding Protein (MBP) fusion to YlbF in *B. subtilis*, and MBP alone as a control (Figure 5A). Cells were harvested near the onset of stationary phase under biofilm conditions, and a two-part purification was carried out from cleared lysates. First MBP-YlbF or MBP alone were purified using an amylose resin gravity column. The MBP tag also has a His tag, which was used to pull out MBP-YlbF or MBP alone from the eluate of the first purification step using magnetic Dyna beads. The resulting samples were subjected to SDS-PAGE (Figure 5B) followed by tandem mass spectroscopy to determine the protein

composition of each. Proteins that were highly enriched in the MBP-YlbF sample compared to the MBP alone were identified as potential interacting partners. YlbF was found to strongly interact with two proteins (Table 1). The first being YmcA, a known interacting partner, and the second being YaaT, the subject of the following paragraphs. That we were able to pull down YmcA with YlbF provided a positive control for our experimental setup since these proteins have been shown to interact.

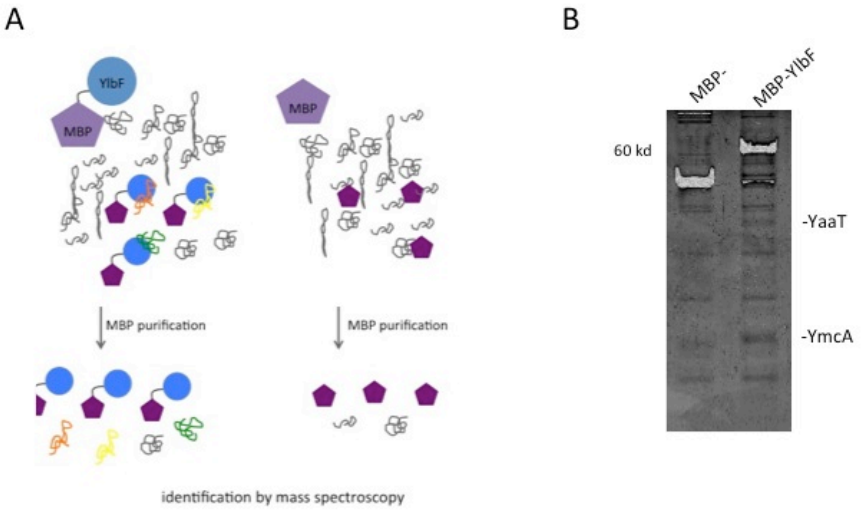

Figure 5: Pull-down with YlbF reveals potential interacting partners.

A. Schematic depicting the MBP pull-down experimental design and concept.

B. SDS-PAGE of proteins pulled down with either MBP-YlbF or MBP alone from *B. subtilis*. This is an example of the types of samples submitted for MS/MS analysis. The bands corresponding to YmcA and YaaT can be seen in the MBP-YlbF sample.

Table 1. List of proteins that co-purify with MBP-YlbF from *B. subtilis* grown under biofilm inducing conditions.

MBP-YlbF (strain AJD17) and MBP (strain AJD18) were purified from 3610 grown under biofilm inducing conditions. Purified MBP-YlbF and MBP along with any associated proteins were separated by SDS-PAGE, and subjected to MS/MS to identify co-purifying proteins. Listed here, are the top nine proteins identified. They are ordered according to the ratio of the peptide counts for each protein between the MBP-YlbF purification and the MBP purification. See Supporting Information methods for details of sample preparation.

Table 1 continued.

Protein	Name	Function	MBP-YlbF	MBP	Coverage
YmcA		unknown	191	1	93
YaaT		unknow	301	2	90.5
PtsG	PTS EIICBA glucose	major glucose permease	61	1	24.5
Fbp	fructose-1,6-bisphosphatase	gluconeogenesis	49	0	39.8
YukB	FtsK/SpoIIIE-like ATPase	YukE secretion	48	0	20.5
Rny	RNase Y	degradation/processing of mRNA	48	0	53.5
FtsA	cell division protein	formation of Z ring	47	1	61.1
Pyc	Pyruvate carboxylase	oxaloacetate metabolism	42	0	29.9
MtnN	methylthioadenosine nucleosidase	methionine salvage	32	0	74.5
TcyA	cysteine ABC transporter	cysteine uptake	32	0	58.2
DnaX	DNA polymerase III	DNA replication	29	1	38.4
YokF	similar to micrococcal nuclease	unknown	29	1	49
CitB	trigger enzyme aconitase	TCA cycle	27	0	24.9
GgaB	glycosyltransferase 2 family	teichoic acid biosynthesis	27	0	16.6
BfmBAA	2-oxoisovalerate dehydrogenase	branched-chain keto acid utilization	26	0	59.1
FolE	GTP cyclohydrolase I	folate biosynthesis	26	0	41.1
RelA	GTP pyrophosphokinase	stringent response	25	1	29.8
ManA	mannose-6-pphosphate isomerase	mannose utilization	47	2	67.3
MurG	UDP-N-acetylglucosamine transferase	cell wall biosynthesis	44	2	73.6
LepA	elongation factor 4	translation	62	3	47.9
GlmS	glutamin-fructose-6-phosphate transaminase	cell wall biosynthesis	80	4	51.5
Buk	butyrate kinase	branched-chain keto acid utilization	60	3	67.5
PonA	class A penicillin-binding protein	cell wall biosynthesis	46	3	36
FolD	methylenetetrahydrofolate dehydrogenase	tRNA modification	43	3	47.3
DegU	two-component response regulator	cell signaling	56	4	94.3

1.5. what is YaaT?

YaaT is a 275 amino acid protein encoded in a multi gene operon near the origin of replication. The gene for YaaT is just downstream of the essential gene *holB*, which encodes the delta subunit of DNA polymerase. It is upstream of *yabA*, which encodes a protein that plays a role in repressing the initiation of DNA replication. The gene for YaaT is conserved amongst Gram-positive bacteria (Figure 6A), even to a higher extent than the genes for *ylbF* and *ymcA*. Mutants lacking YaaT have previously been shown to have a sporulation defect in the lab strain 168. This paper also reported that of a GFP fusion to YaaT localizes to the cell periphery with a possible enrichment at the cell division septum (Hosoya, Asai, Ogasawara, Takeuchi, & Sato, 2002).

1.6. Deletion of *yaaT* reduces biofilm formation and impairs growth in LB

If the interaction between YaaT and YlbF were important for biofilm formation, a mutant lacking YaaT would be expected to have a biofilm phenotype. To investigate this possibility a mutant was created to assess its ability to form a matrix. Due to the presence of essential genes and genes known to have phenotypes in the neighborhood to *yaaT*, a markerless deletion was created using the pminimad system; specifically the entire open reading frame of *yaaT* was removed. The resulting strain indeed is defective for biofilm formation (Figure 6B). Interestingly, while a *yaaT* mutant has a clear phenotype, its ability to form biofilms is not as severely blocked as mutants of *ylbF* or *ymcA*. Strains lacking YaaT form flat colonies that lack wrinkles, but are not shiny indicating the production of some matrix. The colonies also spread out farther than mutants of *ylbF* or *ymcA*.

The mutant lacking *yaaT* also shows growth impairment in liquid LB that is identical to that of mutants lacking *ylbF* or *ymcA* (Figure 6C). Mutants lacking *yaaT* have previously been reported to have a significant sporulation defect, which I have not tried to confirm. Whether or not mutants of *yaaT* are reduced for genetic competence has not been shown.

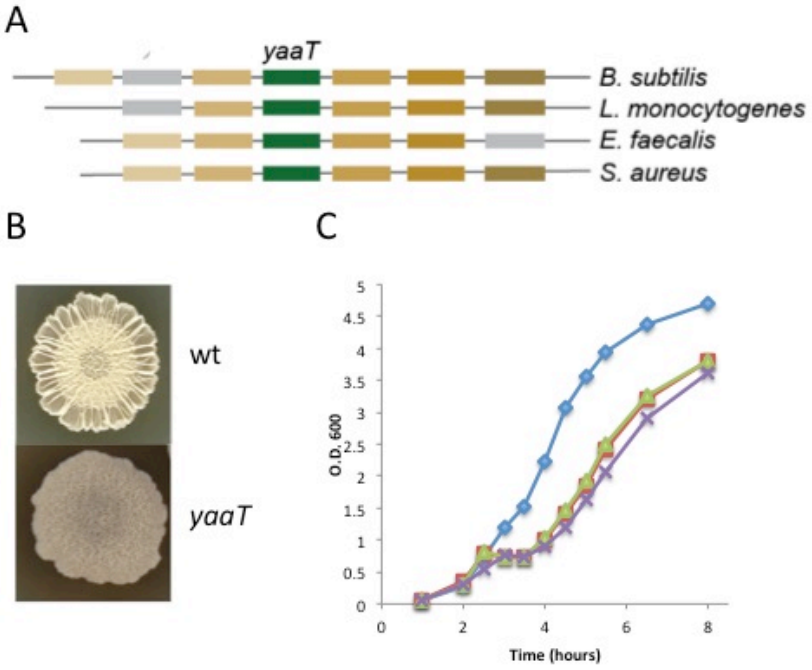

Figure 6. What is YaaT?

A. Schematic showing *yaaT* (green) and its homologs in *Listeria monocytogenes, Staphylococcus aureus*, *and Enterococcus faecalis* with its neighboring genes on the chromosome represented in a contrasting color. The neighboring genes are colored such that matching colors represent genes that are homologs to the respective gene in *B. subtilis*. The percentage identity for the proteins are as follows: YaaT (*L. monocytogenes* 67%, *S. aureus* 63%, *E. faecalis* 59%), and for Rny (*L. monocytogenes* 78%, *S. aureus* 71%, *E. faecalis* 70%).

B. Colony morphology phenotypes of the wild strain (3610) and a mutant of *yaaT* (RL4168) grown for 72 h on MSgg 1.5% agar (all images are at the same scale).

C. Shown are growth curves for the wild type (blue) (RL3852), and the mutants *ylbF* (red) (RL4168), *ymcA* (green) (RL4169), and *yaaT* (purple) in LB medium. The times in panels B, C and D are hours after inoculation of the culture, corresponding to an OD_{600} of ~ 0.05.

1.7. Other proteins that co-purify with YlbF.

In addition to effectively pulling down these two proteins, several additional proteins were enriched in the YlbF-MBP samples. A list can be found on Table 1. While there were different types of proteins present, many of the highly enriched proteins had one of two factors in common. First, many are known to localize to the membrane, PtsG, YukB, Fbp, RNase Y, FtsA, Pyc, TcyA, and others. Several of the proteins have easily identifiable transmembrane domains, and a few that do not have been shown using fluorescent microscopy to localize to the periphery of the cell. Second, many of the proteins of known function play a role in central carbon metabolism, such as PtsG, Fbp, Pyc, CitB, ManA, and others. The possible significance of these interactions is discussed in the next section

Discussion

After it was appreciated that wild strains of *B. subtilis* can form architecturally complex biofilms, a candidate gene approach as well as an unbiased screen were used to identify all of the genes required for matrix production. Many of the identified genes were of known function, including the regulatory genes *sinI, sinR, abrB, spo0A, spo0B,* and *spo0F*. That these genes are required for biofilm formation revealed that the regulation of matrix production is tied into the regulation of sporulation. In addition to regulatory proteins, the proteins that produce the matrix were also identified. The *epsA-O* operon, composed of genes similar to known exopolysaccharide synthesis genes, indeed is responsible for EPS production. The secreted, amyloid-like protein TasA was found to be the protein component of the matrix. Two genes that were completely of unknown function, but required for biofilm formation, are *ylbF* and *ymcA*. They are known to play a role in the regulation of biofilm formation by epistatic analysis with biofilm regulators. We are particularly interested in these genes because they may provide an additional level of matrix production regulation.

In addition to a biofilm defect, mutants lacking YlbF and YmcA are defective for other stationary phase behaviors, such as growth, genetic competence, and sporulation (Tortosa, Albano, & Dubnau,

2000). This is an indication that YlbF and YmcA function more globally than just the regulation biofilm formation at the onset of stationary phase, so a better understanding of their function could provide further insight into how *B. subtilis* recognizes and responds to changes in growth.

Using a pull-down experiment, I found that YlbF interacts with another protein YaaT. Because another group identified this interaction and carried out some biochemical experiments on this complex, further experiments to confirm this interaction were not carried out. While the function of YaaT is not known, the identification of a third interacting partner provides possibly interesting and useful insight. For instance the gene for YaaT is encoded in an operon with several genes for DNA replication and its initiation. Also, YaaT has been shown to localize to the cell periphery even though it does not have a predicted trans membrane domain.

That I also pulled down several membrane proteins with YlbF, indicates that the complex is likely interacting with a protein that localizes to the membrane. It is unlikely that YlbF is directly interacting with all of these with all of the proteins that were pulled-down. What is more likely is that YlbF directly interacts with a specific protein, but then due to that proteins association with the membrane and other membrane proteins leads to the indirect pull-down of several proteins. Often when working with membrane proteins additional considerations are applied to their purification, such as the addition of detergents to buffers. Since it wasn't expected that YlbF would be interacting with membrane proteins, experiments weren't carried out as such. This may in part explain why so many proteins were co-purified with YlbF. To more specifically identify the protein that YlbF is directly interacting with this experiment could be repeated using protocols specific to membrane proteins.

The identification of several proteins involved in either glycolysis or gluconeogenesis is interesting in the context of the growth defect of mutants lacking YlbF. Preliminary evidence indicates that the growth defect corresponds to the depletion of glucose from the media and that mutants are less efficient at transitioning from glucose metabolism to an alternative carbon source. Another protein that co-purifies with YlbF and has previously been shown to interact with glycolytic enzymes is the ribonuclease RNase Y. RNase Y plays a major role in RNA turnover and processing in *B. subtilis*. RNase

Y has also been shown to play a role in the regulation of biofilm formation through the repressor *sinR* (Lehnik-Habrink, Newman, et al., 2011).

Chapter 2

The complex of YlbF-YmcA-YaaT plays a role in biofilm formation in a manner independent of the response regulator Spo0A.

Investigating the role of YlbF and YmcA in the genetic competence of *B. subtilis,* another group of scientists also identified YaaT, and proposed a function of this complex in accelerating the phosphorylation of Spo0A (Carabetta et al., 2013). Because biofilm formation, competence, and sporulation are all regulated by Spo0A, it is reasonable that an effect on the phosphorelay could explain all three phenotypes. However, in this chapter I show that Spo0A is not affected under biofilm conditions in mutants lacking YlbF, YmcA, or YaaT, and thus this complex must have another function that promotes biofilm formation. The conservation of these genes and the growth impairment of these mutants provide further evidence that they must have a function outside of the phosphorelay.

Contributions: Part of this work was published as: (DeLoughery et al., 2016)

DeLoughery, A., Dengler, V., Chai, Y., & Losick, R. (2016). Biofilm formation by Bacillus subtilis requires an endoribonuclease-containing multisubunit complex that controls mRNA levels for the matrix gene repressor SinR. *Mol Microbiol, 99*(2), 425-437. doi: 10.1111/mmi.13240

All of the experiments presented in this section were completed by me. A few strains were previously constructed by Yunrong Chai. Significant intellectual contributions were by Richard Losick and myself. The motivation for this work is based off previous work by Yunrong Chai, Frances Chu, and Daniel Kearns.

2.1. Spo0A activity is not dependent on YlbF, YmcA and YaaT under biofilm-inducing conditions

A possible explanation for the block in biofilm formation observed in mutants lacking YlbF, YmcA or YaaT is that these proteins are needed for efficient phosphorylation of Spo0A, which, in turn, activates the gene for SinI. To investigate this possibility I measured *sinI* expression using a *lacZ* fusion in wild type (biofilm proficient) cells (strain 3610) and cells of mutant derivatives lacking YlbF or YmcA. In liquid, shaking biofilm-inducing medium (MSgg), the pattern of induction of P_{sinI}-*lacZ* in stationary phase was indistinguishable between the mutants and the wild type (Figure 7B) As a control, and as expected, little or no expression was observed in a mutant lacking Spo0A (Figure 7B). Also, the absence of YlbF or YmcA did not affect the levels of SinI protein as judged by immunoblot analysis using anti-SinI antibodies (Figure 7C).

Spo0A~P also contributes to biofilm formation by repressing the gene for AbrB, which is expressed at high levels during exponential phase growth and repressed in stationary phase in a Spo0A-dependent manner. Using a *lacZ* fusion to the promoter for *abrB* (P_{abrB}–*lacZ*), I observed little difference in the pattern of P_{abrB}–*lacZ* expression between the wild type and cells lacking YlbF or YaaT (Figure 7D). *In toto*, the results are inconsistent with the idea that the block in biofilm formation observed in the mutants is due to impaired Spo0A activity.

Figure 7 *YlbF, YmcA and YaaT function independently of Spo0A.*

A. Diagram of the Spo0A-controlled regulatory circuit governing biofilm formation. Dashed line indicates protein-protein interaction and solid lines indicate transcription regulation.

B. Synthesis of β-galactosidase under the control of the *sinI* promoter (P*sinI–lacZ*) in liquid shaking MSgg. Shown is a comparison of synthesis by the wild type (◆)(RL4610) with that of *ylbF* (■)(AJD95), *ymcA* (●)(AJD96) and *spo0A* (▲)(AJD97) mutants (averages plotted for three biological replicates).

C. Comparison of SinI levels by Western blotting carried out as previously described (Chai *et al*., **2008**). Shown is a comparison of levels of the wild type (wt) (RL3852) with that of *ylbF* (RL4168) and *ymcA* (RL4169) mutants in liquid shaking MSgg during stationary phase. A western blot with an antibody against σA, a component of RNA polymerase, was used as a loading control. The same samples were loaded at two concentrations (left 1X, right 0.5X) to confirm linearity of the blot.

D. Same as B except the synthesis of β-galactosidase was under the control of the *abrB* promoter (P*abrB–lacZ*). Wild type (◆)(AJD99) is compared with mutants of *ylbF* (■)(AJD100), *yaaT* (▲)(AJD102), and *spo0A* (×) (AJD101). For this panel the β-galactosidase activity is reported as a percentage of the activity at the initial, 2-h time point.

E. Shown are growth curves for the wild type (■) (RL3852), and the mutants *ylbF* (□) (RL4168), *ymcA* (Δ) (RL4169), and *spo0A* (◆) (RL4620) in LB medium. The times in panels B, C and D are hours after inoculation of the culture, corresponding to an OD600 of ~ 0.05.

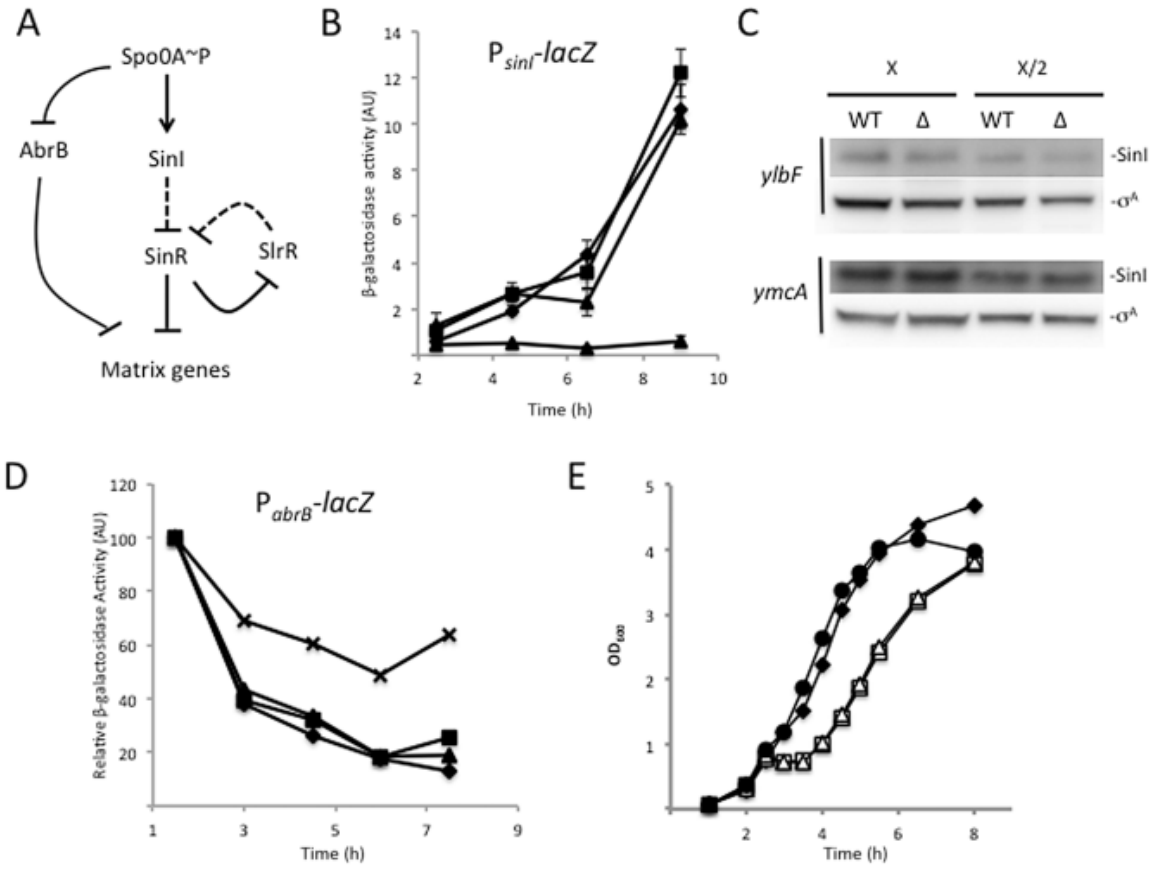

Figure 7 continued.

2.2. Absence of YlbF, YmcA or YaaT, but not Spo0A, impairs growth in Luria-Bertani (LB) medium

Also distinguishing the phenotype of mutant derivatives of the wild strain (3610) lacking YlbF, YmcA and YaaT from a mutant lacking Spo0A is a growth defect in LB medium. Unlike the wild type, cultures of *ylbF*, *ymcA* and *yaaT* mutant cells exhibited a conspicuous and reproducible dip in growth at the onset of stationary phase measured by the OD_{600} in liquid shaking medium. Specifically, the mutants grew at the same rate as the wild type until the cultures reached an OD_{600} of about 1.0 at which point the mutants exhibited a short period of reduced optical density and then resumed growth, ultimately reaching a density comparable with that of wild type (Figure 7E and 8A). Importantly, no such growth impairment was observed for mutants lacking Spo0A or the phosphorelay proteins Spo0B and Spo0F (Figure 8B). Live/dead staining indicated that impaired growth was associated with cell death in *ylbF*, *ymcA* and *yaaT* mutants. Again, no evidence of cell death was observed for a *spo0A* mutant or mutants lacking the other phosphorelay components (figure 8C). Another phenotype that distinguishes the mutants of *ylbF*, *ymcA* and *yaaT* from mutants of components of the phosphorelay is colony morphology on solid LB agar (1.5%). The *ylbF*, *ymcA* and *yaaT* mutants produced colonies that were smaller than those of the wild type, flat and smooth, whereas mutants lacking Spo0A, produced colonies that were very large with a distinct mucoid texture (Figure 8D). The results so far indicate that YlbF, YmcA and YaaT have a function that is distinct from the phosphorelay.

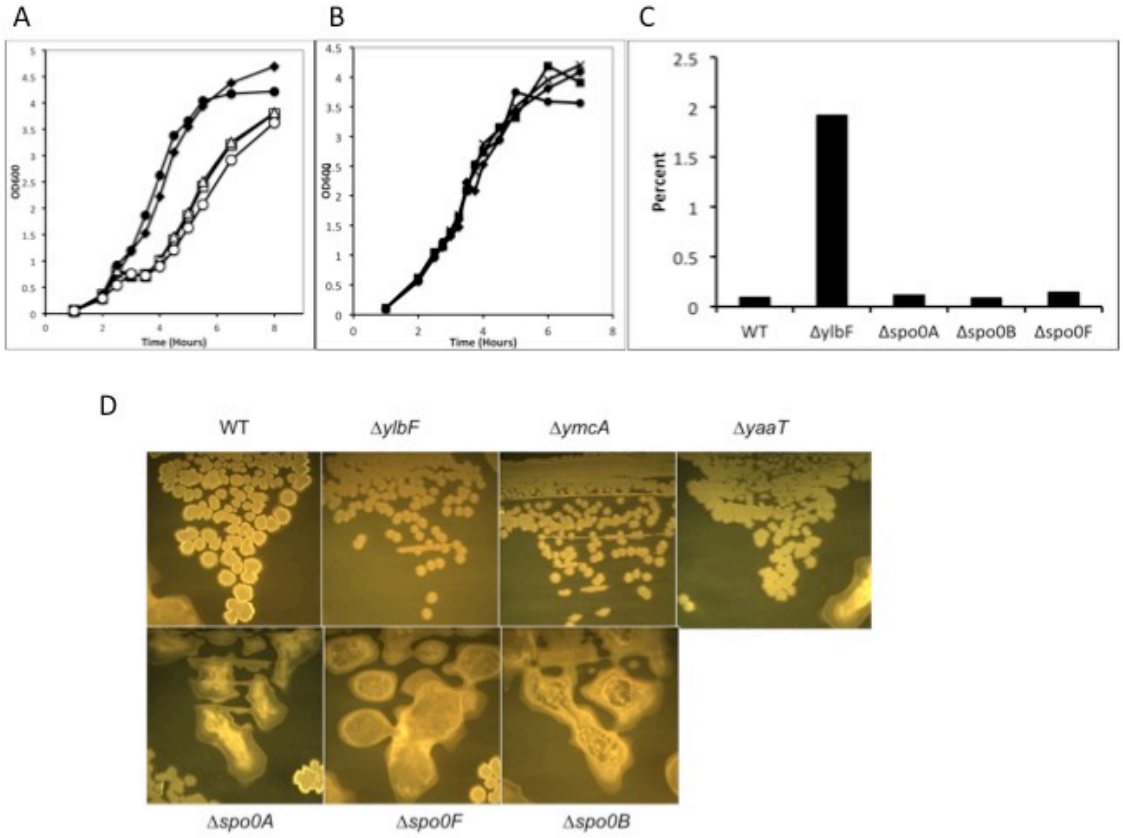

Figure 8. Mutants of *ylbF*, *ymcA*, and *yaaT* do not phenocopy mutants of *spo0A*, *spo0B*, or *spo0F*.

A. Shown are growth curves for the wild type (■) (RL3852), and the mutants *ylbF* (□) (RL4168), *ymcA* (Δ) (RL4169), and *spo0A* (◆) (RL4620) in LB medium. The times in panels B, C and D are hours after inoculation of the culture, corresponding to an OD600 of ~ 0.05.

B. Shown are growth curves for the wild type (■) (RL3852), and the mutants *ylbF* (□) (RL4168), *ymcA* (Δ) (RL4169), and *spo0A* (◆) (RL4620) in LB medium. The times in panels B, C and D are hours after inoculation of the culture, corresponding to an OD600 of ~ 0.05.

C. Live/dead staining used to determine the percentage of dead cells from cultures grown to an OD_{600} of 1.0.

D. Colony morphology of indicated mutants after 24 hours of growth on LB plates

2.3. The genes for *ylbF*, *ymcA*, and *yaaT* are conserved in many species of Gram-positive bacteria that lack Spo0A and the other components of the phosphorelay.

Another indication that YlbF, YmcA and YaaT have a distinct function is based on their gene conservation. Homologous genes to *ylbF*, *ymcA* and *yaaT* can be found in many Gram-positive bacteria, including most of the Firmicutes. Not only are these genes conserved, but also all three have close neighbors that are also conserved. Moreover the genetic synteny of these regions is conserved. These species lack homologous genes to *spo0A* and other genes for components of the phosphorelay.

