CHARACTERIZATION OF A TOXIN-ANTITOXIN SYSTEM IN *E. COLI*

DEVELOPMENT OF DNA CONSTRUCTS, BACTERIAL STRAINS AND METHODOLOGIES TO CHARACTERIZE THE IBS/SIB FAMILY OF TYPE I TOXIN-ANTITOXINS IN *ESCHERICHIA COLI*

BY
SHAHRZAD JAHANSHAHI
B.Sc. University of Toronto

A Thesis
Submitted to the School of Graduate Studies
In Partial Fulfillment of the Requirements
for the Degree
Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2019) McMaster University

(School of Biomedical Engineering) Hamilton, Ontario

TITLE: Development of DNA constructs, bacterial strains and methodologies to characterize the Ibs/Sib family of type I toxin-antitoxins in *Escherichia coli*

AUTHOR: Shahrzad Jahanshahi, B.Sc. (University of Toronto)

SUPERVISOR: Dr. Yingfu Li

NUMBER OF PAGES: xxi, 249

Lay Abstract

Almost all bacteria contain genes that may lead to their growth stasis or death. Normally, these toxins are believed to be neutralized with their cognate antitoxins. In spite of the efforts to understand these toxin-antitoxin (TA) systems, their physiological roles are subject to intense debate. These systems are hard to study mainly because 1) they are only activated under specific conditions and 2) they are low in abundance. Current approaches are not high throughput and sensitive enough. In this thesis, we developed DNA constructs, bacterial strains and methodologies to facilitate the study of these molecules, particularly the Ibs-Sib family. We next employed these tools to gain a fundamental knowledge of their expression under different conditions, which revealed surprising information about the function of these molecules. We believe that future studies can greatly benefit from the tools offered here to tremendously enhance our understanding of these systems and lead to useful applications.

Abstract

Almost all bacteria contain genes that may lead to their growth stasis and death. Normally, these toxins are believed to be neutralized with their cognate antitoxins from a toxin-antitoxin (TA) operon. These modules are also abundant in pathogenic bacteria suggesting a role for them both in normal bacterial physiology and pathogenicity. Their functions have been subject to intense debates. Due to the cell killing capability of the toxin and the gene silencing capability of the antitoxin, they have been utilized for basic research, biotechnology and medical applications. However, further advancements of these applications have been impeded by our limited knowledge of the biology of TAs. Among these TA systems is the Ibs/Sib (A-E) family. Here, we discuss our efforts in characterizing these systems, with a focus on the IbsC/SibC member. Studying them has shown to not be straightforward due to the complexity of their underlying mechanisms and the current approaches being laborious and lacking sensitivity to be applied to these low abundant molecules. We have developed fluorescence-based platforms to take advantage of sensitive and high throughput and resolution techniques such as Fluorescence Assisted Cell Sorting (FACS) to study these molecules instead of relying on traditional culturing methods. While developing these platforms, we gained insights about the biology and regulation of these molecules. To expand this knowledge, we actively pursued investigating the regulation of these molecules at the transcriptional and post-transcriptional levels, both in their native context and in artificial systems. The rest of this thesis summarizes our efforts in solving one of the biggest pieces of the Ibs/Sib puzzle, namely their physiological expressions. With the strategies we have optimized for specific detection of these low abundance molecules, and the knowledge of their biology and regulation presented, we are now at an exciting phase to interrupt the long pause in the study of functions by these molecules and advancement of TA-based applications.

Acknowledgements

I have had an amazing seven years at McMaster and owe my appreciation to many.

First, I would like to express my gratitude to my supervisor, Dr. Yingfu Li, for giving me the opportunity to do research in his lab. He has fostered an amazing lab environment for the growth of scientists and I am happy to have been part of it. I especially appreciate all the resources and freedom he gave me to explore scientific ideas all these years. Our thought-provoking discussions have always been very valuable to me. It has been a pleasure working under his guidance. I would also like to acknowledge my committee members Dr. Elliot and Dr. Soleymani for their ongoing help and guidance. Thank you for always making me feel that I could count on you for support.

In addition, I would like to thank my mentor, Dr. Wendy Mok, who helped me with a great start in the Li lab. She taught me a lot of the fundamental molecular biology techniques. She also helped considerably with the design of my experiments at the beginning. Some of these experiments are presented in chapters 2 and 3 of this thesis.

I feel very grateful for the professional network I have established at McMaster and many friendships that I know will be life-long. I would like to especially acknowledge my bench mates, Zohaib Ghazi and Dingran Chang. It has been a pleasure working alongside you two. I will always remember our deep scientific and philosophical discussions. The past and present members of the Li lab have always been supportive and co-operative. Particularly, I would like to thank Dr. Natalie Schmitt, Alexander Hua, Maria Rey Rincon, Devon Morrison, Dr.

Dingran Chang, Dr. Kha Tram and Rachel Gysbers for their help with proofreading my thesis. I must especially thank Dingran for his faith in me and encouragements in the last stretch. I have met some of the kindest, most interesting and intelligent people in my life here at McMaster. Interactions with them have influenced me both as a scientist and personally. I will miss my time here.

I have also been fortunate to have many good friends outside the lab who always believed in me and were willing to lend a helping hand.

Lastly, I don't believe this journey would have been possible without my parents and two brothers, Siavash and Shahriar. I am incredibly grateful for their ongoing love, advice, faith, and support.

To my parents, Soraya and Nasser, who provided me with the opportunity to be happy and strong and chase my goals

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 5

List of Tables

List of Abbreviations

A260	absorbance at 260
afsRNA	artificial small RNA
Atc	anhydrotetracyclin
Cas	CRISPR-associated
CDI	*C. difficile* infection
cDNA	complementary DNA
CFU	colony forming units
CN_{ct}	copy number calculated from Ct values
CN_{OD}	copy number determined by OD method
CRISPR	clustered regularly interspaced short palindromic repeats
cs-POMs	cancer-specific Protein Output Modifiers
Ct	threshold cycle
DiBAC4(3)	Bis-(1,3-dibarbituric acid)-trimethine oxonal
DNA	deoxyribonucleic acid
eDNA	extracellular DNA
ExPEC	extraintestinal pathogenicity *E. coli*

FACS	fluorescence assisted cell sorting
gDNA	genomic DNA
GFP	green fluorescent protein
GFP AAV	a short-lived unstable green fluorescent protein
GMO	genetically modified organisms
hr	hour
Ibs	induction brings stasis
IGR	intergenic regions
IPTG	isopropyl β-D-1-thiogalactopyranoside
LacR	lac repressor
LB	Lysogeny-Broth
M63	a minimal media for *E. coli* growth
MIQE PCR	minimum information for publication of quantitative real time
mRNA	messenger RNA
nc sRNA	non coding small RNA
NRT	no reverse transcription control
NTC	no template control
OD600	optical density measured at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis

PCD	programmed cell death
PLlacO1	lactose-inducible promoter
PLtetO1	tetracycline-inducible promoter
Popt	a synthetic strong and constitutive promoter
P_{rrnB1}	a growth rate dependent promoter
ppGpp	guanosine 3', 5'-bispyrophosphate
psp	phage shock protein
RBS	ribosome binding site
RNA-seq	next-generation sequencing of cDNA
ROS	reactive oxygen species
RT-qPCR	reverse transcription-quantitative polymerase chain reaction
Sib	short intergenic abundant
S/N	signal to noise ratio
SR	sequencing reads
sRNA	small RNA
TA	toxin-antitoxin
Tat	transactivator of transcription
TAR	transactivation response
TetR	tetracycline repressor
TRD	target recognition domain
tRNA	transfer RNA

UTR	untranslated region
VBNC	viable but non culturable
WT	wild type

Declaration of Academic Achievement

I performed all of the experiments described in this thesis. I produced all of the figures unless otherwise indicated. I wrote the thesis and incorporated the minor revisions by my examination committee. I took the leading role in the conceptualization and design of the experiments and analysis of the data with helpful advice from Dr. Li. Also, Dr. Wendy Mok has contributed to the design of the experiments presented in chapters 2 and 3 of this thesis.

1. Chapter 1: Introduction

1.1 General introduction to toxin-antitoxin systems: Human enemies or unexpected friends?

Almost all bacteria contain genes that inhibit cell growth and may lead to cell death upon uncontrolled or over expression. In normally growing cells, these toxins are co-expressed and neutralized with their cognate antitoxins from a toxin-antitoxin (TA) operon. These modules are present in abundance in bacteria, including the pathogenic ones (e.g. at least 33 TA systems in *Escherichia coli* and more than 60 TA systems in *Mycobacterium tuberculosis*), suggesting a role for these TA systems not only in normal bacterial physiology but also in pathogenicity of bacteria. The functions of these entities have been subject to intense debates. While TA systems were originally linked to control over the genetic material, for example plasmid maintenance, or protection against invading DNA, it is now clear that they are also involved in essential cellular processes like replication, gene expression, and cell wall synthesis. As the result of environmental stimuli/stressors or stochastically, toxin activity is induced to result in the silencing of the above-mentioned biological processes, and subsequently cause death or entry in a dormant state, known as persistence. Since a lot of these biological processes are also targets for disruption by common environmental stressors, such as antimicrobials, their silencing by the toxin is considered a defense mechanism by the cells to evade the disruptive effects of the stress and survive. Surviving or persister cells are held responsible for the recalcitrance of chronic infections. In addition to persistence, TA systems have also been shown to be involved in biofilm formation, which is demonstrated to be the main strategy by pathogenic bacteria to survive the host immune defense and cause chronic

infections. In this chapter, the recent findings on the role of these TA systems in bacterial physiology, survival and pathogenicity will be highlighted. There is increasing evidence that bacterial protein toxins are the chief virulence factors of pathogenic bacteria. We believe that presenting this evidence encourages research on these intriguing systems to unravel the unknowns about these systems such as their mechanisms of action and individual roles. This knowledge would then shed light on bacterial behavior, as well as help us arrive at diagnostic and therapeutic tools, which are in some cases of infection, urgently needed. Scientists have already turned these TA molecules cleverly into versatile tools for a multitude of purposes in basic research, biotechnology and medicine. Some of these applications will be also presented in this chapter.

1.2 Types of TA Systems, their distribution and abundance in bacterial genomes

Toxin-Antitoxin (TA) systems are small genetic modules that consist of 2 components, a stable toxin protein that can have toxic effects on the cells and its labile antitoxin, either RNA or a protein that can counteract the toxicity of its cognate toxin. Depending on the molecular nature of the antitoxin and its mode of action to neutralize the toxin, TAs have been grouped into six types so far: (i) type I, the antitoxin is an antisense RNA which counteracts the toxin by binding to its complementary mRNA and inhibiting its translation; (ii) type II, the antitoxin is a protein which binds to the toxin protein through protein–protein interactions (PPI); (iii) type III, the antitoxin mRNA interacts directly with the toxin protein; (iv) type IV, the antitoxin protein prevents the binding of the toxin protein to its cellular target; (v) type V, the antitoxin protein cleaves specifically the mRNA of its cognate toxin and (vi) the most recently discovered type VI TA system consists of a protein toxin and a protein antitoxin (SocB-SocA TA system). The toxin halts

replication elongation when uninhibited. The antitoxin serves as an adaptor for the protease ClpXP, promoting degradation of the toxin **Figure 1.1**.[1]

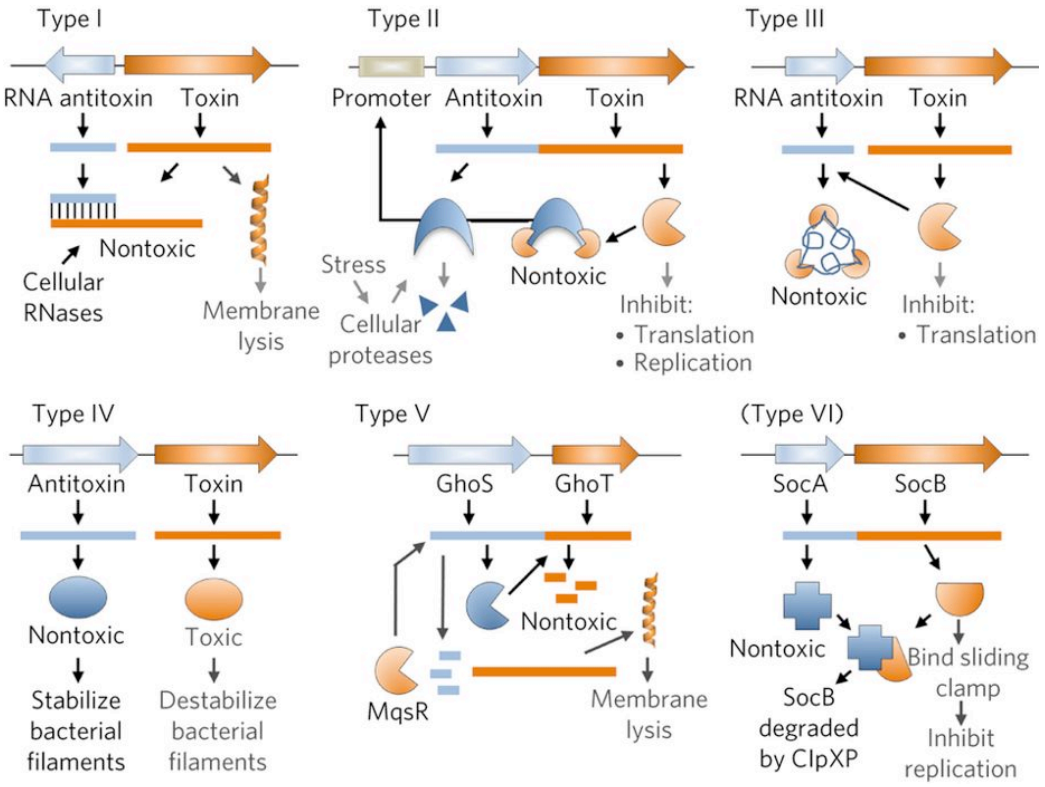


Figure 1.1. Overview of the six classes of toxin-antitoxin systems. The classifications are based on the molecular nature of the antitoxin and its mode of interaction with the toxin. Figure is borrowed from Rebecca Page *et al.* [1] with permission of Nature Chemical Biology . [1]

TA systems exist widely in bacterial genomes, including pathogens. [2] In *Escherichia coli* alone, the reported number of TA systems has recently increased from five to 37 [3]. *Vibrio cholerae*, a human pathogen that causes cholera, contains at least seventeen TA loci. [4] Recently, type I toxin-antitoxin systems were identified for the first time in *Clostridium difficile* (*C. diff*), a major

human enteropathogen and suggested to contribute, together with the well-studied defense system of CRISPR-Cas, to the recurrent infections caused by *C. diff*. [5] TA systems are also often found in multiple copies within genomes.

For example the *ibs-sib* TA pair in *E. coli* exists in 5 copies (*ibs-sib* A-E) with extensive sequence similarity. [6] The surprising abundance of these entities in bacterial genomes raises interesting questions regarding their possible biological roles, and evolution. Why would bacteria continue to harbor genetic elements that can potentially kill them, if not tightly regulated?

1.3 Role of toxin-antitoxin systems in a variety of important physiological functions and ambiguities and controversies surrounding these roles

The role of TA systems in cell physiology is controversial. Some TA systems are considered selfish entities that maintain themselves at the cost of their host.[7] As a consequence, TA systems were originally linked to control over the genetic material, for example conferring protection against DNA loss. TA pairs were first discovered on plasmids, to ensure that the plasmids that carried them were maintained within the cells over generations, by a process called post-segregational killing. This process relies on the differential stability of the toxin and antitoxin. Following division, daughter cells that do not contain the TA-harboring plasmid lack a continuous supply of these molecules but have inherited some of them upon division from the parent cells. Antitoxin molecules in these cells are short-lived, and the long-lived toxin molecules, in the absence of any antitoxin, kill the cells.[4,8,9] This results in survival of only cells that carry the plasmid, hence maintenance of the plasmid containing the TA locus, in cell lines that harbor them. Since this initial discovery, TA systems have been found on the

chromosomes of many bacteria, and have also been shown to increase the stability of important DNA regions such as superintegrons and defense islands. Another control over the genetic material by TA systems is prevention of invading DNA, hence, for example protecting the bacteria against invading phages [10]. Recently, however, it is clear that many TA systems are also involved in essential cellular processes like replication, gene expression, and cell wall synthesis, therefore might be more than just selfish entities (**Figure 1.2** below). [11] Under normal conditions, the antitoxin counteracts the toxin to prevent its activity. However, under nutritional or other types of stress, antitoxin proteins may be degraded, allowing the toxin to exert its activity by inhibiting cellular processes. Toxin activity is induced stochastically or as the result of environmental stimuli, resulting in silencing of the above-mentioned biological processes and changes in cellular growth. As the result, cells often enter in a dormant state or die. This is considered to be a defense mechanism by cells to avoid stress. Moreover, recent research has revealed the long overlooked role of TA systems in bacterial virulence. There is increasing evidence that bacterial protein toxins are the chief virulence factors of pathogenic bacteria and regulate their interactions with mammalian cells. Bacterial toxins use diverse and sophisticated mechanisms to manipulate the host cell functions to the advantage of the bacteria, favoring their spreading, colonization of niches, lifestyle inside eukaryotic cells, and survival. In addition, TA systems are omnipresent on mobile genetic elements in many pathogenic bacteria that are involved in antibiotic resistance and virulence, known as pathogenicity islands. They may function to stabilize these regions. [4,11] It is noteworthy to mention that due to their importance to bacterial adaptation, survival and pathogenicity, these molecules present themselves as great candidates for drug discovery. In addition, due to their cell killing capabilities, they have been used in anticancer, antiviral applications, containment of genetically modified organisms and many other applications, which will be

discussed later in this chapter.

In spite of the growing evidence that TA systems are important for bacterial physiology and pathogenicity, they are still not well studied and there are still mysteries to unravel about their individual functions, due to a few reasons. First, TA systems are ubiquitous, but we often fail to find phenotypes for TA systems deletion mutants. In some cases, even mutants deleted of multiple TA systems have not shown any disadvantage in competition experiments against wild type cells under various stress conditions. It is possible that toxin-antitoxin systems cooperate with other TA systems to deliver some functions. In addition, often times, bacteria are equipped with an arsenal of diverse and sophisticated strategies to fight harsh conditions and persist in the host as pathogens. Because of the uncertain significance and redundancy of mechanisms, they are often not well studied. [4,5] Future studies in which an entire TA system family or all TA loci are deleted will allow us to uncover their biological functions in different bacteria.

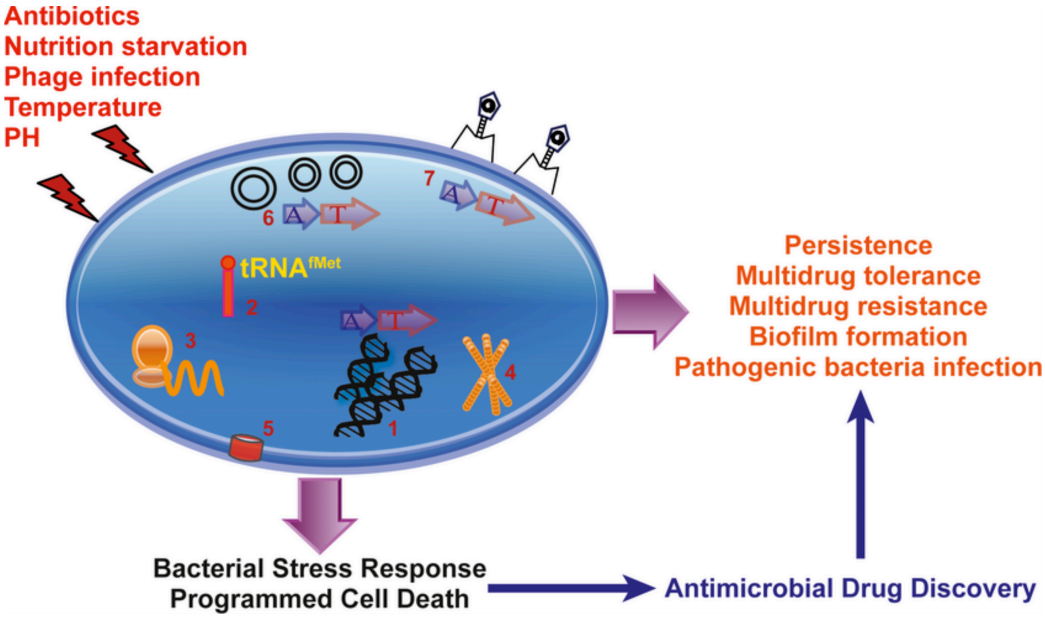


Figure 1.2. An overview of the cellular processes that are inhibited by toxin antitoxin systems, as the result of environmental stressers. The numbers denote: 1. DNA replication, 2. tRNA-related translation, 3. macromolecular synthesis, 4. cytoskeletal polymerization, 5. cell wall disruption, 6. plasmid maintenance, and 7. phage infections. The environmental stressers are highlighted in red. Consequently, cells respond by either death or entry into a dormant state also known as persistence, become multidrug tolerant, form biofilms, survive the host immune response and harsh conditions, and potentially cause chronic infections. Therefore, their inhibition by drugs can potentially make the cells susceptible to the current antibiotics. Figure is borrowed from Yurong Wen *et al.* [1] with permission of Oxford University Press.[11]

Furthermore, until now, many proposed functions for TA systems have been indicated by measuring relative changes in TA gene expression under different stress conditions and by ectopic toxin overexpression. Physiological absolute levels of TA transcription and translation remain an unexplored area. These quantitative studies will contribute tremendously to our knowledge of these systems. Nonetheless, the results from the current studies indicate that TA systems might be very specific with regards to the different environmental cues, stress types (hypoxia, macrophage infection, different categories of antibiotics

etc) and conditions such as mode of growth to which they respond to promote specific adaptations. Interestingly, to further back up this statement, it has been shown that even the multiple homologous TA systems within a single genome in many bacteria, do not cross talk in general. [12] This implies that each may play certain unique physiological roles that contribute to its persistence on the genome. Altogether, these findings emphasize the need to carefully examine toxins in different species, and strains under different physiological and environmental conditions to be able to fully decipher their functions.