2.4. Spo0B and Spo0F, as well as all of the other components of the phosphorelay, do not co-purify with YlbF.

The pull-down experiment to identify potential interacting partners of YlbF revealed that several proteins co-purify with this protein. However, no components of the phosphorelay were co-purified with YlbF (Table 1). This includes Spo0F and Spo0B, which were reported by Carabetta et al (2013) to interact with YmcA in a bacterial 2-hybrid. In addition Carabetta et al. carried out pull-down experiments with mass spectroscopy for all three proteins, YlbF, YmcA, and YaaT, and did not co-purify any components of the phosphorelay in all cases (Carabetta et al., 2013).

Discussion

The response regulator Spo0A, when phosphorylated (Spo0A~P) activates and represses the transcription of hundreds of genes, and has been fairly well studied in the context of sporulation. Spo0A~P also activates biofilm formation through the increased expression of *sinI*, as well as the repression of *abrB*. While not as well understood, Spo0A~P can affect genetic competence. Its major contribution towards competence is likely through repression of *abrB*, which is a known repressor of the master regulator of competence *comK*, but there are likely other contributions. The levels of Spo0A~P needed for each genetic program is different. For instance biofilm formation requires and intermediate level of Spo0A~P, and in fact is repressed by high levels of Spo0A~p. On the other hand sporulation

requires the accumulation of high levels of Spo0A~P (Molle et al., 2003). Moreover, in the lab, we test sporulation, biofilm formation, and competence under very different growth conditions, each with a specific medium, so the extent that the phosphorelay is activated under each varies.

While a defect in Spo0A phosphorylation could potentially explain the competence, biofilm, and sporulation phenotypes of mutants lacking YlbF, YmcA, or YaaT, I do not see a significant defect in the phosphorelay under biofilm inducing conditions. I see wild type levels of *sinI* expression as well as *abrB* expression in MSgg. Mutants also have similar levels of SinI protein compared to wild type. Since the biofilm defect is not explained by a role in the phosphorelay, the complex of YlbF, YmcA, and YaaT must be carrying out another function. This function of YlbF, YmcA, and YaaT towards biofilm formation still remains mysterious and warrants effort towards its elucidation.

While I do not have definitive evidence that YlbF, YmcA, and YaaT are not acting on the phosphorelay, I have shown that the complex must have an additional function, and that it is this additional function that is required for biofilm formation. It is of course possible that the complex of YlbF, YmcA, and YaaT is both acting on the phosphorelay and at the same time carrying out an additional unknown function. It is also possible that the complex is not directly affecting the phosphorelay, but that through an unknown function indirectly affects the phosphorelay under sporulation of competence conditions.

Chapter 3

The complex of YlbF, YmcA, and YaaT interacts with the endoribonuclease RNase Y to control the mRNA levels for the matrix gene repressor SinR

Mutants of the genes *ylbF*, *ymcA* and *yaaT* are blocked in biofilm formation, but the function of YlbF, YmcA, and YaaT remains unclear. In this chapter I provide strong evidence that these proteins are interacting with and affecting the activity of the ribonuclease RNase Y, and that the block in biofilm formation is caused by an increase in the levels of SinR and of its mRNA. My work suggests that *sinR* mRNA stability is an additional posttranscriptional control mechanism governing the switch to multicellularity and raises the possibility that YlbF, YmcA, and YaaT broadly regulate mRNA stability as part of an RNase Y-containing, multi-subunit complex.

Contributions: Part of this work was published as: (DeLoughery et al., 2016)

DeLoughery, A., Dengler, V., Chai, Y., & Losick, R. (2016). Biofilm formation by Bacillus subtilis requires an endoribonuclease-containing multisubunit complex that controls mRNA levels for the matrix gene repressor SinR. *Mol Microbiol, 99*(2), 425-437. doi: 10.1111/mmi.13240

All of the experiments presented in this section were completed by me. A few strains were previously constructed by Yunrong Chai. Significant intellectual contributions were by Richard Losick and myself. The motivation for this work is based off previous work by Yunrong Chai, Frances Chu, and Daniel Kearns.

3.1. SinR levels are higher in *ylbF* and *ymcA* mutants

Because, as I have shown, *ylbF*, *ymcA* and *yaaT* mutations do not act by impairing Spo0A activity or the production of SinI, we next considered the possibility that they exert their effect by influencing the levels of SinR. I carried out immunoblot analysis to investigate the relative levels of SinR in wild-type cells and in mutant cells lacking YlbF and YmcA using antibodies against SinR that we had employed in previous studies (Chai et al., 2008; Chai et al., 2010, 2011). Protein was extracted from stationary phase cells grown under biofilm-inducing conditions and tested for SinR levels at three different concentrations of protein. SinR levels were found to be 1.8 ± 0.07- and 1.7 ± 0.3-fold higher, respectively, in mutants lacking YlbF and YmcA than in the wild type. In contrast, SinR levels were unaltered in a *spo0A* mutant, a result that indicates that the phosphorelay does not influence SinR levels (Figure 9A). For each strain the quantification of the SinR band was normalized to the quantification of the SigA band, used as a loading control. As a control, the band identified as SinR was absent from a lysate of a mutant lacking *sinR* (Figure 9D).

3.2. The increase of SinR protein is sufficient to block biofilm formation.

It is known that even a single extra copy of *sinR* is sufficient to block biofilm formation (Chai et al., 2011). Also, previous work has shown that under biofilm-inducing conditions, four of the six synonymous serine codons (TCN), henceforth called serine-sensitive codons, allow for less efficient translation of the *sinR* mRNA than the other two (AGC or AGT), henceforth called serine-insensitive codons. A strain in which several of the serine-sensitive codons were switched to serine-insensitive codons was found to have slightly higher levels of SinR (approximately 1.6-fold higher) and to be completely defective in biofilm formation, representing independent evidence that biofilm formation is hypersensitive to SinR levels (see Appendix B) (Subramaniam et al., 2013).

To show that the increase in SinR levels in the cells lacking YlbF and YmcA was sufficient to block biofilm formation, the level of SinR was compared between the mutants and a strain harboring mutations causing synonymous switches to serine-insensitive codons (TCA > AGT) in *sinR* that cause a

complete block in biofilm formation (see appendix). The increase in SinR levels in cells lacking YlbF and YmcA was comparable to the increase seen with the synonymous mutations, and the biofilm phenotype defect (flat, unwrinkled colonies that do not spread as much as the wild type) was approximately the same in all cases (Figure 9C, 9D) (Subramaniam et al., 2013).

Figure 9. Mutants of the complex have heightened levels of SinR protein.

A. Comparison of SinR levels by Western blotting using an antibody against SinR. Shown is a comparison of levels of the wild type (wt) (RL3852) with that of *ylbF* (RL4168), *ymcA* (RL4169), and *spo0A* (RL4620) mutants in liquid shaking MSgg during stationary phase. The levels are compared in a dilution series as indicated. Note that the signal intensity for SinR varied from blot to blot, but in each case the wild type and the mutant were included on the same blot. The ratio of SinR/SigA signal was then normalized to wild type for quantification.

B. The densitometry of SinR bands was normalized against the densitometry of σA bands in panel A. The ratio of SinR/σA was then normalized to wild type (average fold change plotted for three biological replicates).

C. Colony morphology phenotypes of wt (3610) and *ylbF* (RL4168), *ymcA* (RL4169), and *sinR* (TCA > AGT) (YC923) mutants grown for 72 hrs on MSgg 1.5% agar plates. (Matrix production is indicated by wrinkly morphology and colony spreading. All images are at the same scale).

D. Comparison of SinR levels between wt (RL3852) and *ylbF* (RL4168), *ymcA* (RL4169) and *sinR*(TCA > AGT) (YC923) mutants using an antibody against SinR.

E. Colony morphology phenotypes of the wild strain (3610), a mutant of *ylbF* (RL4168), a mutant of *sinR* with increased number of serine-sensitive codons *sinR*(AGT > TCA) (YC1173), and double mutant of *ylbF sinR*(AGT > TCA) (AJD205) grown for 72 h on MSgg 1.5% agar (all images are at the same scale). For Panels A and E, a Western blot against σA was used as a loading control.

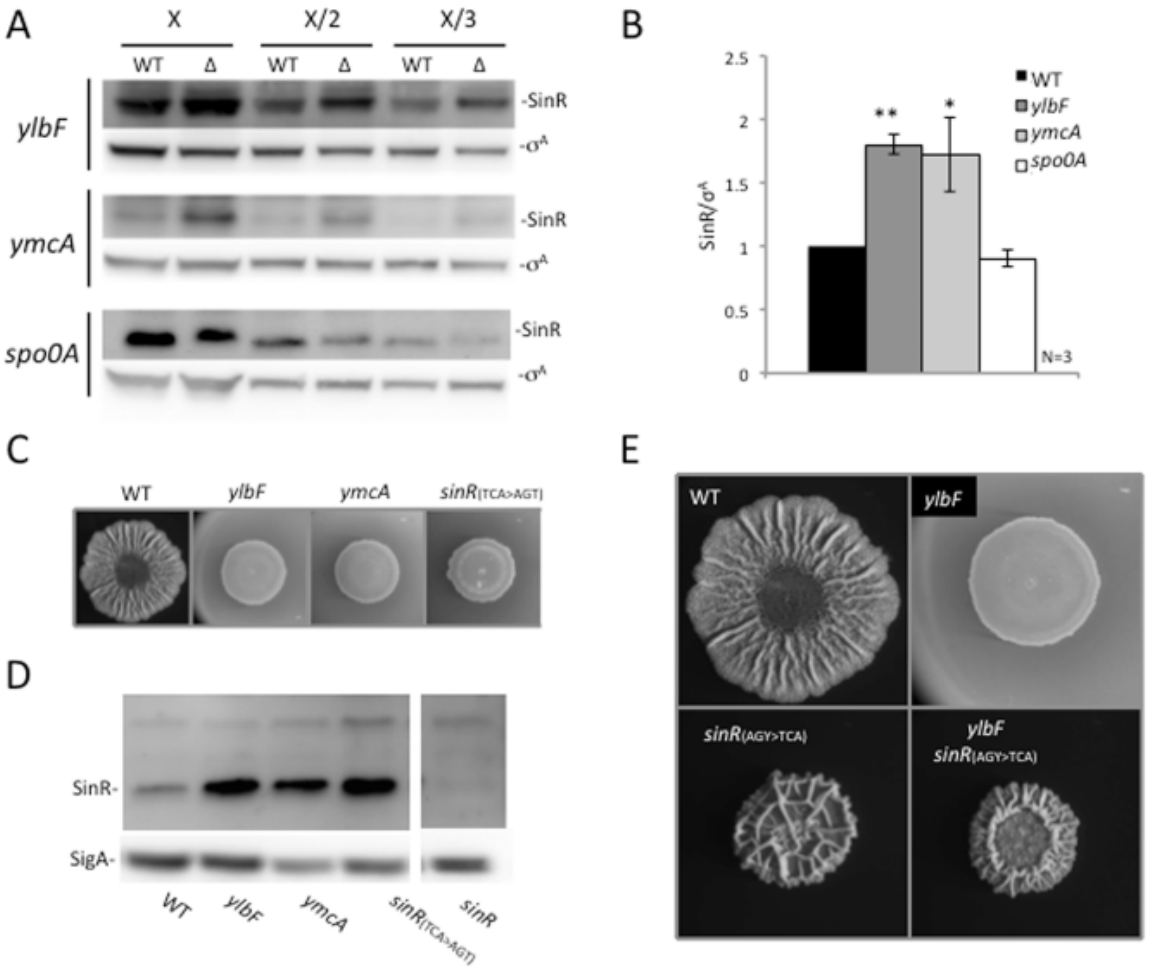

Figure 9 continued.

3.3. A small decrease in SinR levels can restore biofilm formation in mutants lacking YlbF.

As a further test of whether this increased level of SinR is be responsible for the block in biofilm formation, I again turned to the effect of switching serine codons to see if I could restore biofilm formation to mutants of *ylbF*. This time I took advantage of a strain in which serine-insensitive codons in *sinR* were switched to serine-sensitive codons, which was known to result in a small decrease in SinR levels and the production of hyper-wrinkled biofilms (Subramaniam et al., 2013). I therefore introduced the serine-sensitive-codon-containing gene into a YlbF mutant. The results show that the mutant *sinR* gene was able to suppress the biofilm phenotype caused by the *ylbF* mutation, restoring a wrinkled colony phenotype (Figure 9E).

3.4. *sinR* mRNA levels are higher in *ylbF* and *ymcA* mutants

As I have shown, the levels of SinR are higher in cells lacking YlbF and YmcA during stationary phase. One explanation could be an increase in the transcription of *sinR*, which is thought to be transcribed constitutively. To investigate whether the mutations were increasing the transcription of *sinR* I measured expression using a fusion of the promoter immediately upstream of *sinR* to *lacZ*. The results confirmed that *sinR* is transcribed constitutively, revealing little or no effect of the *ylbF* and *ymcA* mutations on expression levels (Figure 10A).

Next, I investigated whether the mutations were influencing the level of the *sinR* mRNA by carrying out Northern blot analysis using a probe against the *sinR* message. The gene for *sinR* is transcribed from three promoters, resulting in three differently sized transcripts (Figure 10B) (Lehnik-Habrink *et a)*. The predominant transcript during exponential phase growth originates from a promoter just upstream of *sinR*. This 400-base transcript contains the *sinR* coding sequence, but not the coding sequence for the adjacent, upstream *sinI* gene. The second transcript, which is 700-base, contains both *sinR* and *sinI* coding regions and originates from a promoter under the control of Spo0A~P just upstream of *sinI* (Lehnik-Habrink *et al*., 2011b). The third transcript originates from within the gene (*yqhG*)

upstream of *sinI* resulting in a transcript of 1.2 kb. This transcript is thought to be constitutively expressed at a low level and is present during both exponential growth and stationary phase. I found that the level of the *sinR*-only transcript was increased in cells lacking YlbF or YmcA during stationary phase, approximately 2.8-fold in the case of YlbF (Figure 10C, 10D). As further control, I again turned to the use of the mutant *sinR* in which serine-sensitive codons were changed to serine-insensitive codons [*sinR*(AGY > TCA)]. Given that the effect of synonymous serine codons is exerted by affecting translation, we expected that the levels of *sinR* mRNA in the synonymous mutant would be unaltered, and this is the case (Figure 10E)

Figure 10. Mutants of the complex have heightened levels of *sinR* mRNA.

A. Comparison of β-galactosidase synthesis under the control of the *sinR* promoter (P_{sinR}–*lacZ*) in liquid shaking MSgg between the wild type (wt) (AJD30) and mutants of *ylbF* (AJD31) and *ymcA* (AJD32) (average plotted for two biological replicates).

B. Schematic depicting the transcription of the *sinI*/*sinR* operon and a Northern blot with a probe against *sinR* mRNA with RNA samples from exponential and stationary phase cells in MSgg medium.

C. Northern blot with a probe against the *sinR* mRNA of samples collected in stationary phase in MSgg medium for the wild type (wt) (RL3852) as well as for mutants of *ylbF* (RL4168) and *ymcA* (RL4169). Also shown is quantification of the densitometry of the *sinR* mRNA normalized to wild type (average plotted for three biological replicates).

D. Northern blot with a probe against the *sinR* mRNA of samples collected in stationary phase in MSgg medium for wt (RL3852) and *sinR*(TCA > AGT) (YC923). Also in included in Panels C and D as loading controls are the two predominant rRNAs (16S and 23S) as visualized by ethidium bromide staining. RNA samples for the Northern blots in C and D were loaded at the indicated dilutions.

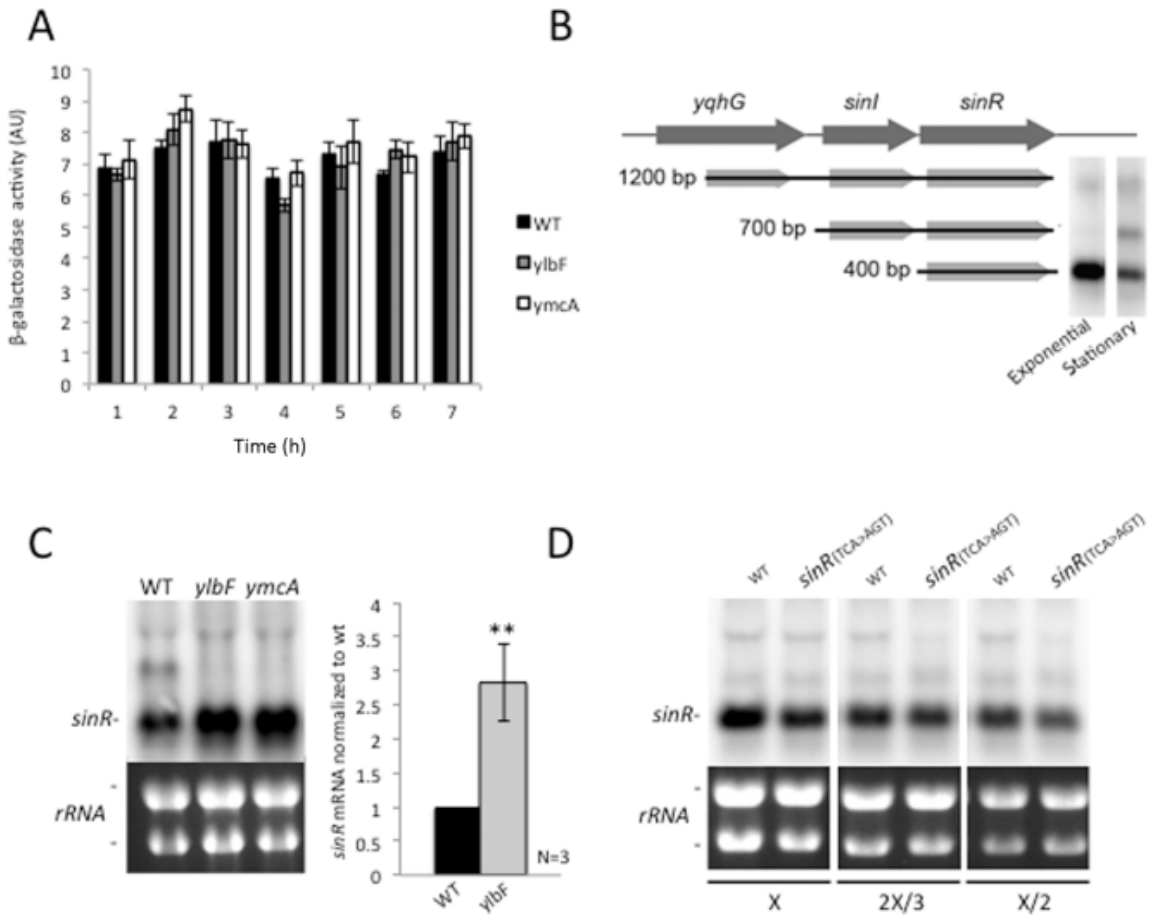

Figure 10 continued.

3.5. How does the absence of YlbF and YmcA cause *sinR* mRNA levels to increase?

We and Dubnau and colleagues (Carabetta et al., 2013) have independently carried out pull-down experiments to identify binding partners for YlbF and YmcA. I used a pull-down with YlbFs, which identified YmcA and YaaT as binding partners (Table 1). Carabetta *et al*. (2013) similarly obtained compelling evidence that YlbF, YmcA and YaaT are in a complex, carrying out their pull-down experiments with all three proteins and carrying out size exclusion chromatography. Interestingly, however, these pull-down experiments revealed several additional proteins, one of which was the endoribonuclease RNase Y. RNase Y was seen in my pull-down experiment with YlbF (Table 1) and in the pull-down experiment with YaaT carried out by Carabetta *et al*. (2013; Table S3 in the supporting materials).

RNase Y provides an intriguing explanation for the increased levels of *sinR* mRNA in the mutants for several reasons. The first, and most importantly, is that RNase Y is known play a general role in mRNA turnover (Durand et al., 2012), but it has already been implicated in destabilizing the message for *sinR* in an unbiased analysis for RNAs whose levels increased in a strain depleted for RNase Y (Lehnik-Habrink *et al*., 2011b). The other reasons are in the following section.

3.6. The gene for RNase Y is near the gene for YmcA and the genes for Rny and YaaT show a striking pattern of co-conservation.

Further indication of a possible connection with RNase Y, is that the gene for RNase Y (*rny*) is near to the gene for YmcA (being separated by five genes), and this synteny is conserved among Gram-positive bacteria that have homologs of *ymcA* (Figure 11A).

Third, the genes for Rny and YaaT are highly co-conserved. Whereas all sequenced members of the *Bacillus* genus have *ylbF*, *ymcA*, *yaaT* and *rny*, only certain members of the distant genus *Actinomyces* show the presence of *rny* and *yaaT* while other members lack these genes. In such species *rny* shows a striking pattern of co-occurrence with *yaaT* (Figure 11B).

3.7. Mutants lacking RNase Y are defective for biofilm formation.

The gene for RNase Y (*rny*) had been believed to be essential and indeed the investigation of Lehnik-Habrink *et al*. (2011b) used a construct in which the levels of RNase Y could be depleted rather than using a mutation. In light of recent work showing that the gene for RNase Y is not required for viability (Figaro et al., 2013), I sought to investigate the effect of a null mutation in the biofilm-proficient, wild strain 3610. Also, because the gene for RNase Y is the upstream member of a two-gene operon, it was important to confirm that the observed effects were not due to a polar effect on the downstream member of the operon. A markerless deletion of the gene for RNase Y (kindly provided by Byoung Mo Koo and Carol Gross from whom we first learned that the gene is not needed for viability) was created in a 3610 background. The resulting strain was in fact completely defective for biofilm formation, producing small, flat colonies (Figure 12A). In addition, the mutation blocked expression of the SinR-repressed promoter for the matrix operon *eps* as judged using a *lacZ* reporter (Figure 12D). Importantly, biofilm formation was restored by the presence of a copy of the gene for RNase Y under the control of its native promoter at an ectopic locus (Figure 12A).

Next, I investigated the effect of the *rny* null mutation on *sinR* mRNA levels. I found that the mutation caused a similar increase in the levels of the *sinR*-only transcript to that seen in a *ylbF* or *ymcA* mutant (Figure 12B). To determine if the biofilm defect could be attributed to increased levels of SinR, we introduced an altered *sinR* enriched in serine-sensitive codons (as employed above for *ylbF*) into an RNase Y mutant. The mutant *sinR* was in fact capable of restoring biofilm formation to the *rny* mutant (Figure 12C).

Finally, I investigated whether the RNase Y mutation would mimic the growth defect exhibited by YlbF and YmcA mutants in LB medium. Indeed, in a 3610 background cells lacking RNase Y showed the same impaired growth in liquid LB as mutants lacking YlbF and YmcA (Figure 12 E).

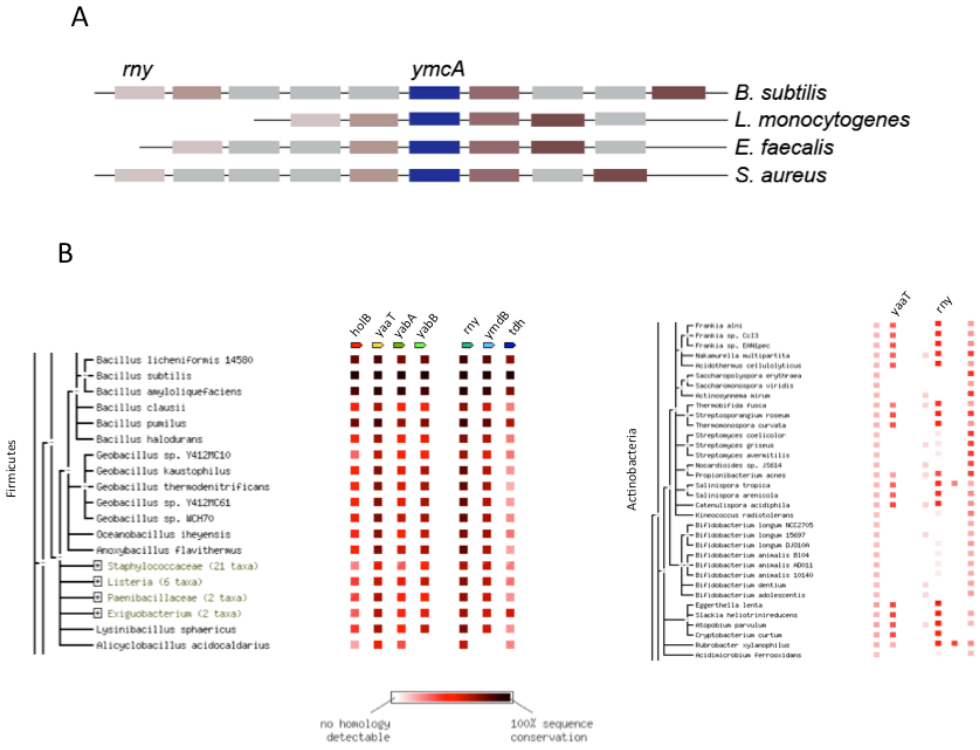

Figure 11.

A. Schematic showing *ylbF ymcA* (blue) and its homologs in *Listeria monocytogenes, Staphylococcus aureus*, *and Enterococcus faecalis* with their neighboring genes on the chromosome represented in a contrasting color.

B. Schematic showing the presence or absence of homologs to *B. subtilis yaaT* and *rny*, as well as their neighboring genes, amongst the Firmicutes (left) and the Actinobacteria (right) The presence of a homolog is indicated by a red box. The darkness of the red indicates the degree of conservation.

Figure 12. An RNase Y mutant phenocopies mutants of the complex.

A. Colony morphology phenotypes of wild type (WT) (3610), an *rny* mutant (AJD204), and an *rny* mutant bearing a complementation construct for *rny* (*amyE::*P_{rny}*-rny*) (AJD211) grown for 72 h on MSgg 1.5% agar plates.

B. Northern blot with a probe against *sinR* of samples collected in stationary phase in MSgg medium for wt (RL3852) and an *rny* mutant (AJD204). As a loading control, the two predominant rRNAs (16S and 23S) were visualized by ethidium bromide staining.

C. Biofilms produced by wt (3610), a mutant of *rny* (AJD204), a mutant of *sinR* with increased serine-sensitive codons *sinR*(AGT > TCA) (YC1173), and a double mutant an *rny* mutant combined with *sinR*(AGT > TCA) (AJD210) grown for 72 h on MSgg 1.5% agar (images at the same scale).

D. Synthesis of β-galactosidase under the control of the *epsA* promoter (P*epsA–lacZ*) in liquid shaking MSgg. Shown is a comparison of synthesis by the wt (◆) (RL4547) with that of *ylbF* (‣) (AJD6), *ymcA* (×) (AJD7), and *rny* (○) (AJD209) mutants. The lines depicting the activity in the *ylbF*, *ymcA*, and *rny* mutants are overlapping.

E. Growth curve in LB medium. Shown are the growth curves for the wild type (◆) (RL3852), and the mutants *ylbF* (●) (RL4168), *ymcA* (▲) (RL4169), and *rny* (■) (AJD204).

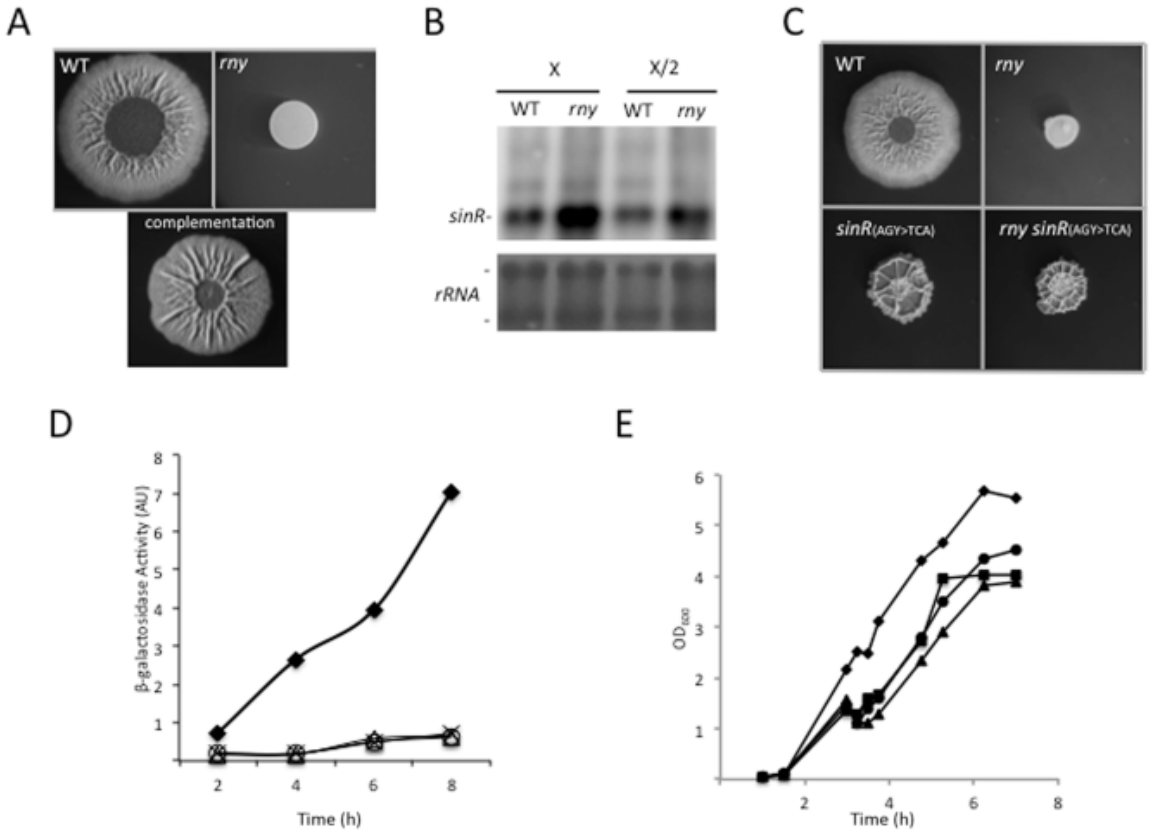

Figure 12 continued.

3.8. YlbF and YmcA interact directly and independently with RNase Y

To further investigate whether YlbF and YmcA interact directly with RNase Y, I carried out bacterial two-hybrid experiments using as 'bait' fusions of YlbF or YmcA to λ CI repressor and as 'prey' a fusion of the cytosolic domain of RNase Y to the N-terminal domain of the α subunit of RNA polymerase (Deighan, Diez, Leibman, Hochschild, & Nickels, 2008). I used the cytosolic domain because RNase Y is an integral membrane protein with a 25-residue long transmembrane domain at its N-terminus, and I reasoned that if YlbF and YmcA interact with RNase Y, it would likely have to be with the cytosolic domain of the endoribonuclease. As a positive control, I used the β-flap of RNA polymerase as bait and σ70 as prey (Deaconescu et al., 2006; Deighan et al., 2008; Kuznedelov et al., 2002). As negative controls I used either λ CI without a bait fused to it tested against the indicated prey proteins or α without a fused prey tested against the indicated bait proteins. As further negative controls, I used λ CI fused to β-flap tested against RNase Y or α fused with σ70 tested against Rny, YlbF and YmcA. The results show that both YlbF and YmcA interacted with the cytosolic domain of RNase Y, with the YlbF interaction being particularly robust (Figure 14). I also observed strong interaction of YlbF with YmcA, in keeping with the finding that these proteins interact with each other, and of RNase Y with itself. Finally, I asked whether we could detect an interaction of YlbF or YmcA with the phosphorelay protein Spo0B, as previously reported (Carabetta et al., 2013). The results failed to reveal such an interaction (Figure 13 and results not shown).

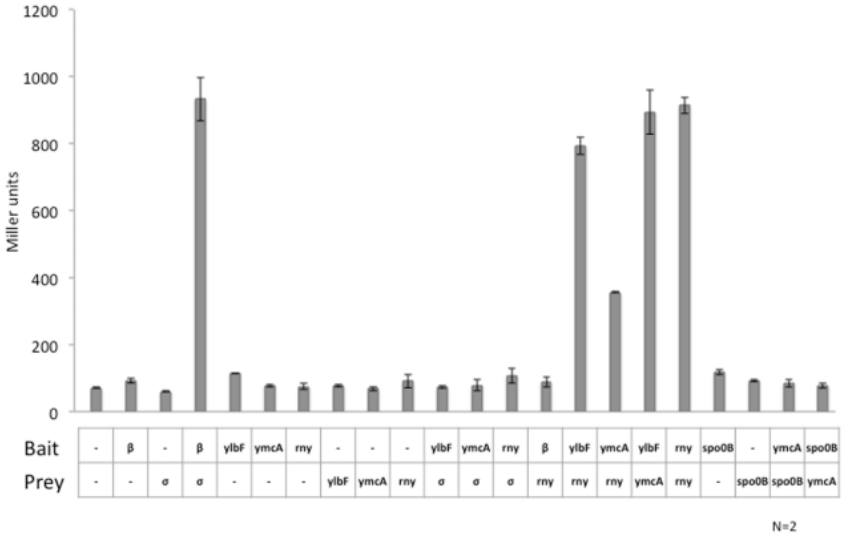

Figure 13. Bacterial two-hybrid assays show that YlbF and YmcA interact directly and independently with RNase Y. The β-galactosidase activity for the indicated pairs of proteins tested in the reporter strain *Escherichia coli* FW102 OL2–62 in which *lacZ* expression is dependent on a direct interaction (average plotted for 2 biological replicates). The strains tested are AJD221-241, for details see supporting information.

3.9. Cleavage of a well-studied target of RNase Y depends on YlbF and YmcA

The regulation of biofilm formation through *sinR* is complicated and affected by many regulatory pathways. To provide strong evidence that the complex of YlbF, YmcA, and YaaT affect the activity RNase Y, I wanted to determine if the RNase Y activity towards a target completely unrelated to biofilm formation, the well studied cleavage of the *cggR-gapA* transcript. The central glycolytic gene repressor *cggR* is co-transcribed with the adjacent downstream gene *gapA* to yield a 2.2 kb, *cggR-gapA* transcript (as well as longer transcript extending further downstream). Previous studies have shown that RNase Y cleaves the 2.2-kb transcript between *cggR* and *gapA*, producing a 1.0-kb transcript that contains most of *cggR* and a 1.2-kb transcript containing *gapA* (Figure 14A) (Commichau et al., 2009; Ludwig et al., 2001). If, as I posit, the activity of RNase Y depends on YlbF and YmcA, then little of the 1.0 kb transcript cleavage should be seen in mutants lacking either protein. Using probes against *cggR* and *gapA* I indeed found that the 2.2-kb *cggR-gapA* transcript is processed into 1.0- and 1.2-kb transcripts in the wild type, and that this cleavage was strongly dependent on YlbF and YmcA (Figure 14B). Thus, a minimum of two previously described targets of RNase Y (*sinR* and *cggR-gapA*) depend on YlbF and YmcA for their cleavage.

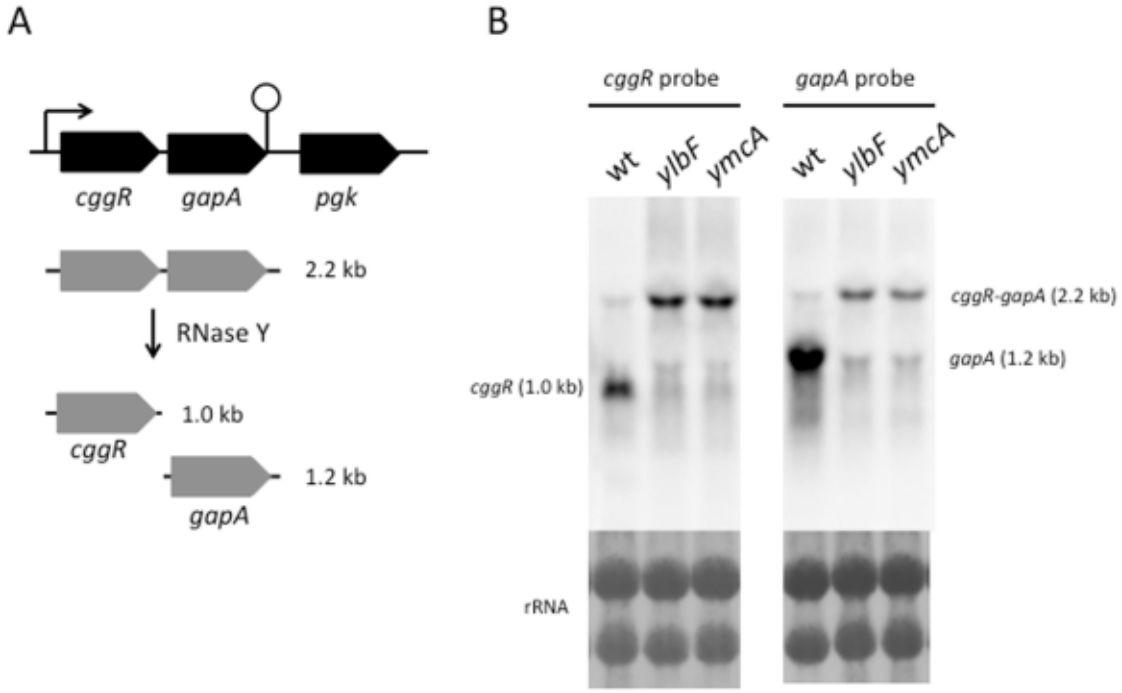

Figure 14. The cleavage of *cggR-gapA* mRNA by RNase Y is dependent on YlbF and YmcA.

A. Schematic depicting the transcription of the *cggR-gapA* operon and transcripts resulting from the cleavage by RNase Y.

B. Northern blot with a probe against the *cggR* mRNA and against *gapA* mRNA of samples collected in stationary phase in MSgg medium for wt (RL3852), and the mutants of *ylbF* (RL4168) and *ymcA*(RL4169). The rRNA bands were visualized by methylene blue staining and imaged with a Sony digital camera.

Discussion

A principal contribution of this investigation is evidence that the dependence of biofilm formation on the YlbF-YmcA-YaaT complex occurs at the level of *sinR* mRNA. I have shown that mutants of the three proteins cause an increase in the abundance of SinR, that this increase is sufficient to explain the block in biofilm formation, and that this increase is due to an increase in *sinR* mRNA.

I have also obtained compelling evidence that the complex controls the levels of *sinR* mRNA through a functional and direct interaction with the ribonuclease RNase Y, which is known to destabilize the *sinR* mRNA (Lehnik-Habrink *et al*., 2011b). This conclusion is supported by the following: First, an *rny* mutant mimics the phenotypes of mutants of the complex, including causing a block in biofilm formation and a growth defect in a rich medium. Also, like mutants of *ylbF* and *ymcA*, a mutant of *rny* has been reported to be impaired for genetic competence and sporulation (Figaro et al., 2013). Second, RNase Y was found to be associated with the complex in independent pull-down experiments carried out using YlbF in the present work and by Carabetta *et al*. (2013) using YaaT. Third, RNase Y and YaaT have similar patterns of subcellular localization. RNase Y is known to be a membrane protein (it has a transmembrane segment at its N-terminus) and it has been reported to localize to the periphery of the cell and to form foci at the division septum (Lehnik-Habrink *et al*., 2011a; Burmann *et al*., 2012)(Burmann, Sawant, & Bramkamp, 2012). Meanwhile, YaaT (which does not have a predicted transmembrane domain) has also been reported to localize to the cell periphery and to the division septum (Hosoya et al., 2002). Fourth, and importantly, bacterial two-hybrid experiments revealed that YlbF and YmcA interact directly and independently with RNase Y. Finally, the processing of a well characterized target of RNase Y, the *cggR-gapA* operon, was found to be dependent on YlbF and YmcA. This is a striking result as the processing of the *cggR-gapA* mRNA is independent of both biofilm formation and the phosphorelay.

In light of these considerations, I propose that YlbF, YmcA and YaaT be renamed RcsA, RcsB and RcsC, respectively, for RNase Y-containing complex subunits A, B and C (Figure 15).

Strictly speaking, we cannot rule out the possibility that the RcsA-RcsB-RcsC (YlbF-YmcA-YaaT) complex both stimulates flux through the phosphorelay and independently causes degradation of mRNAs for SinR and other proteins (such as proteins involved in growth in rich medium). However, a more parsimonious explanation for the results reported by Carabetta *et al*. (2013) is that under certain conditions or in certain genetic backgrounds *rcsA* (*ylbF*), *rcsB* (*ymcA*) and *rcsC* (*yaaT*), mutations raise the levels of mRNAs for metabolic enzymes that influence the activity of the histidine kinases that feed phosphoryl groups into the relay or, conceivably, of negative regulators of the phosphorelay, such as Spo0E, MecA, or the Rap phosphatases.

My work showing that RcsA, RcsB and RcsC function via an interaction with RNase Y raises two important questions. One is whether the complex influences the specificity of RNase Y. RNase Y is known to play a general role in mRNA turnover (Shahbabian *et al*., 2009; Lehnik-Habrink *et al*., 2011b; Durand *et al*., 2012), but also preferentially destabilizes certain mRNAs such as that for *sinR* (Lehnik-Habrink *et al*., 2011b). It will therefore be interesting to carry out RNA-seq experiments with mutants of each of the complex proteins as well as with a mutant lacking RNase Y to identify additional candidates for substrates and to investigate the possibility that a subset of RNase Y targets depend on only one or more of the proteins.

A second question raised by our work is whether the activity of the complex is influenced by physiological or environmental conditions. For example, might the capacity of the complex to degrade *sinR* mRNA be dependent on growth and/or medium conditions? That is, is *sinR* mRNA less stable under biofilm-inducing conditions than during exponential phase growth or in medium that does not support biofilm formation? If so, then the RNase Y-RcsA-RcsB-RcsC complex might serve as an environmental or metabolic sensor that operates at the level of mRNA stability.

SinR is emerging as a hub for at least three inputs in the decision to form a biofilm, all operating at a post-transcriptional level. First, SinR is subject to antagonism by the anti-repressor SinI, whose synthesis is under the control of Spo0A∼P via the phosphorelay (Kearns et al., 2005). Second, *sinR* mRNA is subject to a serine sensing mechanism mediated by serine-sensitive codons, which respond to

the metabolic state of the cell (Subramaniam et al., 2013). Third, as I have now seen, *sinR* mRNA is subject to degradation by a multi-protein complex and conceivably, as we hypothesize earlier, this complex is itself responding to environmental conditions. In this view the formation of a multicellular community is a pivotal decision for *B. subtilis* that is based on the integration of multiple environmental and metabolic inputs that are channeled into the synthesis or activity of SinR.

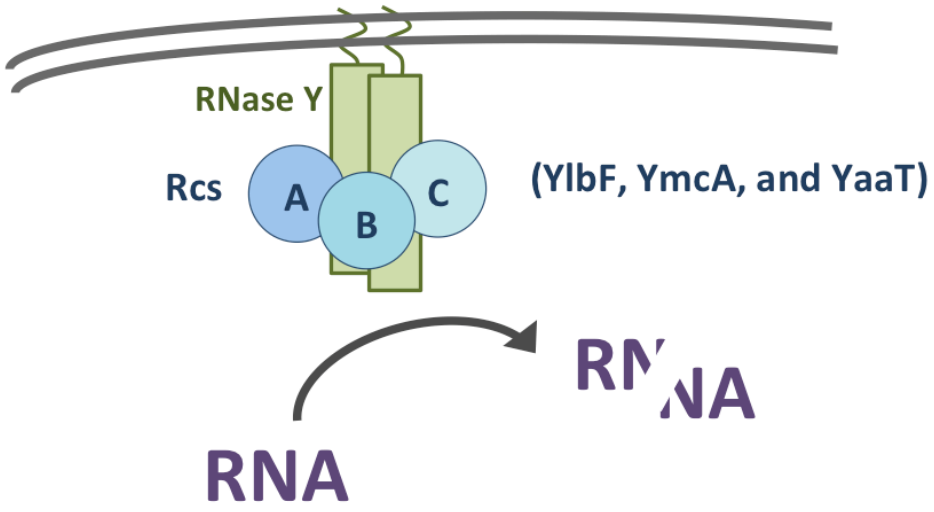

Figure 15. A model for the function of the complex of YlbF, YmcA, and YaaT, supported by the evidence presented here, in which the complex interacts with the ribonuclease RNase Y and affects the cleavage of its RNA targets.

Chapter 4

RNA sequencing reveals a wider role for YlbF in RNA turnover and processing.

After learning that YlbF interacts with RNase Y directly, and that YlbF affects at least two targets of RNase Y, *sinR* and *cggR-gapA*, I became interested in answering two follow up questions. 1. What targets of RNase Y are dependent on YlbF, YmcA, and YaaT? 2. Are YlbF, YmcA, and YaaT regulating the activity of RNase Y under specific conditions? RNA sequencing experiments provide the global information that is needed to answer these questions. RNA-seq may also provide insight into the other phenotypes of mutants of the complex of YlbF, YmcA, and YaaT, such as the growth defect. In collaboration with the laboratory of professor Gene-Wei Li at MIT, I have started to carry out these experiments. My preliminary experiments were not carried out with all of the mutants and the sequencing was not at a sufficient depth to fully answer these questions, but the experiments do provide extremely strong evidence that YlbF is in fact acting with RNase Y. The data also provide novel and interesting information about YlbF, YmcA, and YaaT.

Contributions: This work is unpublished. The technique for mapping transcript ends using RNA-sequencing was developed by Gene-Wei Li. Sequencing library preparations were carried out by myself and Ariel Schieler. Darren Parker contributed significantly to the data analysis.

4.1. RNase Y and RNA turnover in *B. subtilis*.

In bacteria, the levels of RNAs are tightly regulated to meet the needs of the cell. The over abundance of an mRNA can cause the waste of resources by the cell, both the nucleotides of the RNA and the translational machinery and precursors used in translation. Moreover, the over abundance of specific gene products can be toxic to the cell. Microbiologist's understanding of RNA turnover has largely come from studies in *E. coli*, but much of how bacteria carryout RNA turnover remains a mystery. In *E. coli* a majority of RNA turnover is carried out by the essential ribonuclease RNase E, which *B. subtilis* lacks.

In *B. subtilis* there are at least 18 different ribonucleases. The most important of these for mRNA turnover and processing are likely RNase J1, which is essential, RNase Y, and PNPase. RNase J1 is a processive exonuclease that degrades RNAs from the 5' to the 3'. PNPase is a processive exonuclease that degrades RNAs from the 3' end to the 5' end. RNase Y is an endonuclease that likely plays a role in destabilizing RNAs by initiating degradation. RNase Y has been shown to play a role in general mRNA turnover, but is also known to preferentially target specific mRNAs. RNase Y also plays a role in the processing of functional RNAs including the RNA component of RNase P, scRNA, and rRNA.

4.2. Experimental design and a novel approach to RNA sequencing

RNA sequencing has become a powerful tool to determine RNA levels in cells. Gene-Wei Li's group has been developing a method to accurately map the 5' and 3' ends of transcripts with single base pair resolution. In short, purified RNAs from cells are fragmented using chemical fragmentation. The fragmented RNA is then separated by PAGE, and the small fragments, ~15-45 bp, are selected for. Given the degree of fragmentation and the small size selection, fragments that were originally either the 5' end or 3' end are highly enriched. The fragments are then sequenced in a way that allows for the accurate mapping of the fragment ends (see methods). The sequenced ends are then mapped to the genome for visualization. While this method requires more starting material, it provides useful information in the context of RNA processing and degradation .

My first RNA-seq experiment was carried out on samples from wild type and samples from mutants lacking YlbF during exponential growth and early stationary phase in the biofilm medium, MSgg. The sequencing libraries were prepared by Ariel Schieler, in Gene-Wei Li's lab, and were sequenced at the MIT BioMicroCenter. For the initial experiment four samples were multiplexed with several unrelated samples for sequencing. These samples come from wild type 3610 and a mutant lacking YlbF during exponential growth and early stationary phase in liquid MSgg. The YlbF samples have not been sequenced more deeply. In addition, a sequencing experiment with wild type and a mutant lacking YmcA in early stationary phase (OD_{600} of 2.0) and late stationary phase (OD_{600} 4.0) has been carried out. The sequencing for this experiment was deeper allowing for better quantification. The results presented are largely from the YmcA experiment but are reproduced from the YlbF experiment.

4.3. Mutation of components of the YlbF-YmcA-YaaT-Rny complex causes increased mRNA levels of other components. YaaT is expressed on a previously unappreciated transcript with *yabA*.

One striking finding from the RNA-seq experiment is that the levels for some of the transcripts that encode the YlbF-YmcA-YaaT complex are affected by the loss of YlbF or YmcA. For instance, the transcript for *yaaT* increases in the absence of either YlbF or YmcA and the transcript for *ylbF* increases in the absence of YmcA (Table 2). The transcript for *ymcA* is not significantly altered in the absence of YlbF.

While the gene for YaaT is known to be a part of a larger operon, these results reveal a novel transcript that includes just *yaaT* and *yabA* (figure 16). The transcript end mapping provides, with nucleotide resolution, the identity of the 5' and 3' ends of this transcript. Understanding how this transcript arises and what regulates its levels could provide insight into the regulation of the YlbF-YmcA-YaaT complex.

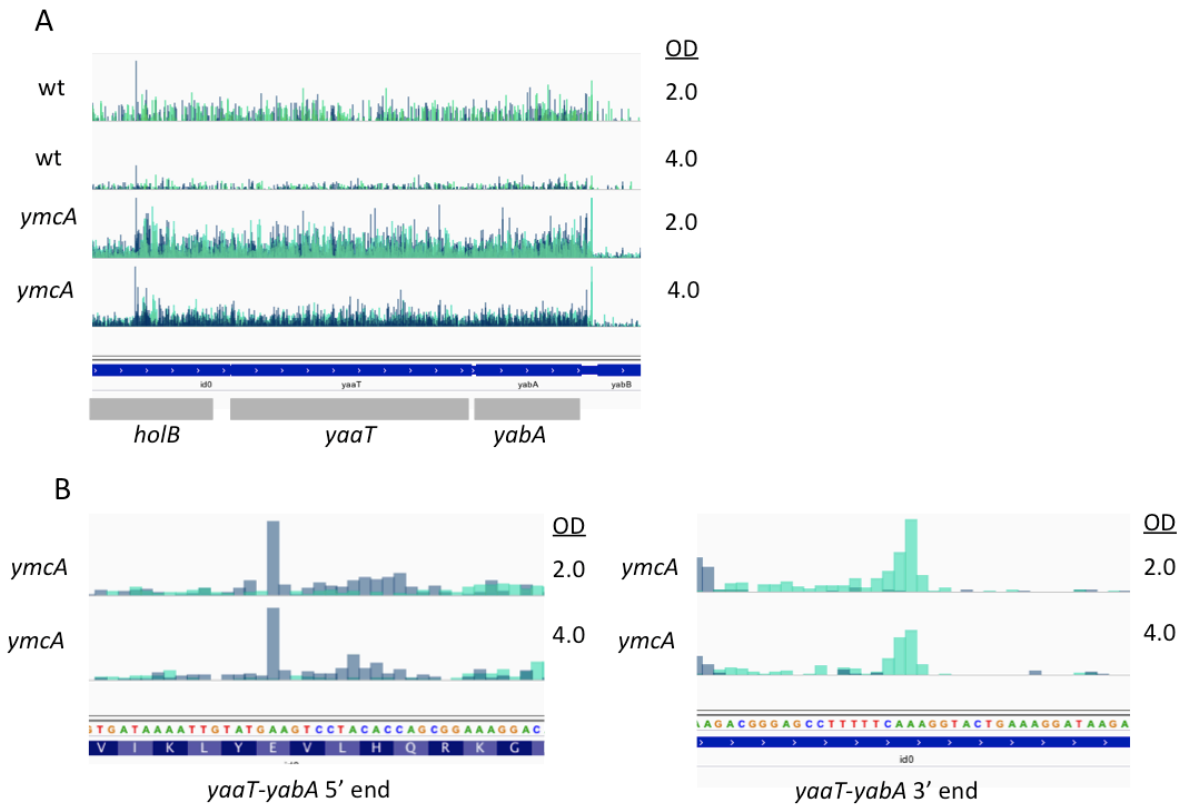

Figure 16. A novel *yaaT-yabA* transcript.

A. A genome view of the genes *holB, yaaT*, and *yabA*, in which the 5' and 3' ends of RNA-sequencing fragments are mapped at each base pair for samples from wild type and a mutant of *ymcA* grown in liquid MSgg to the indicated OD_{600}. The darker blue peaks represent mapped 5' end and the lighter teal peaks represent mapped 3' ends. The data range for all tracks is the same at a maximum of 150.

B. Zoomed in view of the same data as A, to show the precise 5' end and 3'ends of this novel *yaaT-yabA* transcript. The transcript ends are mapped to specific nucleotides.

4.4. Some known targets of RNase Y are affected by YlbF and others are not.

I hypothesize that YlbF, YmcA, and YaaT only affect a subset of the targets of RNase Y, because mutants lacking RNase Y are much more sick than mutants of the complex. A place to start would be to show that some known targets of RNase Y are affected and some other targets are not. RNA-sequencing has not been carried out with a mutant of *rny* in the 3610 background under biofilm conditions. Due to this we do not know the extent of the RNAs targeted by RNase Y, but we can investigate the affect on previously reported targets of RNase Y.

Some transcripts that were significantly increased in the RNase Y depletion experiments were also hyper abundant in mutants lacking YlbF. This includes the *trpC-E* operon, *glmS*, *speD*, and others (Table 3). One transcript that was more abundant under *rny* depletion but not in a *ymcA* mutant is *spx* and one stable cleavage product by RNase Y that is not dependent on YlbF or YmcA is the RNA component of PNPase *rph* (table 2).

The RNA-seq experiment also confirmed that the *cggR-gapA* operon is not processed in mutants lacking YlbF or YmcA (Figure 17A), in accordance with the northern blot analysis in chapter 3. Unlike the northern blot, the RNA-seq allows us to identify the exact cleavage site that is reduced in mutants lacking YlbF. The cleavage site is indeed the known cleavage site of RNase Y, which has previously been identified by primer extension experiments (Figure 17B and C).

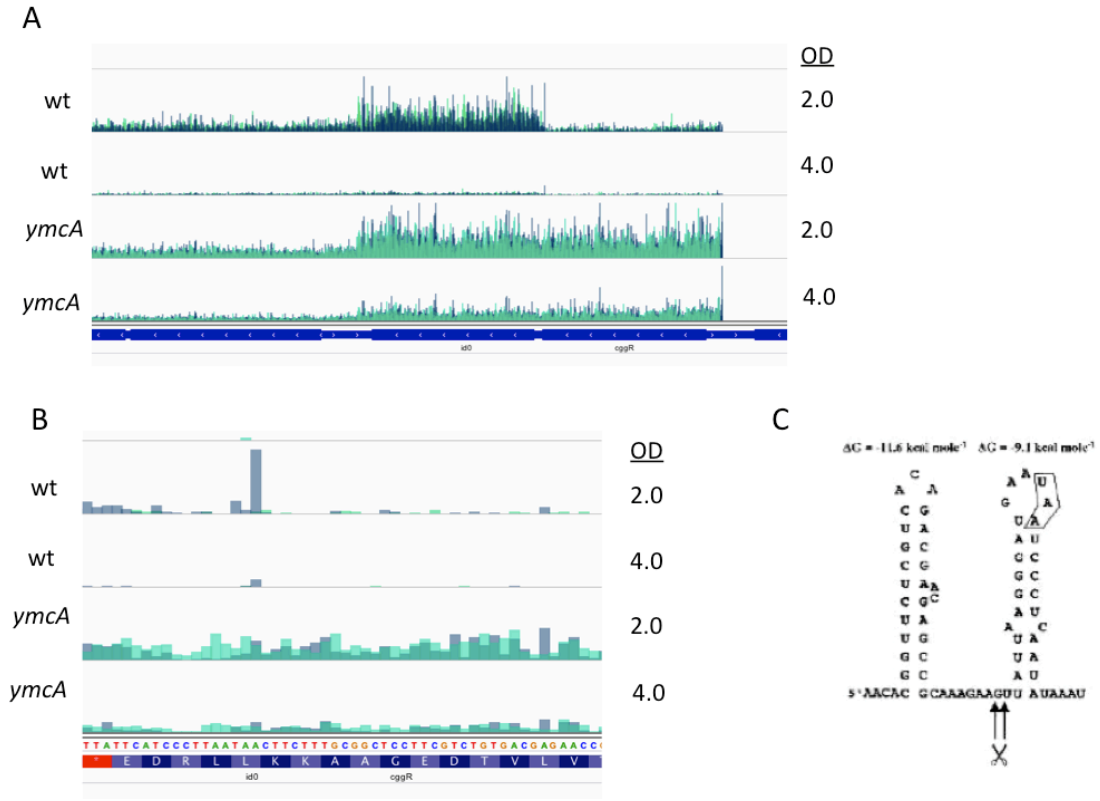

Figure 17. Processing of the *cggR-gapA* transcript analyzed by RNA-seq.