In this section, I will present the increasing evidence that support the principal role of TAs in a variety of important physiological functions such as coping with stress by promoting programmed cell death or by inducing dormant cells, and in pathogenicity and chronic infections. The aim is to encourage further research in these areas, to ultimately find more effective diagnostic and therapeutic tools.

1.3.1 Coping with stress

Many toxin-antitoxin systems including, *mazEF, relBE* in *E. coli*, *relBE* and *parDE* in *Caulobacter crescentus* and *vapC*- and *relE*-homologous systems in *Mycobacterium tuberculosis,* have been shown to be induced by stress. Furthermore, different TAs have been shown to respond to different types of stress. As an example, in *M. tuberculosis*, two out of the 30 functional TA systems are induced specifically during hypoxia while another two are specifically induced during macrophage infection. [12] This indicates that TA systems might respond to specific environmental cues to promote specific adaptations. However, single (in some cases multiple) deletions of TA systems generally do not sensitize the cell to stress. These results may indicate that TA systems cooperate with other TAs or defense mechanisms inside cells to combat stress.

Current hypotheses, as discussed below, propose that TA systems are involved in stress management either by promoting altruistic sacrifice of a large fraction of the population (programmed cell death hypothesis, PCD) or by inducing a dormant stage that allows cells to cope with stress (stasis or persistence). These hypotheses have emerged mainly from the study of 2 *E. coli* K-12 type II systems (*mazEF* and *relBE*). [11-14]

1.3.2 Programmed Cell Death (PCD)

The *mazEF* system was shown by Engelberg-Kulka's group to be responsible for PCD upon numerous unrelated stresses (such as amino acid starvation, high temperature, oxidative stress, thymine starvation and antibiotic treatments). PCD appeared to require the regulatory nucleotide, guanosine 3′, 5′-bispyrophosphate (ppGpp). ppGpp is known to control a lot of the physiological changes that are required to adjust cellular metabolism in response to stress. In this case, it shuts down *mazEF* transcription. As the result of transcription inhibition, and degradation of the previously transcribed MazE antitoxin, which is unstable, the toxin activity will be liberated. The toxin being an mRNA interferase would cause translation inhibition in cells, hence death. [13,14]

Some evolutionary advantages for *mazEF*-mediated death have been proposed: 1) During severe nutritional stress, the death of a subpopulation would provide food for the others [13] 2) *mazEF*-mediated death acts as a defense mechanism that prevents the spread of infecting phages [15] 3) *mazEF*-mediated cell death might preserve the genome by killing the cells that carry genomic defects and mutations. [16]

Another well-studied TA system involved in PCD, is the *relBE* TA. In

prokaryotes, the physiological changes in response to amino acid starvation are defined as the stringent response. These changes are important, as they are believed to appropriately adjust the cellular metabolism to nutritional stress. The regulatory nucleotide ppGpp has been known to control a lot of these changes, as mentioned in the given example above. However, Christensen *et al*. showed that RelE toxin of the *relBE* toxin-antitoxin locus of *E. coli* also acts as a global inhibitor of translation during amino acid starvation. [17]

1.3.3 Stasis/Persistence

A phenomenon related to stress is persistence. It is a stochastic phenomenon that confers a temporary dormant state to a very small fraction of a bacterial population that helps them evade effects of stress (For example drugs to make bacteria multidrug tolerant, etc.). Persistence was first described in the early 1940s while studying the mechanism of penicillin when it was observed that a small population of dormant cells survived the antibiotic treatment. The phenomenon was noted to be different from traditional antibiotic resistance mechanisms. [18,19] These persisters were genetically similar to the rest of the population, not mutants as in bacterial resistance mechanisms, but phenotypic variants of the wild type cells.

Although the molecular basis of persistence is still largely unknown, the main model of persister formation holds TA pairs primarily responsible for this process, as they create dormancy. [20] Toxins inhibit various cellular processes including replication, cell wall assembly, and protein translation. [21] Since bactericidal antibiotics kill cells by interference with these cellular functions, their inhibition by toxins prevents bacterial killing and leads to the formation of persister cells [22]. Persistence is an important phenomenon to investigate as recent findings suggest

that persisters are responsible for the recalcitrance of chronic infections. [23]

Below, we will discuss the role of toxin-antitoxins in persister cell formation in different bacteria and the chronic infections caused by them.

In the model organism, *E. coli* K-12, type I *tisAB/itsR1* system has been shown to play a role in persister formation [24]. Lewis and collaborators showed that treating cells with the antibiotic ciprofloxacin induces TisB expression to consequently cause persistence. Also, the frequency of persister cells was drastically reduced (10- to 100-fold) in a strain that did not produce the TisB toxin (Δ*tisAB* strain). Additionally, HipA and ShpAB toxins have been found to increase multidrug tolerance in *E. coli* (and also *Salmonella*), by promoting the persistence phenotype in the population of cells. [4,25,26]

Burkholderia cenocepacia is a pathogen that can cause severe lung infections in immune-compromised people, for example cystic fibrosis patients. [27,28] These microorganisms are often very difficult to eradicate due to the presence of persister cells. Van Acker and collaborators investigated whether TA modules played a role in the tolerance of *B. cenocepacia* toward different antibiotics. They showed that nine toxin encoding genes were up regulated after treatment with tobramycin, but none after treatment with ciprofloxacin. Their results indicated that most, but not all TA modules in *B. cenocepacia* contributed to its persistence and that this contribution depended on the mode of growth and the antibiotic used.

The major human enteropathogen, the multidrug resistant *C. difficile*, causes recurring and severe infections, named CDI for short. *C. difficile* must cope with multiple stress factors inside the host and seems to have a multitude of sophisticated strategies to do so. As CDI progresses, *C. difficile* initiates a sporulation cycle, which has been demonstrated to contribute to persistence of *C. difficile* in the patient. However, the *C. difficile* persister cells are still not well studied. The roles of Toxins A and B in CDI are well established. But only

recently, type I toxin-antitoxin (TA) systems in this pathogen have been identified. *C. difficile* encodes at least four TA systems. The ability of these systems to silence metabolism suggests that they might also be contributing to the formation of metabolically dormant cells or persisters, hence have direct implications in recurrent infections. Further research is necessary to assess the role of TAs in CDI to better understand *C. difficile* pathogenesis. [5]

TA systems are not only involved in cell physiology and persister cells in human pathogens but also in phytopathogens. For example, the gram-negative bacterium *Xylella fastidiosa* is a phytopathogen that causes diseases in many economically important crops (e.g. grapes, citrus, etc.) worldwide. A citrus-pathogenic strain of *X. fastidiosa,* upon treatment with an inhibitory concentration of copper, widely used in agriculture to limit the spread of plant pathogenic bacteria and fungi, forms persister cells [29]. 12 out of 65 TA systems are shown to be induced in these persister cells with *mqsRA* being the most induced. [30] The MqsRA system was first reported in *Escherichia coli* and was the first TA system shown to be directly involved in persister cell and biofilm formation. [20] Overexpression of MqsR in *X. fastidiosa* led to a higher percentage of cells surviving under copper stress as the result of becoming persisters, which was confirmed by the increased number of elongated cells. In *X. fastidiosa*, as well as in several other bacterial models, persistence seems to represent an important mechanism employed by the bacterial cells to survival in harsh environmental conditions, and regrow and recolonize after stress is ended. [20] It is also noteworthy, as the authors in this study suggest, that for the toxin to cause the persistence phenotype and for stress survival, the levels of the TA components must be tightly controlled to ensure the right toxin-antitoxin balance. [31]

As illustrated by the examples above, bacterial cells utilize TA systems to respond to external stress by initiating programmed cell arrest, and persister cell formation.

1.3.4 Biofilm formation and chronic infections

Closely related to bacterial persistence is the ability of free-floating bacterial cells to adhere to an inert or living surface and form structured communities enclosed within their self-produced polymeric matrix, called biofilms. It is widely accepted, for example through the work of Lewis *et al.*, that biofilms provide a host for the cells to survive unfavorable conditions and spread subsequently, therefore, they are involved in the recalcitrance of infections. Upon antibiotic treatment, the majority of cells are killed, except persister cells. The host immune system is supposed to clear these cells, but fails to reach them as they are embedded within the biofilm. These cells may subsequently repopulate the biofilm, spread and cause symptoms. [22,32] Biofilm formation is demonstrated to be the main strategy by pathogenic bacteria to survive the host immune defense. [33]

Most bacteria in the environment live in biofilms due to the adaptive advantages described above. [30,34,35]. Examples of biofilm-associated infections include *E. coli* (where they are best studied), *Staphylococcus aureus, M. tuberculosis, V. cholera, Lactobacillus acidophilus, Gardnerella vaginalis* and the phytopathogen *X. fastidiosa*. [4,20,31,36]

Mechanisms of biofilm formation are poorly understood. Cell death and lysis are known to be essential in biofilm formation as lysed cell parts contribute to the assembly of the matrix. Biofilm communities consist of viable non-persister cells, dead cells and persister cells [22,23], but the molecular basis which regulates the cell fate during biofilm formation remains unclear. In this context, few studies have reported involvement of TA systems in biofilm formation, suggesting that TA systems might regulate cell death and the formation of persister cells. [23] The mechanisms by which TAs impact biofilm formation are generally not reported.

Some reported mechanisms for biofilm formation by toxin-Antitoxins in different bacteria and the chronic infections caused by them are discussed below.

In *E. coli*, the effects of several TA systems on biofilm formation have been reported. The MqsRA TA system, as described above, was the first TA system to be directly related to persister cell formation. [37] It was also one of the first TA systems reported to have a direct role in biofilm formation in *E. coli*. [38] The MqsR toxin is an RNase and a motility quorum-sensing regulator (hence the name MqsR) that forms a type II TA system with MqsA antitoxin. [39] MqsR is shown to be directly associated with biofilm formation, as it is induced in biofilms, and its deletion decreases biofilm formation in *E. coli.* [40] The MqsA antitoxin regulates (reduces) biofilm formation by regulating (repressing) the expression of the stress response regulators RpoS and CsgD. When bound to the promoter regions of *rpoS* (sigma factor S) and *csgD*, the antitoxin represses the production of c-di-GMP, curli and cellulose (regulatory and structural elements that promote biofilm formation), leading to biofilm repression in *E. coli*. [41,42] Hence, MqsA is considered a biofilm formation inhibitor. Degradation of MqsA antitoxin by Lon protease, triggered by oxidative stress, permits a switch to the biofilm mode of growth. MqsR toxin acts as a transcriptional de-repressor, and activates the genes repressed by the antitoxin. [43]

Furthermore, in *E. coli*, a toxin called Hha controls cell death and biofilm dispersal by activation of prophage lytic genes and several proteases. [44]

Further evidence for the involvement of TA systems in biofilm formation comes from the *E. coli* MG1655 Δ5 strain, which has five of the most-studied TA systems deleted (Δ5, MazF/MazE, RelE/RelB, ChpB, YoeB/YefM and YafQ/DinJ). [45,46] Decreased biofilm formation is reported in this strain relative to wild type cells and it is shown to be caused by the overexpression of YjgK, a protein that represses the production of fimbriae involved in both biofilm

attachment and dispersal. [45] Moreover, the authors demonstrate that the five TA systems affect cell death differently during biofilm formation. [46] Some were principal mediators of cell death as opposed to some that were only back up death systems. Some mediated cell death in liquid media, some in biofilm formation and some in both, and some required certain conditions to mediate death. [3]

HipBA has also been shown to influence *E. coli* biofilm formation in the presence of several drugs and in the absence of antibiotic pressure. Zhao and collaborators show that HipA impacts biofilm formation by cell death and releasing Extracellular DNA (eDNA) in the biofilm. This is proven to not be the mechanism of biofilm formation by MazF, for example. The results from their study collectively suggest that there are different mechanisms for the effects of TA systems on biofilm formation. The mechanisms and roles of toxins vary from species to species, strain to strain and need to be studied case by case. [3]

Another bacteria that forms biofilms is *X. fastidiosa*. The phytopathogen, for pathogenicity, relies on its ability to colonize the host plants by moving through the xylem vessels, and to subsequently form biofilms. [47] The MqsR toxin is shown, as in *E. coli*, to be capable of inducing biofilm formation, and repressing cell movement in *X. fastidiosa* as well. [31] The formations of biofilm and cell movement are conflicting processes in this bacterium. [47] It has been demonstrated that the MqsRA TA system plays a key role in alternation between sessile and motile growth in *X. fastidiosa*. In *X. fastidiosa*, deletion of the *mqsR* toxin gene reduces biofilm formation, and its overexpression induces biofilm formation. Deletion of the *mqsA* antitoxin gene, on the other hand, induces biofilm formation. These results indicate that excess of free toxin in the cell relative to the antitoxin, favors biofilm formation. [31] Because biofilms confer protection against antimicrobial agents [35], the free toxin in the cell may increase in response to stress, to induce biofilm formation as a defense mechanism. The authors suggest that careful regulation of expression seems to happen inside cells to ensure the

most beneficial ratio of the antitoxin to toxin. Altogether, the results of this work indicate that the MqsRA TA system has a key role in life style, pathogenicity, survival and adaptation of *X. fastidiosa*, and its interaction with the host plant. Further studies on MqsRA system may elucidate the genetic mechanisms by this TA system[31]

Another pathogen, *V. cholerae* colonizes the human intestine following ingestion of contaminated food and water. It is also found in the aquatic environment. To survive in both aquatic and host environments, *V.cholerae* is capable of coping with different harsh conditions during its transition to and growth within a niche. It is resistant to different stressors such as oxidative compounds and bile, which have antimicrobial activity. [48,49] It can form biofilms that are important for its resistance to such stressors, hence longer survival. [50] The role of the RelBE-family of TA systems in biofilm formation has been examined in this pathogen under host-like stress conditions. Deletion analyses of each *relBE* locus indicated that RelBE systems were involved in biofilm formation and resistance to reactive oxygen species (ROS). Interestingly, all *relBE* loci were induced when standard conditions were used to induce virulence and two of the *relBE* mutants displayed a colonization defect. Although further studies are needed to characterize the mechanisms of action by these systems, the results here reveal that RelBE systems are important for *V. cholerae* physiology and survival. [4] It has also been suggested that in *V. cholera,* RelBE pairs are also involved in stabilizing the massive superintegron cassette array with antibiotic resistance, metabolic and virulence functions. [51] For other bacteria, environmental stressors such as the ones studied here can promote transcription of toxin-antitoxin loci resulting in proteins to help bacteria survive the stress better. [52]

1.3.5 Concluding remarks regarding the physiological roles of TA in bacteria

Herein, through example studies, we demonstrated that TA systems are involved in general bacterial physiological processes such as the general stress response, persistence, and biofilm formation and have a direct impact on the pathogenicity of bacteria. A surprising correlation was recently discovered between the number of TA modules in the genome of bacteria and their virulence capacity [53]. More recently, research has found that different TA systems provide different isolates of extraintestinal pathogenicity *E. coli* (ExPEC) with competitive advantage in colonizing their niche environments. Some TA systems were shown to be critical for bladder colonization whereas others were associated with kidney colonization.[54] Likewise, in *Salmonella Typhimurium,* and *V. cholera,* TA systems have been shown to provide these pathogens with a virulence advantage to survive in mouse host organs. [55]

Considering the wealth of evidence, some presented in this chapter, to support the importance of TA systems in bacterial physiology and pathogenicity and chronic infections that affect our health and economy, we believe that careful examination of the functions by these systems and deciphering of their mechanisms of action, can teach us a lot about the physiology of bacteria and help us arrive at diagnostic and therapeutic tools, some of which are urgently needed. Scientists have already turned these molecules cleverly into versatile tools for a multitude of purposes in basic research, biotechnology and medicine. Some of these applications will be discussed below.

1.4 Current applications of the Toxin-Antitoxin systems

Some bacterial TA systems are functional when expressed in eukaryotic cells and this has led to several innovative applications. Here, I will describe how bacterial TA systems have been used for cell ablation for the purpose of containment of genetically modified organisms (GMO, e.g. yeast and plants) and for the purpose of research in developmental biology. I will also describe how they have helped the current molecular techniques to achieve unusual genetic modifications and also for the engineering of high expression eukaryotic cell lines. Finally, I will examine how TA systems have been employed in antiviral, more specifically anti HIV and HCV, and anticancer gene therapies.

1.4.1 Cell ablation for the containment of genetically modified yeast

With the increasing use of genetically modified organisms (GMOs), appropriate containment strategies are required to prevent their accidental release into the environment. Containment strategies also help with the controlled release of the GMOs into the environment as required by certain biotechnological applications such as bioremediation, bioleaching and biopesticides.

One method of containment involves engineering of the GMOs with a controllable lethal function that is only activated in the presence of a specific environmental cue. TA systems have been used for this purpose in several cases. The general design of such engineered systems is shown in **Figure 1.3** below and operates as follows. The antitoxin gene and the toxin gene are cloned within the same yeast vector under the control of two different promoters. The promoter of the antitoxin gene is designed so that in a controlled environment (such as in a

fermenter), and in the presence of a specific signal (e.g. a certain media constituent or nutrient), it will drive the expression of the antitoxin. The toxin, however, is repressed under these controlled conditions. With this design, if the GMO escapes from these controlled conditions into the environment, and in the absence of the specific signal, transcription of the antitoxin will be repressed while the toxin will be actively transcribed and kill the escaped organism. [56]

Thus far, the *E. coli relBE* [57] and *kis-kid* [58] and the *Streptococcus pyogenes* ε-ζ[59] TA systems have been shown to be functional in the yeast *Saccharomyces cerevisiae* for the purpose of containment.

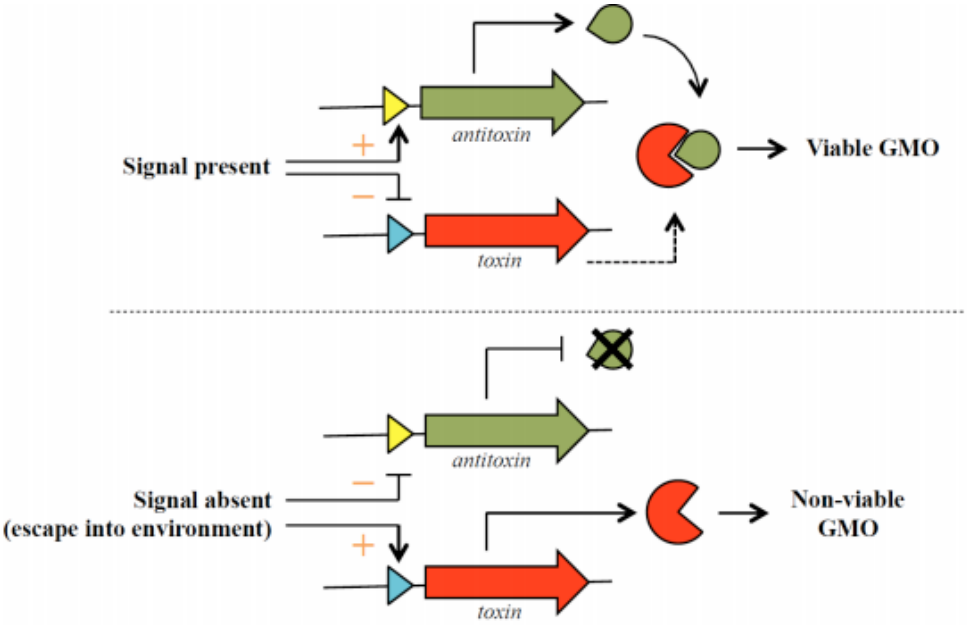


Figure 1.3. General design of a containment system for genetically modified yeast, using a bacterial toxin-anti toxin system. Figure is borrowed from Chew Chieng Yeo et al. [56] The antitoxin gene and the toxin gene are cloned within the same yeast vector under the control of two different promoters. The promoter of the antitoxin gene is designed so that in a controlled environment (such as in a fermenter), and in the presence of a specific signal (e.g. a certain media constituent or nutrient), it will drive the expression of the antitoxin. The toxin, however, is repressed under these controlled conditions. With this design, if

the GMO escapes from these controlled conditions into the environment, and in the absence of the specific signal, transcription of the antitoxin will be repressed while the toxin will be actively transcribed and kill the escaped organism.

1.4.2 Cell ablation for the biological containment of genetically modified plants

Bacterial toxins have also been used to prevent the accidental spread of transgenic crops.

One approach has been through the specific ablation of the plant's reproductive organs, to cause sexual sterility. Barnase, a small RNase produced by *Bacillus amyloliquefaciens* is one of the earliest toxins used since the 1990's to successfully engineer male, female, or bisexual sterility in different transgenic plants. The bacteria also synthesizes the antitoxin which is a small protein called Barstar. Placing the toxin under the control of a tissue-specific promoter (e.g. reproductive organs, plant anther/or pistil) will enable its specific expression and destruction of those organs, therefore causing sterility. Also, when required, fertility of the plant can be restored by the controlled expression of the antitoxin. [60,61]

Barnase-Barstar is not a TA system as the genes for Barnase and Barstar are not in an operon. Nonetheless, it is still illustrative of the potential for a bacterial regulatory module to be used for biological containment of crops. Independent expressions of the toxin and antitoxin modules from different loci, present no barrier, but even permit fine control over the expression levels of the toxin and antitoxin, in a way that is responsive to a particular need. The use of a "true" bacterial TA system, for the purpose of specific cell ablation in plants still remains to be carried out.

1.4.3 Cell ablation in higher eukaryotes for developmental biology research

The Kis-Kid TA system has been used to successfully ablate a specific cell type in a developing zebrafish. To the aim of targeting primordial germ cells, Kid-encoding mRNA was fused to the 3'-UTR of the zebrafish *nos1* gene to direct the expression of the Kid toxin preferentially to the primordial germ cells. To protect the rest of the cells from the potential leaky expression of the toxin, the mRNA of the kis antitoxin gene was fused to the globin 3'–UTR. The constructs were co-injected into one-cell-stage embryos.