A. A genome view of the genes *cggR and gapA* in which the 5' and 3' ends of RNA-sequencing fragments are mapped at each base pair for samples from wild type and a mutant of *ymcA* grown in liquid MSgg to the indicated OD_{600}. The darker blue peaks represent mapped 5' end and the lighter teal peaks represent mapped 3' ends. The data range for all tracks is the same at a maximum of 250.

B. Zoomed in view of the same data as A, to show the precise 5' end of the transcript produced by cleavage with RNase Y.

C. Figure from a paper that identified the precise cleavage location of the *cggR-gapA* operon(Ludwig et al., 2001).

Table 2. Change in abundance of mRNA in mutants lacking *ymcA* under biofilm conditions. Some of the genes that are increased significantly in cells lacking YmcA are shown. The first ratio given is the change in read counts between a mutant of *ymcA* and wild type of the number of reads that map to that gene normalized to the number of reads that map to *sigA* for each sample. The second ratio, comes from Lehnik-Habrink et al 2011 and shows the ratio of RNA abundance between cells that have been depleted for RNase Y and wild type cells.

Genes that increase in this work (ratio)

	This work *ΔymcA*	Lehnik-Habrink et al RNase Y depletion
trpE	10.8	13.1
trpB	9.6	9.9
trpF	26.2	11.4
trpC	27.2	12.3
trpD	23.3	12.7
trpA	1.9	2.8
speD	2.4	3
cggR	5.1	2.5
hisC	1.7	2.5

unchanged in this work

	This work *ΔymcA*	Lehnik-Habrink et al RNase Y depletion
spx	1.1	2.5

4.5. Turnover of many riboswitches depends on YlbF.

Riboswitches are regulatory elements associated with mRNAs, generally in the 5' untranslated region. They fold into a secondary structure that then regulates either transcription or translation of that mRNA. In *B. subtilis* most of the riboswitches regulate transcription by either allowing RNAP to transcribe through the riboswitch, to the downstream genes, or by causing termination of transcription at the riboswitch. Whether the riboswitch allows transcription or not is generally determined by the riboswitch binding to a small molecule or tRNA. Under conditions that promote termination, this system rapidly produces large amounts of the prematurely terminated transcript, from the transcription start site to the riboswitch, that need to be degraded.

A specific example in *B. subtilis* is a class of riboswitches known as S-boxes. S-boxes bind to an important metabolite S-adenosylmethionine (SAM) and when bound by SAM cause transcriptional termination. *B. subtilis* has eleven known S-boxes. When SAM is present in the cell, transcription terminates at these S-boxes producing high levels of the prematurely-terminated transcripts, which then remain bound to SAM. Turnover of these transcripts is important for recycling the nucleotides of the transcript but also for releasing the metabolite SAM. It has been shown the RNase Y plays a major role in initiating the turnover of riboswitches in *B. subtilis*.

I was curious to see if riboswitch turnover is also dependent on YlbF and YmcA. The RNA-seq results show that many riboswitches are not turned over in mutants lacking YlbF or YmcA. Figure 18 shows the genome view of six highly affected riboswitches and the quantitative effect on these as well as others is shown in Table 3. Since not all of the riboswitches are expressed under this condition and since we don't have sequencing information from mutants of *rny* under these conditions, it is difficult to make global conclusions about the role of YlbF and YmcA towards riboswitch turnover, but these results indicate that they play a profound role. To highlight one case, the riboswitch for *thiC* is extremely abundant in mutants of *ymcA*. RNA-seq analyzes RNA levels at a population level, which makes it

difficult to determine the concentration for a single cell. However, the abundance of the *thiC* riboswitch is similar to the abundance of tRNAs in mutants lacking *ymcA* (not shown).

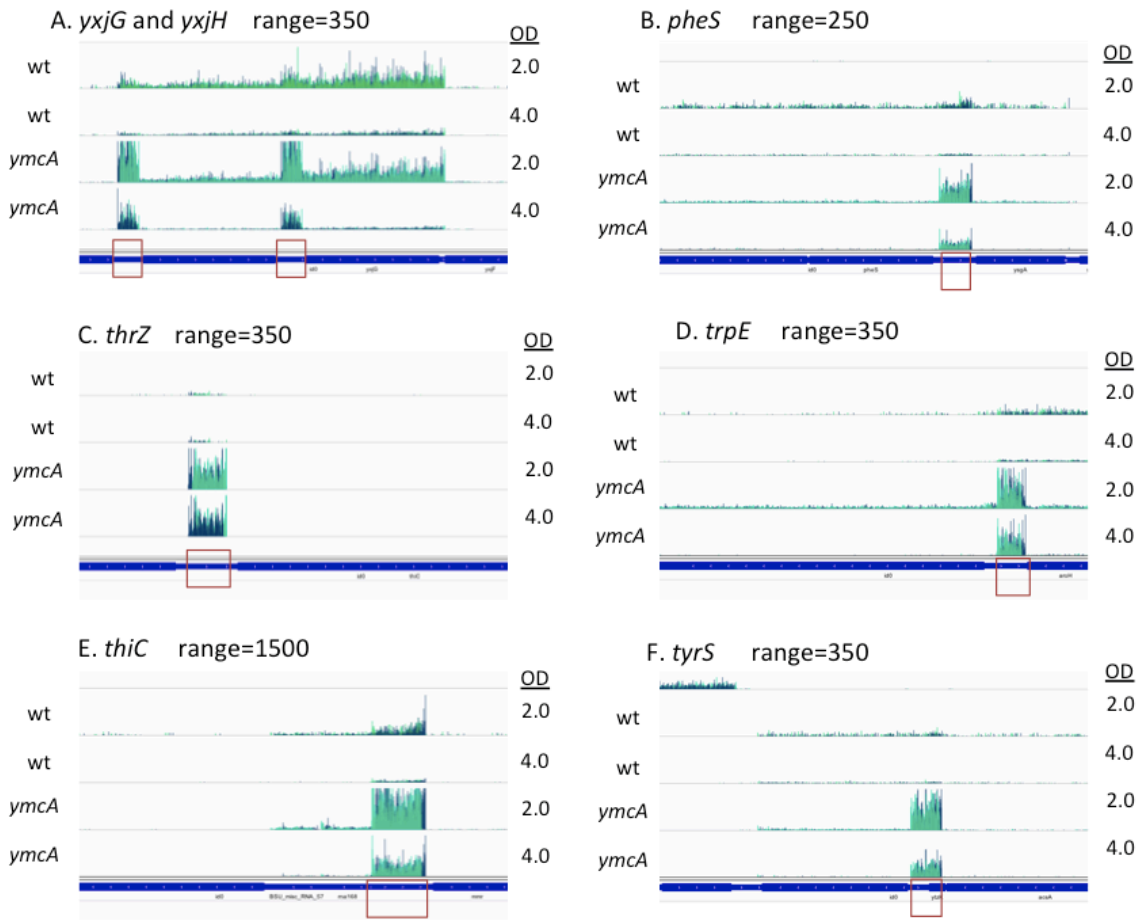

Figure 18. Riboswitches are hyper abundant in mutants lacking *ymcA*.

A genome view of the indicated genes in which the 5' and 3' ends of RNA-sequencing fragments are mapped at each base pair for samples from wild type and a mutant of *ymcA* grown in liquid MSgg to the indicated OD_{600}. The darker blue peaks represent mapped 5' end and the lighter teal peaks represent mapped 3' ends. The data range for all tracks is the same within a panel and set at the indicated maximum. The red boxes indicate the location of the riboswitch.

Table 3. Change in abundance of some riboswitches in mutants lacking *ymcA* under biofilm conditions that are expressed and significantly increase. The riboswitches are listed by type. The ratio given is the change in read counts for that riboswitch between a mutant of *ymcA* and wild type at two time points (an OD_{600} of 2.0 and an OD_{600} of 4.0). The number of reads that map to that riboswitch are normalized to the number of reads that map to *sigA* for each sample. The last four colums give the normalized number of read counts that mapped to each riboswitch from the described sample.

	ratio OD 2.0	ratio OD 4.0	#reads 2.0 wt	#reads 4.0 wt	#reads 2.0 ymcA	#reads 4.0 ymcA
S- box (bind SAM)						
metE	2.7	3.4	54,430	21,008	147,511	71,697
yjdJ	3.5	4.7	12,530	4,078	43,477	19,069
mtnk	2.2	2.5	14,807	6,101	32,736	14,970
mtnW	1.8	3.3	16,310	3,228	29,808	10,804
yoad	**5.1**	1.9	4,628	3,795	23,829	7,352
yxjG	4.2	**8.4**	9,079	1,688	37,908	14,167
yxjH	2	3.6	20,587	3,849	41,759	13,944
metI	0.6	0.8	18,686	8,792	11,581	7,172
T- box (bind tRNA)						
ilvB	1	3.95	81,146	11,668	85,701	46,069
ileS	3.7	4	10,190	5,278	38,204	21,122
pheS	2.74	**5.9**	6,962	1,441	19,131	8,559
thrS	2.7	3.8	3,818	1,032	10,241	3,903
thrZ	3.6	**7.7**	14,314	3,903	10,241	3,903
trpS	6.3	6.3	4,322	1,895	27,244	11,972
tyrS	**15.3**	**14.9**	2,844	1,384	43,558	20,601
TRAP						
trpE	**7.2**	**13.3**	2,776	1,026	19,753	13,656
gly- box (bind glycine						
gcvT	3.3	3.3	3,667	2,432	12,020	8,126
L- box (bind lysine)						
lysC	1.8	4.2	11,468	4,425	20,886	18,738
Thi- box (bind thiamine)						
tenA	**5.5**	4	2,452	5,278	13,495	21,122
thiC	**15.1**	**16.3**	7,135	5,822	107,712	95,077
bind manganese						
ykoy	3.6	4.2	6,914	3,505	24,564	14,783

Discussion

In bacteria, RNA, in particular mRNAs have very short half-lives. One implication of this is that bacteria are able to very quickly change patterns of gene expression under changing conditions. RNA turnover also plays a role in determining protein levels and in recycling nucleotides. In most cases it is the initiation event of RNA degradation that ultimately determines RNAs stability. In *B. subtilis*, cleavage of RNA by the endoribonuclease RNase Y is often that initiating event. In fact, once cleaved, the RNAs are generally degraded so quickly that the degradation intermediates cannot be detected by northern blotting or RNA-seq. Since RNase Y initiates the degradation of many RNAs, an understanding of the regulation of its activity and the control of its specificity would provide a lot of information about the regulation of RNA degradation in general. We now have shown that YlbF, YmcA, and YaaT are required for full activity of RNase Y, and RNA-seq could explain to what extent this complex affects RNase Y. Furthermore sequencing will allow for the identification of other targets of YlbF, YmcA, and YaaT. As we have shown, it is an effect on *sinR* that results in the biofilm defect of these mutants, but the specific cause of the growth defect has not been identified.

Most cleavage events by RNase Y result in the degradation of both RNA products, but there are a few examples where one of the RNA products is stable. A well-studied example of this is the *cggR-gapA* transcript in which RNase Y cleaves in between *cggR* and *gapA* leading to the degradation of the 5' end containing *cggR* but the stabilization of the 3' end containing *gapA*. This type of processing event allows for uncoupling the expression of various proteins encoded together in a polycistronic operon. The processing of the *cggR-gapA* operon is dependent on YlbF and YmcA. With deeper sequencing it will be interesting to see if we can identify additional and novel cleavage sites by RNase Y and to see if those are also affected by YlbF, YmcA, and YaaT.

One surprising finding of this study is that the transcript levels of *yaaT* are increased in the absence of YlbF or YmcA, and that the transcript for *ylbF* is increased in the absence of YmcA. A possible explanation for this is that there is some sort of feedback regulation of these genes. So in the

absence of one, the others are overexpressed in attempts to compensate. Understanding how reduced RNA turnover leads to an increase in the expression of genes for this complex could provide a lot of insight into the biology of these genes and how they are regulated. Based on microarray experiments under many conditions, it has not been seen that the levels of the mRNAs for these genes significantly change.

The role of YlbF, YmcA, and YaaT in the turnover of riboswitches was also a striking feature of this experiment. Since not all of the riboswitches are expressed under the conditions tested here, and we don't have sequencing information for mutants of *rny* under, it is difficult to make global conclusions about the role of YlbF and YmcA in riboswitch turnover, but these results show that they play a significant role. To highlight one case, the riboswitch for *thiC* is extremely abundant in mutants of *ymcA*. RNA-seq us used to analyze RNA levels for a population, which can make it difficult to understand the implications of the results for a single cell and or from a biological perspective. One way to consider the biological significance for the *thiC* riboswitch is to compare the total reads mapped to it with the reads mapped to the housekeeping sigma factor (*sigA*), normalized by length. In the sample from a mutant of *ymcA* at an OD_{600} of 2.0 there are ~17 reads per base pair of *sigA* and roughly 585 reads per base pair of the *thiC* riboswitch. The biological implications of the failure to turn over riboswitches have not really been studied. One reason their turnover is important is probably for the recycling of nucleotides, but there are probably more important reasons. One might be the fact that the riboswitches are still able to bind their ligands, if not degraded. For instance all of the S-box riboswitches that are not degraded can bind to SAM, an important and costly metabolite, and the T-box riboswitches can bind to both charged and uncharged tRNAs. The binding of these metabolites by riboswitches would likely interfere with a variety of cellular processes. The presence of the riboswitch could also bind ligand that would otherwise bind the same riboswitch as it is being transcribed, thus reducing the effectiveness of regulating transcription by riboswitches. Mutants of YlbF and YmcA provide a background for studying the biological implications of reduced ability to turnover riboswitches.

We are very interested in carrying out more RNA-sequencing experiments that will allow us to understand what RNAs are affected in the absence of RNase Y, and of those which are affected in the absence of YlbF, YmcA, and YaaT. These experiments may also shed light onto whether or not YlbF, YmcA, and YaaT are acting under specific conditions to regulate RNA turnover. Finally, these experiments could provide information towards understanding what RNase Y recognizes in RNA. Additional targets of RNase Y, plus the information of their specific cleavage sites, may help solve this problem. We are also interested in whether there is a specific sequence that correlates with RNase Y targets that are affected by YlbF, YmcA, and YaaT and those that are not.

Summary

The work presented here was motivated by a question: what is the role of YlbF, YmcA, and YaaT in biofilm formation in *B. subtilis*? I have uncovered a function for these proteins. The complex of YlbF, YmcA, and YaaT bind to the ribonuclease, RNase Y. This interaction is required for the full activity of RNase Y towards its role in RNA turnover. For biofilm formation, the message for *sinR*, a target of RNase Y, is more abundant in mutants of the complex. This over abundance of mRNA results in higher levels of SinR protein, which ultimately blocks biofilm formation. There have been two papers published since beginning this study that altered my work in the direction of this solution. The first was a paper from the laboratory of Jörg Stülke that identified *sinR* as a target of RNase Y(Lehnik-Habrink, Newman, et al., 2011). The second was the very timely finding that RNase Y is not an essential gene. This was published by the group of Ciarán Condon (Figaro et al., 2013), but was brought to our attention during a conversation with Carol Gross and Byoung-Mo Koo, who provided us with the strain. Without a mutant lacking RNase Y, confirming the connection between this protein and the complex would have been much more difficult.

I had expected YlbF, YmcA, and YaaT to be playing a role in the known regulatory pathway for biofilm formation, but these findings have proven an alternative. This complex is acting in a separate pathway to regulate matrix production, and it is acting at the level of mRNA stability. Any solution to the problem would have provided insight and satisfaction, but coincidentally, the function of the complex has

turned out to be very interesting and a part of a field that is growing quickly. Preliminary RNA-sequencing experiments have shown that YlbF, YmcA, and YaaT play a more general role in RNA turnover, and learning this has raised many questions that I will contine to pursue answers for.

One is, to what extent does YlbF, YmcA, and YaaT play a role in RNA turnover. Based on the fact that mutants of RNase Y are more impaired for growth than mutants of YlbF, YmcA, or YaaT, it is likely that RNase Y has more targets than the complex, and our preliminary RNA-seq data provides evidence towards this. Carrying out RNA-seq with all of the mutants including one of RNase Y will allow us to answer this question on a global scale. Furthermore, while mutants of *ylbF* and *ymcA* are identical in all of the phenotypes tested, mutants of *yaaT* have an intermediate phenotype. Sequencing experiments may provide an answer for why this is. Also, there are additional phenotypes of YlbF, YmcA, and YaaT mutants, such as growth impairment, that are presumably caused by an effect on other RNAs beside *sinR*. One possibility is the lack of processing of the *cggR-gapA* operon, but identifying all of the genes affected in mutants of the complex may provide more information on the cause of the growth defect. In mutants lacking YlbF or YmcA, this transcript is not cleaved as it is in wild type. The cleavage in wild type leads to the degradation of the *cggR* portion of the transcript. CggR, the central glycolytic genes repressor, is named such due to its role in regulating carbon metabolism and the growth defect of YlbF, YmcA, and YaaT mutants has already been connected to a switch from glucose metabolism to an alternative carbon source.

Another question of interest is whether or not the complex of YlbF, YmcA, and YaaT are regulating specific RNA degradation by RNase Y. To answer this question, RNA-sequencing will carried out with samples from different phases of growth and in different media. For biofilm formation, samples from exponential growth will be compared to samples from early and late stationary phase. Other conditions that can be tested include samples groown in LB or samples from cells that have been treated with certain stressors. The results of these experiments may provide information as to whether YlbF, YmcA, and YaaT are required for turnover of specific targets under all conditions or if they are only required for turnover under specific conditions. RNA-seq experiments under a variety of conditions could

also identify more targets of RNase Y and the complex because different genes will be expressed under different growth conditions. If the complex of YlbF, YmcA, and YaaT are acting to regulate the stability of RNAs under specific conditions, it will be interesting to determine if they are doing so based on oa specific signal from the cell or its environment.

Finally, sequencing experiments may provide some insight into the recognition sequences or features of RNase Y and the complex. The sequencing will likely identify novel cleavage sites for RNase Y with nucleotide resolution, which microarray experiments cannot do. The sequencing results also allow for the investigation of untranslated regions, such as riboswitches and non-coding RNAs that the microarray experiments do not include. One challenge towards identifying specific cleavage sites is that most RNAs that are cleaved by RNase Y are degradaded so quickly that we cannot see them by northern blotting or RNA-seq. One strategy to identify more cleavage sites is to stabilize the cleavage products so that they can be sequenced. One way to do this is to delete the ribonuclease PNPase, which is responsible for turning over the cleavage product that is at the 5' of the original cleavage site. Combining a PNPase mutant with a mutant of RNase Y produces cells that are barely viable, and thus it is difficult to compare this strain to a normally growing wild type. However, combining the PNPase deletion with mutants of YlbF or YmcA results in viable cells that are not significantly impaired for growth. If more examples of cleavage sites can be identified it will be easier to look for common sequences or secondary structures. For YlbF, YmcA, and YaaT comparing RNase Y targets that are affected to those that aren't might provide information on what the complex is recognizing. One area that might be particularly fruitful in this type of analysis is with the riboswitches, because, within one type, they are fairly similar yet only some of them are affected. For instance, the eleven S-box riboswitches have a lot of similarites both in sequence and in secondary structure, but they are not all affected by the complex. Precisely determining which of these are affected and which are not would allow for this type of analysis.

There are more detailed questions about YlbF, YmcA, and YaaT that remain. A variety of genetic and biochemical approaches could be used to better understand the interactions of these proteins in the complex and of the complex with RNAse Y. What is the stoicheometry of the complex? Can all three

proteins interact with RNase Y at the same time? What regions of the proteins or residues are required for these interactions? Evaluating the conservation of these genes and their amino acid sequence could provide candidates towards this. For instance, RNase Y and YaaT are highly co-conserved, and are conserved more widely than YlbF and YmcA. Are there changes in the RNase Y sequence that highly correlate with the presence or absence of YlbF and YmcA amongst these species? If so, this could indicate that those residues are important for the interaction. A similar type of analysis can be carried out with YlbF, YmcA, and YaaT as well to look for the co-conservation of specific residues.

Of course, a major unanswered question is how specfically, at a biochemical level, are YlbF, YmcA, and YaaT affecting RNA stability. Are they able to directly bind to RNA?, in which they could be playing the role of recruiting RNAs to RNase Y to be degraded. Are they binding to RNase Y and somehow changing its affinity towards specific targets? There are a variety of biochemical techniques that could be used to see if the proteins are able to bind to RNA, such as gel shift assays, filter binding assays, and RNA-precipitation with sequencing to identify binding targets. Another interesting way to investigate this question is based on the localization of these proteins. It is known that RNase Y localizes to the membrane, which implies that RNA degradation is happening there. It will be interesting to see if YmcA, YlbF, and YaaT also localize to othe membrane and to see if this pattern changes based on the conditions. It is also possible to use super-resolution microscopy to look at the localization of RNAs. It would be interesting to see if RNAs localize to the membrane as they are being degraded, and if the complex could affect this.

Another aspect of the biology of YlbF, YmcA, and YaaT that is interesting is their role in other species. For instance, in *S. aureus* mutants of RNase Y and of YlbF are viable. It would be interesting to see if mutants of the complex have any phenotypes and it would also be interesting to identify the targets in *S. aureus*. A Tn-seq experiment to look for genes important for *S. aureus* growth in a rich medium indicated that mutants of the homolgs of the complex are outcompeted by wild type ((Valentino et al., 2014)). The discovery of the role of YlbF, YmcA, and YaaT has opened up many further lines of inquiry

that will be interesting towards understanding the role of this complex at a global level and its effects on a broader level.

Bibliography

Aguilar, C., Vlamakis, H., Losick, R., & Kolter, R. (2007). Thinking about Bacillus subtilis as a multicellular organism. *Curr Opin Microbiol, 10*(6), 638-643. doi: 10.1016/j.mib.2007.09.006.

Arnold, T. E., Yu, J., & Belasco, J. G. (1998). mRNA stabilization by the ompA 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *Rna, 4*(3), 319-330.

Babitzke, P., & Kushner, S. R. (1991). The Ams (altered mRNA stability) protein and ribonuclease E are encoded by the same structural gene of Escherichia coli. *Proc Natl Acad Sci U S A, 88*(1), 1-5.

Bai, U., Mandic-Mulec, I., & Smith, I. (1993). SinI modulates the activity of SinR, a developmental switch protein of Bacillus subtilis, by protein-protein interaction. *Genes Dev, 7*(1), 139-148.

Banse, A. V., Chastanet, A., Rahn-Lee, L., Hobbs, E. C., & Losick, R. (2008). Parallel pathways of repression and antirepression governing the transition to stationary phase in Bacillus subtilis. *Proc Natl Acad Sci U S A, 105*(40), 15547-15552. doi: 10.1073/pnas.0805203105.

Beauregard, P. B., Chai, Y., Vlamakis, H., Losick, R., & Kolter, R. (2013). Bacillus subtilis biofilm induction by plant polysaccharides. *Proc Natl Acad Sci U S A, 110*(17), E1621-1630. doi: 10.1073/pnas.1218984110

Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R., & Kolter, R. (2001). Fruiting body formation by Bacillus subtilis. *Proc Natl Acad Sci U S A, 98*(20), 11621-11626. doi: 10.1073/pnas.191384198

Branda, S. S., Gonzalez-Pastor, J. E., Dervyn, E., Ehrlich, S. D., Losick, R., & Kolter, R. (2004). Genes involved in formation of structured multicellular communities by Bacillus subtilis. *J Bacteriol, 186*(12), 3970-3979. doi: 10.1128/jb.186.12.3970-3979.2004

Braun, F., Le Derout, J., & Regnier, P. (1998). Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of Escherichia coli. *Embo j, 17*(16), 4790-4797. doi: 10.1093/emboj/17.16.4790

Burmann, F., Sawant, P., & Bramkamp, M. (2012). Identification of interaction partners of the dynamin-like protein DynA from Bacillus subtilis. *Commun Integr Biol, 5*(4), 362-369. doi: 10.4161/cib.20215

Carabetta, V. J., Tanner, A. W., Greco, T. M., Defrancesco, M., Cristea, I. M., & Dubnau, D. (2013). A complex of YlbF, YmcA and YaaT regulates sporulation, competence and biofilm formation by accelerating the phosphorylation of Spo0A. *Mol Microbiol, 88*(2), 283-300. doi: 10.1111/mmi.12186

Celesnik, H., Deana, A., & Belasco, J. G. (2007). Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal. *Mol Cell, 27*(1), 79-90. doi: 10.1016/j.molcel.2007.05.038

Chai, Y., Chu, F., Kolter, R., & Losick, R. (2008). Bistability and biofilm formation in Bacillus subtilis. *Mol Microbiol, 67*(2), 254-263. doi: 10.1111/j.1365-2958.2007.06040.x

Chai, Y., Norman, T., Kolter, R., & Losick, R. (2010). An epigenetic switch governing daughter cell separation in Bacillus subtilis. *Genes Dev, 24*(8), 754-765. doi: 10.1101/gad.1915010

Chai, Y., Norman, T., Kolter, R., & Losick, R. (2011). Evidence that metabolism and chromosome copy number control mutually exclusive cell fates in Bacillus subtilis. *Embo j, 30*(7), 1402-1413. doi: 10.1038/emboj.2011.36

Chen, L. H., Emory, S. A., Bricker, A. L., Bouvet, P., & Belasco, J. G. (1991). Structure and function of a bacterial mRNA stabilizer: analysis of the 5' untranslated region of ompA mRNA. *J Bacteriol, 173*(15), 4578-4586.