These embryos were raised to adulthood without any apparent somatic defects, but showed primordial germ cell loss. The sterile males could induce females to lay eggs but not fertilize the eggs because of their undeveloped gonads. Interestingly, all the adults developed from germ cell-ablated embryos were sterile males. Therefore, it was concluded that in zebrafish, the germ line is essential for the development of females but not male somatic tissues with the exception of their gonad. [62]

1.4.4 In molecular biology: Genetic manipulation

TA systems have enabled disruption of genes that are, otherwise, resistant to disruption. The clever trick played here is that a full copy of the gene of interest is provided to the cells on a "helper" plasmid prior to any disruptions. It is named, " helper" as it provides the cells with the functional back up of the gene, in the case that the fitness of the cells might be compromised due to possible loss of function. Conventional methods of gene disruption using a resistance gene are subsequently carried out in cells that are host to the "helper". The "helper" plasmid also

contains a toxin gene under regulation of an inducible promoter. Once successful replacement of the gene is confirmed, the strains are grown in a medium containing the inducing agent, which will drive the expression of the toxin. This will exert a strong selection pressure for the strains to lose the "helper". This strategy was successfully employed to delete several genes resistant to deletion, in *P. pastoris*, and increased their targeting efficiencies. The *mazF* toxin in particular was used under regulation of a methanol inducible promoter. [63]

1.4.5 In molecular biology: Gene expression manipulation

TA systems have been used as tools for overproduction of heterologous proteins in eukaryotic cells. How strongly a transgene is expressed in transfected mammalian cell lines depends on its integration site on the chromosome (i.e. a highly transcribed region vs. not). The integration occurs mostly at random. Screening and identification of the clones that highly express the gene of interest are laborious and time consuming. Using the Kis-Kid TA system, Nehlsen *et al.* developed a method to efficiently select and enrich for mammalian cells that highly express the gene of interest. [64] The method is based on the transcriptional coupling of the gene of interest with the *kis* antitoxin gene. Cells expressing the cognate toxin, Kid, are transfected with a plasmid encoding the gene of interest together with the antitoxin. It was shown that in the presence of toxin expression, transfectants increased their transgene expression. This is possibly because toxin expression in these cells poses a strong selection pressure for highly expressing clones created by random integration of the transgene-kis cassette. Cells with reduced expression levels are likely eliminated.

1.4.6 Gene therapy: Antiviral

Great potential has been shown for the toxins of TA systems as antiviral agents. Especially those that act as ribonucleases can be particularly useful for the control of RNA viruses. One concern is that while bacterial endoribonuclease toxins usually cleave specific RNA sequences, they are not cell-specific. A clever strategy to restrict the action of these toxins to the infected cells, involves relying on viral regulatory elements for activation of toxin effect. This tactic has been implemented in developing anti HIV and HCV therapies.

An essential regulatory protein, in the early stage of HIV-1 infection, is called Tat (transactivator of transcription). The Tat protein functions by binding to the transactivation response (TAR) sequence and inducing the production of various HIV-1 proteins for infection purposes. An anti HIV therapy, as depicted below in **Figure 1.4**, uses a construct where the *E. coli* MazF toxin gene is placed under the regulation of a TAR sequence in a retroviral vector. In HIV infected cells, Tat binding to the TAR promoter sequence will drive the expression of the toxin gene, hence killing the cells. [65] The *E. coli* MazF functions as an endoribonuclease that cleaves mRNA, including its own, specifically at ACA codons. To prevent self-cleavage, the *E. coli mazF* gene, which harbors nine ACA codons, was engineered to be void of ACA sequences without altering the amino acid sequences, to maintain its toxic effect. The HIV RNA, containing over 240 ACA sequences, is conveniently prone to degradation by MazF.

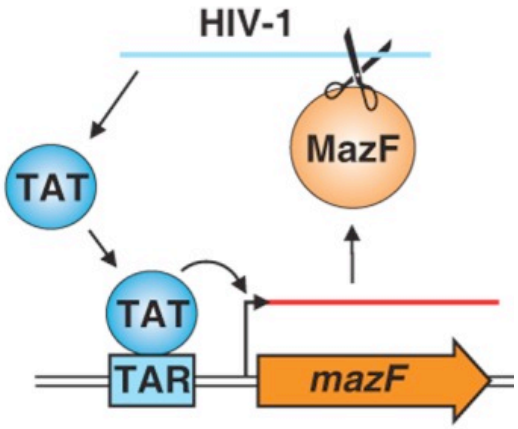


Figure 1.4. An anti-HIV strategy uses the mazF toxin of the bacterial MazEF toxin-antitoxin module. Figure is borrowed and modified from Unterholzner SJ et al. [66] It places the toxin under regualtion of HIV specific regulatory elements, namely the TAT (transactivator of transcription) protein which binds to the transactivation response (TAR) sequence to drive expression of the toxin gene in HIV infected cells and their selective killing.

It was shown that the transduction of a human T lymphoid line with a plasmid containing the TAR- MazF construct prior to its infection with HIV-1, stopped the replication of the virus. Interestingly and fortunately, the levels of induced MazF did not seem to cause serious cell damage, in spite of its ability to cleave cellular RNA. Therefore, cellular growth remained as normal.

Furthermore, the persistence and the *in vivo* safety of the MazF-transduced autologous CD4+ T cells were tested by first transducing the cynomolgus macaque primary CD4+ T cells with a TAR-MazF encoding plasmid, then infusing them into the autologous monkeys, and subsequently monitoring several parameters related to their persistence and safety, for more than half a year. The levels of MazF-transduced T cells in the peripheral blood were still significantly detected throughout the entire experimental period. They were also detected in the lymphoid tissues and the spleen. Also, no lesions or antibodies against MazF were detected. Moreover, the MazF-transduced T cells harvested from the monkeys

more than half a year post-infusion were still able to inhibit the replication of the virus. Together, these results confirm the persistence and safety of the MazF-mediated antiviral therapy. [67]

Using a similar strategy to that described above for HIV, an anti HCV therapy relies on HCV specific proteases (NS3 in particular) to cleave the protease cleavage site that is cleverly used to artificially fuse the MazF toxin to the MazE antitoxin (**Figure 1.5**). As a result, only in HCV infected cells can the toxin be liberated from the toxin-antitoxin complex to exert its ribonuclease activity against the virus. The dosage of MazF needs to be fine-tuned to avoid harm to human cells while eradicating the virus. Optimizing the MazF dosage will make this antiviral approach more feasible. [68]

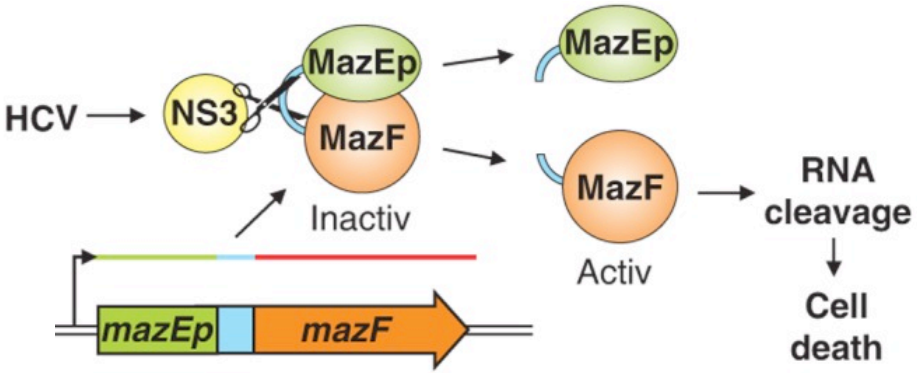


Figure 1.5. An anti-HCV strategy uses the bacterial MazEF toxin-antitoxin module. Figure is borrowed and modified from Unterholzner SJ *et al.* [66] It relies on an HCV specific protease NS3, to cleave the protease cleavage site that is cleverly used to artificially fuse the MazF toxin to the MazE antitoxin. As a result, only in HCV infected cells can the toxin be liberated from the toxin-antitoxin complex to exert its ribonuclease activity against the virus.

1.4.7 Gene therapy: Anti-cancer

TA systems are excellent candidates for anticancer therapies, as they offer not only cytotoxicity by the toxin to target the cancerous cells, but potential selectivity too so that the healthy cells are spared from the toxic treatments. Finding both cytotoxicity and selectivity in anticancer agents has proved to be very difficult and has slowed down progress towards an effective treatment. Using TA modules, selectivity can potentially be achieved by the recue capabilities or so-called cytoprotective actions of the antitoxin to protect the healthy cells.

A controlled cytotoxic or cytoprotective action in cancerous or healthy cells respectively, can be achieved by fine-tuning the production of the toxin relative to the antitoxin.

As an example, ectopic expression of the Kid toxin has been shown to induce death in eukaryotic cells (HeLa cell lines). The Kis antitoxin, however, protects the cells from this effect. [58] This offers the possibility of regulating survival and death in human cells, simply by controlling the relative amounts of Kid and Kis differently in different cell types (cancerous vs. healthy). But how are these therapies able to distinguish between cancer and normal cells? Synthetic systems are under development to link the toxin and antitoxin to different sensors. These sensors are called cancer-specific Protein Output Modifiers (cs-POMs). They are regulatory elements that alter the intracellular concentration of the T-A molecules in human cells in response to the presence or absence of certain cancer specific stimuli. In the presence or absence of an oncogenic signal, they output cytotoxic or cytoprotective ratios of the toxin to antitoxin and therefore evoke an appropriate response that kills cancer cells but protects healthy cells from damage (**Figure 1.6**a below).

An example of such design is illustrated in **Figure 1.6**b and c. In tumor cells, the oncoprotein, named E6, binds to and induces the polyubiquitination of specific target proteins. As the result, the polyubiquitinated proteins become subject to

proteosomal degradation (**Figure 1.6**b). Fusion of such proteins (i.e. targeted by E6) to the Kis antitoxin (**Figure 1.6**c) provides a strategy for the elimination of the antitoxin in cells that carry the oncoprotein E6. It will be polyubiquitilated and therefore degraded. This will liberate the toxins to induce death in cancer cells, but not in other cells that lack the oncoprotein E6 [69].

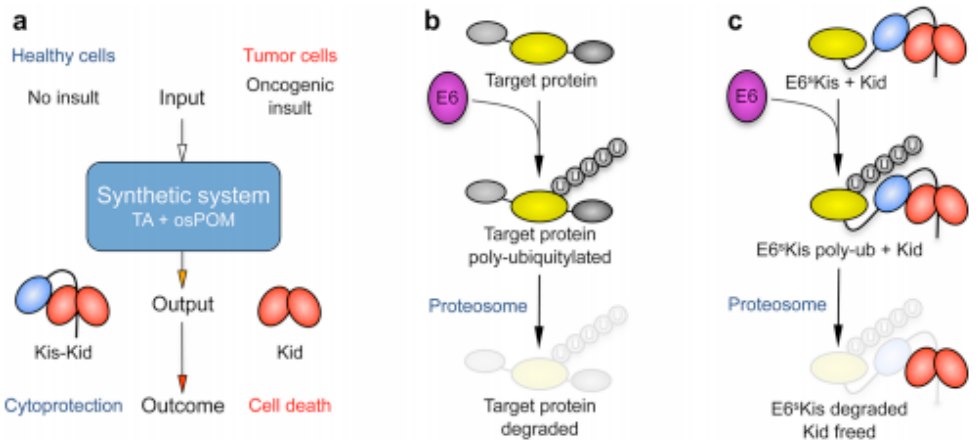


Figure 1.6. An anti-cancer strategy utilizing the Kis-Kid toxin-antitoxin module. (a) General scheme of regulatory elements that output a cytotoxic outcome in response to an oncogenic signal and a cytoprotective outcome in the absence of an oncogenic signal. (b) In tumor cells, the oncoprotein, named E6, binds to and induces the polyubiquitination of specific target proteins. As the result, the polyubiquitinated proteins become subject to proteosomal degradation. (c) Fusion of such proteins (that upon binding to the oncoprotein E6 are subject to degradation) to the Kis antitoxin provides a strategy for the elimination of the antitoxin in cells that carry the oncoprotein E6. This will liberate the toxins to induce death in cancer cells, but not in other cells that lack the oncoprotein E6. The figure is borrowed from Preston MA *et al.* [69] with permission from the American Chemical Society, copyright (2016).

Similar observations have been reported earlier for the *E. coli* RelE toxin in a human osteosarcoma cell line [70] and *E. coli* MazF toxin in human T-Rex-293 cells [71].

These findings altogether demonstrate that TA systems, through highly specific ablation of target eukaryotic cells, offer themselves as potentially important tools in containment of GMOs, developmental studies, and even treatment of cancer and viral infections.

1.5 Gaps in our knowledge of TA systems prevent advancements in their applicability

A large number of chromosome-encoded TA systems have been discovered in a broad variety of bacterial species. However, the investigation of these toxin-antitoxin systems, especially the type I is still in its infancy. Why would bacteria continue to carry well-conserved genes, in many instances in multiple copies that are potentially harmful to them if not tightly regulated? These toxin-antitoxin systems are predicted to have significant, as yet largely unexplored biological relevance, except for a few that were described in previous sections. Many other aspects pertaining to their properties, expression, regulation and mechanisms of action also remain unclear.

Only a few studies actually report mechanisms for TA actions. They were described in the earlier sections. These studies collectively suggest that there are different mechanisms for the effects of different TA systems. In addition, the mechanisms and roles of toxins vary species to species, and strain to strain, emphasizing the need to study them case-by-case. In other words, each toxin has to be individually studied in different organisms in which it exists and under a series of different conditions that might trigger the activity of the toxin.

1.6 The Ibs/ Sib TA systems

The focus of this dissertation is on the Ibs/Sib family of type I TA systems. The *ibs-sib* locus is broadly distributed and conserved across species, including human pathogens. These bacteria containing the *ibs-sib* locus could have anywhere from one to seven copies of the gene pair.[72] In *E. coli,* the bacteria under investigation in this paper, five highly similar copies of *ibs-sib (A-E)* have been identified. The *ibsC toxin-sibC* antitoxin pair has mainly been the focus of my studies, while some analyses of the other homologous copies have also been performed and summarized in the last chapter of this dissertation. The *ibs-sib* gene pairs are oriented such that the *sib* antitoxin gene is directly antisense to the coding region (18-19 amino acids long) of the *ibs* toxin gene including its ribosome-binding site. This orientation may have consequences for regulation by small RNA.

Whilst some information on structural requirements, and kinetics of interaction are known about the *ibs/sib* TA family, their biological functions, expression, regulatory and action mechanisms, like those of many other chromosome-encoded type I TA systems, are very unclear. At the beginning of this study, very little was known about the levels of expression of these genes, how their expression was regulated and what their biological functions were. It has been unknown whether they are important for the adaptability and survival of the bacteria or if they are merely selfish genomic elements. IbsC shares characteristics with other type I toxins, which have been implicated in biofilm and persister cell formation. Therefore, it is possible for its production to be beneficial for the survival of the bacterial population under adverse growth conditions. To begin to address some of these ambiguities pertaining to the significance of IbsC/SibC, details related to their expression and regulation had to be elucidated. In addition, we found some of the traditional approaches in the study of these molecules could be replaced by

newer, more sensitive and high throughput techniques to facilitate the study of theses molecules. This present work lays the groundwork necessary to reveal the true role of the enigmatic IbsC/SibC and its related homologues.

1.7 Objectives of the present study and overview of chapters

In chapter 2, we present some platforms we have created to facilitate the studies of the IbsC-SibC and potentially other TA pairs.

In the first part of chapter 2, we summarize our efforts in developing a platform to study persistence and the potential involvement of TAs in the phenomenon. With the large body of evidence suggesting the role of TA in persistence and some of our own observations regarding IbsC causing reversible growth stasis, we hypothesized that IbsC is involved in persistence. To study persistence, however, is not straightforward due to a) the complex underlying mechanisms, and redundant strategies and conditions that cause persistence b) the current approaches that rely on plating and colony counting to identify and isolate persisters, are not only laborious (especially having to study multitude of conditions to elucidate the one(s) that cause persistence and involve the toxin under study) but may lack sensitivity to pick up persisters as they are rare events. To facilitate the study of persistence, we have developed a fluorescence-based platform to isolate and identify persisters among a population of wild type cells based on their distinct fluorescence behavior.

In the second part of chapter 2, we present a platform we have developed to study interactions between SibC and IbsC. SibC RNA as a natural gene-silencing molecule with all the attractive traits of an RNA molecule, appealed to us with its great potential to be turned into a generic gene-silencing tool. By studying the

interactions between the SibC molecule (or its mutants) and its known target, IbsC we envisioned the core essential sequence and structural features of the molecule could be deciphered. Subsequently, replacing the target recognition domain (TRD) for IbsC with the TRD for any molecule of interest, could give rise to a generic gene-silencing tool. To study the interactions between the toxin and antitoxin, we previously relied on growth assays (i.e. death/ rescue for unsuccessful and successful interactions respectively). However, to simplify the approach and for improved accuracy, we decided to replace the death phenotype with fluorescence. As such, successful interactions between SibC or its derivatives and IbsC would result in a distinct fluorescence behavior that could be measured. Developing these fluorescence-based platforms would allow us to take advantage of sensitive, high throughput and resolution techniques such as fluorescence assisted cell sorting (FACS) and time resolution microscopy to study these molecules instead of traditional methods that rely heavily on culturing.

While developing the platforms above we gained some insights into the biology and regulation of these molecules. For example, we learned that toxicity of the IbsC molecule was not an all or none phenomenon determined by certain sequence and structural requirements that we had previously elucidated. It was also dose-dependent. However, the dosage at which the molecule became toxic remained to be discovered. Furthermore, the rescue capabilities of the antitoxin molecule were shown to depend on the concentration of the toxin and antitoxin. But, the optimal antagonizing ratio remained to be uncovered. Chapter 3 describes our efforts to unravel this information. To expand our knowledge of these systems, we actively pursued studying their regulation at the RNA and DNA levels. We believe many TA-based applications will benefit from this knowledge. Scientists have reported that many proposed TA-based applications are delayed in their advancements because they require accurate knowledge of dosage and control over it.

In the first component of this project, we engineered *E. coli* strains with the capability to produce different levels of the IbsC toxin and SibC antitoxin. Using Reverse Transcription-quantitative PCR (RT-qPCR), we showed that we could accurately measure and have control over the toxin and antitoxin levels in these cells. By assessing growth under different levels of toxin and antitoxin induction, we deciphered the toxic RNA dosage and [antitoxin RNA]/[toxin RNA] for successful rescue. Some other details about the regulation of these molecules at the RNA level will be discussed. In addition, we believe the strains we have developed here, can be utilized in the persistence studies described before. In combination with the aforementioned fluorescence platforms, they will enable us to establish links between toxin expression and persistence. Moreover, we can utilize these strains to establish links between *ibsC-sibC* expression and other bacterial behaviors, as well as for any other application that requires fine control over production and measurement of these molecules.

In the second component of this project, we examined the regulation of *ibsC* and *sibC* genes at the DNA levels. We studied the promoter of the antitoxin and a regulatory region upstream of the toxin core promoter at its 5' Un-Translated Region (5' UTR) to gain insights into regulation of expression from these genes in their native context. Our lab had previously studied the toxin promoter. We believe that studying these regions, could add to our limited collection/tool box of promoters to drive expression of genes differently for different purposes.

In chapter 4, we solve one of the biggest pieces of the IbsC/SibC puzzle, namely their expression levels. Because the IbsC-SibC TA pair, as a type 1 TA pair, relies on RNA-RNA interactions to achieve regulation, the levels of the toxin mRNA and its antitoxin RNA must be carefully tuned to ensure the most beneficial antitoxin sRNA/toxin mRNA ratio for a desired function by these molecules.[31] Therefore, it is important to measure the levels of these RNA pairs as a function of developmental stage or environmental conditions in order to fully elucidate

their functions.[73] However, it is an extremely difficult task because of their low copy numbers[74] and technical challenges for detecting low-abundant RNAs. Conventional techniques, such as cloning, northern blot, and microarray, are not sensitive enough to detect low-abundant RNAs[75,76,77]. Therefore, in spite of great interest in the knowledge of expression from these genes, it remains an unexplored area, consequently impeding the growth of knowledge on physiological functions of these molecules. In this chapter, we summarize our efforts in optimizing an RT-qPCR based approach for the detection of these low-abundance molecules. Using this approach, for the first time to our knowledge, we detected IbsC toxin mRNA produced by the native chromosomal *ibsC* gene. The reliability of these results was confirmed by RNA-seq. We then used this method to measure the abundance of SibC and the IbsC mRNA in wild type *E. coli* cells at different stages of growth and showed that their transcription is growth phase dependent. Additionally, we estimated average copy numbers of SibC and the IbsC mRNA per cell. Towards investigating the potential role of the toxin in stress, we measured expression of these molecules in a slowly growing, stressed population of nutrient deprived cells. Altogether, the effective RNA detection method, the knowledge on the expression of the chromosomal *ibsC-sibC* genes acquired by this method, and the engineered cells containing additional, regulatable *ibsC-sibC* genes, will lend themselves useful to finally interrupt the long pause in the study of this toxin/antitoxin pair. The tools, knowledge and techniques developed here will open the avenue to new discoveries pertaining to the biological relevance of these molecules.

In chapter 5 of this dissertation, we examined the other homologous *ibs/sib* gene copies. As previously mentioned, *ibs-sib* pairs exist in 5 repeats in the MG1655 genome. There are even different repeats of these genes across different species of *E. coli*. There is interest in the literature as to why there are multiple copies of these loci. Fozo offers multiple hypotheses as follows.[78]

We know that the *ibs* genes possess nearly canonical ribosome binding sites. In addition, their ribosome binding sites and coding sequences are well conserved across species. This suggests that the sequences are maintained for a function. However, observations, both by others and us, suggest that their expression levels are low. One possibility could be that the different Ibs proteins have different functions. However, due to their incredible similarity, they are unlikely to have different targets, but may interact with the same target differently (or interact with the same target under different growth/stress conditions). Alternatively, the multiple copies of the *ibs* genes may exist to allow the bacterial cell to fine tune the Ibs production. This could be achieved if the different genes were expressed differently and showed variation in gene expression patterns. Another possibility could be that the presence of multiple copies ensures sufficient production of these molecules. This hypothesis assumes that the Ibs proteins have a redundant function. It also assumes that the cell may require a specific level of total Ibs protein production from the multiple *ibs* genes while tightly controlling expression from each gene to avoid the toxicity associated with each Ibs overproduction.

To test these hypotheses, careful evaluation of the transcriptional and translational control of each gene, with respect to how quickly and strongly it is expressed under a given condition is necessary. To begin these investigations, we decided to detect and measure endogenous Ibs and Sib RNA levels. A challenge in studying these RNAs individually was the great sequence similarity. Within each individual sequence, we found the region that was most dissimilar to the other four copies. Targeting these unique regions, we designed RT-qPCR primers and confirmed the specificity of our RT-qPCR reactions for amplification of individual *ibs/sib* copies by Sanger sequencing. We also designed nanostring probes and next gen sequencing primers, with the same binding sequence as our RT-qPCR primers, for potential experiments in the future utilizing these high

throughput methods. We validated them theoretically by a team of bioinformaticians at Nanostring and technicians at Illumina sequencing.

Utilizing the RT-qPCR approach, we performed RNA expression analyses from *ibsA-E* and *sibA-E* genes under two different growth stages (i.e. early log and late log, 0.3 and 0.8 respectively) to confirm the above hypotheses about the expression patterns from these genes.

Taken together this thesis details our efforts from developing platforms, strains and methods for the study of the Ibs/Sib TA systems, to their actual characterizations. We investigate the regulation, expression and potential functions of these systems with a special focus on the IbsC-SibC pair. Theses studies contribute tremendously to our biological understanding of these TA systems. They are additionally important due to the potential by these molecules to be turned into antimicrobial agents and synthetic biology tools (e.g. gene silencing molecules). We need to understand these systems well to ultimately develop useful TA-based applications.

2 Chapter 2: Developing DNA Constructs And Cellular Platforms To Facilitate The Study Of The IbsC/SibC As Antimicrobial Agents And For Future Sib sRNA-Based Knockdown Tools

2.1 Introduction

RNA molecules act as regulators in all organisms in which they have been studied. In bacteria, they are named small RNAs (sRNA). A subclass of these sRNAs silence gene expression by base pairing with the messenger RNAs (mRNAs) of their targets, which encode short, hydrophobic and toxin proteins. This base pairing changes the transcript's stability and/or translation. They are classified as type I toxin-antitoxins (TA).

As discussed in depth in the previous chapter, advances in characterization of some of the type I TA pairs have suggested functions for the small toxic proteins at the cell membrane, and given insights into how they are regulated by sRNAs.[79] Consequently, these molecules have been considered as versatile tools for a multitude of purposes in basic research and biotechnology.

Herein, I would like to briefly describe the potential by both the type I IbsC toxin and its cognate SibC sRNA antitoxin to be turned into such molecules, antibacterial agents specifically. However, to fully exploit their potential for the purpose of developing practical applications, further characterizations of these molecules are required. As current technical challenges are major factors delaying the study of these molecules, this chapter summarizes our efforts to develop platforms to facilitate these characterizations. Below, we will describe for the toxin and the antitoxin individually, the potential to be turned into antibacterial agents, required characterizations and the platforms we have developed to perform these characterizations.

2.1.1 The type I IbsC toxin, a potential drug target for combating persistence and chronic infections

As previously mentioned, for many type I toxins, the functions of endogenous levels of the proteins are still obscure. However, hypotheses of their physiological roles may be drawn from their well-studied type II counterparts (and few type I). One of these hypothesized roles is in persistence, where the production of toxins in a sub population of cells causes them to enter a temporary dormant, multidrug tolerant state and become culprits of recalcitrant chronic infections.[79] This is of a particular interest to us because persistence has been underappreciated for some time as a mechanism for bacteria to evade antibiotics, in spite of the fact that it poses a major challenge for the treatment of infectious diseases, as most common antibiotics kill only growing bacterial cells. [80]

Based on recent studies that have associated persister formation with dormancy and TA systems and our own observations of IbsC causing reversible growth stasis (discussed in chapter 4), we hypothesize that induction of *ibsC* leads to persister cell formation and therefore, survival under times of stress such as antibiotic exposure. Sufficient evidence for our hypothesis, will present IbsC as a potential antibiotic target for *E. coli* and perhaps other more clinically relevant pathogens. We believe that answers to two smaller questions can raise sufficient evidence for our hypothesis: 1) Do persisters exhibit higher toxin expression relative to wild type cells? 2) Are the cells with higher levels of the toxin expression, more likely to form persisters?

However, there are challenges associated with the identification and isolation of persister cells. The extremely low percentage of persister cells in a bacterial population and complex pathways involved in persister formation have delayed the study of persistence. The traditional methods such as plating and colony

counting after antibiotic treatment to isolate the surviving persister population do not account for so called viable but not culturable (VBNCs) cells. Therefore, these approaches should usually be accompanied by microscopy examinations, more specifically live/dead stainings, to validate persister enumerations by plating. To facilitate isolation and identification of the rare persister cells for the purpose of investigating the potential role of the IbsC toxin in persistence, we developed constructs and engineered *E. coli* strains here to link persistence to fluorescence. Subsequently, high throughput, sensitive, accurate and high resolution (single-cell fluorescent) techniques such as Fluorescence Assisted Cell Sorting (FACS) and Time-lapse microscopy can be used to substitute antibiotic sensitivity assays for identification and isolation of these persister cells. We believe our efforts to facilitate the study of persistence will take us steps closer to understanding the mechanisms of bacterial persistence and ways to combat them.

2.1.2 The type I SibC antitoxin, a potential antimicrobial agent and useful tool in synthetic biology

The potential for the SibC antitoxin as an antibacterial agent and a versatile tool in synthetic biology, lies in its gene silencing capabilities. The discovery of important and diverse functional RNAs (riboswitches and non-coding sRNAs) in cells which act by interacting with other nucleic acids, proteins and small molecules, has inspired scientists to design synthetic RNA-based switches of gene expression and synthetic counterparts for novel functions including sensing, and regulatory. [81,82] In addition to diverse regulatory mechanisms, there are unique properties of RNA that make it an attractive design substrate. Important features that make them advantageous relative to proteins and support their scaling into large genetic system designs include: better secondary structure and folding

predictability, more compact genetic footprints, faster timing, less energy and resource load on host since they do not require translation, and flexibility to undergo change to generate new functions, [82] as James Collins, a Boston university biomedical engineer once said, these RNA-based components can be cut and pasted together into genetic circuits, "just as an electronic engineer uses a soldering gun to put together electronic components on a circuit board ". [14]

Great applications from environment and agriculture to health and medicine have been offered by these large-scale genetic system designs based on catalytic and antisense mediated activities of RNAs in nature. However, a main limitation on the field is the availability of new and diverse functional RNA components, which form the basic building blocks for the construction of these complex genetic circuits. There are computational and molecular evolution approaches to generate new functional components. But, these components can also simply be derived from naturally occurring elements. The RNA component can be "harvested" from its native surrounding and used in a synthetic genetic context to output the desired activity. Also, the native function of the naturally occurring RNA element can be changed to new functions. [82] This is the future that I envision for the SibC molecule and potentially its other homologs as these genes exist in multiple copies within the genome of the organisms hosting them.

SibC, as a natural, antisense, gene-silencing sRNA, has an exciting potential for generation of new functional RNA-based genetic parts that can be expanded into more complex genetic circuits. To exploit this potential, however, further characterization of the molecule is required.

Mechanistic insights into the action of SibC are lacking. The minimal sequence and structural feature of the SibC sRNA that are essential for its effective gene silencing need to be discovered by studying the interactions and monitoring the activity of the wild type and mutant (SibC-derived) molecules against the target

(i.e. IbsC mRNA). Traditionally, these studies are carried out by performing death/rescue assays which are time consuming and laborious. Herein, we developed constructs and strains to facilitate the study of the interactions by replacing the death/rescue phenotypes with simple fluorescence.

We believe the knowledge from these studies will feed directly into the design of efficient artificial sRNA (afsRNA) scaffolds derived from the *E. coli* sRNA, SibC. Substituting the original target recognition sequence for IbsC for example, with a new sequence recognizing mRNA of any essential gene should lead to effective silencing of that essential gene and thus bacterial cell death. Virulence and resistance genes are also plausible gene targets for fighting bacteria. Consequently, this system could be adapted to fight against more medically relevant bacteria such as *C. difficile*. Furthermore, the designed afsRNAs could be used as gene-silencing tools in more elaborate ways such as in detailed-studies of gene function *in vivo*, and metabolic engineering of bacterial strains for higher performance.

Ultimately, transferrable capabilities into other organisms and extendable architecture into more complex genetic devices are some of the desired features that we envision for these molecules. Such engineered riboregulators may lend further insight into mechanistic actions of the endogenous SibC molecule. In subsequent sections, we will discuss some of the insights we gained into the biology and regulatory mechanisms of these molecules, while developing and optimizing the described platforms for the study of IbsC and SibC molecules. These insights were followed up by thorough investigations in the subsequent chapters to expand our biological and regulatory knowledge of these molecules.

2.2 Materials and Methods

2.2.1 Oligonucleotides and reagents

All PCR primers and oligonucleotides used in this study were chemically synthesized by Integrated DNA Technologies (Coralville, IA, USA). Oligonucleotides that were longer than 40 nucleotides were purified by 10% (8 M urea) denaturing PAGE (polyacrylamide gel electrophoresis) prior to usage.

Kanamycin, chloramphenicol, Ampicillin, isopropyl β-D-1-thiogalactopyranoside (IPTG) and Anhydrotetracyclin (Atc) used in this study were purchased from Sigma Aldrich (Oakville, ON, Canada).

For molecular cloning, the High-Fidelity PCR enzyme mix, Klenow fragment (exo-), and T4 DNA ligase were purchased from Fermentas (Burlington, ON, Canada). Restriction enzymes (EcoRI and BamHI) were obtained through either Fermentas or New England Biolabs (Pickering, ON, Canada). Plasmid Mini-prep kits were purchased from either Qiagen (Mississauga, ON, Canada) or Promega (Madison, WI, USA).

2.2.2 Syntheizing constructs by annealing, Klenow and crossover PCR

P_{opt}-*gfp* construct

The P_{opt} promoter sequence was derived from the Anderson promoter collection (iGEM, Berkley, accession number BBa_J23100). P_{opt} is a strong constitutive promoter. Two complementary oligonucleotides were annealed to create the P_{opt} sequence. Annealing involved combining the oligonucleotides at equimolar concentrations, heating them at 90^{o}C for 1 min, followed by cooling at ambient

temperature for 10 min. The promoter was then digested with XhoI and EcoRI restriction enzymes and cloned into pNYL-MCSII at these restriction sites. *gfp* was cloned downstream of this promoter at EcoRI and BamHI sites (Figure 2.2A, Left).

This construct served as a positive control for our promoter-reporter assays.

P_{opt}-*gfpAAV* construct:

gfpAAV has the following mutations relative to wild type *gfp*: a) Three amino acid substitutions: S2R, T65G and S72A b) An extra (GSGC) tag before the stop codon that is absent in wild type *gfp.* These mutations make the GFP AAV protein less stable with a lower half-life relative to wild type GFP. Therefore it is used to report the transient state of fluorescence, since no fluorescence build up from the previous states will be present. To generate the *gfpAAV*-encoding gene, we used the wild type GFP as a template and designed two primer sets to introduce the mutations into the original template by PCR. We generated the N-terminus and C-terminus coding regions separately in two different PCR reactions and fused the two parts of the molecule by crossover PCR. The oligonucleotides to synthesize *gfpAAV* are presented in Supplementary Table 2.1and the modifications relative to wild type GFP are highlighted. The primers were also designed so to introduce the EcoR1 and BamH1 restriction sites into the molecule. Following synthesis, digestion and purification, *gfpAAV* was cloned at the EcoRI and BamH1 sites of the pNYL-MCSII plasmid harboring the P_{opt}-*gfp* construct to replace the wildtype *gfp* as shown in Figure 2.2A, middle.This construct also served as a control in our experiments.

$P_{rrnB140}$-*gfpAAV* construct:

PrrnB (-40+1) is a growth rate dependent promoter used in our studies to drive the

expression of the fluorescence reporter gene only at times of rapid growth, so to allow for distinguishing between actively growing cells and non-growing cells (e.g. persisters and non persisters). The E. coli rrnB operon can be found online (gene bank accession: J01695). Complementary oligonucleotides were annealed to construct this promoter. They are presented in Supplementary Table 2.1. The annealing procedure was performed as described for Popt. The oligonucleotides also included the XhoI and EcoR1 restriction sites. Following synthesis, digestion and purification, the PrrnB140 sequence was cloned into the pNYL-MCSII plasmid harboring the Popt-gfpAAV, to replace the Popt promoter (Figure 2.2A, Right).

P_{LtetO1}-*ibsC-gfp*:

We fused a wild type *gfp* into the *ibsC* sequence between codons 12 and 13, to make a fluorescent construct for assessing interactions between IbsC and SibC by means of fluorescence. To express the molecule inside the cells, we placed it under regulation of the tetracycline inducible promoter P_{LtetO1} and cloned it into the pBSΔBADkan plasmid. The construct was either expressed from the plasmid directly or integrated into the genome and expressed chromosomally. Klenow was performed to fuse the P_{LtetO1} promoter to the 5' UTR of *ibsC* and its first 12 codons. Annealing was performed to generate the rest of the *ibsC*-encoding region. PCR was performed to amplify the wild type GFP without a stop codon or ribosome-binding site. Subsequently, a series of crossover PCRs were performed to fuse all the components. By PCR, we incorporated PmeI and BamHI restriction sites into the P_{LtetO1}-*ibsC-gfp* construct, which was subsequently cloned into the pBSdeltaBADkan plasmid at the same sites.

2.2.3 Strains and plasmids

The cloning experiments were performed in *E. coli* DH5αZ1, provided by the laboratory of Hermann Bujard. [83] This strain is a variant of the *E. coli* DH5αZ cloning strain which was derived from *E. coli* K-12. The *E. coli* DH5αZ1 has been engineered to constitutively express a tetracycline repressor (TetR) and a lac repressor (LacR). Therefore, it allows for inducible expression of genes under regulation of tetracycline and IPTG inducible promoters where these inducing agents are provided to the cells extracellularly. For example, the P_{LtetO1} promoter used in our experiments is tightly repressed by TetR, but it can be induced over 2500-fold in the presence of anhydrotetracycline (Atc).

The pNYL-MCSII plasmid is derived from pZE21-MCS1. This plasmid contains a tetracycline-inducible promoter (P_{LtetO1}) and the ColE1 origin of replication (high plasmid copy numbers). Our group has also removed the consensus ribosome-binding site (RBS) present on pZE21-MCS1 and restored its multiple cloning site (Swanson and Li, unpublished data). [83,84]

The pNYLcat vector is derived from the pNYL-MCS11 plasmid where the marker for kanamycin resistance was replaced with one for chloramphenicol resistance.

The pBSdeltaBADkan is depicted in **Figure 2.1** B. Encoding the *polB* and *araC* genes, which are also encoded by the chromosome; it is used for homologous recombination of genes into the chromosome. The pKD46 plasmid encodes the phage λ Red recombinases and is ampicillin resistant. For more information regarding these plasmids refer to the work by Campbell & Brown and Datsenko & Wanner. [85,86]

2.2.4 Cloning

Molecular cloning steps, including PCR, restriction digestion, and ligation were conducted following established protocols provided by suppliers. The synthetically prepared promoter-reporter constructs were cloned into the XhoI, EcoRI and BamHI sites in pNYL-MCS11 as previously described. Plasmids carrying the constructs were transformed into *E. coli* strain DH5αZ1 by electroporation. Cloned constructs were confirmed by DNA sequencing at Mobix Lab (McMaster University). They were then amplified by primers containing the PmeI and BamHI restriction sites and cloned into the pBSΔBADKan plasmid at these sites for subsequent integration into the chromosome described below.

For the SibC-IbsC/GFP interaction screens, *sibC* was placed under the control of a lactose-inducible promoter (P_{LlacO1}). The P_{LlacO1}-encoding sequence was synthesized by annealing two complementary oligonucleotides while *sibC* was synthesized using Klenow extension of oligonucleotides with overlapping regions. P_{LlacO1} was subsequently fused to the *sibC* sequence by crossover PCR. These inserts were digested with XhoI and BamHI restriction enzymes and were ligated into pNYLcat at the same sites. The P_{LlacO1}-*sibC* ligation products were transformed into *E. coli* DH5αZ1, and successful clones were confirmed by sequencing before being transformed into the *E. coli* DH5αZ1*ibsCgfp* (see below for the procedure to create this strain).

2.2.5 Integration of promoter-reporter constructs and ibsC-gfp hybrid into chromosome with pBSdeltaBADkan and pKD46

For the integration of any construct in this chapter into the chromosome of *E. coli* DH5αZ1, pBSdeltaBADKan and pKD46 were used according to the protocols by Campbell & Brown and Datsenko & Wanner.[85,86] The constructs, as previously

described, were either PCR amplified from a plasmid or synthesized by annealing, Klenow and crossover PCR. They were then purified, and subsequently ligated into pBSdeltaBADkan at PmeI and BamHI sites. Ligation products were transformed into the *E. coli* DH5αZ1 and successful clones were confirmed by sequencing. Using the construct-harboring pBSdeltaBADkan plasmids as template, a DNA sequence of about 2.5 kb encoding 500 bp from polB, the kanamycin resistance marker, the desired construct, and 500 bp from *araC* was amplified by PCR. 50 ng of this linear DNA was transformed into *E. coli* DH5αZ1 harbouring pKD46 (which encodes the phage λ Red recombinase). Transformants were then plated on LB agar supplemented with kanamycin and cultured at 37^{o}C in order to select for integrants with the desired construct and the kanamycin resistance marker integrated at the araBAD locus as well as for the loss of the temperature sensitive plasmid. Loss of the ampicillin harboring pKD46 plasmid was confirmed by no growth in ampicillin containing media. Successful integrants were confirmed by colony PCR in which the polB-kanR-construct-araC fragments were amplified. The amplicons were further validated by sequencing.

Using this approach, *ibsC-gfp* under the control of a tetracycline inducible promoter (P_{LtetO1}) was introduced into the chromosome of *E. coli* DH5αZ1, which endogenously expresses the tetracycline repressor. We named this strain *E. coli* DH5αZ1*ibsC-gfp*. It was subsequently used in SibC-IbsC interaction screens.

2.2.6 Growth assays/cell cultures

Bacteria were routinely cultured in LB. When indicated, 50μg/mL of kanamycin or 25μg/ mL of chloramphenicol was added. In the SibC-IbsC interaction assays, expression of *sibC* and *ibsC* were induced with 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) and 200 ng mL^{-1} of anhydrotetracycline (Atc) unless otherwise indicated.

Growth curves were used to assess toxicity of the *ibsC-gfp* construct. The cells were grown in LB-Kan overnight. Subsequently, they were subcultured by 400-fold dilution in fresh LB-Kan, in the presence and absence of Atc and IPTG (for induced and uninduced *ibsC* and *sibC* conditions respectively). They were grown for multiple hours and growth was monitored hourly by measuring the optical density at 600 nm (OD_{600}) for 0.2 mL of each culture using a 96 well microplate reader. Growth curves were generated. For fluorescence measurements, cells were taken after a stated number of hours post subculture and washed in PBS.

In addition, the cells carrying the different promoter-GFP reporters were cultured to different growth phases (i.e. stationary overnight and active growth at OD_{600}=0.3) to examine the behavior by the constructs under these conditions. These strains were grown overnight (16 hours) at 37^{o}C with shaking at 260 rpm in LB broth supplemented with Kanamycin. Following overnight growth, the cells were diluted by 200-fold in 3 mL of fresh LB-Kanamycin media. These cultures were incubated at 37^{o}C with shaking for 3 hours. All assays were carried out in triplicates. After 3 hours, 1 ml from each sample was pelleted at 5000rpm for 5 min, and suspended in Phosphate Buffered Saline (PBS). Overnight cultures were also prepared in the same manner. OD_{600} for 0.2 ml of each sample was measured using a 96 well microplate reader. Hundred microliters of this sample were subsequently used for fluorescence measurements.

We also performed Knock-In growth assays (Supplementary Figure) to isolate the cells that had successfully integrated the correct construct into their genome and lost the recombinant plasmid for downstream studies. We screened the engineered cells for Kan resistance to select for the construct integration into the chromosome and Amp susceptibility to counter select for the strains that still carried the recombinant plasmid. Glycerol stocks of these clones were prepared.

2.2.7 Fluorescence assays

The fluorescence of each sample was measured using a Tecan Safire microplate reader (Männedorf, Germany) at excitation wavelength of 488 ± 5 nm and emission wavelength of 509 ± 5 nm. Fluorescence signals were normalized by the growth of each sample. The assays were run in triplicate.

2.2.8 Microscopy

SibC activity against IbsC-GFP was also assessed by wide-field fluorescence microscopy. Post subculture and induction, cells were grown to OD_{600} of about 0.2. Cells were subsequently mixed with glycerol at a 1:1 ratio. Samples were spread onto a slide, covered, sealed by clear nail polish and scanned.

2.2.9 Western Blots

Western blots were performed according to the protocol described in the thesis by Wendy Mok. [87]

2.2.10 RT-qPCR

Fluorescence knockdown experiments were accompanied by Reverse Transcription-quantitative PCR (RT-qPCR) to assess the interactions between the SibC sRNA and IbsC-GFP mRNA. The primer sequences are shown in Supplementary Table 2.1.

2.3 Results

2.3.1 Developing an E. coli strain to investigate the potential role of the IbsC toxin in persister cell formation

2.3.1.1 Overview of the design

To examine whether IbsC causes persistence in *E. coli* cells, we designed a construct where *gfp,* as the reporter, was placed under control of a growth rate dependent promoter ($P_{rrnB140}$) (**Figure 2.1**A). Utilizing this promoter, *gfp* is highly expressed in cells at times of rapid growth, for example, when the cells are in their exponential phase of growth. At times of slow or no growth, for example, when the cells are in their stationary phase of growth or undergoing persistence, little or no *gfp* will be expressed. This will provide a way to monitor the metabolic state of cells by means of fluorescence.

Bartlett *et al.* found that the core promoter region of P_{rrnB1} (-41 to +1 with respect to the transcription initiation site, containing the -10 and -35 hexamers for recognition by the sigma 70 subunit of RNA polymerase), coined $P_{rrnB140}$ here, is sufficient for growth rate-dependent control of transcription. [88] Therefore, this is the promoter we used for our work here.