Chen, Y., Cao, S., Chai, Y., Clardy, J., Kolter, R., Guo, J. H., & Losick, R. (2012). A Bacillus subtilis sensor kinase involved in triggering biofilm formation on the roots of tomato plants. *Mol Microbiol, 85*(3), 418-430. doi: 10.1111/j.1365-2958.2012.08109.x

Chu, F., Kearns, D. B., Branda, S. S., Kolter, R., & Losick, R. (2006). Targets of the master regulator of biofilm formation in Bacillus subtilis. *Mol Microbiol, 59*(4), 1216-1228. doi: 10.1111/j.1365-2958.2005.05019.x

Chu, F., Kearns, D. B., McLoon, A., Chai, Y., Kolter, R., & Losick, R. (2008). A novel regulatory protein governing biofilm formation in Bacillus subtilis. *Mol Microbiol, 68*(5), 1117-1127. doi: 10.1111/j.1365-2958.2008.06201.x

Commichau, F. M., Rothe, F. M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., . . . Stulke, J. (2009). Novel activities of glycolytic enzymes in Bacillus subtilis: interactions with essential proteins involved in mRNA processing. *Mol Cell Proteomics, 8*(6), 1350-1360. doi: 10.1074/mcp.M800546-MCP200

Deaconescu, A. M., Chambers, A. L., Smith, A. J., Nickels, B. E., Hochschild, A., Savery, N. J., & Darst, S. A. (2006). Structural basis for bacterial transcription-coupled DNA repair. *Cell, 124*(3), 507-520. doi: 10.1016/j.cell.2005.11.045

Deana, A., Celesnik, H., & Belasco, J. G. (2008). The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature, 451*(7176), 355-358. doi: 10.1038/nature06475

Deighan, P., Diez, C. M., Leibman, M., Hochschild, A., & Nickels, B. E. (2008). The bacteriophage lambda Q antiterminator protein contacts the beta-flap domain of RNA polymerase. *Proc Natl Acad Sci U S A, 105*(40), 15305-15310. doi: 10.1073/pnas.0805757105

Deikus, G., & Bechhofer, D. H. (2011). 5' End-independent RNase J1 endonuclease cleavage of Bacillus subtilis model RNA. *J Biol Chem, 286*(40), 34932-34940. doi: 10.1074/jbc.M111.287409

DeLoughery, A., Dengler, V., Chai, Y., & Losick, R. (2016). Biofilm formation by Bacillus subtilis requires an endoribonuclease-containing multisubunit complex that controls mRNA levels for the matrix gene repressor SinR. *Mol Microbiol, 99*(2), 425-437. doi: 10.1111/mmi.13240

Durand, S., Gilet, L., Bessieres, P., Nicolas, P., & Condon, C. (2012). Three essential ribonucleases-RNase Y, J1, and III-control the abundance of a majority of Bacillus subtilis mRNAs. *PLoS Genet, 8*(3), e1002520. doi: 10.1371/journal.pgen.1002520

Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Brechemmier-Baey, D., & 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), 2141-2152. doi: 10.1093/nar/gki505

Figaro, S., Durand, S., Gilet, L., Cayet, N., Sachse, M., & Condon, C. (2013). Bacillus subtilis mutants with knockouts of the genes encoding ribonucleases RNase Y and RNase J1 are viable, with major defects in cell morphology, sporulation, and competence. *J Bacteriol, 195*(10), 2340-2348. doi: 10.1128/jb.00164-13

Gilet, L., DiChiara, J. M., Figaro, S., Bechhofer, D. H., & Condon, C. (2015). Small stable RNA maturation and turnover in Bacillus subtilis. *Mol Microbiol, 95*(2), 270-282. doi: 10.1111/mmi.12863

Hamon, M. A., & Lazazzera, B. A. (2001). The sporulation transcription factor Spo0A is required for biofilm development in Bacillus subtilis. *Mol Microbiol, 42*(5), 1199-1209.

Hamon, M. A., Stanley, N. R., Britton, R. A., Grossman, A. D., & Lazazzera, B. A. (2004). Identification of AbrB-regulated genes involved in biofilm formation by Bacillus subtilis. *Mol Microbiol, 52*(3), 847-860. doi: 10.1111/j.1365-2958.2004.04023.x

Hosoya, S., Asai, K., Ogasawara, N., Takeuchi, M., & Sato, T. (2002). Mutation in yaaT leads to significant inhibition of phosphorelay during sporulation in Bacillus subtilis. *J Bacteriol, 184*(20), 5545-5553.

Hui, M. P., Foley, P. L., & Belasco, J. G. (2014). Messenger RNA degradation in bacterial cells. *Annu Rev Genet, 48*, 537-559. doi: 10.1146/annurev-genet-120213-092340

Jiang, M., Shao, W., Perego, M., & Hoch, J. A. (2000). Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis. *Mol Microbiol, 38*(3), 535-542.

Kearns, D. B., Chu, F., Branda, S. S., Kolter, R., & Losick, R. (2005). A master regulator for biofilm formation by Bacillus subtilis. *Mol Microbiol, 55*(3), 739-749. doi: 10.1111/j.1365-2958.2004.04440.x

Khemici, V., Poljak, L., Toesca, I., & Carpousis, A. J. (2005). Evidence in vivo that the DEAD-box RNA helicase RhlB facilitates the degradation of ribosome-free mRNA by RNase E. *Proc Natl Acad Sci U S A, 102*(19), 6913-6918. doi: 10.1073/pnas.0501129102

Kolter, R., & Greenberg, E. P. (2006). Microbial sciences: the superficial life of microbes. *Nature, 441*(7091), 300-302. doi: 10.1038/441300a

Kuznedelov, K., Minakhin, L., Niedziela-Majka, A., Dove, S. L., Rogulja, D., Nickels, B. E., . . . Severinov, K. (2002). A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. *Science, 295*(5556), 855-857. doi: 10.1126/science.1066303

Laalami, S., Bessieres, P., Rocca, A., Zig, L., Nicolas, P., & Putzer, H. (2013). Bacillus subtilis RNase Y activity in vivo analysed by tiling microarrays. *PLoS One, 8*(1), e54062. doi: 10.1371/journal.pone.0054062

Lehnik-Habrink, M., Newman, J., Rothe, F. M., Solovyova, A. S., Rodrigues, C., Herzberg, C., . . . Stulke, J. (2011). RNase Y in Bacillus subtilis: a Natively disordered protein that is the functional

equivalent of RNase E from Escherichia coli. *J Bacteriol, 193*(19), 5431-5441. doi: 10.1128/jb.05500-11

Lehnik-Habrink, M., Schaffer, M., Mader, U., Diethmaier, C., Herzberg, C., & Stulke, J. (2011). RNA processing in Bacillus subtilis: identification of targets of the essential RNase Y. *Mol Microbiol, 81*(6), 1459-1473. doi: 10.1111/j.1365-2958.2011.07777.x

Ludwig, H., Homuth, G., Schmalisch, M., Dyka, F. M., Hecker, M., & Stulke, J. (2001). Transcription of glycolytic genes and operons in Bacillus subtilis: evidence for the presence of multiple levels of control of the gapA operon. *Mol Microbiol, 41*(2), 409-422.

Mackie, G. A. (1998). Ribonuclease E is a 5'-end-dependent endonuclease. *Nature, 395*(6703), 720-723. doi: 10.1038/27246

Marincola, G., Schafer, T., Behler, J., Bernhardt, J., Ohlsen, K., Goerke, C., & Wolz, C. (2012). RNase Y of Staphylococcus aureus and its role in the activation of virulence genes. *Mol Microbiol, 85*(5), 817-832. doi: 10.1111/j.1365-2958.2012.08144.x

Mathy, N., Benard, L., Pellegrini, O., Daou, R., Wen, T., & 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), 681-692. doi: 10.1016/j.cell.2007.02.051

McDowall, K. J., & Cohen, S. N. (1996). The N-terminal domain of the rne gene product has RNase E activity and is non-overlapping with the arginine-rich RNA-binding site. *J Mol Biol, 255*(3), 349-355. doi: 10.1006/jmbi.1996.0027

McDowall, K. J., Lin-Chao, S., & Cohen, S. N. (1994). A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J Biol Chem, 269*(14), 10790-10796.

McLoon, A. L., Kolodkin-Gal, I., Rubinstein, S. M., Kolter, R., & Losick, R. (2011). Spatial regulation of histidine kinases governing biofilm formation in Bacillus subtilis. *J Bacteriol, 193*(3), 679-685. doi: 10.1128/jb.01186-10

Melefors, O., & von Gabain, A. (1991). Genetic studies of cleavage-initiated mRNA decay and processing of ribosomal 9S RNA show that the Escherichia coli ams and rne loci are the same. *Mol Microbiol, 5*(4), 857-864.

Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., Gonzalez-Pastor, J. E., Liu, J. S., & Losick, R. (2003). The Spo0A regulon of Bacillus subtilis. *Mol Microbiol, 50*(5), 1683-1701.

Mudd, E. A., Krisch, H. M., & Higgins, C. F. (1990). RNase E, an endoribonuclease, has a general role in the chemical decay of Escherichia coli mRNA: evidence that rne and ams are the same genetic locus. *Mol Microbiol, 4*(12), 2127-2135.

Newman, J. A., Hewitt, L., Rodrigues, C., Solovyova, A. S., Harwood, C. R., & Lewis, R. J. (2012). Dissection of the network of interactions that links RNA processing with glycolysis in the Bacillus subtilis degradosome. *J Mol Biol, 416*(1), 121-136. doi: 10.1016/j.jmb.2011.12.024

Norman, T. M., Lord, N. D., Paulsson, J., & Losick, R. (2013). Memory and modularity in cell-fate decision making. *Nature, 503*(7477), 481-486. doi: 10.1038/nature12804

Ono, M., & Kuwano, M. (1979). A conditional lethal mutation in an Escherichia coli strain with a longer chemical lifetime of messenger RNA. *J Mol Biol, 129*(3), 343-357.

Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D. H., . . . Belasco, J. G. (2011). An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in Bacillus subtilis. *Mol Cell, 43*(6), 940-949. doi: 10.1016/j.molcel.2011.07.023

Richards, J., Mehta, P., & Karzai, A. W. (2006). RNase R degrades non-stop mRNAs selectively in an SmpB-tmRNA-dependent manner. *Mol Microbiol, 62*(6), 1700-1712. doi: 10.1111/j.1365-2958.2006.05472.x

Subramaniam, A. R., DeLoughery, A., Bradshaw, N., Chen, Y., O'Shea, E., Losick, R., & Chai, Y. (2013). A serine sensor for multicellularity in a bacterium. *Elife, 2*, e01501. doi: 10.7554/eLife.01501

Taraseviciene, L., Miczak, A., & Apirion, D. (1991). The gene specifying RNase E (rne) and a gene affecting mRNA stability (ams) are the same gene. *Mol Microbiol, 5*(4), 851-855.

Tortosa, P., Albano, M., & Dubnau, D. (2000). Characterization of ylbF, a new gene involved in competence development and sporulation in Bacillus subtilis. *Mol Microbiol, 35*(5), 1110-1119.

Tzeng, Y. L., Zhou, X. Z., & Hoch, J. A. (1998). Phosphorylation of the Spo0B response regulator phosphotransferase of the phosphorelay initiating development in Bacillus subtilis. *J Biol Chem, 273*(37), 23849-23855.

Valentino, M. D., Foulston, L., Sadaka, A., Kos, V. N., Villet, R. A., Santa Maria, J., . . . Gilmore, M. S. (2014). Genes Contributing to Staphylococcus aureus Fitness in Abscess- and Infection-Related Ecologies. *mBio, 5*(5). doi: 10.1128/mBio.01729-14

Vanzo, N. F., Li, Y. S., Py, B., Blum, E., Higgins, C. F., Raynal, L. C., . . . Carpousis, A. J. (1998). Ribonuclease E organizes the protein interactions in the Escherichia coli RNA degradosome. *Genes Dev, 12*(17), 2770-2781.

Yao, S., Richards, J., Belasco, J. G., & Bechhofer, D. H. (2011). Decay of a model mRNA in Bacillus subtilis by a combination of RNase J1 5' exonuclease and RNase Y endonuclease activities. *J Bacteriol, 193*(22), 6384-6386. doi: 10.1128/jb.05939-11

Yarchuk, O., Jacques, N., Guillerez, J., & Dreyfus, M. (1992). Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *J Mol Biol, 226*(3), 581-596.

Appendix A: Materials and Methods

This appendix contains the materials and methods used in chapters one through four. It also contains a list of strains used, a list of primers used, and a list of plasmids used in chapters one through four.

A.1 bacterial growth and media

For general purposes, *B. subtilis* strains PY79, 3610, and their derivatives were grown in Luria-Bertani (LB) medium (10 g tryptone, 5 g yeast extract, and 5 g NaCl per liter broth) at 37°C. In liquid, cultures were grown shaking. Antibiotics were added to the media at the following concentrations for *B. subtilis* strains: 10 μg ml^{-1} of tetracycline, 100 μg ml^{-1} of spectinomycin, 10 μg ml^{-1} of kanamycin, 5 μg ml^{-1} of chloramphenicol, 1 μg ml^{-1} of erythromycin, and 50 μg ml^{-1} spectinomycin, and for *E. coli*: ampicillin (100 μg ml−1), chloramphenicol (20 μg ml−1), and kanamycin (10 μg ml−1). The defined medium, MSgg, was used to induce biofilm formation either as a liquid or supplemented 1.5% Bacto-agar (Difco) for plate. MSgg medium composition: 5 mM potassium phosphate (pH 7), 100 mM MOPS (pH 7), 2 mM $MgCl_2$, 700 μM $CaCl_2$, 50 μM $MnCl_2$, 50 μM $FeCl_3$, 1 μM $ZnCl_2$, 2 μM thiamine, 0.5% glycerol, 0.5% glutamate, 50 μg ml^{-1} tryptophan, 50 μg ml^{-1} phenylalanine and 50 μg ml^{-1} threonine.

A.2 strain construction

Strains were constructed using standard genetic cloning protocols for *B. subtilis* (Harwood & Cutting, **1990**). Antibiotic marked fragments of DNA were introduced into PY79 by transformation (Wilson and Bott, **1968**)and into 3610 by SPP1 phage transduction (Yasbin and Young, **1974**). The markerless deletion of *yaaT* was created using the pminimad system as previously described (Arnaud *et al*., **2004**; Patrick and Kearns, **2008**; Chen *et al*., **2009**). An insertional deletion of *rny* was obtained from Byoung Mo Koo and the marker was removed through a cre recombinase system. Synonymous mutants of *sinR* were generated at the native locus using synthetic DNA fragments followed by allelic exchange as previously described (Subramaniam *et al*., **2013**). The complementation strain of *rny* was created by integrating *rny* with its native promoter into the chromosome at *amyE*. Reporter strains were created by

fusing *lac*Z to the promoter for the gene of interest and integration into the chromosome at *amyE* by cloning the promoter into the plasmid pDG268 (Antoniewski *et al*., **1990**). The P_{sinR}-*lacZ* was integrated into the chromosome at an alternative *amyE* site inserted into *bkdB* near the terminus (strain AHB290 a gift from A. Camp) also described in (Chai *et al*., **2011**). The promoter for *sinR* was amplified using the primers PsinRF1 and PsinRR1 and contains 190 bp directly upstream of the *sinR* open reading frame. Strain genotypes, full construction details, and a list of plasmids used are available in the supplemental materials.

Construction of rny markerless deletion in 3610

An SPP1 phage lysate was created from BKE16960 and used to transduce *rny::erm* into 3610. At the same time the plasmid pBKACL-cre, purified from recA+ *E. coli*) was transformed into PY79. The plasmid pBKACL-cre was moved from PY79 into 3610 *ymdA::erm* using transduction and grown at 30°C. Transductants were restreaked on LB Spec 1mM IPTG plates and grown at 30°C overnight. The resulting strain was streaked onto LB plates and grown at 42°C to promote the loss of pBKACL-cre. Multiple colonies were then grown on LB plates and tested to ensure both sensitivity to spec and erm. Deletion was confirmed by PCR.

Construction of yaaT markerless deletion in 3610

A. 3 growth curves and live/dead staining

Cultures were grown overnight at 25°C to ensure that they did not reach stationary phase. Cultures were back diluted to an OD600 0f 0.05. Growth of the cultures was measured by the OD600 at indicated time points. The live/dead staining was carried out on cells collected from early stationary phase cultures in LB, an OD600 of ~1.2. Live/dead staining was carried out according to the manufacturers protocols (LIVE/DEAD BacLight Bacterial Viability Kit, Life Technologies) and cells were imaged on an Olympus workstation BX61 (Olympus, Shinjuku Japan). For each sample at least 2,000 cells were counted from fields of view containing 100-250 cells. For each strain biological duplicates were analyzed and the standard deviation is reported.

A.4 biofilm assays

B. subtilis cells were first grown in LB broth at 37° C to mid-exponential phase. For formation of biofilm colonies, 2 µl of the cells was spotted onto MSgg medium solidified with 1.5% agar. Plates were incubated at 23° C for 3-4 days before analysis. All images were taken using either a Nikon CoolPix 950 digital camera or using a SPOT camera (Diagnostic Instruments, USA). Assays for the *b*-galactosidase activities were described previously (*5*).

A.5 β-galactosidase assays

The following protocol was adapted from a previously established protocol adapted to a plate reader (Kearns *et al*., **2005**). Cultures were grown overnight in LB shaking at 25°C. The cultures were diluted back to an OD600 of 0.05 in liquid MSgg. Samples were collected at the indicated times for the analysis of β-*galactosidase* activity. One milliliter of cells were collected by centrifugation and resuspended in 500 µL of Z buffer (40 mM NaH2PO4, 60 mM Na2HPO4, 1 mM MgSO4, 10 mM KCl and 38 mM β-mercaptoethanol). To 96 well plates, 50 µL of each resuspension as well as 50 µL of a five- and 25-fold dilution were added to 50 µL of Z buffer containing lysozyme (0.2 mg ml−1 final concentration). Plates were incubated at 37°C for 25 min. Reactions were started by the addition of 20 µL of 4 mg ml−1 2-nitrophenyl β-D-galactopyranoside in Z buffer. The absorbance at 420 nm was read on a plate reader every minute for 1 h. The reported activity was calculated according to the following equation: dilution factor × VmeanOD420/OD600.

A.6 western blotting

For Western blots, cleared whole cell lysates were run on 12.5% SDS-PAGE nupage gels (Life Technologies, Carlsbad, CA, USA) and transferred to a PVDF membrane (Millipore, Billerica, MA, USA) by electro blotting (Bio-Rad, Hercules, CA, USA). Bands were compared with the Benchmark Prestained Protein Ladder (Life Technologies). Samples were normalized to total protein concentration determined by Bradford assay (Bio-Rad). Rabbit anti-SigA polyclonal antibodies were used at a concentration of 1:100,000. Rabbit anti-SinR polyclonal, affinity purified antibodies were used at a concentration of 1:1500. Rabbit anti-SinI polyclonal antibodies were used at a concentration of 1:5000.

They were visualized with goat anti-rabbit antibodies conjugated to Horseradish Peroxidase at a concentration of 1:10,000 (Bio-Rad) and were developed with SuperSignal West Dura chemiluminescent substrate (Thermo, Waltham, MA, USA). For full details for sample preparation and imaging, see previous work (Subramaniam *et al*., **2013**). For Fig. 2B, the quantification represents the average of biological triplicates and the standard deviation is reported. For each replicate the *ylbF*, *ymcA* and *spo0A* mutants were normalized to wild type.

A.7 northern blotting

Samples for Northern blot analysis were prepared from cells grown in liquid shaking cultures of MSgg at an OD600 of 2.0–2.5. Preparations of total RNA samples from *B. subtilis* were created using a Trizol extraction as previously described (Cheung *et al*., **1994**). Samples were run on 1.5% agarose gels containing 20 mM guanidine thiocyanate (Goda and Minton, **1995**). For a loading control, rRNA was visualized by ethidium bromide staining. RNA was transferred to Hybond-N + membranes (Amersham Biosciences, Piscataway, NJ, USA) using a capillary transfer. Transcripts were detected using Digoxigenin (DIG) probes (Roche Diagnostics, Indianapolis, IN, USA). A probe against SinR was created by in vitro transcription with T7 RNA polymerase (Roche Diagnostics) using PCR-generated DNA fragments as a template. The specific primers are detailed in the supplemental information. Hybridization and signal detection were carried out according to the manufacturer's instructions (DIG RNA labeling and detection kit, Roche Diagnostics) with the exception that the *sinR* probe hybridization was carried out at 74°C. For *cggR* and *gapA* the probe was hybridized at 70°C. Bands were compared with the RNA Molecular Weight Marker III (Roche Diagnostics). For Fig. 3B quantification was carried out by measuring the densitometry of the *sinR*-only transcript in *ylbF* and wild type. The average of 3 biological replicates is reported with the standard deviation. As loading controls rRNA was visualized either by ethidium bromide staining with UV imaging or methylene blue (0.25% methylene blue, 0.3 M sodium acetate pH 5.2) staining and imaging with a Sony digital camera (Sony Corp., model NEX-5, Tokyo, Japan) as indicated in figure legends.

A.8 gel imaging and quantification

Densitometry analysis was performed using ImageJ software (http://imagej.nih.gov.ezp-prod1.hul.harvard.edu/ij/) Rectangles were drawn around distinct bands and the average pixel intensity was measured. The same size of the rectangle is used to quantify bands that are being compared, and the density of the background of the same size is subtracted from the measurement. For statistics, the sample sizes are indicated in the legends. The error bars represent the standard deviation. Two-tailed *P*-values were determined based on unpaired *t*-tests. In images statistical significance is indicated as follows; '*' for $P \leq 0.05$, '**' for $P \leq 0.01$, and '***' for $P \leq 0.001$.

A.9 pull down and mass spectroscopy with MBP-YlbF

A fusion of MBP-YlbF and MBP alone were purified from *B. subtilis* grown under biofilm inducing conditions using an adaptation of the pMAL protein fusion and purification system (New England Biolabs, Ipswitch MA). Overnight cultures of strain AJD17, harboring the gene for a fusion of MBP and YlbF under the control of an IPTG inducible promoter, and AJD18, harboring the gene for MBP alone under the control of an IPTG inducible promoter, were diluted 1:100 in 500 ml of MSgg supplemented with 100 μM IPTG, 2 cultures for each strain. One cultures for each strain was grown to and OD600 of 0.9 and the other to an OD600 of 1.2. The cultures for each strain from the two ODs were combined 1:1 by volume, and the cells were pelleted by centrifugation (4,000 × g for 20 minutes at 4°C). We used both ODs because we weren't sure if we would see a biologically relevant interaction before the onset of matrix production or after. The cells were washed in 50 ml of pMAL column buffer (20 mM Tris-HCl pH 7.4, 200 mM NaCl2) and collected by centrifugation (4,000 × g for 20 minutes at 4°C). The cells collected from both ODs for each strain were combined in this step. The pellets were resuspended in 5 ml of pMAL column buffer containing lysozyme from egg whites (20 μg ml-1)(Sigma) and protease inhibitors (complete Tablets, Mini EDTA-free, EASYpack, Roche). The cells were incubated at room temperature for 20 minutes at 37°C and on ice for 40 minutes to allow for the lysozyme to act. The lysates were then frozen in liquid nitrogen and allowed to thaw 3 times to help with lysis. Finally, the cells were sonicated to further disrupt the cells. A cleared lysate was created by centrifugation for 25 minutes at 4°C

(18,000 ×g for 25 minutes at 4°C). An overview of the MBP purification strategy employed is as such. It should be noted that the MBP-tag also has a His-tag and both tags were used in the purification procedure. First, the MBP- YlbF fusion and MBP-tag alone were concentrated using an amylose resin (High Flow) that binds the MBP-tag (New England Biolabs). The cleared lysates were incubated overnight with 300 µL of the resin, in batch, gently shaking at 4°C. The resin was then collected on a column. In this step the MBP-YlbF/MBP associated resin was lightly washed, using 1.5 ml pMAL column buffer. The MBP-YlbF/MBP and associated proteins were eluted into 1 ml of elution buffer (column buffer + 10mM maltose). The eluted protein was dialyzed into binding buffer (50 mM sodium phosphate buffer pH 8.0, 300 mM NaCl2, 0.02% Tween). The MBP-YlbF/MBP and associated proteins were further purified using the Dynabead His-Tag Isolation and Pull Down system (Life technologies). Resuspended Dynabeads were added, 50 µL, to eppendorf tubes. The beads were collected with a magnet and the supernatant was removed. The protein solutions were added to the beads and mixed. The lysate bead mixtures were incubated at room temperature for 10 minutes with gentle shaking. The beads were collected using a magnet for 2 minutes. The beads were washed 4 times in 300 µL of binding buffer. Each time the beads were resuspended, lightly vortexed, and then collected with a magnet. The proteins associated with the beads were eluted directly into 50 µL of a 1X SDS-PAGE loading buffer (NuPAGE LDS sample buffer, Life Technologies). Samples were run on an SDS-PAGE gel (Life Technologies)(see materials and methods western blotting). Each lane was cut from the gel and separated into 7 fragments. The fragments along with an equal size piece of the gel from a lane with no protein were submitted for MS/MS analyses at the Mass Spectroscopy and Proteomics Resource Core at Harvard University FAS Center for Systems Biology (52 Oxford St, Cambridge, MA) by microcapillary reverse-phase HPLC nano-electrospray tandem mass spectrometry (µLC/MS/MS) on a LTQ- Orbitrap mass spectrometer (Thermo). Details can be found at http://proteomics.fas.harvard.edu/. For each strain only one sample was analyzed by MS/MS.

A.10 bacterial two-hybrid

Each ‘Bait’ protein was cloned with a N-terminal fusion of λCI repressor (residues 1–239 of λCI). The ‘bait’ protein was cloned into the NotI and BamhI sites of pAC-CI using isothermal assembly. For each bait the entire ORF was used except for RNase Y, which 25 amino acid truncation of the N-terminus was used to remove a transmembrane domain. The ‘Prey’ protein was cloned with a N-terminal fusion to the N-terminal domain of αRNAP (αNTD residues 1–248). The prey was cloned into the BamHI and NotI sites of pBRα using isothermal assembly. For a list of plasmids and primers see supporting information. Bait and Prey fusions to be tested were cotransformed into the reporter strain FW102 OL2–62, which contains a *lacZ* reporter. To test for an interaction, the reporter strain was grown in liquid shaking LB with appropriate antibiotics overnight at 37°C. The next morning, the cultures were back-diluted 1:100 into LB (kan20, amp100, cm20) supplemented with 20 nM IPTG. One milliliter of each culture was collected after 3 h of shaking at 37° and the OD600 was measured. The pellet was resuspended in 1 ml of buffer Z (40 mM NaH2PO4, 60 mM Na2HPO4, 1 mM MgSO4, 10 mM KCl, and 38 mM β-mercaptoethanol). To this 100 µl of SDS (2.5%) and 50 µl of chloroform were added to lyse cells. The samples were briefly vortexed. To each sample, 200 µl of ONPG (4 mg ml−1) was added for 5 min. The reaction was stopped by the addition of sodium bicarbonate (500 µl of 1 M). The samples were centrifuged on a benchtop centrifuge at 13 K rpm for 5 min. The absorbance at 420 nm was determined for each of the supernatants. The miller unites were calculated as follows: [*OD*420 / (*OD*600 × 1 ml × 5 min)] × 1000. For each pair tested two biological replicates. The standard deviation is reported.

A.11 RNA-sequencing

Cells were harvested from MSgg liquid shaking cultures at an OD_{600} of approximately 0.4 (15 ml of culture) for exponential growth samples and at an OD_{600} of approximately 2.25 (10 ml of culture) for stationary phase samples. RNA was isolated and DNAse I treated using the RNeasy Minikit according to manufacturers protocol with the exception that collected cells were treated with lysozyme, 20 µg/ml, for 10 minutes at 37°C. For 5 µg of each sample, rRNA removal was carried out using the Illumina Ribo-zero rRNA removal kit (bacteria) according to manufacturers protocol. RNA was precipitated with 2 µl of

glycogen blue (thermo), 500 µl of 10 mM Tris pH 7.0, 55 µl 3 M NaOAc (pH 5.5), and 550 µl of cold isopropanol at -80°C for 30 minutes. RNA was pelleted by centrifugation, 13,000 rpm for 30 minutes at 4°C. Pelleted RNA was washed with cold 80% ethanol. Pellets were careful dried by removal of supernatant and standing at room temperature. The pellet was resuspended in 50 µl of 10 mM Tris pH 7.0. For each sample, 40 µl, were fragmented using the Ambion RNA fragmentation reagents (Life Technologies) by first incubating the samples at 95°C for 2 minutes then adding 4.4 µl of 10×fragmentation buffer at 95°C for 1.5 minutes. The fragmentation was halted by the addition of 5 µl of stop buffer. The samples were diluted with 500 µl of 10 mM Tris pH 7.0 and precipitated, pelleted, and washed as described previously. The samples were resuspended in 10 µl of 10 mM Tris pH 7.0 and run on a 15% TBE-Urea PAGE gel (Life Technologies) at 200 V for 65 minutes. For this the samples were prepared with 2× urea loading dye (Life technologies) and denatured at 95°C for two minutes. The gel was stained with SYBR gold and the RNAs from ~15 bp to 45~bp were excised. The RNA was retrieved from the gel, by first spinning the excised slices through a 0.7 ml eppendorf tube with a whole in the bottom created by a 16 gauge needle by centrifugation at 12,000×g for 3 minutes. To this, 500 µl of 10 mM Tris pH 7 was added and the mixture was shaken at 70°C for 10 minutes on a thermomixer (Eppendorf). Gel pieces were removed by centrifugation through a Spin-X cellulose acetate column. RNA was then precipitated, pelleted, and washed as previously described. Samples were resuspended in 15 µl of 10mM Tris pF 7. RNA fragments were dephosphorylated using T4 PNK without ATP (NEB) and with the RNase inhibitor Superasein (Life Technologies) according to manufacturers directions. Samples were again precipitated, pelleted, and washed as previously described and resuspended in 11 µl of 10 mM Tris pH 7. The entire samples were denatured at 80°C for 2 minutes. The linker-1 (see oligos) was ligated to the RNA fragments using T4 ligase 2 truncated (NEB) according to manufacturers protocol at 25°C for 2.5 hours. The ligated product was precipitated, pelleted, and washed as previously described and resuspended in 6 µl of 10 mM Tris pH 7. The samples were run on 10% TBE-Urea gel (Life Technologies) at 200 V for 50 minutes. Fragments in the size range of 35 – 65 bp were excised and gel purified, precipitated, pelleted, and washed as previously described. Ligated products were then reverse

transcribed using Superscript III (Life Technologies) according to manufacturers directions in 40 µl reactions with the primer oCJ485. The reaction was quenched by addition of 2.3 µl of 1 M NaOH for 15 minutes at 95°C to remove RNA. To these, 23 µl of 2×TBE Urea loading dye was added, and the samples were run on a 10% TBE-Urea gel (Life Technologies) at 200 V for 80 minutes, each sample split into two lanes. RT product was excised from the gel, crushed, gel purified, precipitated, pelleted, and washed as previously described, with the exception that the precipitation was carried out with Tris pH for DNA. The samples were resuspended in 15 µl of 10 mM Tris pH 8. The entire RT samples were circularized using Circ Ligase (Epicentrene) according to manufactures protocols at 60°C for 1 hour. The reaction was inactivated by incubation at 80°C for 10 minutes. PCR enrichment was carried out using Phusion polymerase (NEB) according to manufactures protocols with uniquely barcoded linkers for each samples (see primers). Each reaction had a total volume of 17 µl and used 1 µl of the circularization reaction as a template per reaction. Four reactions were carried out with each sample with a varying number of cycles, 6, 8, 10, and 12. The reactions were run on a non-denaturing 80% TB gel (Life Technologies) at 180 V for 45 minutes. Products from a cycle that produced enough of the correct product without too much background were excised, gel purified, precipitated, pelleted, and washed as described. The samples were resuspended in 10 µl of 10 mM Tris pH 8 and submitted to the MIT BioMicroCenter for quality control using a bioanalyzer and then sequenced using Illumina HISEQ 2000.

Table A1 list of strains

Strain	Genotype	Reference/construction
3610	Wild type	(Branda *et al.*, 2001)
PY79	Laboratory strain	
RL3852	3610 *epsH::tet*	(Kearns *et al.*, 2005)
RL4618	3610 *ylbF::mls*	(Branda *et al.*, 2004)
RL4619	3610 *ymcA::spec*	(Branda *et al.*, 2004)
AJD82	3610 *ΔyaaT* (markerless)	This study
RL4620	3610 *spo0A::kan*	(Chu *et al.*, 2008)
AJD145	3610 *spo0F::kan*	Transduction from RL4257 into 3610ka)
AJD147	3610 *spo0B::kan*	Transduction from RL4259 into 3610

RL4547	3610, *epsH::tet*, P_{epsA}-*lacZ::amyE cm*	(Kearns *et al.*, 2005)
AJD6	3610, *ylbF::mls*, P_{epsA}-*lacZ::amyE cm*	Transduction from RL4547 into RL4168
AJD7	3610, *ymcA::spec*, P_{epsA}-*lacZ::amyE cm*	Transduction from RL4547 into RL4169
RL4610	3610, *epsH::tet*, P_{sinI}-*lacZ::amyE cm*	(Chu *et al.*, 2008)
AJD95	3610, *ylbF::mls*, P_{sinI}-*lacZ::amyE cm*	Transduction from RL4610 into RL4168
AJD96	3610, *ymcA::spec*, P_{sinI}-*lacZ::amyE cm*	transduction from RL4610 into RL4169
AJD97	3610, *spo0A::kan*, P_{sinI}-*lacZ::amyE cm*	Transduction from RL4610 into RL4620
AJD99	3610, *epsH::tet*, P_{abrB}-*lacZ::amyE spec*	transduction from RL4608 into RL3852
AJD100	3610, *ylbF::mls*, P_{abrB}-*lacZ::amyE spec*	Transduction from RL4608 into RL3852
AJD101	3610, *spo0A::kan*, P_{abrB}-*lacZ::amyE spec*	Transduction from RL4608 into RL4620
AJD30	3610, *epsH::tet*, P_{sinR}-*lacZ::amyE cm*	This study
AJD31	3610, *ylbF::mls*, P_{sinR}-*lacZ::amyE cm*	This study
AJD32	3610, *ymcA::spec*, P_{sinR}-*lacZ::amyE cm*	This study
YC923	3610, TCA>AGT in serine residues 16, 18, and 33 in sinR in YC285	(Subramaniam *et al.*, 2013)
YC1173	3610, AGC/T>TCA in serine residues 32 and 76 in sinR in YC285	(Subramaniam *et al.*, 2013)
YC285	a kan resistance gene inserted to a locus downstream of sinR in 3610 chromosome	(Subramaniam *et al.*, 2013)
AJD205	3610, *ylbF::mls* in YC1173	This study
BKE16960	168 *ymdA::erm*	From Byoung-Mo Koo
AJD204	3610, *Δrny* (markerless)	This study
AJD209	3610, *Δrny*, P_{epsA}-*lacZ::amyE cm*	This study
AJD210	3610, *Δrny* in YC1173	This study
AJD17	3610, *ylbF::mls*, P_{spank}-*MBP-ylbF::amyE spec, epsH::tet*	This study
AJD18	3610, P_{spank}-*MBP::amyE spec*, *epsH::tet*	This study
RL3855	3610, *sinIR::spec*	(Kearns *et al.*, 2005)
RL4608	P_{abrB}-*lacZ::amyE spec*	(Chu *et al.*, 2008)
FW102O_L2-62	*E. coli* Bacterial 2 hybrid reporter F plasmid with p*lac*O_L2–62-*lacZ*	(Deaconescu *et al.*, 2006)
AJD221	FW102O_L2-62 pACCI, pBRαNTD	This study
AJD222	FW102O_L2-62 pACCI-βflap, pBRαNTD	This study
AJD223	FW102O_L2-62 pACCI, pBRαNTD-σ^{70}	This study
AJD224	FW102O_L2-62 pAC-βflap-CI, pBRσ^{70}-αNTD	This study
AJD225	FW102O_L2-62 pACCI-YlbF pBRαNTD	This study
AJD226	FW102O_L2-62 pACCI-YmcA pBRαNTD	This study

AJD227	FW102O_L2-62 pACCI-Rny pBRαNTD	This study
AJD228	FW102O_L2-62 pACCI-Spo0B pBRαNTD	This study
AJD229	FW102O_L2-62 pACCI pBRαNTD-YlbF	This study
AJD230	FW102O_L2-62 pACCI pBRαNTD-YmcA	This study
AJD231	FW102O_L2-62 pACCI pBRαNTD-Rny	This study
AJD232	FW102O_L2-62 pACCI pBRαNTD-Spo0B	This study
AJD233	FW102O_L2-62 pACCI-βflap pBRαNTD-Rny	This study
AJD234	FW102O_L2-62 pACCI-YlbF pBRαNTD-σ^{70}	This study
AJD235	FW102O_L2-62 pACCI-YmcA pBRαNTD-σ^{70}	This study
AJD236	FW102O_L2-62 pACCI-Rny pBRαNTD-σ^{70}	This study
AJD237	FW102O_L2-62 pACCI-Rny pBRαNTD-Rny	This study
AJD238	FW102O_L2-62 pACCI-YlbF pBRαNTD-YmcA	This study
AJD239	FW102O_L2-62 pACCI-YmcA pBRαNTD-Rny	This study
AJD240	FW102O_L2-62 pACCI-YmcA pBRαNTD-Spo0B	This study
AJD241	FW102O_L2-62 pACCI-Spo0B pBRαNTD-YmcA	This study

A.12 list of plasmids and primers

Primers: all lister 5'->3'
yaaT markerless deletion:
AJDPr73 5'-AACAGCTATGACCATGATTACGCC**AAGCTT**GATACAATGGCAATATCCTGGAAG-3' **HindIII**
AJDPr74 5'- GCTTATCCCTCCTGCAACATTAAC -3'
AJDPr75 5'- GTTAATGTTGCAGGAGGGATAAGCCGAGGTGTGGAACCTTGGAT -3'
AJDPr76 5'- CGTTGTAAAACGACGGCCAGT**GAATTC**GATAAGCCTTCATCATTGCAAAAATC -3' **EcoRI**
Rny complementation:
AJDPr146 5'-ATCGGGATCCGCGATCTTCTTTTGGATGAATTG-3'
AJDPr147 5'-ATCGGAATTCCAAAAAAATAAAGTGATGCTTAG-3'
MBP-YlbF
AJDPr22 5'-ATGTAAGCTTTAAGGAGGATTTATAGATGCACCATCATCATCATCATTCTTC-3'
AJDPr23 5'-TCGTGCTAGCGGACACTTTACATCCGCA-3'
MBP
AJDPr24 5'-TCGTGCTAGCTCATCCACTTCCAATATTGGATTGGAA-3'
Northern probe SinR
VD30 5'-TTGATTGGCCAGCGTATTAAACAAT-3'
VD31 5'-GAAATT**AATACGACTCACTATAGG**GAGACTCCTCTTTTTGGGATTTTCTCCAT-3' **T7 promoter**
Northern probe SinI
VD32 5'-TGGTGATTTAATGGCAAATGA-3'
VD33 5'-GAAATT**AATACGACTCACTATAGG**GAGAGAAAGGATTTACGGTATGACTT-3' **T7 promoter**
Northern Probe YqhG

VD34 5'-AAAAGGATAAGTTATTGTCACTAGG-3'
VD35 5'-GAAATT**AATACGACTCACTATAGG**GAGAAGCTTGAAATATTATTTTGCAGATA-3' **T7 promoter**
PsinR-lacZ
PsinRF1 5'- TGGCGAATCCGCAAAACAAGAGCACTTTGAATTGGA -3'
PsinrR1 5'- GCCAGGATCCTCCTTGTGATATTATAGCACATTCAGA -3'
Northern probe cggR:
cggRF: GATGTTTTAGGTTTGACTCTTTTGGAAAAG
cggRR: GAAATT**AATACGACTCACTATAGG**GACTTTAAATCTTCTAAAGGC **T7 promoter**
Northern probe gapA:
gapAF: CGATGCGCTAACCACGATGTTA
gapAR: GAAATT**AATACGACTCACTATAGG**GACCATAACCATTGTAGAAAG **T7 promoter**
Primers for bacterial 2 hybrid:
pACYlbFF AGACGTTTGGC**GCGGCCGC**AATGTATGCGACGATGGAATCC **NotI**
pACylbFR GATGCAAGATCTGTAAGGTAA**GGATCC**TTAGGACACTTTACATCCGCAGC **BamHI**
pACymcAF AGACGTTTGGC**GCGGCCGC**AATGACGCTCTACTCAAAAAAAGACATTG **NotI**
pACymcAR :GATGCAAGATCTGTAAGGTAA**GGATCC**TTAGAGAGAACAGCTGTTATTTGAATC **BamHI**
pACrnyF AGACGTTTGG**CGCGGCC**GCAAAAACCATTGCCGAAGCGAAAATTG **NotI**
pACrnyR GATGCAAGATCTGTAAGGTAA**GGATCC**TTATTTTGCATACTCTACGGCTC **BamHI**
pACspo0BF: AGACGTTTGGC**GCGGCCGC**AATGAAGGATGTTTCAAAAAATCA **NotI**
pACspo0BR: GATGCAAGATCTGTAAGGTAA**GGATCCC**TAGTCCAACCCAATTTCAATC **BamHI**
pBRylbFF: GAGAAACCAGAG**GCGGCCGC**AATGTATGCGACGATGGAATCC **NotI**
pBRylbFR TGCGCCGGCGTAGA**GGATCC**TTAGGACACTTTACATCCGCAGC **BamHI**
pBRymcAF GAGAAACCAGAG**GCGGCCGC**AATGACGCTCTACTCAAAAAAAGACATTG **NotI**
pBRymcAR TGCGTCCGGCGTAGA**GGATCC**TTAGAGAGAACAGCTGTTATTTGAATG **BamHI**
pBRrnyF GAGAAACCAGAG**GCGGCCGC**AAAAACCATTGCCGAAGCGAAAATTG **NotI**
pBRrnyR TGCGTCCGGCGTAGA**GGATCC**TTATTTTGCATACTCTACGGCTC **BamHI**
pBRspo0BF GAGAAACCAGCG**GCGGCCGC**AATGAAGGATGTTTCAAAAAATCA **NotI**
pBRspo0BR TGCGTCCGGCGTAGA**GGATCC**CTAGTCCAACCCAATTTCAATC **BamHI**

Plasmids:
-**Pminimad2**
-**pDG364**
-**pDR110**
-**Pminimad*ΔyaaT*** mls, Amp This study
- **pBKACL-cre**: Cre recombinase under Pspank, Spec Amp, From Byoung-Mo Koo
-***amyE::P$_{spank}$-MBPylbF*** Spec, Amp This study
Coding sequence for MBPYlbF was amplified from plasmid PYC and cloned into pdr111 at the NheI and EcorI sites using primers AJDPr22 and AJDPr23.
-***amyE::P$_{spank}$-MBP*** Spec, Amp
Coding sequence for MBPYlbF was amplified from plasmid PYC and cloned into pdr110 at the NheI and EcorI sites using primers AJDPr22 and AJDPr24.
-***amyE::P$_{rny}$-rny*** Spec, Amp This study
The promoter and coding sequence of Rny was cloned into pDG364 at the XhoI and EcoRI sites using primers AJDPr146 and AJDPr147
-***anyE::P$_{sinR}$-lacZ*** Cm, Amp This study
The promoter of *sinR* was cloned in front of *lacZ* at the EcoRI and HindIII sites using primers PsinRF1 and PsinrRR1
pACλCI Cm^R (Dove *et al.*, 1997)
P$_{lacUV5}$-directed synthesis of the λCI protein
pACλCI-β-flap Cm^R (Deighan *et al.*, 2008)
P$_{lacUV5}$-directed synthesis of the λCI protein fused via three

pBRα Amp^R (Dove *et al.*, 1997)
P_{lacUV5}/P_{lpp} -directed synthesis of the full length α subunit of *E. coli* RNAP
pBRα-σ^{70} Amp^R (Deighan *et al.*, 2008)
P_{lacUV5}/P_{lpp}- directed synthesis of the λCI protein fused directly to *E. coli* σ^{70} region 4 (residues 528-613 of σ^{70}). The σ^{70} moiety also carries the D581G substitution
pACλCI-YlbF, Cm^R, this study
The YlbF open reading frame (primers pACYlbFF and pACYlbFR) was cloned into pACλCI at the BamHI and NotI sites using isothermal assembly.
pACλCI-YmcA Cm^R, this study
The YmcA open reading frame (primers pACYmcAF and pACYmcAR) was cloned into pACλCI at the BamHI and NotI sites using isothermal assembly.
pACλCI-Rny Cm^R, this study
The Rny cytoplasmic domain open reading frame (primers pACRnyF and pACRnyR) was cloned into pACλCI at the BamHI and NotI sites using isothermal assembly.
pACλCI-Spo0B Cm^R, this study
The Spo0B open reading frame (primers pACSpo0BF and pACSpo0BR) was cloned into pACλCI at the BamHI and NotI sites using isothermal assembly.
pBRα-YlbF Amp^R, this study
The YlbF open reading frame (primers pBrYlbFF and pBrYlbFR) was cloned into pBRα at the BamHI and NotI sites using isothermal assembly.
pBRα-YmcA Amp^R, this study
The YmcA open reading frame (primers pBRYmcAF and pABRYmcAR) was cloned into pACλCI at the BamHI and NotI sites using isothermal assembly.
pBRα-Rny Amp^R, this study
The Rny cytoplasmic domain open reading frame (primers pBRRnyF and pBRRnyR) was cloned into pACλCI at the BamHI and NotI sites using isothermal assembly.
pBRα-Spo0B Amp^R, this study
The Spo0B open reading frame (primers pBRSpo0BF and pBRSpo0BR) was cloned into pACλCI at the BamHI and NotI sites using isothermal assembly.

Appendix B: A serine Sensor for Multicellularity in a Bacterium

A Serine Sensor for Multicellularity in a Bacterium

Arvind R. Subramaniam[1, 2], Aaron DeLoughery[1], Niels Bradshaw[1], Yun Chen[1], Erin O'Shea[1, 2, 3, 4], Richard Losick*[1], and Yunrong Chai*[1, 5]

Abstract

We report the discovery of a simple environmental sensing mechanism for biofilm formation in the bacterium *Bacillus subtilis* that operates without the involvement of a dedicated RNA or protein. Certain serine codons, the four UCN codons, in the gene for the biofilm repressor SinR caused a lowering of SinR levels under biofilm-inducing conditions. Synonymous substitutions of these UCN codons with AGC or AGU impaired biofilm formation and gene expression. Conversely, switching AGC or AGU to UCN codons upregulated biofilm formation. Genome-wide ribosome profiling showed that ribosomes paused longer at UCN codons than at AGC or AGU during biofilm formation. Serine starvation recapitulated the effect of biofilm-inducing conditions on ribosome pausing and SinR production. As serine is one of the first amino acids to be exhausted at the end of exponential phase growth, ribosome pausing at serine codons may be exploited by other microbes in adapting to stationary phase.

Introduction

Bacteria constantly monitor their environment and internal physiological state so that they can adapt to changing conditions. A wide variety of sensing mechanisms are deployed for this purpose, including dedicated protein sensors, such as histidine kinases, which mediate changes in gene expression by controlling the phosphorylation of cognate response regulators in response to environmental cues (*1*). Bacteria also sense changes in their environment and physiology by means of dedicated RNAs, such as the highly structured, leader RNA for the tryptophan operon, which controls the transcription of downstream genes in the operon by a mechanism involving ribosome stalling at tryptophan codons (*2*). Here we report the discovery of an unusually simple mechanism of environmental sensing involved in the process of biofilm formation by the bacterium *Bacillus subtilis* that does not require a dedicated RNA or protein.

Biofilm formation involves a switch from planktonic growth as individual cells to the formation of complex, multicellular communities in response to environmental cues (*3*). In *B. subtilis*, these communities are embedded in a self-produced matrix consisting of polysaccharide and an amyloid-like protein, which are specified by the *epsA-O* and the *tapA-sipW-tasA* operons, respectively (*4*, *5*). The transition to multicellularity is governed in part by four histidine kinases (KinA, KinB, KinC and KinD) that control the phosphorylation of the response regulator, Spo0A, a master regulator of post-exponential phase gene expression (Figure 19A) (*6*, *7*). Recent studies suggest that KinA and KinB respond to impaired respiration (*8*), whereas KinC responds to membrane perturbations and KinD to unknown chemical signals (*9–12*). Once phosphorylated, Spo0A turns on *sinI*, a gene encoding a small protein antagonist of the biofilm-specific regulatory protein SinR (*5*, *13*). SinR, which is produced constitutively, is a repressor of the matrix operons, *epsA-O* and the *tapA-sipW-tasA*, as well as other biofilm-related genes (*5*, *14*, *15*). SinR is also a repressor of the gene for SlrR (*16*), which together with SinR sets up a self-reinforcing, double-negative feedback loop for matrix gene expression (Figure 19A) (*17*, *18*). A special feature of SinR of relevance to this investigation is that the expression of matrix genes is hypersensitive to small perturbations in the level of the protein (*19*). This hypersensitivity is attributed to

molecular titration of SinR by SinI and cooperativity among SinR molecules bound to tandem target sequences at regulatory sites for the matrix operons (*19*, *20*).

In the current study, we present evidence for the existence of a novel cellular sensing mechanism controlling biofilm formation. Rather than relying on regulation by a dedicated RNA or protein, ribosomes pause at certain serine codons, resulting in lower SinR levels, which as a consequence, contribute to derepression of SinR-controlled genes. We propose that specific serine codons in the *sinR* gene act as a simple sensor for monitoring, and triggering a response to, serine depletion under biofilm-inducing conditions.

Results

Switching synonymous serine codons in *sinR* affects biofilm formation.

In a genetic screen to identify suppressor mutations that rescued the biofilm-defective phenotype of a *B. subtilis* mutant (Methods), we unexpectedly recovered a variant that contained a 'silent' mutation that resulted in a switch from one serine codon to a synonymous codon in *sinR*. This observation prompted us to ask whether switching other serine codons might also influence biofilm formation. Serine is specified by six codons: AGC, AGT, TCA, TCC, TCG and TCT (where T is U in the mRNA). We noticed that the *sinR* coding sequence has a slightly higher frequency of the four TCN serine codons as compared to the average frequency of TCN codons in the *B. subtilis* genome (Figure 1B, P =0.22, N=12). To test whether this bias towards TCN codons has an effect on biofilm formation, we systematically replaced the three TCA codons in *sinR* (Figure 19 and 20) with each of the other five serine codons. Replacing the TCA codons with AGT or AGC resulted in flat, featureless colonies on solid biofilm-inducing medium, indicating severely impaired biofilm formation (Figure 19C). In contrast, replacing the TCA codons with either TCC or TCT had little or no effect on colony morphology whereas switching to TCG increased the wrinkled appearance of the colonies (Figure 19C), which is indicative of robust biofilm formation.

Next, we asked whether switching serine codons was altering the level of SinR protein in liquid biofilm-inducing medium. Immunoblot analysis with anti-SinR antibodies revealed slightly yet

consistently higher SinR levels in the strain with the AGT variant of *sinR* when compared to either the wild-type strain with three TCA codons or the TCG variant of *sinR* (Figures 19D, 19E).

SinR is highly similar (85% identity) to SlrR, which also plays a critical role in biofilm formation and whose gene (*slrR*) is under the direct negative control of SinR (*16*). Because SlrR cross reacts with the anti-SinR antibodies, we were also able to detect SlrR in our immunoblot analysis. Strikingly, the levels of SlrR were almost perfectly anti-correlated with those of SinR, with the differences in the SlrR protein levels among the *sinR* synonymous variants being much higher than the corresponding differences in SinR protein levels (Figure 19F). Because repression by SinR is ultrasensitive to SinR levels (*19*), small differences in SinR protein levels among *sinR* synonymous variants might be sufficient to cause large differences in the levels of expression of SinR-repressed genes such as *slrR*. Consistent with this idea, an *eps-lacZ* transcriptional fusion reporter for the SinR-repressed *epsA-O* matrix operon showed that the four TCN *sinR* variants had 3- to 19-fold higher *b*-galactosidase activity than the AGT and AGC variants (Figure 19G).

Figure 19. Switching synonymous serine codons in *sinR* affects biofilm formation.

(A) Regulatory circuit controlling biofilm formation in *B. subtilis*. **(B)** Top: Serine codon usage in the *sinR* coding sequence. Number within parenthesis indicates the frequency of the corresponding codon in *sinR*. Bottom: Average serine codon usage across 4153 protein-coding sequences in the *B. subtilis* genome. Number within parenthesis indicates the relative frequency of each codon in the genome. **(C)** Colony morphology for the wild-type strain and the indicated *sinR* synonymous variants grown on solid biofilm-inducing medium. Three TCA codons in the wild-type sequence of *sinR* were switched to each of the other five serine codons. The wild-type (WT) *sinR* sequence was replaced by the *sinR* synonymous mutant at the native *sinR* locus of the strain 3610. **(D)** SinR protein level during entry into biofilm formation (OD_{600} = 2) measured using an anti-SinR antibody that also cross-reacts with SlrR, a protein that is 85% identical to SinR. Western blot against the RNA polymerase subunit SigA was used as the loading control. Whole cell lysates were loaded at different dilutions (indicated as X, X/2, and X/3). **(E)** Densitometry of SinR bands in (C) after normalization by SigA. **(F)** Top panel: Western blot against SinR and SlrR using anti-SinR antibody. Bottom panel: Densitometry ratio of the SlrR and SinR bands in the top panel. Error bars represent standard error over three replicate Western blots. The SlrR/SinR ratio for each blot was normalized such that the wild-type strain had a ratio of 1. **(G)** Matrix gene expression monitored using a P_{epsA}-*lacZ* transcriptional reporter inserted at the chromosomal *amyE* locus. *b*-galactosidase activity was measured at OD_{600} = 2 in liquid biofilm-inducing medium. Error bars represent standard error of three measurements.

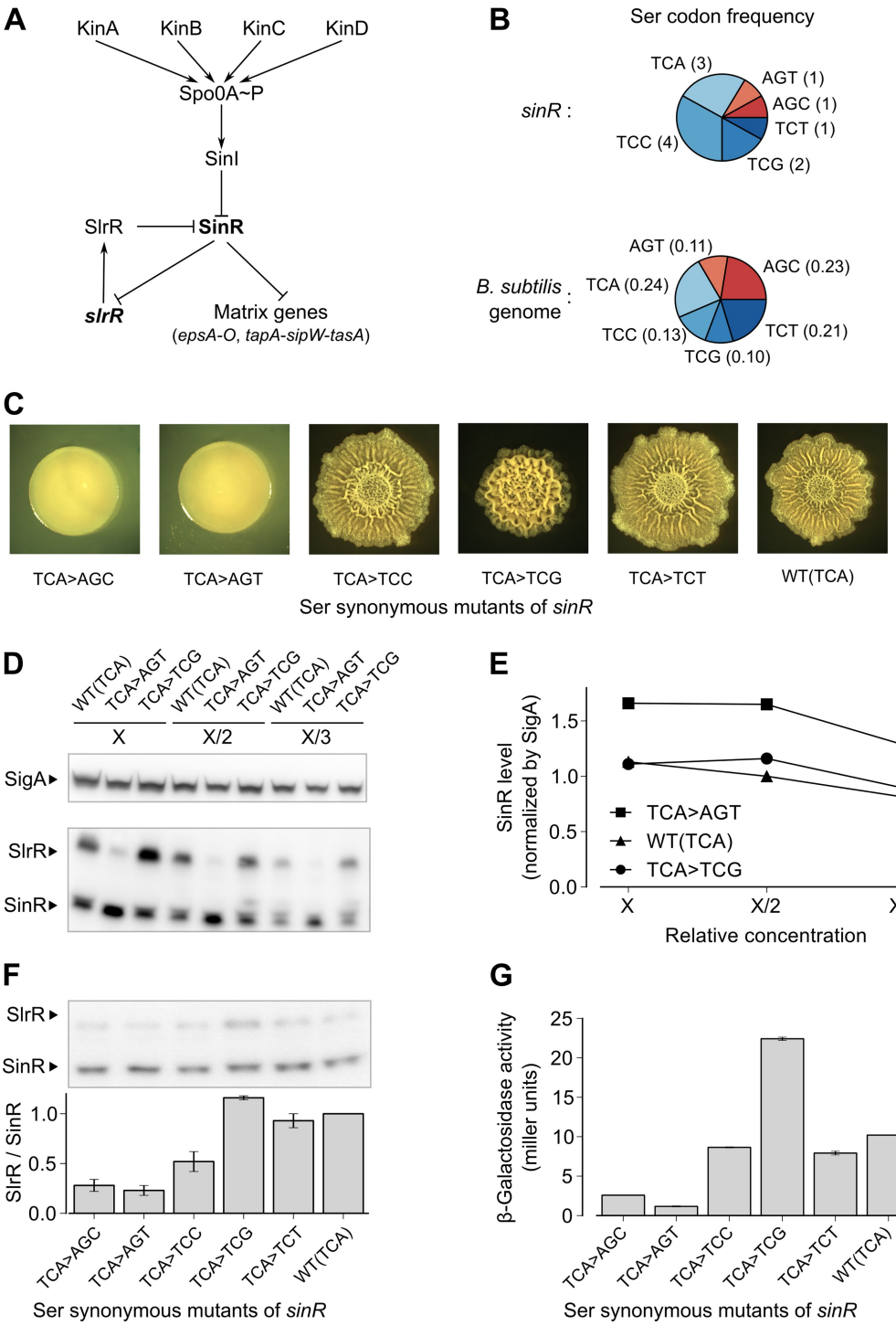

Figure 19 continued.