As a control for the growth rate dependent activity of $P_{rrnB140,}$ a synthetic strong, constitutive promoter (from the Anderson promoter collection, iGEM, Berkeley; accession number BBa_J23100), coined P_{opt}, was also fused to the reporter genes (**Figure 2.1**A).

To reliably use fluorescence as a measure of the metabolic state of the cells at the time of measurement, a mutated *gfp* coined *gfpAAV*, with a short half-life was

used. As a result of its short half-life, this molecule does not accumulate in the cells and reliably reports only the transient state of the cells. In other words, measured fluorescence levels will be merely a reflection of the promoter activity. When driven by the growth rate dependent promoter, it will show the transient metabolic state of the cells; fluorescence will not be due to GFP accumulation from before the time of measurement.

As a control for the behaviour of the mutated *gfp*, wild type *gfp* was also placed under regulation of the promoters (**Figure 2.1**A). Promoters and reporters were separately cloned into the pNYL-MCS11 plasmid for each promoter-reporter construct. By PCR, the whole promoter-reporter construct was amplified and subsequently cloned into the pBSdeltaBADkan plasmid (**Figure 2.1**B), which shared a homologous region with a portion of the *E. coli* genome (**Figure 2.1** B&C). Amplifying the promoter-reporter construct together with the surrounding homologous region from the plasmid allowed for the integration of the constructs into the genome of *E. coli* by recombination (**Figure 2.1** C).

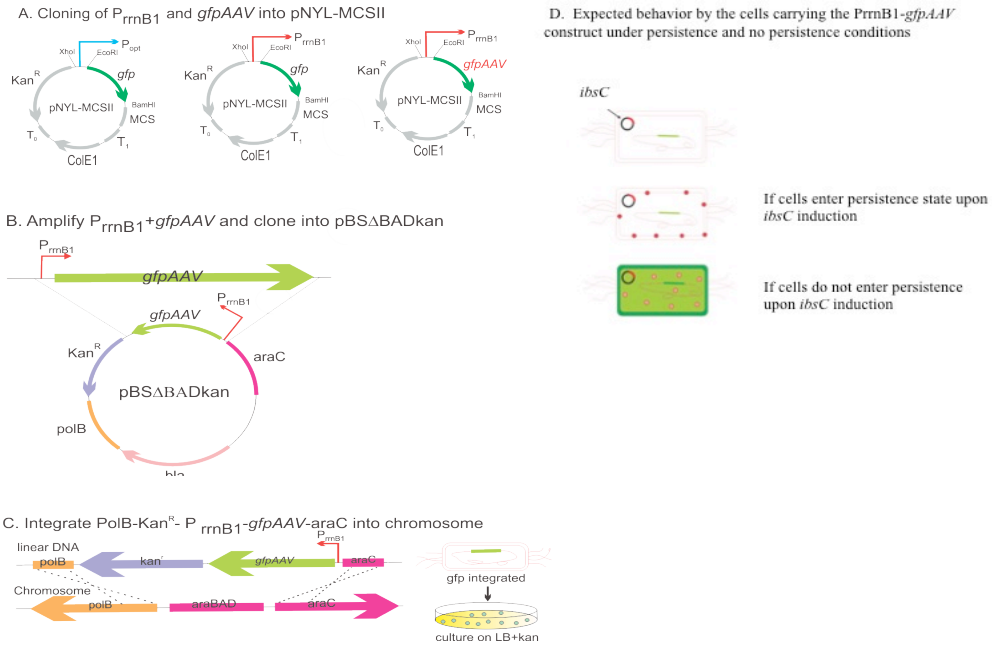


Figure 2.1. Overview of the fluorescence-based platform to investigate the potential role of IbsC in persistence. (A) $P_{rrnB140}$*gfpAAV* is expected to report the transient metabolic state of the cells due to the growth rate dependence activity of the promoter and short half-life of the GFPAAV protein. Therefore, it can be used to distinguish between persister and non persister cells. As controls, we also created constructs that contained the wild type *gfp* and a strong constitutive promoter, Popt. (B) Constructs in A were amplified and cloned into the pBSdeltaBADKan plasmid which shared a homologous region with a portion of the *E. coli* genome for the homologous recombination of these constructs into the genome in (C). (D) If IbsC is provided to cells and causes persistence, the cells carrying the $P_{rrnB140}$*gfpAAV* in their genome will turn dim. Otherwise, they will be fluorescent. This strategy would allow for accurate and easy isolation and identification of persister cells.

Having developed these constructs and strains, they can now be used to study the potential role of IbsC in persistence. As a first step towards studying persistence we may need to isolate persister cells and measure IbsC levels. These strains can be used to isolate persisters simply based on their fluorescence levels, instead of relying on colony counting approaches. In addition, if *ibsC* is supplied to these cells on a plasmid, and causes persistence, the cells will turn dim as the result of the growth rate dependent promoter being unable to drive expression of the fluorescent reporter under dormant conditions (**Figure 2.1**D). The dim persister cells can subsequently be distinguished from fluorescent non-persister cells by accurate, sensitive, high throughput and resolution techniques such as FACS and time-lapse microscopy. Establishing a link between *ibsC* expression levels and persistence will provide sufficient evidence for its involvement in persistence.

2.3.1.2 Confirmation of expected performance by the Constructs

To confirm that the promoter and *gfp* constructs behaved as expected, we performed fluorescence assays as shown in Figure 2.2. The results demonstrated that the only construct that could distinguish between the metabolically active and

non-active cells based on their growth rates (i.e. cells at exponential vs. stationary phase) was the $P_{rrnB140}$ -*gfpAAV*.

The need for the growth rate dependent promoter is evident as neither of the fluorescence reporters driven by the strong constitutive promoter P_{opt} can discern between the two metabolic states of the cells.

The results clearly demonstrate the instability of the GFPAAV relative to wild type GFP as expected. When placed under regulation of the strong constitutive P_{opt} promoter, the wild type GFP fluoresces much more highly than the GFPAAV. In spite of its lower fluorescence levels relative to wild type GFP, the need for this short-lived GFP for the purpose of probing the metabolic state of the cells becomes apparent when these two fluorescent reporters are placed under regulation of a growth rate dependent promoter. Although regulated by a growth rate dependent promoter, the *gfp* in the PrrnB-*gfp* construct leads to higher fluorescence in the stationary phase when cells experience lower growth rates than in the log phase. We speculate this is due to the accumulation of the highly stable wild type GFP reporter in these cells from the earlier stages of growth. When replaced by the transient and unstable *gfpAAV* reporter, the fluorescence behavior reverses to our favor. $P_{rrnB140}$-*gfpAAV* is the only construct to accurately report the metabolic state of the cells (i.e. fluorescence when cells are actively growing during the log phase and not fluorescent when cells are not growing as in stationary phase). Fluorescence assays here were validated by western blot analyses as shown in the Supplementary Figure 2.

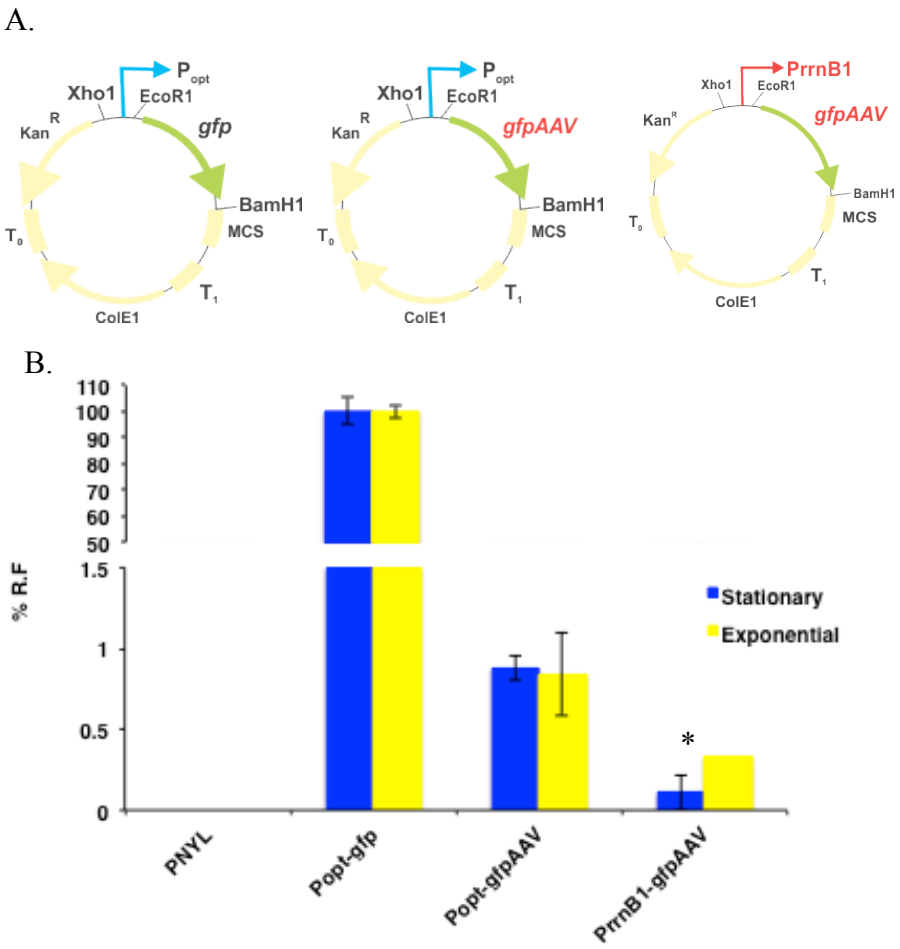


Figure 2.2. Fluoresecnce behavior by the promoter-reporter constructs designed to ultimately study persistence. Top. Different promoter-reporter constructs. PrrnB140gfpAAV is expected to report the transient metabolic state of the cells due to the growth rate dependence activity of the promoter and short half-life of the GFPAAV protein. Therefore, it can be used to distinguish between persister and non persister cells. As controls, we also created constructs that contained the wild type *gfp* and a strong constitutive promoter, Popt. Bottom. Fluorescence performance of the strains carrying the different promoter-fluorescence reporters on their chromosome, taken at stationary phase from overnight cultures in blue and at exponential phase from cultures grown 3 hours post subculture. PrrnB1-gfpAAV is the only construct that can distinguish between the metabolic states of the cell. The two-sample t test calculates a P-value of 0.026 for the difference between the the two metabolic states tested and measured by PrrnB1-gfpAAV. The data are averages of three repeats.

An improvement to the current platform would be to increase the fluorescence signal from the $P_{rrnB140}$ *gfpAAV* construct. As previously mentioned, we only used the -40 to +1 region of the growth rate dependent promoter. The extended promoter containing the -70 to +1 region, coined P70, including the upstream (UP) element will increase transcription by interacting with the alpha subunit of RNA polymerase and the binding sites for the positive transcription factor FIS. Using P70 instead of $P_{rrnB140}$ would probably enhance fluorescence. In addition, providing the constructs to the cells on a high copy plasmid would probably increase the fluorescence levels relative to the current design where they are expressed from the chromosome. Nonetheless, the $P_{rrnB140}$ *gfpAAV* construct has proven to be useful to distinguish between the different metabolic states of the cells, even in the current design.

2.3.1.3 Isolation of persister cells

As a step towards establishing a link between persistence and TA systems, we need to elucidate the conditions under which persistence happens. As previously mentioned, one of the factors delaying the study of persistence are complex pathways involved in persister formation. It should be noted that IbsC might play a physiological role in persistence only under certain stress conditions and/or at certain levels of expression.

Therefore, *ibsC* induction may need to be accompanied with different temperature, nutrient, oxidative, acidity, osmotic and antibiotic stress conditions to (elucidate the right conditions for the toxin to) cause persistence. Furthermore, the expression of the toxin may require tuning. Choice of the antibiotic can also

be challenging because the toxin under study and its mechanisms of actions are poorly understood. This is relevant because different classes of antibiotics act against defined and different targets and one of the proposed models by which a toxin causes persistence in response to an antibiotic is by decreasing the activities of the targets of the antibiotic. The basis for this proposed model lies in the fact that bactericidal antibiotics require active targets to exert their function. Therefore, choice of the antibiotic for these experiments can vary based on the target for the toxin inside bacteria or its mechanism of action. To tackle this problem, a couple of different antibiotics from unrelated classes (e.g. cell wall synthesis inhibitors, DNA-damaging, protein synthesis inhibitors) should be tried. In addition to the choice of antibiotic, its mode of administration may also be important to cause persistence. For example, stepwise administration (i.e. an initial exposure to a low dose of an antibiotic followed by a higher dose rather than being exposed to a high dose from the beginning) has been shown to yield higher levels of TisB -induced persisters.[24]

Here, we performed antibiotic sensitivity assays to isolate persister cells in response to Kanamycin and Ampicillin treatments. Antibiotic treatment of microbial populations typically results in a biphasic killing pattern. It leads to an initial phase of rapid killing of the bulk population. However, beyond a certain threshold (concentration or duration in our case), a killing plateau is observed, as only persister cells remain viable. Colony Forming Units (CFUs) vs. treatment duration are shown in **Figure 2.3**.

To generate these curves, overnight cultures of *E. coli* were diluted 1:1000 in fresh media in two separate flasks and grown to the OD_{600} of 0.2 and 0.6. At these OD_{600} levels, they received antibiotic treatment. Kanamycin was supplied to one flask at a final concentration of 50 ug/ml and Ampicillin to the other flask at a concentration of 25 ug/ml to yield persisters. Every hour, post antibiotic treatment, samples were drawn from the

cultures, appropriately diluted and plated to count the viable CFUs as a measure of surviving cells. Viable CFU counts were consequently plotted against treatment duration.

We were only able to isolate persister cells in response to Ampicillin, in a growth phase dependent manner. As shown in **Figure 2.3**A, the cells that were grown to the OD_{600} of 0.2 prior to antibiotic treatment, never generated a second phase. By hour 2, no CFUs could grow on culture plates, indicating that no cells had survived the antibiotic treatment. Therefore, no persisters were able to be isolated under these conditions. Cells in **Figure 2.3**B were grown to a later log phase (OD_{600} of 0.6) prior to antibiotic treatment. This lead to a biphasic killing curve, hence formation of persisters.

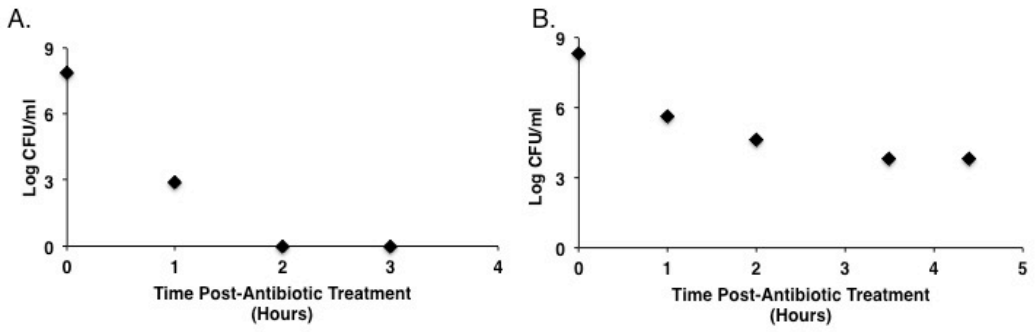


Figure 2.3.Pesister cell isolation through conducting antibiotic sensitivity assays and creating antibiotic killing curves. A. *E. coli* cells were grown to the OD600 of 0.2 prior to receiving Ampicillin treatment at a final concentration of 25 ug/ml. No persisters were able to be isolated under these conditions. B. *E. col*i cells were grown to an OD_{600} of 0.6 prior to receiving Ampicillin treatment. Persisters formed under these conditions. Persister formation seems to exhibit growth phase dependency.

The *E. coli* strains developed in this project that carry the $P_{rrnB140}$-*gfpAAV* construct are expected to exhibit unique fluorescence behaviors under persistence

conditions. Therefore, the traditional colony counting approach used here to isolate and identify persister cells can be simply substituted with more advanced fluorescence-based techniques such as FACS to facilitate the study of persistence. Together with toxin expression analyses, we can investigate whether there is a link between these persisters and TA systems.

2.3.2 Developing a platform to study IbsC-SibC interactions

2.3.2.1 Developing a hybrid *ibsc-gfp* encoding sequence and strains that express it

To study the interaction between IbsC and SibC, we have designed a system to monitor the activity of wild type or mutant SibC molecules against IbsC by means of fluorescence.

Previously, we used growth assays to monitor and assess the sRNA-mRNA interactions. We placed the *ibsC* gene under the control of a tetracycline-inducible promoter (P_{LtetO1}) and knocked it into the genome of *E. coli* DH5αZ1. Its expression was induced with anhydrotetracycline (Atc). *sibC* and its derivatives were cloned downstream of a lactose-inducible promoter (P_{LlacO1}) on a plasmid coined pNYL-cat (derived from pNYL-MCS11) which was subsequently transformed into these cells (**Figure 2.4**). The expression of *sibC* was induced with isopropyl β-D-thio-galactopyranoside (IPTG). Both of these inducing factors were provided to cells in their common Luria-Bertani (LB) medium and extracellularly. The toxin and antitoxin were either induced individually or together. The effects of their production on bacterial growth could be measured 6 hour (h) post-induction and used to assess their activities and interactions. The ability of wild type or mutant SibC to suppress the expression of wild-type *ibsC*

and prevent the growth suppression associated with the overproduction of the toxin could be examined through growth assays carried out in these *E. coli* DH5αZ1 strains.

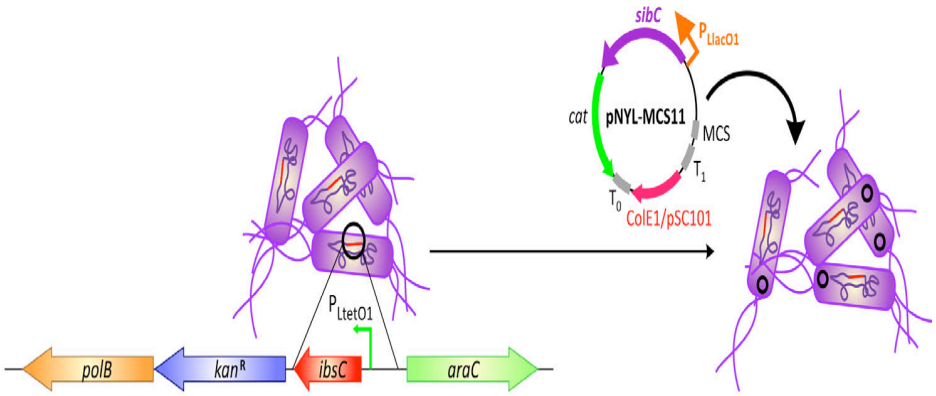


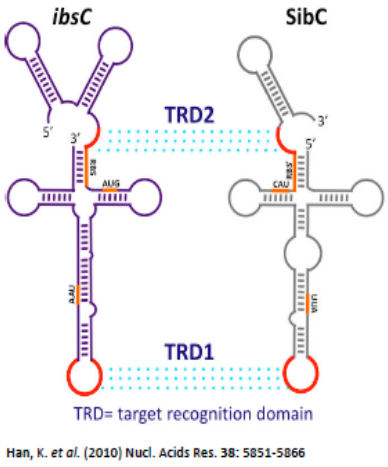
Figure 2.4. The design of an assay platform to examine the SibC sRNA-IbsC mRNA interactions in *E. coli* by growth assays. *ibsC* or *ibsC-gfp* is placed under the control of a tetracycline-inducible promoter (P_{LtetO1}) and knocked it into the genome of *E. coli* DH5αZ1. *sibC* and its derivatives are cloned downstream of a lactose-inducible promoter (P_{LlacO1}) on a plasmid coined pNYL-MCS11 which is subsequently transformed into these cells. The illustration was provided by Wendy Mok.

However, to simplify the analytical techniques (e.g. eliminate the need for protein purification for complementary western blot analyses) and measurements and for improved accuracy, we decided to replace the growth/death phenotype with fluorescence for monitoring the activity of SibC against IbsC. This would allow for accurate and high-throughput screening strategies that will be discussed later. In addition, the use of fluorescence would allow for studies in which IbsC is produced at sub-lethal levels.

To achieve this, we synthetically fused the *gfp*-encoding sequence into the *ibsC* sequence by cross over PCR. The location on *ibsC,* for *gfp* insertion was chosen,

so that the toxicity of the molecule would be lost but its interaction with SibC would remain intact. In a recent study published by Han and colleagues, it was demonstrated that the recognition between IbsC mRNAs and SibC sRNAs is dependent on two target recognition domains (TRDs), coined TRD1 and TRD2. [89] RNA interactions at these two sites allow for the formation of a "kissing complex" intermediate, followed by extension of base pairing to the neighbouring regions to form a complete stable duplex (**Figure 2.5**A). Therefore, we kept away from these two sites. Instead, we chose to insert the *gfp* sequence within a stretch of hydrophobic amino acids previously shown by our lab to be important for toxicity. [90] The arrow on **Figure 2.5**B points to the insertion site. From here on, I will refer to this hybrid molecule as *ibsC-gfp*.

A



B.