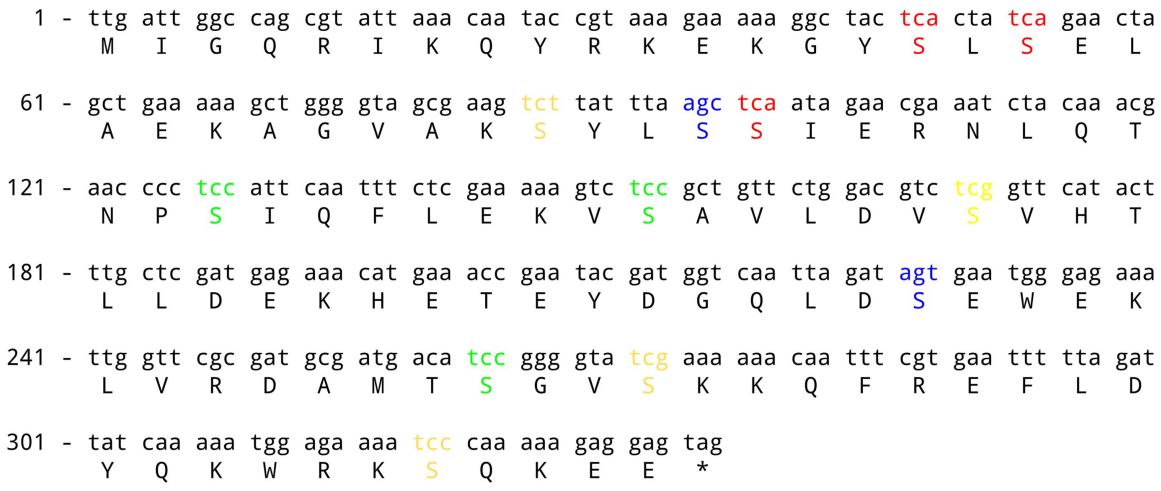

Figure 20. *sinR* coding sequence. The three TCA codons (switched in Figure 1) are highlighted in red. The three TCC codons and the two AGC/AGT codons (switched in Figure 1–figure supplement 2) are highlighted in green and blue respectively. The remaining serine codons are shown in yellow.

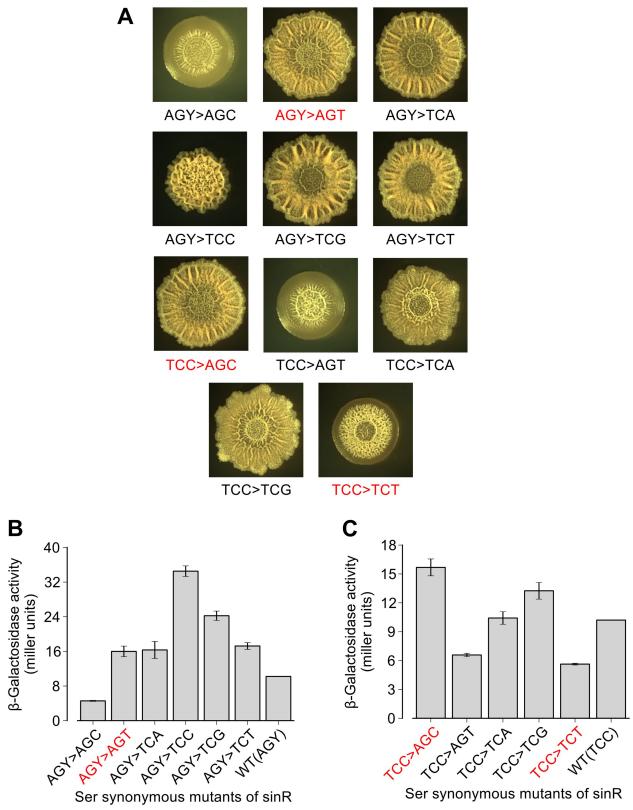

Figure 21. Effect of TCC and AGC/AGT synonymous substitutions in the *sinR* gene on colony morphology and biofilm reporter activity. (A) Colony morphology for the indicated *sinR* synonymous variants grown on solid biofilm-inducing medium. Either three TCC codons or two AGY codons in the wild-type sequence of *sinR* were switched to remaining serine synonymous codons. The wild-type (WT) *sinR* sequence was replaced by the *sinR* synonymous variant at the native *sinR* locus of the strain 3610. Colony morphology of the wild-type strain is shown in Figure 1. **(B, C)** Matrix gene expression monitored using a P_{epsA}-*lacZ* transcriptional reporter inserted at the chromosomal *amyE* locus. Strains were grown in liquid biofilm-inducing medium and *b*-galactosidase activity was measured at an OD_{600} = 2. Error bars represent standard error of three measurements. The synonymous variants highlighted in red do not follow the hierarchy between TCN and AGY codons seen for the six TCA synonymous variants in Figure 1.

To test the generality of the observed hierarchy between the synonymous variants of *sinR*, we generated an additional set of eleven *sinR* synonymous variants in which we replaced either three TCC codons or two AGC/AGT codons (Figure 20) with their synonymous counterparts. Eight of these variants conformed to the hierarchy described above, namely, the four TCN variants behaved oppositely to the two AGC/AGT variants in colony morphology and in *eps-lacZ* reporter expression (Figure 21). The three variants that did not conform to the hierarchy could potentially reflect alterations to the mRNA sequence context near the mutation rather than the effect of a synonymous substitution *per se*. Taken together, the above results suggest that serine synonymous codons in the *sinR* coding sequence have a stereotypical effect on biofilm formation that is primarily determined by the differential usage of the four TCN and the two AGC/AGT codons.

Entry into biofilm formation is accompanied by codon-specific ribosome pausing.

What is the mechanism by which serine codon usage affects SinR protein levels and biofilm formation? Synonymous codon changes can alter the synthesis of the encoded protein through changes in the translation initiation rate, mRNA levels or the ribosome elongation rate (*21*). However, the effect of synonymous codon usage on the initiation rate and mRNA levels is context-specific; only codons near the AUG start site affect translation initiation (*22*), whereas only codons that are located in certain regions of secondary structure or at ribonuclease cleavage sites affect mRNA levels (*23*). Our observation that synonymous codon replacements at multiple locations along *sinR* have a stereotypical effect on biofilm formation argues against such context-specific mechanisms (except for the three exceptional cases noted above).

To test the alternative hypothesis that serine codon usage might alter the ribosome elongation rate, and given that the ribosome elongation rate at a codon varies inversely with the average ribosome density at that codon, we measured ribosome density on mRNAs at single-codon resolution using the ribosome profiling method (*24–26*). We grew *B. subtilis* in liquid biofilm-inducing medium, harvested cells either during exponential phase growth (OD_{600} = 0.6) or during stationary phase when biofilm

formation is induced (OD_{600} = 1.4), and performed deep-sequencing of ribosome protected mRNA fragments and size-matched total mRNA fragments. Ribosome profiling yielded 3.75 and 2.55 million sequencing reads aligning to annotated protein-coding sequences for the exponential phase sample and the biofilm entry sample, respectively. The number of reads aligning to a single codon on individual mRNAs was too low for accurate quantification of ribosome density, and was not sufficient to directly detect ribosome pausing on the *sinR* transcript. However, we reasoned that the global pattern of ribosome pausing across all mRNAs should reflect ribosome pausing on individual transcripts such as *sinR*. Therefore, we calculated the median A-site ribosome density for each of the 61 sense codons across the 1556 protein-coding sequences in the exponential phase sample and the 1148 sequences in the biofilm entry sample that had an average coverage greater than one sequencing read per codon (Figure 2–figure supplement 1 and methods). This analysis reproduced the previous observation (*27*) of ribosome pausing 8 to 11nt downstream of Shine-Dalgarno-like trinucleotide sequences both during exponential phase and during biofilm entry (Figure 22 C and D).

During exponential phase, median A-site ribosome density varied over a two-fold range, with no systematic difference between serine codons and the remaining codons (Figure 22A). By contrast, during biofilm entry, median A-site ribosome density was significantly higher at serine and cysteine codons as compared to the remaining codons (Figure 22B), suggesting that ribosomes selectively pause at these codons during biofilm entry. Notably, the ribosome density at serine codons was not uniform: the four TCN codons had 1.9 to 2.1-fold higher ribosome density whereas the AGC and AGT codons had only 1.1 and 1.3-fold higher ribosome density respectively, relative to the median value across 61 sense codons. Further, this difference in ribosome density between TCN codons and the AGC/AGT codons was essentially identical when computed separately for codons located in the first half or in the second half of each gene (Figure 22 E and F), a finding that underscores the statistical robustness and the context independence of the observed difference.

Figure 22. Entry into biofilm formation is accompanied by codon-specific ribosome pausing.

Genome-wide median A-site ribosome density and A-site total mRNA density for 61 sense codons (excluding start and stop codons) **(A)** during exponential phase growth (OD_{600} = 0.6), and **(B)** during stationary phase when biofilm formation is induced (OD_{600} = 1.4). The six serine (red) and two cysteine (green) codons are highlighted. Genome-wide ribosome density and total mRNA density were measured by deep-sequencing of ribosome-protected mRNA fragments and total mRNA fragments respectively, of a *B. subtilis* 3610 derivative RL3852 (D*epsH*) grown in liquid biofilm-inducing medium. **(C)** Median ribosome density across all protein coding sequences was computed for the 60 nt region around each of six Shine-Dalgarno-like trinucleotide sequences (*27*) for the exponential phase sample (**C**) and the biofilm entry sample (**D**). **E.** Each gene was conceptually divided into two equal halves and the A-site ribosome density and A-site mRNA density was computed separately for codons located either in the first half (**E**) or in the second half (**F**) of each gene.

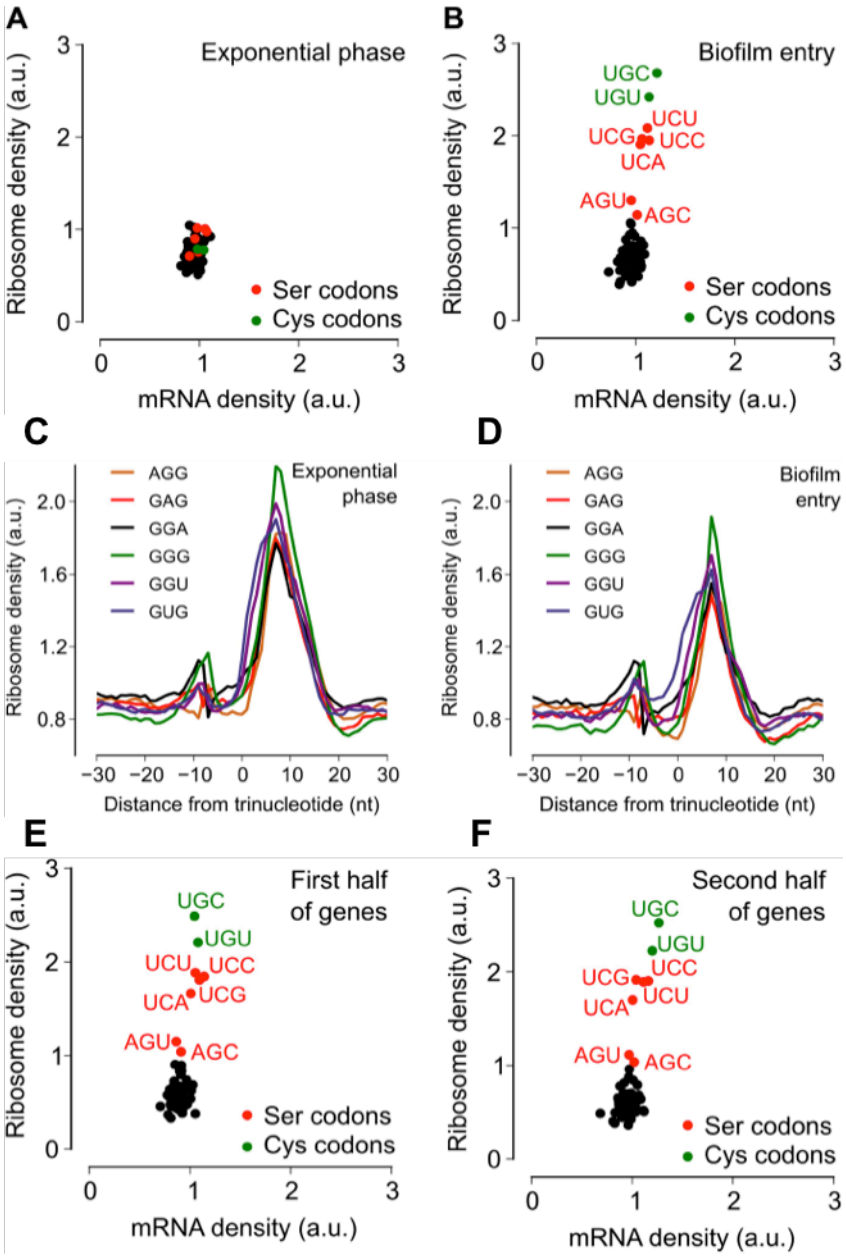

Figure 22 continued.

Serine starvation causes ribosome pausing and inhibits SinR synthesis in a codon-specific manner.

Because ribosomes paused at serine and cysteine codons only during biofilm formation and not during exponential phase growth, we hypothesized that pausing was caused by the depletion of intracellular pools of these two amino acids during biofilm entry rather than by any intrinsic feature of the mRNA (*27*) or the nascent polypeptide (*28*). Our hypothesis is also supported by the previous observation that synonymous codon usage can have a starvation-specific effect on protein levels (*29*). Further, serine is a precursor metabolite for the biosynthesis of cysteine (*30*); hence cysteine depletion was likely the result of a decrease in intracellular serine concentration. Consistent with this hypothesis, we also observed ribosome pausing at both serine and cysteine codons during serine starvation of a *B. subtilis* serine-auxotrophic mutant (Figure 23C). Importantly, serine starvation resulted in ribosome pausing at only the four TCN serine codons but not at the AGC and AGT codons, matching the hierarchy seen during biofilm entry (Figure 22B). Serine starvation also resulted in differential levels of production of SinR-YFP protein fusions bearing different synonymous serine codons, whereas serine-rich growth resulted in identical levels of fusion protein production from these variants (Figures 23A, 23B). Finally, the addition of excess serine or cysteine (but not any of the other 18 amino acids) blocked biofilm formation in wild type cells as judged after 48 hours of growth in biofilm-inducing medium (Figure 24).

Given that both biofilm entry and serine starvation resulted in ribosome pausing at serine and cysteine codons, we asked whether these two apparently different conditions invoke the same gene expression program in *B. subtilis*. The fold-change in average ribosome density of *B. subtilis* genes was positively correlated between biofilm entry and serine starvation (Figure 3D, $R^2 = 0.27$, $P = 10^{-5}$, 1724 genes). However, a subset of 68 genes was induced at least 10-fold higher upon biofilm entry than during serine starvation (indicated by red markers in Figure 23D, Table B1). This subset included genes for anaerobic metabolism such as *lctEP*, *nasDE*, and *cydAB*, and is consistent with the recently proposed role of impaired respiration in biofilm formation (*8*). We observed that sulfur metabolism genes were also

enriched in this subset, possibly indicating a stronger response to cysteine depletion during biofilm entry than during serine starvation.

In toto, the results with the serine auxotroph support the inference that ribosome stalling observed during biofilm formation is due to a drop in intracellular serine levels. Our efforts to measure intracellular serine levels directly during growth in minimal, biofilm-inducing medium (MSgg) have been unsuccessful. Hence, we cannot rule out the less likely possibilities that biofilm entry causes serine and cysteine to be sequestered away from protein synthesis or that aminoacylation rate of the corresponding tRNAs decreases without changes in the intracellular serine and cysteine pools.

Figure 23. Serine starvation causes ribosome pausing and inhibits SinR synthesis in a codon-specific manner. (A, B) Three *sinR* synonymous variants were synthesized with ten serine codons switched to AGC, TCA or TCG. The variants were expressed as SinR-YFP fusions from the *amyE* locus under the control of a *lac* promoter in a 3610-D*serA* serine auxotroph strain growing in serine-limited medium. Black arrow around 300 min indicates the onset of serine starvation caused by depletion of exogenously-added serine in the growth medium. Cell density (A) and the corresponding SinR-YFP protein level (B) were monitored using a 96-well spectrophotometer. **(C)** Genome-wide median A-site ribosome density for 61 sense codons (excluding start and stop codons) during serine starvation (vertical axis) and serine-rich growth (horizontal axis) of a serine auxotrophic strain. **(D)** Fold-change in average ribosome density for individual genes upon biofilm entry (vertical axis) or serine starvation (horizontal axis). Genes that were preferentially up-regulated at least 10-fold upon biofilm entry in comparison to serine starvation are highlighted in red (68 genes, Table 1). Only genes with a minimum of 100 ribosome profiling reads in at least one of the samples were included in this analysis (1724 genes) and the reported log2 fold-changes are median-subtracted values across this gene set

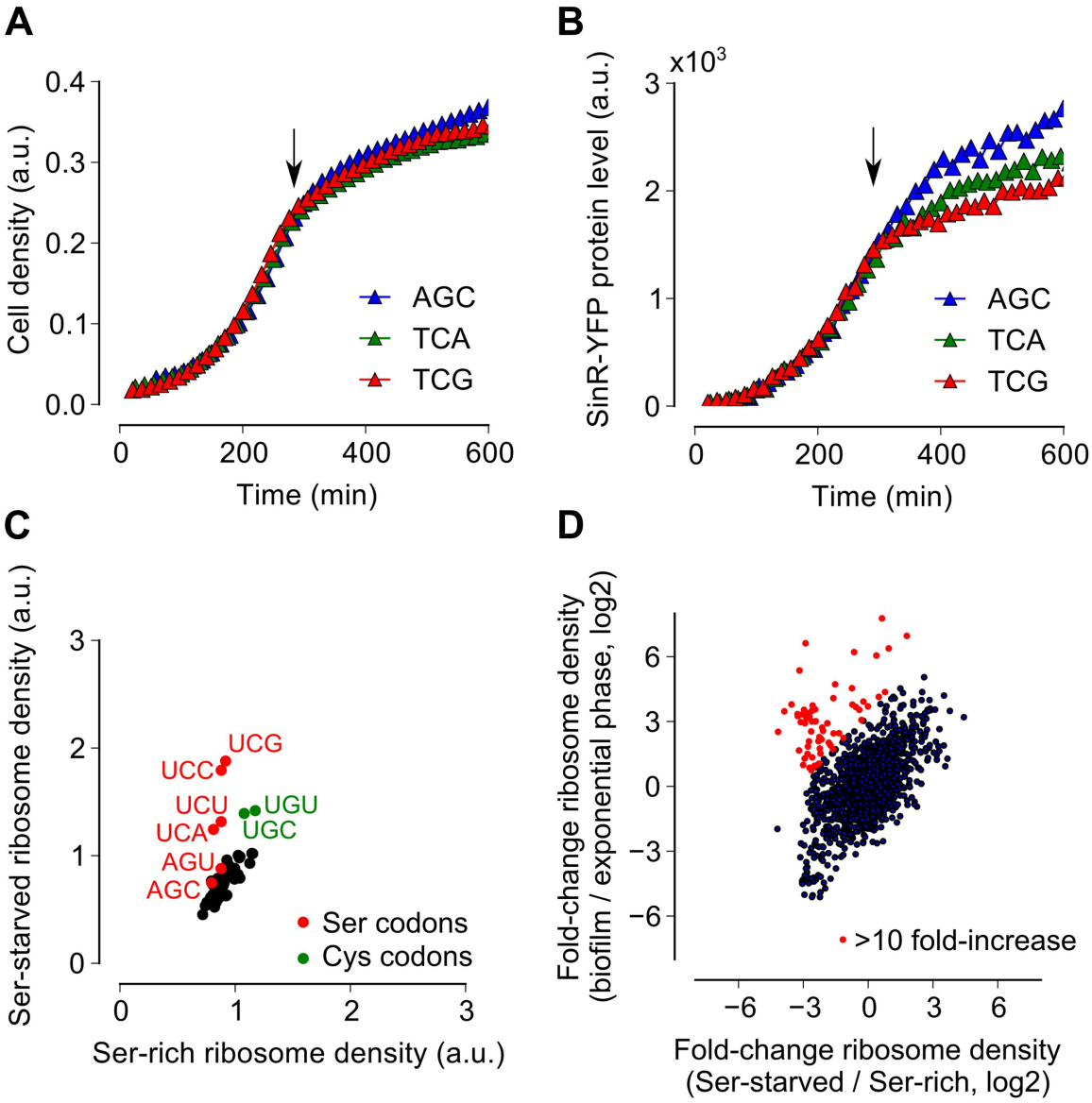

Figure 23 continued.

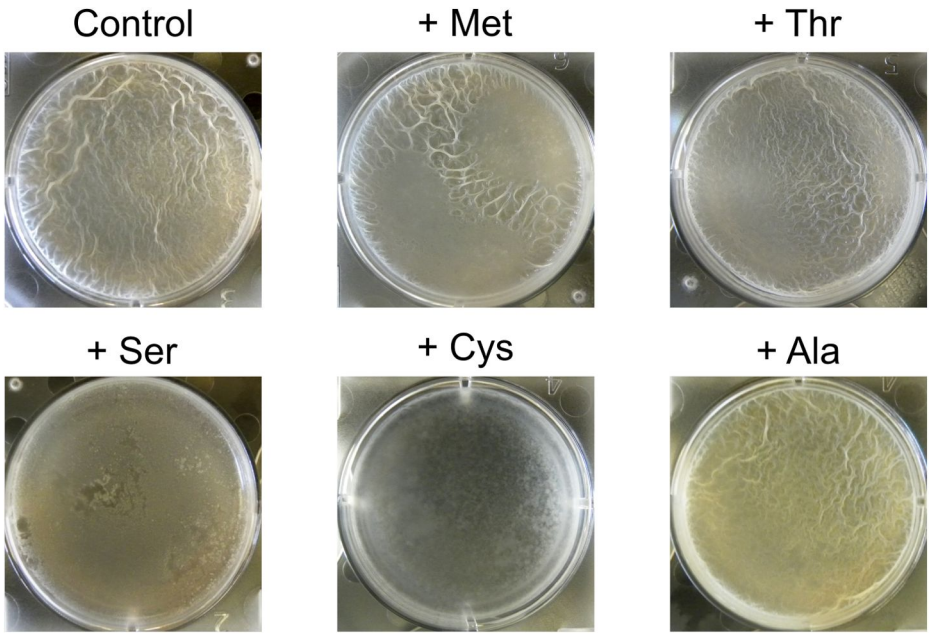

Figure 24. Addition of excess serine or cysteine blocks pellicle formation by *B. subtilis*. Single amino acids were added at 300 μg ml^{-1} to liquid MSgg medium. Biofilm formation of 3610 was assayed visually by pellicle formation at the air-liquid interface 48 hours after inoculation. Serine and cysteine were found to block pellicle formation out of all twenty amino acids tested.

Serine codon bias in biofilm-regulated genes reflects their expression during serine starvation.

Here we found that ribosomes pause at TCN serine codons (UCN in mRNA) and thereby modulate production of SinR and entry into biofilm formation. Based on our genome-wide measurements of ribosome density, we expect that significant ribosome pausing should occur during biofilm entry not only on the *sinR* mRNA, but also on other mRNAs that are enriched for any of the four UCN codons. For example, we found that the four TCN codons are over-represented in two nucleotide biosynthesis genes, *pyrAA* and *purB* (Figure 25A). These two genes are also transcriptionally down regulated during biofilm entry (Figure 25B). A high frequency of TCN codons in these genes might serve to reinforce their transcriptional down regulation by causing ribosome pausing. Conversely, TCN codons are under-represented in the genes encoding lactate dehydrogenase, *ldh* (*31*) and a master regulator of post-exponential phase gene expression, *spo0A* (*13*) (Figure 25A). The two genes are transcriptionally up regulated upon biofilm entry (Figure 25B). The low frequency of TCN codons in these genes might represent a mechanism for optimizing the production of their protein products by minimizing ribosome pausing during biofilm entry. Consistent with this idea, replacement of AGC/AGT codons by TCN codons in *spo0A*, whose protein product positively regulates biofilm entry (by turning on the synthesis of the SinR antagonist SinI), resulted in defective biofilm formation (Figure 25C) in contrast to the stimulatory effect of replacing AGC/AGT codons with TCN codons in *sinR* (Figure 21).

Figure 25. Serine codon bias of biofilm-regulated genes reflects their expression under serine starvation. (A) Relative serine codon fraction in genes for nucleotide biosynthesis (*pyrAA*, *purB*), lactate dehydrogenase (*ldh*) and a sporulation regulator (*spo0A*). Numbers in parantheses indicate the number of serine codons in each gene. Relative fraction of serine codons across the *B. subtilis* genome is shown for comparison. **(B)** Fold-change (expressed in log2 units) in average ribosome density upon biofilm entry for the four genes shown in A. **(C)** Colony morphology of a wild-type strain and two *spo0A* synonymous variants grown on solid biofilm-inducing medium. Seven AGC/AGT codons in wild-type *spo0A* were replaced by either 7 TCC codons or 3 TCC and 4 TCG codons and inserted at the chromosomal *spo0A* locus. Both the wild-type *spo0A* and the synonymous *spo0A* variants were inserted with a downstream selection marker. **(D)** Left: Codon Adaptation Index (CAI) for the four genes shown in A. Right: Distribution of CAI values for 4153 protein-coding sequences of *B. subtilis*.

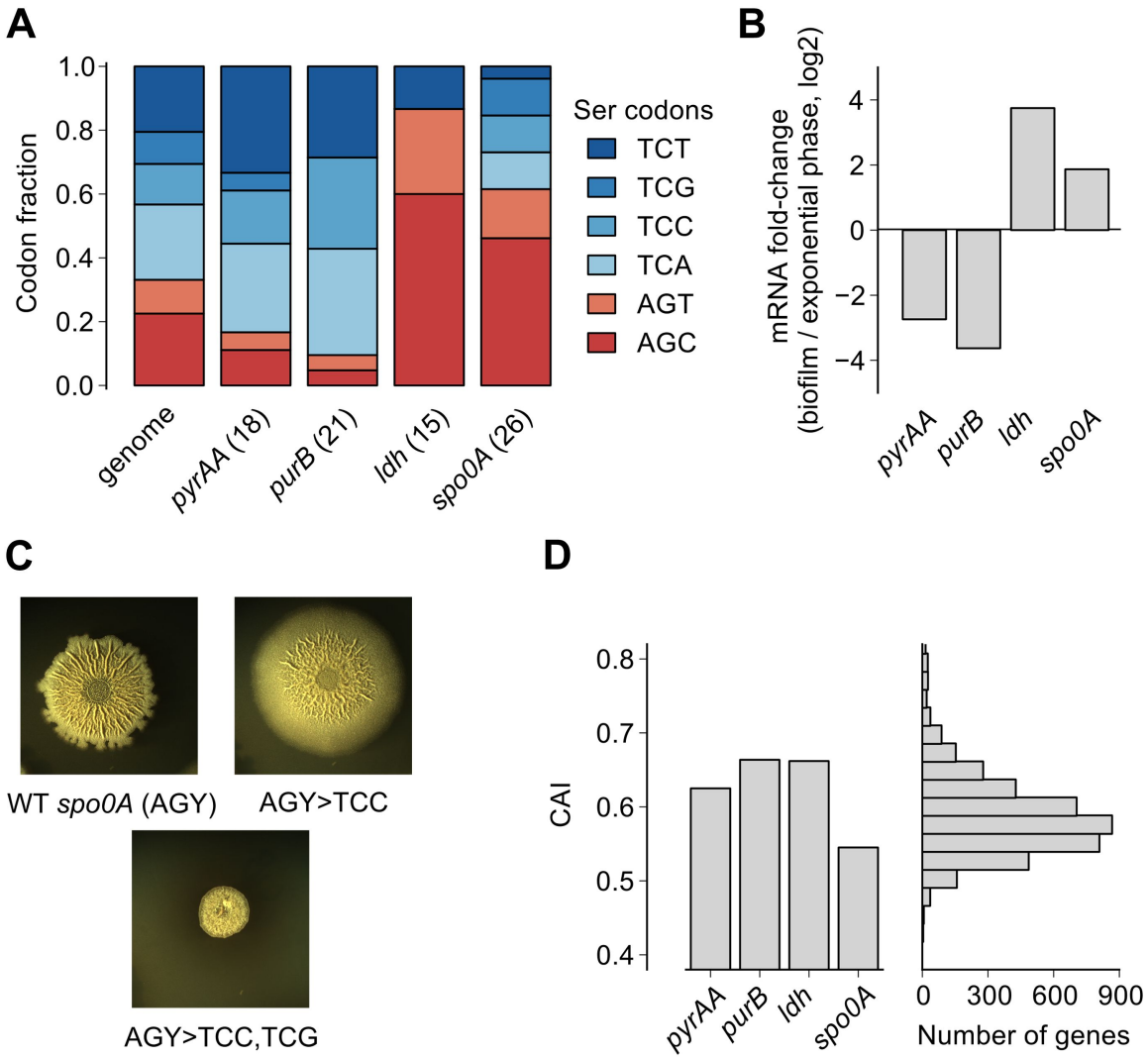

Figure 25 continued.

Discussion

Together, our results implicate serine depletion as an environmental cue that contributes to promoting biofilm formation in *B. subtilis* together with other cues that are sensed by the histidine kinases KinA–D. Serine depletion is sensed through a remarkably simple mechanism based on ribosome pausing at TCN serine codons in the gene for a regulatory protein, SinR, whose repressive effects are highly sensitive to small changes in the level of the protein. We presume that ribosome pausing at TCN lowers SinR levels simply by slowing the rate of ribosome movement along the mRNA (elongation). However, it is possible that ribosome pausing at serine sensitive codons during biofilm entry could be followed by downstream events such as ribosome rescue (*32*) or mRNA decay (*33*) that might also contribute to lowering the levels of the SinR protein.

The serine sensing mechanism uncovered here operates through over-representation of the TCN serine codons in the *sinR* gene without the necessity for any other dedicated protein or RNA for sensing serine depletion. By contrast, transcriptional attenuation, a widespread mechanism among bacteria for sensing amino acids that also relies on ribosome pausing, involves a translation-transcription coupling mechanism that is mediated by highly structured mRNAs and leader peptides (*2*).

We note that the biased usage of the four TCN serine codons, which act as starvation sensors during biofilm formation, is not evident from widely-used phenomenological measures of codon bias such as the codon adaptation index (Figure 25D), which primarily reflects codon preferences during exponential growth (*34*, *35*). The difference in ribosome pausing between the four TCN codons and the two AGC/AGT codons under biofilm-inducing conditions is likely mediated by differences in concentration of the corresponding aminoacylated tRNAs (*36*, *37*), as was recently observed in serine-starved *E. coli* (*29*). Interestingly, the hierarchy between TCN codons and AGC/AGT codons in *B. subtilis* during serine starvation is similar to the one in *E. coli* even though copy numbers of the corresponding tRNA genes have diverged significantly between these two organisms (*38*). Despite different tRNA gene copy numbers, it is possible that the relative abundances of the serine tRNA

isoacceptors are similar between the two organisms or that their relative abundances might be regulated in the same manner in response to nutrient deprivation (*39*).

Serine is one of the first amino acids to be completely consumed from the culture medium when either *B. subtilis* or *E. coli* cells are grown in complex rich medium (*36–38*).

Indeed, ribosome pausing has been observed at serine codons during growth of *E. coli* in Luria-Bertani broth (*27*). Thus the role of synonymous serine codons as starvation sensors discovered here in the specific context of biofilm formation might be a general regulatory strategy in microbes for adapting to nutrient depletion at the end of exponential phase growth. It is noteworthy that depletion of specific amino acids affects developmental transitions in several eukaryotic cells (*39–41*). It will be interesting to test whether a codon-based sensing mechanism, similar to the one found here in bacterial biofilm development, also plays a role in eukaryotic cells during amino acid depletion.

Acknowledgements

We thank P. Cluzel, A.L. Sonenshein, T. Norman, and J. Weissman for valuable comments during manuscript preparation and T. Gao for performing some experiments. Preliminary work by AS in the Cluzel lab was supported by a start-up grant from Harvard University.

References

1. A. H. West, A. M. Stock, Histidine kinases and response regulator proteins in two-component signaling systems, *Trends in Biochemical Sciences* **26**, 369–376 (2001).

2. T. M. Henkin, C. Yanofsky, Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions, *BioEssays* **24**, 700–707 (2002).

3. R. Kolter, E. P. Greenberg, Microbial sciences: The superficial life of microbes, *Nature* **441**, 300–302 (2006).

4. S. S. Branda, J. E. González-Pastor, S. Ben-Yehuda, R. Losick, R. Kolter, Fruiting body formation by Bacillus subtilis, *PNAS* **98**, 11621–11626 (2001).

5. D. B. Kearns, F. Chu, S. S. Branda, R. Kolter, R. Losick, A master regulator for biofilm formation by Bacillus subtilis, *Mol. Microbiol.* **55**, 739–749 (2005).

6. M. Jiang, W. Shao, M. Perego, J. A. Hoch, Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis, *Molecular Microbiology* **38**, 535–542 (2000).

7. H. Vlamakis, Y. Chai, P. Beauregard, R. Losick, R. Kolter, Sticking together: building a biofilm the Bacillus subtilis way, *Nat Rev Micro* **11**, 157–168 (2013).

8. I. Kolodkin-Gal *et al.*, Respiration control of multicellularity in Bacillus subtilis by a complex of the cytochrome chain with a membrane-embedded histidine kinase, *Genes Dev.* **27**, 887–899 (2013).

9. D. López, M. A. Fischbach, F. Chu, R. Losick, R. Kolter, Structurally diverse natural products that cause potassium leakage trigger multicellularity in Bacillus subtilis, *PNAS* **106**, 280–285 (2009).

10. M. Shemesh, R. Kolter, R. Losick, The Biocide Chlorine Dioxide Stimulates Biofilm Formation in Bacillus subtilis by Activation of the Histidine Kinase KinC, *J. Bacteriol.* **192**, 6352–6356 (2010).

11. Y. Chen *et al.*, A Bacillus subtilis sensor kinase involved in triggering biofilm formation on the roots of tomato plants, *Molecular Microbiology* **85**, 418–430 (2012).

12. P. B. Beauregard, Y. Chai, H. Vlamakis, R. Losick, R. Kolter, Bacillus subtilis biofilm induction by plant polysaccharides, *PNAS* (2013), doi:10.1073/pnas.1218984110.

13. V. Molle *et al.*, The Spo0A regulon of Bacillus subtilis, *Molecular Microbiology* **50**, 1683–1701 (2003).

14. F. Chu, D. B. Kearns, S. S. Branda, R. Kolter, R. Losick, Targets of the master regulator of biofilm formation in Bacillus subtilis, *Molecular Microbiology* **59**, 1216–1228 (2006).

15. Y. Chai, R. Kolter, R. Losick, A Widely Conserved Gene Cluster Required for Lactate Utilization in Bacillus subtilis and Its Involvement in Biofilm Formation, *J. Bacteriol.* **191**, 2423–2430 (2009).

16. F. Chu *et al.*, A novel regulatory protein governing biofilm formation in Bacillus subtilis, *Mol. Microbiol.* **68**, 1117–1127 (2008).

17. Y. Chai, T. Norman, R. Kolter, R. Losick, An epigenetic switch governing daughter cell separation in Bacillus subtilis, *Genes Dev.* **24**, 754–765 (2010).

18. T. Norman, N. Lord, J. Paulsson, R. Losick, Memory and modularity in cell-fate decision making, *Nature* (accepted for publication).

19. Y. Chai, T. Norman, R. Kolter, R. Losick, Evidence that metabolism and chromosome copy number control mutually exclusive cell fates in Bacillus subtilis, *The EMBO Journal* **30**, 1402–1413 (2011).

20. Y. Chai, F. Chu, R. Kolter, R. Losick, Bistability and biofilm formation in Bacillus subtilis, *Mol. Microbiol.* **67**, 254–263 (2008).

21. J. B. Plotkin, G. Kudla, Synonymous but not the same: the causes and consequences of codon bias, *Nat. Rev. Genet.* **12**, 32–42 (2010).

22. G. Kudla, A. W. Murray, D. Tollervey, J. B. Plotkin, Coding-Sequence Determinants of Gene Expression in Escherichia Coli, *Science* **324**, 255–258 (2009).

23. J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, S. N. Cohen, Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays, *PNAS* **99**, 9697–9702 (2002).

24. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling, *Science* **324**, 218–223 (2009).

25. E. Oh *et al.*, Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo, *Cell* **147**, 1295–1308 (2011).

26. N. T. Ingolia, G. A. Brar, S. Rouskin, A. M. McGeachy, J. S. Weissman, The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments, *Nat. Protoc.* **7**, 1534–1550 (2012).

27. G.-W. Li, E. Oh, J. S. Weissman, The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria, *Nature* **484**, 538–541 (2012).

28. C. A. Charneski, L. D. Hurst, Positively Charged Residues Are the Major Determinants of Ribosomal Velocity, *PLoS Biol.* **11**, e1001508 (2013).

29. A. R. Subramaniam, T. Pan, P. Cluzel, Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria, *PNAS* **110**, 2419–2424 (2013).

30. Y. Gagnon *et al.*, Clustering and co-transcription of the Bacillus subtilis genes encoding the aminoacyl-tRNA synthetases specific for glutamate and for cysteine and the first enzyme for cysteine biosynthesis., *J. Biol. Chem.* **269**, 7473–7482 (1994).

31. H. C. Ramos *et al.*, Fermentative Metabolism of Bacillus subtilis: Physiology and Regulation of Gene Expression, *J. Bacteriol.* **182**, 3072–3080 (2000).

32. K. C. Keiler, P. R. H. Waller, R. T. Sauer, Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA, *Science* **271**, 990–993 (1996).

33. C. S. Hayes, R. T. Sauer, Cleavage of the A Site mRNA Codon during Ribosome Pausing Provides a Mechanism for Translational Quality Control, *Molecular Cell* **12**, 903–911 (2003).

34. P. M. Sharp, W. H. Li, The Codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res.* **15**, 1281–1295 (1987).

35. S. G. Andersson, C. G. Kurland, Codon preferences in free-living microorganisms., *Microbiol. Rev.* **54**, 198–210 (1990).

36. J. Elf, D. Nilsson, T. Tenson, M. Ehrenberg, Selective Charging of tRNA Isoacceptors Explains Patterns of Codon Usage, *Science* **300**, 1718–1722 (2003).

37. K. A. Dittmar, M. A. Sorensen, J. Elf, M. Ehrenberg, T. Pan, Selective charging of tRNA isoacceptors induced by amino-acid starvation, *EMBO Rep* **6**, 151–157 (2005).

38. T. M. Lowe, S. R. Eddy, tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence*Nucl. Acids Res.* (1997) (available at http://gtrnadb.ucsc.edu/Baci_subt/, http://gtrnadb.ucsc.edu/Esch_coli_K12/).

39. R. H. Doi, I. Kaneko, B. Goehler, Regulation of a serine transfer RNA of Bacillus subtilis under two growth conditions., *Proc Natl Acad Sci U S A* **56**, 1548–1551 (1966).

40. P. Liebs, K. Riedel, J. P. Graba, D. Schrapel, U. Tischler, Formation of some extracellular enzymes during the exponential growth of Bacillus subtilis, *Folia Microbiol.* **33**, 88–95 (1988).

41. B. M. Prüss, J. M. Nelms, C. Park, A. J. Wolfe, Mutations in NADH:ubiquinone oxidoreductase of Escherichia coli affect growth on mixed amino acids., *J. Bacteriol.* **176**, 2143–2150 (1994).

42. G. Sezonov, D. Joseleau-Petit, R. D'Ari, Escherichia coli Physiology in Luria-Bertani Broth, *J. Bacteriol.* **189**, 8746–8749 (2007).

43. F. T. Marin, Regulation of development in Dictyostelium discoideum: I. Initiation of the growth to development transition by amino acid starvation, *Developmental Biology* **48**, 110–117 (1976).

44. M. S. Sundrud *et al.*, Halofuginone Inhibits TH17 Cell Differentiation by Activating the Amino Acid Starvation Response, *Science* **324**, 1334–1338 (2009).

45. J. Wang *et al.*, Dependence of Mouse Embryonic Stem Cells on Threonine Catabolism, *Science* **325**, 435–439 (2009).

46. J. Sambrook, *Molecular Cloning: A Laboratory Manual, Third Edition* (Cold Spring Harbor Laboratory Press, ed. 3rd, 2001).

47. A. Wach, PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in S. cerevisiae, *Yeast* **12**, 259–265 (1996).

48. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.* **10**, R25 (2009).

49. N. T. Ingolia, L. F. Lareau, J. S. Weissman, Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes, *Cell* **147**, 789–802 (2011).

Materials and Methods

Bacterial strains and media

For ribosome profiling during biofilm formation, a 3610-D*epsH* strain (RL3852) was used to ensure dispersed growth in liquid media (*5*). For serine starvation experiments, a serine-auxotrophic 3610-D*serA* strain (YC865) was used. A list of strains, plasmids, and oligonucleotides used in this work are summarized in Supplementary file 1.

For general purposes, *B. subtilis* strains PY79, 3610, and their derivatives were grown in Luria-Bertani (LB) medium (10 g tryptone, 5 g yeast extract, and 5 g NaCl per liter broth) at 37°C. Antibiotics were added to the media at the following concentrations for *B. subtilis* strains: 10 μg ml^{-1} of tetracycline, 100 μg ml^{-1} of spectinomycin, 10 μg ml^{-1} of kanamycin, 5 μg ml^{-1} of chloramphenicol, and 1 μg ml^{-1} of erythromycin. Minimal MSgg medium was used as the liquid biofilm-inducing medium. The same medium with 1.5% Bacto-agar (Difco) was used as the solid biofilm-inducing medium. MSgg medium composition: 5 mM potassium phosphate (pH 7), 100 mM MOPS (pH 7), 2 mM $MgCl_2$, 700 μM $CaCl_2$, 50 μM $MnCl_2$, 50 μM $FeCl_3$, 1 μM $ZnCl_2$, 2 μM thiamine, 0.5% glycerol, 0.5% glutamate, 50 μg ml^{-1} tryptophan, 50 μg ml^{-1} phenylalanine and 50 μg ml^{-1} threonine. For overnight growth of serine auxotrophic 3610 strains, MSgg medium was supplemented with serine to a final concentration of 5 mM. For serine starvation experiments in which YFP fluorescence was measured (Figures 3A, 3B), MSgg medium was supplemented with 800 μM serine and 400 μM serine methyl-ester (Sigma). Serine methyl-ester was added to ensure slow growth under serine starvation conditions.

Strain construction

General methods for molecular cloning followed published protocols (*42*). SPP1 phage-mediated transduction was used to transfer antibiotic-marked DNA fragments between different strains (*5*). Long-flanking PCR mutagenesis was applied to generate insertional deletion mutations (*43*). Synonymous

switches in *sinR* and *spo0A* were generated by using synthetic DNA fragments (Genewiz, USA) or by applying site-directed mutagenesis (Roche). Sequences of the primers used in constructing mutant *sinR* alleles are described in Supplementary file 1. Incorporation of synonymous substitutions into the *sinR* or *spo0A* gene at the native locus was done by allele exchange and followed a method described previously (*19*).

Biofilm assays

B. subtilis cells were first grown in LB broth at 37° C to mid-exponential phase. For formation of biofilm colonies, 2 µl of the cells was spotted onto MSgg medium solidified with 1.5% agar. Plates were incubated at 23° C for 3-4 days before analysis. All images were taken using either a Nikon CoolPix 950 digital camera or using a SPOT camera (Diagnostic Instruments, USA). Assays for the *b*-galactosidase activities were described previously (*5*).

Genetic screen for suppressor mutants

Following a previously-published protocol (*17*), we set up a genetic screen to search for spontaneous mutations that suppressed the defective biofilm phenotype of a *B. subtilis ΔslrR* mutant (YC131). The defective biofilm phenotype of the *ΔslrR* mutant is manifested as an inability to form robust floating pellicles (*16*, *17*). Briefly, the *ΔslrR* strain was inoculated into liquid MSgg medium in 6-well plates and incubated at 30°C. After 48 hours, pellicle formation was examined visually. In some wells, robust pellicles appeared possibly due to a second, suppressor mutation elsewhere in the genome. Cells from those wells were picked and streaked out on fresh LB agar plates to isolate single colonies. Cells from the single colonies were then tested for altered colony morphology on solid MSgg medium to confirm the suppressor phenotype. Similar genetic screens in previous studies (*5*, *17*) had established that mutations in the *sinR* gene could suppress the defective biofilm phenotype of the *ΔslrR* mutant. Therefore, we isolated genomic DNA from the putative suppressor mutants, amplified the *sinR* gene by PCR, and then sequenced the *sinR* locus. Once a mutation in the *sinR* gene was confirmed by sequencing, the same mutation was reconstituted in the wild type background (3610) following a previous protocol (*17*), and assayed for alteration in colony morphology on solid MSgg medium.

Bacterial growth for ribosome profiling

For ribosome profiling during biofilm formation, fresh colonies were inoculated into 8 ml of MSgg liquid medium and grown for 12 hours at 30°C, 200 rpm. Saturated cultures were diluted 1:1000 into 200 ml aliquots of fresh MSgg medium and shaken in a 1L flask at 30°C, 200 rpm. For exponential-phase ribosome profiling (Figure 2A), cultures were harvested at OD_{600} = 0.6. For ribosome profiling during biofilm entry (Figure 2B), cultures were harvested at OD_{600} = 1.4. For the serine repletion experiment (Figure 4), serine was added to a final concentration of 2.5 mM at OD_{600} = 1.4 and harvested after 30 min at 30°C, 200 rpm. For the serine starvation experiment (Figures 3B, 3C), pre-cultures were grown in MSgg medium supplemented with 5 mM serine and then diluted into 200 ml of the same medium. At an OD_{600} = 0.6, the cultures were filtered and re-suspended either in MSgg medium (starvation) or in MSgg medium with 5mM serine (control), and harvested after 60 min at 30°C, 200 rpm (Figure 3D).

Western Blotting

Cultures were grown with shaking at 37°C, and 14 ml culture aliquots were harvested at an OD_{600} between 2.0 and 2.5. Cell pellets were collected by centrifugation and washed once with 10 ml of lysis buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA pH 7.4). Pellets were resuspended in 1.2 ml lysis buffer and incubated with 20 µg/ml of lysozyme (Sigma) for 1 hour on ice. Cells were further lysed by sonication. Cell debris was removed by centrifugation (14,000 rpm, 30 min, 4°C). The concentration of total protein in the lysates was determined by a Bradford assay (Bio-Rad). Samples for SDS-PAGE were prepared in Laemmli buffer normalized to equal protein concentration. Samples were ran on an NuPAGE 12% gel (Invitrogen, 1.0mm, Bis/Tris, 200V, ~ 50 min) and transferred to a PVDF membrane (Millipore) at 100 V for 1 hour. The membrane was blocked in 5% milk-TPBS for 1 hour, and then incubated with anti-SinR antibody (1:2500, polyclonal) and anti-SigA (1:100,000, polyclonal) overnight. The membranes were washed 3 times in TPBS for 5 minutes each. Blots were incubated with goat anti-rabbit secondary antibody conjugated to Horseradish peroxidase (1:10,000, Bio-Rad). Blots were washed 3 times in TPBS, and developed with SuperSignal West Dura chemiluminescent substrate (Thermo) and imaged on a gel-

doc (Bio-Rad). Densitometry analysis of Western blot images was performed using ImageJ software (NIH, http://rsbweb.nih.gov/ij/). Rectangles were drawn around each distinct band and the average pixel intensity in this rectangle was calculated, followed by subtraction of background pixel intensity from an identical rectangle drawn in a region without any band. These background subtracted values were used for calculating SinR/SigA, SlrR/SigA and SinR / SlrR ratios in Figure 1.

YFP fluorescence measurements

A previously published measurement protocol was used (*33*). Fresh colonies were inoculated into 1 ml of MSgg liquid medium with 5 mM serine and grown overnight in deep 96-well plates at 30°C, 1400 rpm. In the morning, saturated cultures were diluted 1:100 into 1 ml of MSgg medium with 800 µM serine and 400 µM serine methyl-ester and shaken at 30°C, 1350 rpm for 3 hours. Three aliquots of 150 µl from each culture was pipetted into 3 wells of three 96-well plates (3799, Costar). Wallac Victor2 plate reader (PerkinElmer) was used to monitor cell density (absorbance at 600 nm) and YFP expression (fluorescence, excitation 504 nm and emission 540 nm). Each plate was read every 15 min using a robotic system (Caliper Life Sciences) and shaken at 1000 rpm in between readings (Variomag Teleshake shaker). 30°C and 60% relative humidity was maintained throughout the experiment.

Ribosome profiling and mRNA quantification

Ribosome profiling protocol was adapted from published literature (*24–26*) with minor modifications as described below. Briefly, 200 ml of bacterial culture was harvested by filtration. The filter was immediately inserted into a 50 ml conical tube, flash frozen in liquid nitrogen, and stored at -80°C until further processing. Frozen cells were re-suspended in 8 ml of polysome resuspension buffer (20 mM Tris pH 8.0, 10 mM $MgCl_2$, 100 mM NH_4Cl, and 100 µg ml^{-1} Chloramphenicol). Re-suspended cells were pelleted by centrifugation (3000 g, 4°C, 5 min) and the supernatant was discarded. The cell pellet was re-suspended in 500 µl of polysome lysis buffer (1X polysome resuspension buffer, 5 mM $CaCl_2$, 0.4 % TritonX-100, 0.1 % NP-40, and 100 U/ml RNase-free DNase [04716728001, Roche]), and transferred to an ice-cold 1.5 ml tube containing 500 µl of 0.2-0.3 µm acid-washed glass beads (G1277, Sigma). Cells were lysed by vortexing at maximum speed on a vortexer in a 4°C room (Vortex Genie 2,

10 ´ 30 s with 1 min cooling on ice in between). The lysate was clarified by centrifugation (20,000 g, 4°C, 10 min) and the supernatant was transferred to a fresh 1.5 ml tube. 500 µg of total RNA (A_{260} units) was digested (25°C, 1400 rpm, 60 min, 150 µl volume) with 2 U/µg of Micrococcal nuclease (LS004797, Worthington). The digestion was quenched with 1.5 µl of 0.5 M EGTA, loaded on top of a 10-50 % sucrose gradient and ultra-centrifuged in a SW41 rotor (35000 rpm, 4°C, 150 min). Monosomes collected by gradient fractionation (Biocomp Instruments).

For total RNA extraction, 100 µl of polysome lysate was mixed with 400 µl of RNA extraction buffer (0.3 M sodium acetate, 10 mM EDTA, pH 4.5) and the aqueous phase was extracted twice with phenol-chloroform and once with chloroform. RNA was precipitated with an equal volume of isopropanol. The pellet was washed with 70 % ice-cold ethanol and re-suspended in 100 µl of 10 mM Tris pH 7.0. 10 µg of total RNA was DNase-treated and mRNA enriched using the Microbe Express kit (Ambion). mRNA was fragmented by heating at 95°C with a bicarbonate buffer (*24*) for 20 min.

Collected monosome fractions were purified using the same phenol-chloroform method as used for total RNA extraction above. Monosomes and fragmented mRNA were then used for small RNA sequencing library preparation. Size selection, dephosphorylation, polyadenylation, reverse-transcription, circularization and PCR amplification were performed using the same protocol as in (*24*). An rRNA subtraction step was carried out between the circularization and PCR amplification steps using the same protocol as in (*25*). Typically, several samples were multiplexed for sequencing in a Illumina HiSeq sequencer such that at least 1 million reads were obtained for each sample.

Deep-sequencing data analysis

Deep-sequencing data analysis was carried out in Bash and Python programming languages, and performed on the Harvard research computing Odyssey cluster. Main steps are summarized below and shown schematically in Figure 2–figure supplement 1.

1. Each 50nt single-end read was polyA-trimmed by identifying 10 or greater number of adjacent adenines and discarding all nucleotides starting from -1nt of the polyA run. The first 5¢nt of the read was also discarded since its identity was ambiguous in several reads. PolyA-trimmed reads

were first aligned against all non-coding RNAs in the *B. subtilis* genome using bowtie aligner [ver. 0.12.7, (*44*)].

2. The non-coding RNA sequences were downloaded from NCBI (NC_000964.frn).
3. Reads that did not align to non-coding RNAs were aligned against the whole *B. subtilis* genome using bowtie aligner. The *B. subtilis* genome was downloaded from NCBI (NC_000964.fna). Only reads that had less than three mismatches with the reference genome were considered for further analysis.
4. Reads that aligned to the *B. subtilis* genome were further trimmed by 8nt from each end to approximate the ribosome A-site coordinate. The remaining sequence was normalized by its length and assigned to the corresponding genomic coordinate, and this value was designated as the read density at this genomic coordinate during further downstream analyses.
5. Average ribosome and mRNA density for a single gene was calculated by summing the read density between the start and stop codon, and then normalizing by the length of the gene.
6. Fold-change in ribosome density for a single gene between two samples was calculated by taking the log2 of the ratio of average ribosome density between the two samples for that gene. The median value of this log2 fold-change across all genes that received a minimum of 100 reads in at least one of the two samples was then subtracted from the fold-change value for each gene. This median-subtracted log2 fold change in reported throughout this work. We note that the average ribosome density on any gene is directly proportional to the corresponding mRNA level in the absence of specific translational regulation. Hence fold-changes in ribosome density (such as the one shown in Figure 3D) primarily reflect fold-changes in mRNA level.
7. To calculate the ribosome and mRNA density at individual codons,
    a. The start codon was treated as a single separate codon irrespective of its identity.
    b. Only genes with average ribosome density of at least 1 read per codon were considered.
    c. The first nucleotide of each codon was assigned as the start of the A-site of the ribosome based on the peak in ribosome density at the start and stop codons (see below). For each

of the 64 codons and the start codon, read density at the first nucleotide of the codon was averaged across all occurrences of that codon in a single gene and then normalized by the average ribosome density for that gene. Hence a codon without over- or under-representation of ribosome or mRNA density will have a density value equal to 1.

d. Genome-wide ribosome and mRNA density was calculated as the median of the individual gene read density from Step (b) across all genes that pass the threshold of Step (a).

e. Ribosome profiling measurement resulted in a high ribosome density at start and stop codons. It is unclear whether this increase is a true biological signal or caused by the measurement protocol (*45*). Hence these codons were excluded from the ribosome density plots shown in Figures 2A, 2B and 3C. However, the increased ribosome density at these codons resulted in a concomitant decrease in ribosome density at the remaining codons due to normalization by the average ribosome density for each gene (which included the start and stop codons).

Codon usage analysis

Codon Adpatiaton Index (CAI) was calculated according to the original prescription of Sharp and Li (*34*). For this calculation, 68 genes that had an annotation as 'ribosomal' were used as the reference set of highly expressed genes. The annotations file used for this analysis was downloaded from NCBI (NC_000964.ptt). The P-value for the higher frequency of TCN codons in *sinR* was calculated by assuming a binomial distribution of TCN and AGY codons. The P-value represents the binomial probability that there are 10 or more TCN codons in the *sinR* gene (12 serine codons total) given the genome-wide frequency of serine codons (0.66 for TCN codons and 0.34 for AGC/AGT codons).