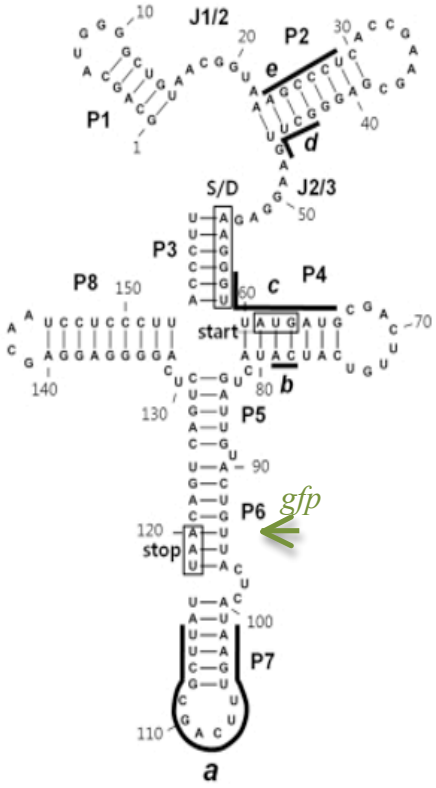


Figure 2.5. Design of an *ibsc-gfp* construct to examine the SibC sRNA- IbsC mRNA interactions in *E. coli* by fluorescence assays. A. Predicted secondary structure of the IbsC toxin and SibC antitoxin and initial sites of interaction, namely TRD1 and TRD2. B. Secondary structure of the IbsC toxin and the location on *ibsC* for *gfp* insertion, shown by the green arrow. This insertion site is away from the two previously determined TRDs and within a stretch of hydrophobic amino acids that were previously shown to be important for toxicity of the molecule. It was strategically chosen, so to rid the molecule of its toxicity but maintain its interactions with SibC. The secondary structure figures are borrowed from Han, *et al.*[89]

2.3.2.2 Confirming loss of toxicity of IbsC-GFP by growth assays

To test for the loss of toxicity of IbsC-GFP, growth assays were performed. Results from these assays indicated that cells with induced levels of *ibsC-gfp* showed normal levels of growth similar to control and those with uninduced levels of *ibsC-gfp*, testifying to the loss of toxicity by the molecule (Figure 2.6).

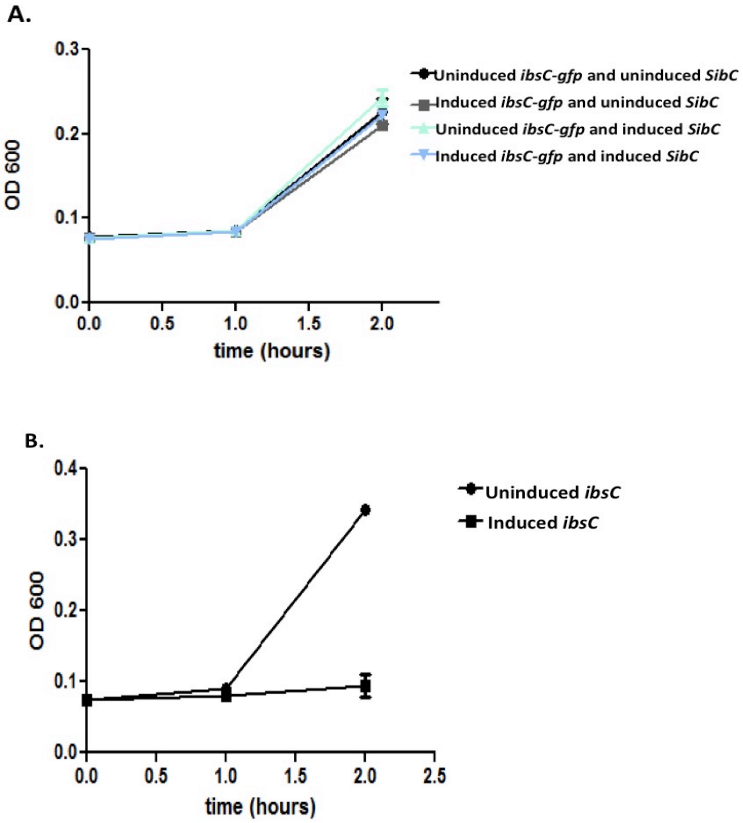


Figure 2.6. **Growth assays to show loss of toxicity by the IbsC-GFP construct.** A. Growth of pNYLSibC-DH5αZ1*ibsC-gfp* under induced and uninduced conditions indicates loss of toxicity by the IbsC-GFP construct. B. Growth of the DH5αZ1*ibsC* strain, carrying the wild type *ibsC* gene, under induced and uninduced conditions to illustrate growth stasis under toxic conditions.

2.3.2.3 Confirming maintained interaction between the IbsC and SibC

To test for maintained interaction between the IbsC and SibC RNAs, we performed TECAN fluorescence scanning, wide-field fluorescence microscopy and Real-time PCR (RT-PCR). The rational used to interpret the results from all of these experiments was that maintained interaction between the two molecules would result in repressed levels of IbsC-GFP (protein and potentially mRNA) and hence lowered levels of fluorescence detected by TECAN scanning and microscopy, and perhaps lowered levels of mRNA detected by RT-PCR in cells where both molecules are induced relative to the ones where only *ibsC-gfp* is induced.

As predicted, the results from these experiments suggested lowered levels of fluorescence and IbsC-GFP mRNA, hence maintained interactions (Figure 2.7A&B). The results were reproducible. However, the small differences among the experimental samples or between them and the controls indicated low signal to noise (S/N) ratios.

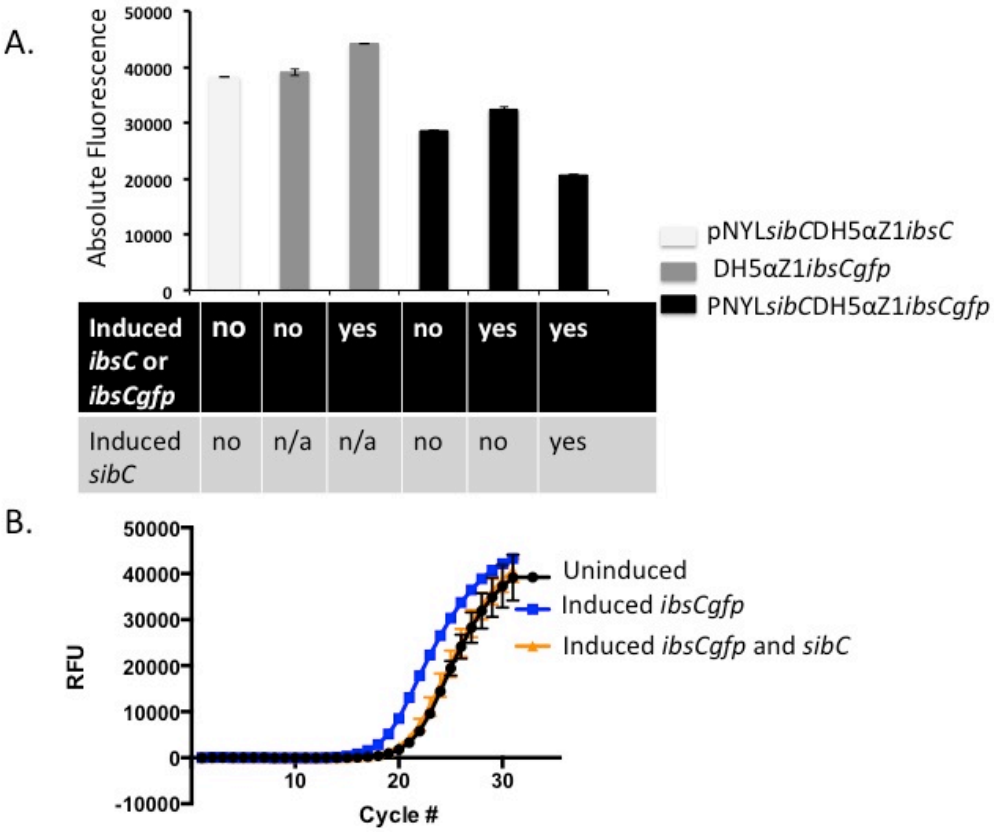


Figure 2.7. Fluorescence knockdown assays and RT-qPCR experiments to show maintained interaction between SibC and IbsC-GFP. A. Results from TECAN scanning to compare the levels of cell fluorescence, emitted by induced levels of translated IbsC-GFP, in the presence and absence of induced SibC antitoxin. These cells are engineered to all carry the *ibsC/gfp* gene, under regulation of a tetracycline (Atc) inducible promoter in their chromosome. An exception is the pNYL*sibC*DH5αZ1*ibsC* which carries the wild type *ibsC* gene as control. *sibC* is supplied to some of these cells, as indicated in the strain name, under regulation of an Isopropyl β-D-1 thiogalactopyranoside (IPTG) inducible promoter on the pNYLMCS11 plasmid. Fluorescence is used as a measure of effectiveness of the SibC antitoxin in interacting with the IbsC-GFP mRNA and repressing its translation. Comparing the first three bars, as expected, higher levels of fluorescence are observed in cells in which *ibsC-gfp* is induced, relative to the ones in which the gene is un-induced or not present. Also, comparing the last two bars, lower levels of fluorescence are observed by IbsC-GFP in the presence of induced levels of SibC, implying maintained interaction between the two molecules and translational repression by SibC. However, we observe high levels of auto fluorescence. Therefore S/N ratios needed to be corrected by improving the design as discussed in subsequent chapters. B. Amplification Curves generated by performing Reverse Transcription-qPCR (RT-qPCR) on IbsC-GFP mRNAs in the pNYL*sibC*DH5αZ1*ibsCgfp* strain under *ibsCgfp* and *sibC* induced and uninduced conditions . The blue curve represents the IbsC-GFP mRNA levels in cells to which Atc has been provided in their growth media for *ibsC-gfp* induction only. The orange curve represents the IbsC-GFP mRNA levels in cells to which Atc and IPTG have been provided in their growth media for induction of both the *ibsC-gfp* and *sibC*. Finally, the black curve represents the IbsC-GFP mRNA levels in cells in which neither the *ibsC-gfp* nor *sibC* has been induced (control). As indicated by the left shift by the blue curve relative to the orange and black curves, higher levels of the

IbsC-GFP complementary DNA (cDNA) (and therefore mRNA) were originally present in the cells with induced levels of only the *ibsCgfp* relative to i) the cells where the antitoxin is also induced and ii) control cells where neither genes are induced. These results suggest that induction of *sibC* in cells has an impact in lowering of the IbsC-GFP mRNA levels through certain regulatory mechanisms. Hence, the IbsC-GFP molecule must have maintained its ability to interact with SibC.

2.3.2.4 An alternative expression platform

We speculated that expression of the construct from the chromosome was too weak to allow for fluorescence at desirable levels well above the auto fluorescence. To resolve the low S/N ratio, we decided to modify the design of our platform to express the *ibsC-gfp* construct in the cells from a plasmid. As such, *E. coli* DH5αZ1 cells were co-transformed with a plasmid, coined pBSdeltaBAD into which the *ibsC-gfp* construct was integrated downstream of the P_{LtetO1} promoter and with the pNYL-cat plasmid harbouring the *sibC* gene under regulation of the P_{LlacO1} promoter (**Figure 2.8** below).

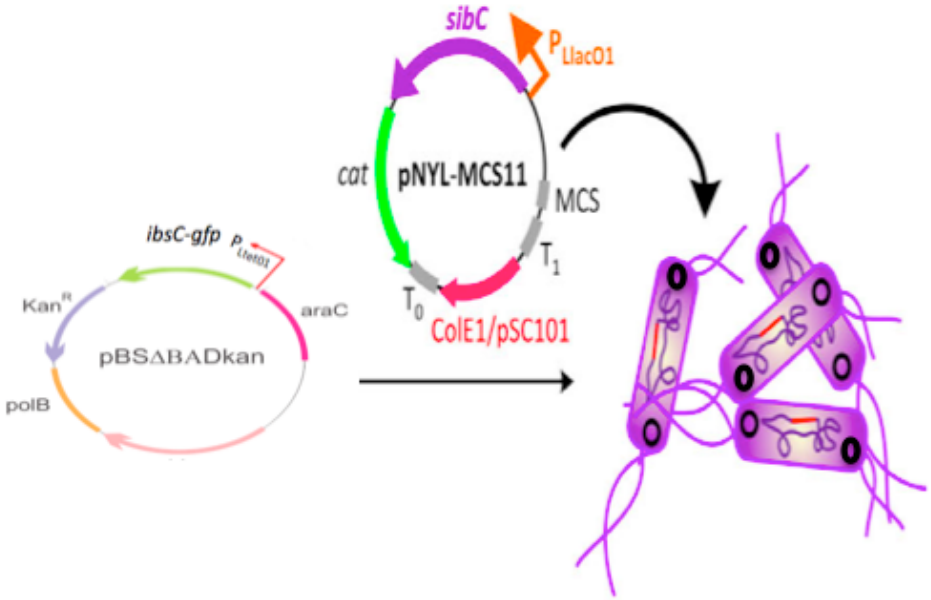


Figure 2.8. The improved design of the IbsC/GFP-SibC interaction assay platform. The *ibsC-gfp* was cloned and expressed from a plasmid to achieve higher levels of its production inside the cells, and therefore higher signal to noise

(S/N) ratios. (i.e. a higher difference between fluorescence and negative controls). Everything else remained the same as before. Figure was provided by Wendy Mok.

As shown in the **Figure 2.9** below, cells with maximally-induced levels of *ibsc-gfp* produced fluorescence levels well above background and more than 10-fold higher than that by cells which expressed the construct from their chromosome (Figure 2.7). In addition, a dose-dependent increase in fluorescence was observed with increasing levels of the inducing agent, Atc. Finally, fluorescence from un-induced cells indicated some leaky expression from the plasmid.

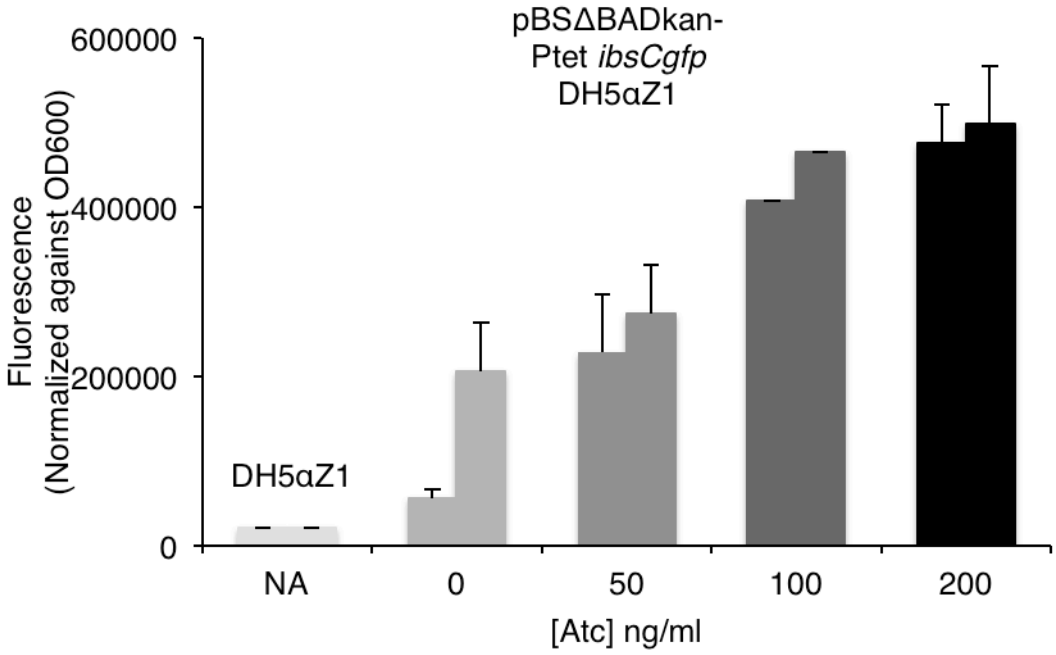


Figure 2.9. Results from TECAN scanning to measure the levels of cell fluorescence from pBSΔBADkan-Ptet *ibsCgfp* DH5αZ1 cells under different induction levels. *E. coli* DH5αZ1 is used as control. The cells here are engineered to express the *ibsCgfp* gene from a plasmid, coined pBSdeltaBAD, under regulation of a tetracycline (Atc) inducible promoter. These are referred to as pBSΔBADkan-Ptet *ibsCgfp* DH5αZ1 cells. Atc is supplied to these cells at different concentrations to examine the effect of different levels of expression on fluorescence. Cells with maximally-induced levels of IbsC-GFP (with 200 ng/ml Atc), produced fluorescence levels well above background and more than 10 fold higher than that by cells which expressed the construct from their chromosome in

previous experiments (Figure 2.7 above). In addition, a dose-dependent increase in fluorescence was observed with increasing levels of Atc. Finally, fluorescence from un-induced cells indicated leaky expression from the plasmid.Data are averages of three technical replicates and repeated for two biological replicates as represented by the double bars.

Growth curves generated for these cells also revealed interesting information about their behavior. First, growth had severely slowed down in cells expressing the *ibsC-gfp* construct from the pBSdeltaBADKan plasmid, indicating that toxicity by the construct was not lost but manifested itself in a dose dependent fashion. Previously, low expression levels from the chromosome masked the toxicity of the construct. However, the cells expressing the same *ibsCgfp* construct from the pBSdeltaBADKan plasmid (coined pBSΔBADkan-Ptet *ibsCgfp* DH5αZ1 strain) were shown to have an extremely long lag phase but still capable of growing to OD_{600} of 1 after about 12.5 hours of growth (shown in **Figure 2.10**A, as opposed to wild type *E. coli* DH5αZ1 cells which take about 8 hours). Even this type of slow growth was conditional to relatively low dilution (less than 200-fold as opposed to 400-fold typically done) of overnight cultures at the time of subculture. Later, these cells were transformed with pNYLcat*sibC* to study interactions between the toxin and antitoxin. Growth curves under induced and uninduced conditions are shown in **Figure 2.10**B. Impaired growth was observed in these cells even in the absence of *ibsC-gfp* induction by Atc, which confirmed leaky expression of the construct from plasmid. Finally, SibC induction by IPTG was able to rescue the cells from lethal effects of un-induced levels of IbsC-GFP but not its induced levels. This indicated to us that rescue capabilities of the SibC antitoxin might be concentration dependent.

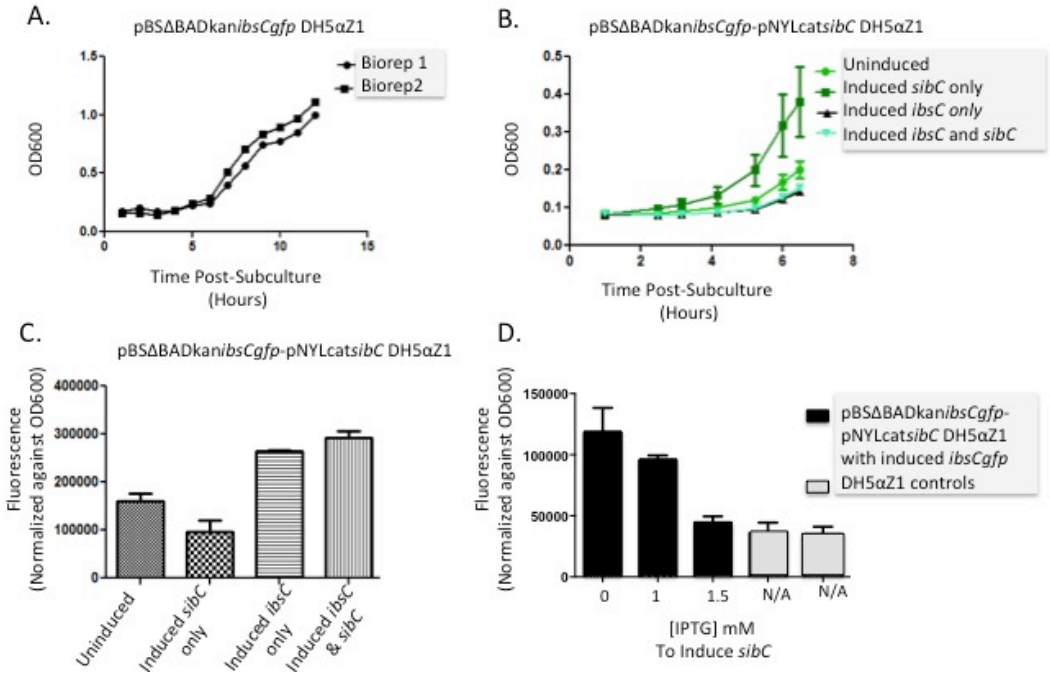


Figure 2.10. Characterization of the fluorescence-based platform to study IbsC-SibC interactions. A. Growth curves generated for *E. coli* DH5αZ1 cells transformed with pBSdeltaBAD-*ibsC-gfp*. Growth is severely slowed down in cells expressing the *ibsC-gfp* construct from plasmid, indicating that toxicity by the construct is not lost. The cells have an extremely long lag phase but are capable of growing to OD600 of 1 after about 12.5 hours of growth (as opposed to wild type *E. coli* DH5αZ1 cells which take about 8 hours). B. Growth curves generated for *E. coli* DH5αZ1 cells co-transformed with pBSdeltaBAD-*ibsC-gfp* and pNYL-cat- *sibC*. Atc and IPTG are provided to cells extracellularly to induce the expression of *ibsC-gfp* and *sibC* respectively. Impaired growth is also observed in the absence of *ibsC-gfp* induction by Atc, which confirms leaky expression of the construct from plasmid also seen in **Figure 2.9** above. Also, SibC induction by IPTG is able to rescue the cells from lethal effects of un-induced levels of IbsC-GFP but not its induced levels. C. Fluorescence knockdown experiments of cells whose growth curves are presented in B, to investigate the repressive effects of SibC. The findings are in agreement with those from the rescue assays in B. As illustrated by the first two bars, SibC induction by IPTG was able to lower the fluorescence levels in cells with un-induced levels of IbsC-GFP but not with induced levels, as illustrated by the last two bars. These data, suggest that perhaps the ability of SibC to antagonize IbsC-GFP is dependent on the concentrations of the toxin and anti-toxin. D. More fluorescence knockdown experiments to test the concentration dependence of SibC effects by providing different levels of SibC induction. Cells, except for *E. coli* DH5αZ1 controls were all induced with 200ng/ml AtC to express the *ibsC-gfp*. A stepwise reduction in fluorescence is observed with increasing levels of SibC induction (i.e. [IPTG]mM). Interactions between wildtype SibC and IbsC seemed to be maintained as SibC induction is able to return fluorescence levels to those of the control. Therefore, this platform can successfully be used to study the interactions between IbsC and SibC. Results from these fluorescence knockdown experiments were reproduced with different levels of toxin induction in **Supplementary Figure 2.3**.

To further investigate the repressive effects of SibC, fluorescence knockdown

experiments were performed. **Figure 2.10**C summarizes the findings and they agree with those from the rescue assays in **Figure 2.10**B. As illustrated by the first two bars, SibC induction by IPTG was able to lower the fluorescence levels in cells with un-induced levels of IbsC-GFP but not with induced levels, as illustrated by the last two bars. These data, suggested that perhaps the ability of SibC to antagonize IbsC-GFP is dependent on the concentrations of the toxin and antitoxin induced by Atc and IPTG respectively in these experiments.

To confirm and further investigate the concentration dependence of SibC effects, we repeated the fluorescence knockdown experiments but provided different levels of IbsC and SibC induction this time (**Figure 2.10**D above & **Supplementary Figure 2.3** below).

In Supplementary Figure 2.3, we showed that induction of sibC with 1mM IPTG was only sufficient to repress uninduced levels of fluorescence as shown previously in figure 2.10C. It was not enough to effectively repress induced levels of fluorescence. Higher concentrations of SibC, induced with a higher dose of 1.5 mM IPTG were required to effectively repress induced levels of fluorescence with 100ng/ml (**Supplementary Figure 2.3**) and 200ng/ml Atc as shown in **Figure 2.10**D) to those of the negative control. Therefore, activity of SibC against IbsC-GFP was shown to be concentration dependent.

As described, the developed platform could successfully be used to evaluate the interactions between SibC and IbsC. However, there were some challenges. Since the construct had maintained partial toxicity at certain concentrations, one of the challenges faced when performing these experiments was that the cells grew differently under different inducing conditions. By using different dilution factors at the time of subculture for the different inducing conditions, we were able to have the cultures grow similarly and to a larger OD600 over a shorter duration (data not presented). This was at the cost of perhaps introducing another form of

variation by introducing different concentrations of mature cells to the different cultures. Nonetheless, the results were interesting. We observed surprisingly lower fluorescence (almost 10-fold reduction) in the cells with induced levels of *ibsc-gfp* as compared to the previous results from the same cells taken at lower OD_{600} values. In addition, no sign of discrepancy in fluorescence was observed between cells with and without SibC.

The results were surprising because at higher ODs and with a higher fraction of matured cells, one would expect higher levels of fluorescence. A plausible explanation could be that later into growth, a higher rate of chromosomal SibC production lowers fluorescence and masks the effects of the plasmid-encoded SibC. This is likely, especially since in Fozo's work, IbsC mRNA expressed from a medium copy plasmid was only observed in cells that had the ibsC/sibC encoding region of their chromosome deleted. This showed that chromosomal expression of *sib* genes was sufficient to repress the Ibs mRNA expressed from the plasmid. Similarly, a plasmid encoding *sib* (with *ibs* on the opposite strand and capable of being expressed) could only be transformed into cells that had the chromosomal *sib* intact. [79]. In addition, in chapter 4 of this dissertation we show that SibC is indeed more highly expressed at later stages of growth relative to the earlier ones. Therefore, we decided that studies from here on should ideally be carried out in *ibsC/sibC* knockout strains of *E. coli.* However, the current platform is still useful to study the interactions between the toxin and antitoxin at low OD_{600} values plausibly because chromosomal SibC cannot be produced at sufficient levels to interfere with the effect of the plasmid-encoded SibC.

It should also be mentioned that some biological variations in terms of growth, fluorescence intensity and repression were observed from time to time with the double-plasmid system. It is possible that some colonies developed mutations that caused any or all of the following changes: escape from toxicity, diminished fluorescence and poor interaction between IbsC and SibC. Different colonies

showing different behaviors could be restreaked to confirm behavior and for clonal isolation. They could be sequenced, and retransformed to investigate whether any mutations had developed. We believe that these variations would be minimized in an ideal system where an *ibsCgfp* construct with no toxicity could be created in an *ibsC-sibC* deleted background. Until then, our current platform has proven to be useful for studying the interactions between the toxin and antitoxin, especially at lower OD_{600} levels as presented.

2.4 Discussion

Both the type I IbsC toxin and its cognate SibC sRNA antitoxin have the potential to be turned into antibacterial agents. However, to fully exploit their potential for the purpose of developing molecular tools, further characterizations of these molecules are required. This chapter summarizes our efforts to develop platforms to facilitate these characterizations, as current technical challenges are major factors delaying the study of these molecules.

The potential for the IbsC toxin to cause persister cell formation, hence serve as an antibacterial agent is backed up by our observations of the IbsC protein causing reversible growth stasis and based on evidence from other toxin proteins of its class. To investigate this potential, we engineered molecular constructs and *E. coli* strains with distinct fluorescence behaviors under persistence versus non-persistence states. Fluorescence in these cells is based on the growth rate-dependent activity of a reporter that drives the expression of a short-lived mutant fluorescent protein. At times of rapid growth, such as in non-persister cells, the reporter is expressed and the cells fluoresce; whereas at times of no growth or dormancy, as in persister cells, the reporter is not expressed and the cells turn dim.

The mutant fluorescent reporter, due to its short half-life, helps to report the transient metabolic state of the cells. In this design, fluorescence is the basis for identification and isolation of persister cells. Therefore, accurate, sensitive, high throughput and resolution (single-cell based) techniques such as FACS and Time-lapse microscopy can substitute traditional plating and colony counting.

Persistence has thus far been neglected despite being the main cause of recurring infections. We believe our efforts here will facilitate its study by overcoming technical challenges such as rare persister cell counts, viable but non culturable cells which cannot be picked up by plating and the need for high throughput experimentation to elucidate the complex pathways underlying persistence.

A link between the IbsC toxin and persistence remains to be explored under different stress conditions and at different expression levels. The techniques we develop in later chapters to analyze toxin expression and the strains we develop to allow for controlled production of different levels of the toxin and antitoxin, together with the strains presented in this chapter, will help us establish a link between persistence and toxin expression levels.

In addition to the toxin, the sRNA antitoxin has great potential to be turned into antibacterial agents and potentially useful tools for synthetic biology due to its gene silencing capabilities and its unique properties as RNA that make it advantageous relative to a protein. Through studying the interactions between the toxin and antitoxin, the sequence and structural moieties that are essential for activity of the antitoxin can be deduced. Using these essential moieties and replacing the IbsC-binding regions of the molecule with a complementary sequence against another target, we proposed SibC-based gene silencing molecules could be derived. To exploit this potential, we developed a fluorescence-based platform for further characterization of the molecule. Our objective was to substitute traditional means of studying the interaction between

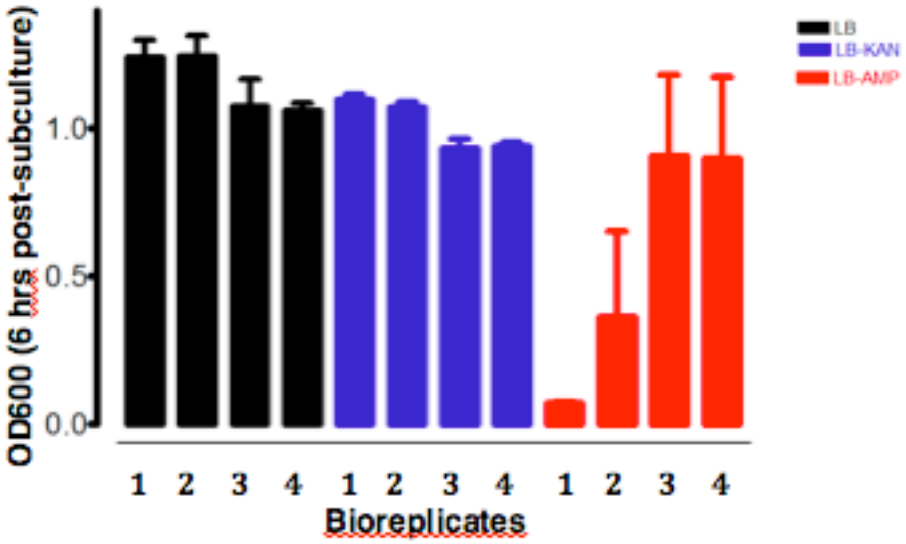
the toxin and antitoxin by growth assays with simple fluorescence. While developing our platform, Han and colleagues developed a SibC-derived molecules to silence lacZ. [91] This provided a proof of concept to our vision. We still continued to optimize our platform, as it would prove useful for further characterization of the SibC and IbsC molecules.

In this platform, we fused a GFP protein into the *ibsC* molecule at a region to rid the molecule of its toxicity, but maintain its interaction with the antitoxin based on findings of previous studies. While optimizing the platform to maximize the fluorescence S/N ratio, we learnt that toxicity of the hybrid molecule is dosage dependent. When expressed from the chromosome under regulation of an Atc-inducible promoter, unlike WT IbsC, the hybrid *ibsc-gfp* is nontoxic. However, when expressed from a plasmid (to increase the fluorescence S/N ratio), it became partially toxic. The cells could still grow to a desired OD_{600} for fluorescence assays to be performed, and eventually to an OD_{600} of 1, but they had a long lag phase. In addition, these fluorescence assays showed that the antitoxin had maintained its interaction with the toxin; therefore these assays could be used successfully to monitor and assess activity of WT SibC and its derivatives against a potential target. However, we found that these assays were OD_{600} dependent. In other words, at earlier stages of growth, the fluorescence differences between successful and unsuccessful interactions of IbsC and SibC were maintained, however at later stages of growth all the differences were abolished. We hypothesized this effect might be due to interference from chromosomal levels of antitoxin which are expressed at higher levels later into growth. In later chapters, we indeed prove that this is the case. Therefore, to remove this condition for the assay to be useful, a further optimization to the platform would be to knock out the chromosomal copies of the toxin and antitoxin. Another interesting observation during the optimization process was that the ability of the antitoxin to repress the toxin was dosage dependent. This was the first time we observed the

dosage-dependency of toxicity by the IbsC molecule and rescue by the antitoxin.

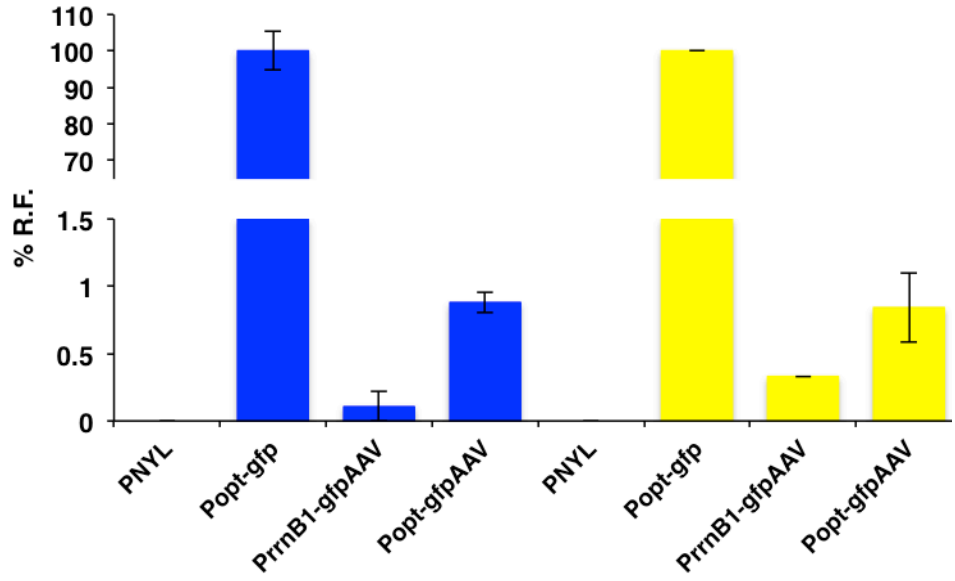
Although our current platform has shown to allow for successful study of interactions under certain conditions, further optimization of the platform as illustrated above would require further knowledge of the biology and regulation of these molecules. In the subsequent chapters we investigate some of these aspects.

2.6. Supplementary Figures

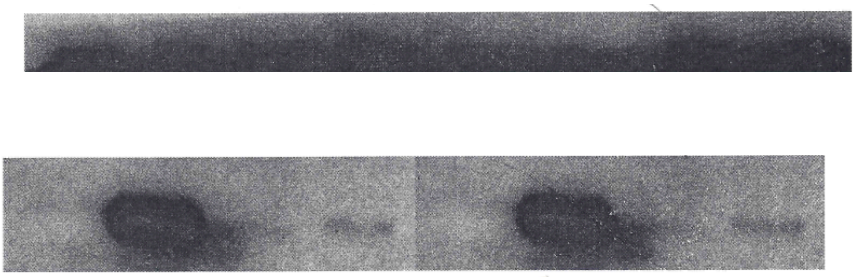


Supplementary Figure 2.1. Screening colonies for Kan resistance and Amp susceptibility to isolate successful knock-in clones. These traits would indicate successful integration of the promoter-reporter-Kan constructs into their chromosome and loss of the Ampicillin resistant plasmid contacting the recombinase. Glycerol stocks of these strains, such as bio replicate 1 here, were prepared for downstream investigations.

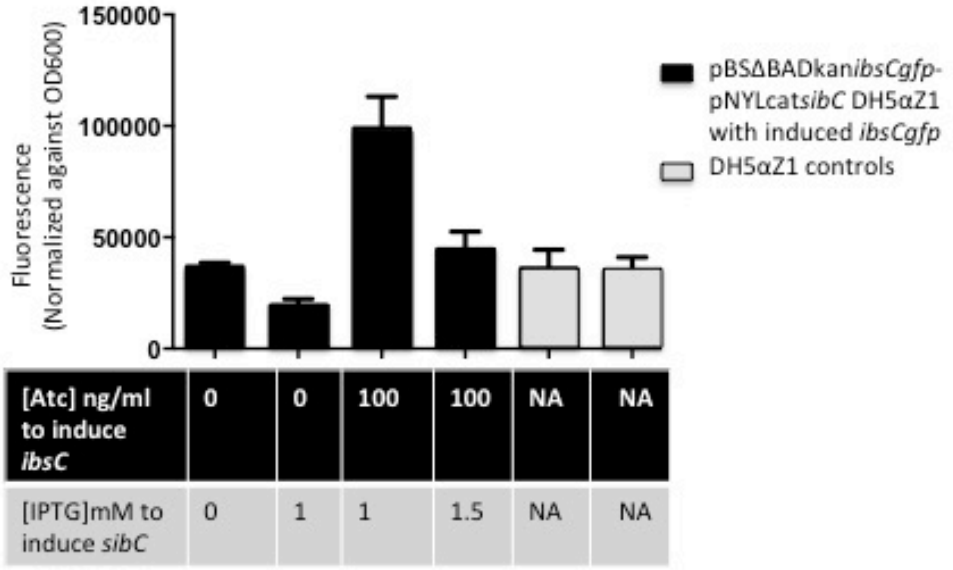
A.



B



Supplementary Figure 2.2 Behavior by different promoter-fluorescent reporter constructs designed to study persistence. A. Fluorescence performance of strains carrying different promoter-fluorescence reporters taken at stationary phase from overnight cultures in blue and at exponential phase from cultures grown 3 hours post subculture in yellow. B. Western blots confirming fluorescence measurements. (Top) Anti-GroEl, (Bottom) Anti-GFP



[Atc] ng/ml to induce *ibsC*	**0**	**0**	**100**	**100**	**NA**	**NA**
[IPTG]mM to induce *sibC*	0	1	1	1.5	NA	NA

Supplementary Figure 2.3. Fluorescence knockdown experiments to show rescue capabilities of SibC antitoxin against IbsC-GFP toxicity are concentration dependent. Also see **Figure 2.10** B, C and D.

Name	Sequence 5'→3'	Purpose
Popt-AN- Fwd	TCATCATCTCGAGTTGACGGCTAGCTCAGTCCTAGGTACAGTGCTA GCGAATTCCTAGTAG	Synthesis of P_{opt} by annealing
Popt-AN-Rev	CTACTAGGAATTCGCTAGCACTGTACCTAGGACTGAGCTAGCCGTC AACTCGAGATGATGA	Synthesis of P_{opt} by annealing
PrrnB(-40+1)-AN-Fwd	CTAGTAGCTCGAGTGGTGGCGCATTATAGGGAGTTATTCCGGCCTGACAAGAGGAGAATTCTCATCAT	Synthesis by annealing
PrrnB(-40+1)-AN-Rev	ATGATGAGAATTCTCCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTCGAGCTACTAG	Synthesis by annealing
GFPAAV-Fwd1	TCATCATGAATTCTAAGGAGGTTTCTTTAATGGCTAGAAAAGGAGAAGAACT	Synthesis of gfpAAV N-terminus by PCR
GFPAAV-Rev1	CGGGTATCTAGCAAAGCACTGAACACCATATCCGAAAGTAGTGAC	Synthesis of gfpAAV N-terminus by PCR
GFPAAV-Fwd2	GTCACTACTTTCGGATATGGTGTTCAGTGCTTTGCTAGATACCCG	Synthesis of gfpAAV C-terminus by PCR
GFPAAV-Rev2	CTAGTAGGGATCCTTAATTAAACTGCTGCAGCGTAGTTTTCGTCGTTTGCTGCAGGCCTTTTGTATAGTTCATCCATGCCA	Synthesis of gfpAAV C-terminus by PCR
IbsC-GFP Fwd	ACTTCAAAATTAGACACAAC	RT-qPCR
IbsC-GFP Rev	AAGGGTAAGGGAGGATTGCT	RT-qPCR

Supplementary Table 2.1. Oligonucleotides and primers used in this study. Underlined are XhoI and EcoR Restriction sites. Highlighted in grey is the Ribosome Binding Site (RBS) and highlighted in yellow are the start and stop codons. Highlighted in green is the "AAV tag". The three mutated codons in *gfpAAV* relative to wild type *gfp* are highlighted in red.

3 Chapter 3: Investigating Multiple Factors That Influence The Regulation And Function Of The IbsC/SibC Toxin-Antitoxin Pair In *E. coli*

3.1 Introduction

To understand the regulatory mechanisms of the antitoxins, it is important to study their interactions with the cognate toxins. While developing a fluorescence-based platform to study the interactions between the IbsC toxin and SibC antitoxin described in chapter 2, some aspects about the toxicity of the IbsC molecule and its regulation by the antitoxin revealed themselves to us. Hence, we embarked on a journey, described in this chapter, to follow these clues to expand our knowledge of the biology and regulation of these systems. We believe this knowledge will feed directly into improving the design of the platforms and protocols developed, by others and us, to study these molecules and advance toxin-antitoxin (TA)-based applications.

Previously, our lab has explored some characteristics of the IbsC toxin at the sequence level. Mok and colleagues deduced the sequence requirements to confer IbsC toxicity. [90] However, through the studies presented in chapter 2, we learned that the toxicity of IbsC is not an all or none phenomenon solely dependent on its sequence and structure but also depends on the dosage of the molecule. Therefore, we set to determine the RNA expression levels that could cause growth stasis. This work is important because based on our own observations, and those of others in the literature [78,92], it seems that the cells might utilize multiple levels of regulation to fine-tune the levels of toxin expression according to cellular requirements. Depending on their abundance level, these molecules exert different effects on the cells, ranging from death, to reversible growth stasis, to membrane damage. It is known that hydrophobic proteins, such as IbsC, target cellular membranes and at high enough levels, can cause membrane damage and growth stasis. However, since the majority of studies have thus far focused on studying the effects of the IbsC molecule at overexpressed levels, our aim here is to

understand how lower levels of IbsC affect *E. coli*.

Towards this same goal as ours, Fozo and colleagues utilized a strain in which the *sibC* antitoxin promoter was deleted (ΔP*sibC* strain). This led to detectable levels of the IbsC toxin mRNA, in the absence of repression by the antitoxin. While no major growth defects were observed relative to the wild type cells, increased levels of phage shock protein (psp) transcripts were measured, signifying membrane damage. In addition, the $DiBAC_4(3)$ dye uptake assays showed higher proton motive force dissipation in the ΔPsibC strain relative to the wild type population (7.5% vs. 1.0% respectively; Fontaine and Fozo, unpublished observations). Thus, even low levels of IbsC were shown to cause membrane damage but no growth impairment. [78,92] Therefore, to truly understand function of these molecules, an accurate knowledge of their dosage and a reliable method to track them inside cells is necessary.

In this project, we engineered strains with the capability to express different and controlled levels of the IbsC toxin and SibC antitoxin. We utilized these strains to study the effects of IbsC and SibC RNA expression levels on some of the functions by these molecules and cellular behavior. We optimized a Reverse Transcription-quantitative PCR (RT-qPCR)-based approach, described in chapter 4 in depth, to accurately measure expression from the *ibsC* and *sibC* genes in these cells. By simultaneously performing growth assays, we could establish links between levels of toxin RNA and growth defects inside cells. We believe these engineered strains can help us discover potentially different functions by these molecules at different levels of expression.

In addition, since the interactions between the IbsC toxin and SibC antitoxin is regulated by RNA-RNA interactions, cells must tightly control the ratio between the antitoxin and toxin RNAs to exhibit a desired function by these systems. In chapter 2, we observed for the first time that the rescue capability of the antitoxin

was concentration dependent. Earlier studies also report similar observations. For example, in a study by Fozo and colleagues,[92] they found that *ibs/sib*-harboring multi-copy plasmids could only be introduced into cells with an intact chromosomal copy of the corresponding *sib* gene or those in which the plasmid-encoded Sib RNAs were expressed at high levels. This indicated that Sib RNAs had to be produced at sufficiently high levels (i.e. either from the chromosome or a plasmid) to be able to repress the Ibs mRNA expressed from the plasmid, hence allow for its presence inside the cells. However, no systematic expression studies on these systems have ever been performed. Therefore, by utilizing the engineered strains mentioned above, and performing simultaneous gene expression analyses and growth assays, we decided to elucidate the optimal ratio of [SibC] to [IbsC] at which SibC is most effective against IbsC. This investigation is further justified by our general knowledge that, in bacteria, transcription and translation happen simultaneously, in the same location. Consequently, for successful gene silencing by natural or artificial sRNAs (afsRNA) and perhaps a desired phenotype to appear, the expressed sRNA must bind the target mRNA before the ribosome. Therefore, the ratio of cellular [sRNAs] to their target [mRNAs] is a key parameter. With the ribosome existing as the most abundant molecule inside the cells, one must maximize the expression of sRNAs by adjusting some design parameters (e.g. promoter strength, plasmid copy number, etc.) to ensure sufficient expression of the sRNA for competition with the ribosome.[93] We believe that the knowledge of natural [sRNA]/[mRNA] ratios, will provide insights to guide decisions about these design parameters.

In addition to developing an artificial expression platform to study the effects of varying levels of RNA expression on function by the toxin and antitoxin, we decided it was also important to study the native promoters that drive the expression of these molecules in their natural biological context. Our lab has previously studied the *ibsC* promoter. [87] Therefore, in this chapter we summarize

our efforts to characterize the *sibC* promoter. The importance of these studies is multiple folds. The information that can be obtained from studying the IbsC and SibC molecules in their natural biological context will tremendously contribute to the existing knowledge of cellular biology and regulation of TA molecules by their promoters. This knowledge can further guide the rational design of effective artificial sRNA (afsRNA) molecules. In addition, by elucidating the positive and negative regulatory regions of this promoter, we could potentially isolate DNA regions that could serve as elements to add to the current available repertoire of promoters for fine-tuning their power to drive expression of their downstream genes, for genetic engineering applications.

3.2 Materials and methods

3.2.1 Media/Reagents/Plasmids/Strains

Information pertaining to the cloning and culture reagents and media, plasmids and strains used in this project can be found in the material and methods section of Chapter 2.

3.2.2 Synthesis and molecular cloning of the ibsC/sibC encoding regions and their regulatory regions

We used the *E. coli* K12 or unmodified DH5αz1 genome as the template to perform culture/colony PCR to amplify the *ibsC/sibC* encoding regions and the entire Intergenic Regions (IGRs) beyond their 5' and 3' sites. This construct was

subsequently used to amplify the *ibsC/sibC* encoding regions and only desired subsets of regions beyond their 5' and 3' sites. Primers and the constructs synthesized are shown in **Supplementary Table 3.1**. PmeI and BamH1 restriction sites were incorporated into these constructs by PCR, for their subsequent cloning into the pBSdeltaBAD plasmid. These plasmids were transformed into electrocompetent DH5αz1 cloning strains. The cells were plated in LB-Kan solid media. Colony Forming Units were counted and reported for the different strains carrying the different constructs and used as a measure of the significance and regulatory role of the regions under study. The plasmids were sequenced and retransformed to confirm the results.

3.2.3 Engineering of E. coli strains capable of producing controllable and different levels of IbsC and SibC RNAs

To achieve regulatable levels of *ibsC* and *sibC* expression in *E. coli* DH5αZ1, they were placed under the control of the P_{LtetO1} and P_{LlacO1} promoters respectively.

P_{LtetO1} is obtained by combining the strong and highly repressible P_L of phage lambda with tet operators. Tet repressor (TetR) and anhydrotetracycline (Atc) can tightly regulate activity of P_{LtetO1}. DH5αZ1 is engineered to express a Tet repressor, which can repress P_{LtetO1} up to 5000-fold. This is the greatest range of regulation reported for *E. coli* promoters. In addition, P_{LtetO1} can be differentially induced with varying levels of Atc. Therefore, it will serve our purpose of achieving different levels of RNA expression. P_{LlacO1} is constructed following the same strategy, but uses the lac operator sequences. It is also a strong promoter that can be repressed by the lac repressor, which is constitutively expressed by *E. coli* DH5αZ1. It can be regulated over a greater than 600 fold range. [94]

We annealed two complementary oligonucleotides to separately generate the P_{LlacO1} and P_{LtetO1} sequences. XhoI and BamH1 restriction enzymes were incorporated into the P_{LtetO1} and P_{LlacO1} sequence respectively. Klenow extension of oligonucleotides with overlapping regions was performed to create the *ibsC/sibC*-encoding region. Subsequently, the promoters were fused to the *ibsC/sibC*-encoding region by crossover PCR. The inserts were purified, digested and ligated into pNYLcat at the same sites. For details pertaining annealing, klenow, crossover PCR, and cloning, refer to the materials and methods section of chapter 2. The ligated plasmid was transformed into *E. coli* DH5αZ1, propagated, purified and sequenced to confirm successful clones. Successful clones were subsequently used for gene expression and growth analyses.

To achieve another level of control over *ibsC* expression, the gene was also integrated into the genome of *E. coli* DH5αZ1, under regulation of P_{LtetO1}, following protocols by Datsenko&Wanner and Campbell &Brown [95] and previously described in detail in chapter 2. These cells were subsequently transformed with pNYLcat*sibC* previously constructed and described in chapter 2.

3.2.4 Gene expression and growth analyses

Reverse Transcription-quantitative PCR (RT-qPCR) was performed to analyze gene expression. Details pertaining to the entire RT-qPCR procedure can be found in Chapter 4 of this dissertation. Cells were grown and growth curves were generated as previously described in chapter 2.

3.3 Results

3.3.1 Effects of RNA abundance on functions by IbsC and SibC

3.3.1.1 Expression platforms to achieve controlled induction of the *E. coli ibs-sibC* gene expression

To achieve controlled induction of different expression levels of *ibsC* and *sibC*, we utilized two *E. coli* strains previously engineered by our lab[87,96] ,with an *ibsC* gene and a *sibC* gene, either placed on a plasmid (Figure 3.1a below) or knocked in the chromosome (Figure3.1C).

The first strain, referred to as DH5αZ1PE, is a DH5αZ1 strain containing a P_{LtetO1}-regulated *ibsC* and a P_{LlacO1}-regulated *sibC* placed on the pNYL-MCSII plasmid in the same locus. As a result, the production of IbsC and SibC RNAs can be induced by anhydrotetracycline (Atc) and Isopropyl β-D-1 thiogalactopyranoside (IPTG), respectively. For the experiments with this strain, Atc was used at 2.5, 5 and 10 ng/ml to induce the *ibsC* expression, while 1 mM IPTG was used to induce the *sibC* expression.

The second strain, named DH5αZ1*ibsC,* contains a $P_{\mathbf{LtetO1}}$-regulated *ibsC* knocked into the chromosome of DH5αZ1 and a $P_{\mathbf{LlacO1}}$-regulated *sibC* on pNYL-MCSII. For the experiments with this strain, we used 1 mM IPTG for *sibC* induction but increased Atc concentrations to 100 and 200 ng/ml for *ibsC* induction, considering that the chromosomal expression is less efficient.

RNA samples were prepared from these cells as well as wild type DH5αZ1 and MG1655. The levels of IbsC and SibC RNAs were measured by RT-qPCR and

normalized against the measured IbsC RNA levels in MG1655. The obtained relative "RNA Abundance" values are plotted in Figure 3.1b and Figure 3.1d, respectively.

Four observations are made with the data: (1) The engineered cells exhibited much higher SibC and IbsC RNA levels relative to the two wild-type cells. (2) Higher Atc concentrations resulted in higher levels of the IbsC RNA. (3) The addition of 1 mM IPTG resulted in increased SibC levels. (4) The level of the IbsC RNA was significantly reduced upon induced expression of SibC by IPTG, consistent with the proposed role of SibC (suppressing the levels of IbsC).

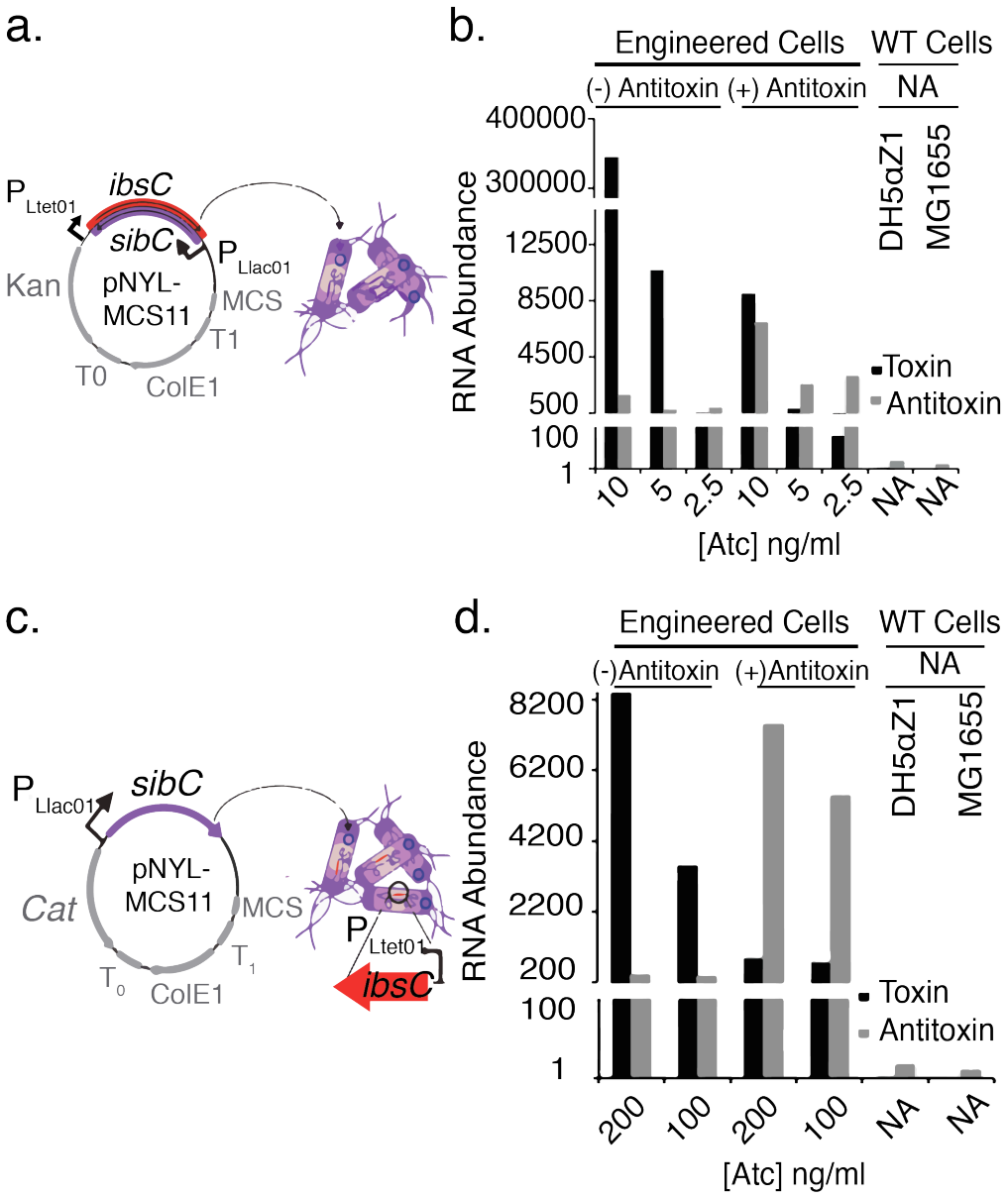


Figure 3.1. Controlled production and reliable measurement of IbsC toxin and SibC antitoxin RNAs in engineered *E. coli* stains. (a) Strain of DH5αZ1PE where *ibsC* and *sibC* are placed in the plasmid pNYL-MCSII. PLtetO1: tetracycline inducible promoter. PLlacO1: IPTG inducible promoter. (c) Strain of DH5αZ1ibsC containing a PLtetO1-regulated knocked-in ibsC in the chromosome of DH5αZ1 and a PLlacO1-controled sibC in pNYL-MCSII. (b) and (d) represent the abundance of IbsC and SibC RNAs in DH5αZ1PE and DH5αZ1ibsC normalized to the level of IbsC RNA in MG1655, in the presence of the indicated concentrations of Atc with and without 1 mM IPTG. Panel a and c schematics were borrowed. [87]

These results suggest that expression of the toxin and antitoxin in these biological

systems can be controlled and reliably tracked.

Interestingly, despite similar levels of antitoxin induction, higher levels of the antitoxin were observed in cells in which the toxin levels were higher (Figure 3.1b, 10 vs. 5 and 2.5 bars and Figure 3.1d, 200 vs. 100 ng/ml [Atc])). This was consistent even in the absence of any antitoxin induction. Since the genomic copies of the toxin and antitoxin are intact in these cells, we hypothesize that this effect is an attempt by the cells to counteract the higher levels of the toxin by production of more antitoxin from the genomic copy, hence restore the balance between toxin and antitoxin levels. By deleting the genomic copies of these molecules in these strains, we can test this hypothesis. Similar levels of antitoxin induction should result in similar levels of the antitoxin being measured by RT-qPCR in these genomic knock out strains, resulting in an even higher level of control over expression of these molecules. However, for the purpose of our current investigations, the level of control we already have over expression of these molecules is sufficient.

3.3.1.2 Linking varying IbsC mRNA levels to growth to determine IbsC toxic RNA levels

With the ability to induce and measure different levels of toxin and antitoxin RNA expression in these two cell lines, we were next interested in determining the RNA dosage at which the toxin caused growth defects. We knew that overexpression of the toxin clearly had dramatic growth consequences for the cell, but to determine the minimum toxic dosage, we induced the RNA expression of the toxin to different extents in these cells, in the presence or absence of antitoxin expression, as previously described. As such, we were able to establish progressive growth impairment conditions as shown in **Figure 3.2**.

Growths from these cells were monitored hourly by measuring their OD600s. The growth differences were found to be the most drastic at about 7.5 hours (460 mins) of growth (**Figure 3.2** a&b). RNAs from cells at this time point were isolated for RT-qPCR analysis. Interestingly, all of these cells resumed normal growth at 20 hours (see last data point in **Figure 3.2**a), indicating that growth impairments due to overexpression of the toxin are reversible.

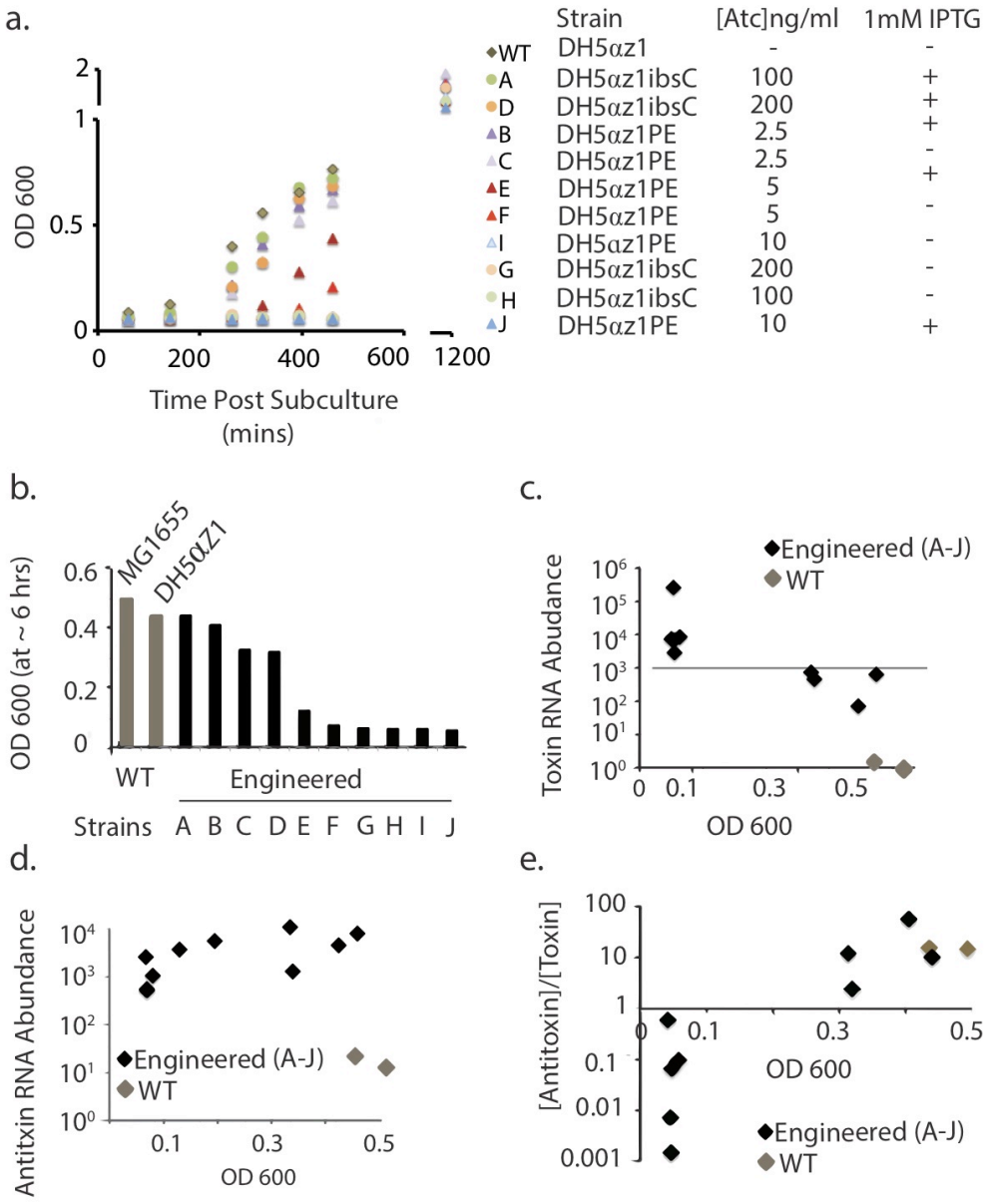


Figure 3.2. Growth impairment vs. measured IbsC toxin and SibC antitoxin RNA levels in the engineered *E. coli* cells under various Atc and IPTG conditions. RNA expression analyses were performed by RT-qPCR. (a) Growth curves for DH5α (diamonds), DH5αZ1PE (filled circles) and DH5αZ1ibsC (filled triangles) under 10 induction conditions are shown as A-J on the right. (b) OD600 at the 6-hour time point. (c) Measured IbsC toxin RNA abundance vs. OD600. (d) Measured SibC RNA abundance vs. OD600. (E) SibC/IbsC ratios vs. OD600.

Plotting IbsC RNA levels in these cells (relative to WT cells) vs. OD_{600} showed that higher levels of IbsC are generally associated with smaller OD_{600} values (**Figure 3.2** c). There appeared to be a threshold value of 1000, above which the cells experienced severe growth defects and below which they grew almost normally. It is interesting that these cells can tolerate 1000-fold higher levels of IbsC RNA than the wild-type cells with little or no impact on their growth.

3.3.1.3 SibC levels and growth

We also examined the SibC RNA levels in these cells in relation to cell growth measured by OD600. The plot of relative SibC levels in these cells (also compared to WT cells) against OD600 showed that higher levels of SibC were generally associated with better growth (**Figure 3.2**d). It is worth mentioning that the lower variation in antitoxin levels between samples relative to that measured for toxin levels is due to the fact that we used only one concentration of the inducing agent IPTG to induce *sibC* expression. We believe that the observed variation is partially due to expression from the chromosomal copy of *sibC*. As previously observed in

Figure 3.1 3.1, the same level of induction lead to higher levels of expression in cells with higher toxin concentration. We believe this is an attempt by the cells to restore physiological levels of the toxin. Presumably, the cells would permit that through an existing, yet unknown, sensing mechanism (e.g. up-regulation of certain transcription factors or by auto regulation).

3.3.1.4 [SibC]/[IbsC] protective ratio

Overexpression of the Ibs proteins has dramatic growth consequences for the cell, but this can be prevented by expression of the corresponding antisense Sib RNA. To determine the sufficient levels of SibC relative to IbsC at which protection by the antitoxin could be conferred, SibC to IbsC ratios in these cells were calculated from the measured IbsC and SibC RNA levels and plotted in **Figure 3.2**e, against OD600 numbers. The plot of SibC/IbsC vs. OD600 revealed that the cells, under the tested conditions (e.g. the OD600 at which they were studied, etc.), grew relatively normally when SibC/IbsC was greater than 1:1, and below that, cell growth was severely inhibited. These results linking IbsC and SibC RNA levels to growth impairments were reproduced in an independent experiment (Supplementary Figure 3.1 and Supplementary Figure 3.2)

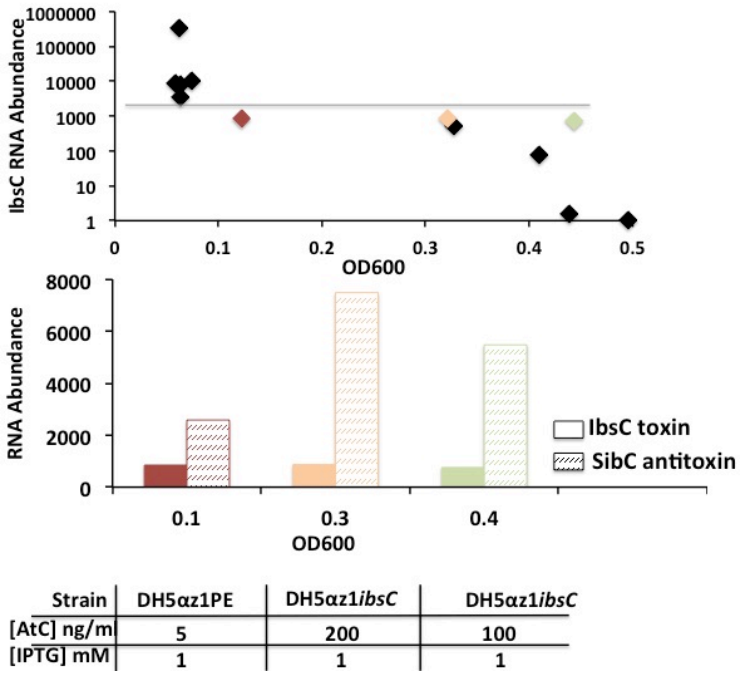
3.3.1.5 Interesting exceptions to the general IbsC and SibC RNA abundance patterns linked to growth impairments

There were a few exceptions to the general trends described above that we believe reveal interesting clues about the regulation of these molecules.

As for toxin levels in relation to growth impairment, higher toxin RNA levels generally led to higher levels of growth impairment as previously described and illustrated in **Figure 3.2**c. However, one exception to this trend was observed and illustrated in **Figure 3.3** below, in burgundy. Despite having similar levels of toxin mRNA to the ones highlighted in green and orange, all below the threshold to cause toxicity, the sample in burgundy experienced major growth impairments. To understand what might have been the cause, we compared their antitoxin

levels. As illustrated in **Figure 3.3**, middle panel, the sample in burgundy showed much lower levels of the antitoxin relative to the others. This indicates that the growth fate of a cell is determined not only by toxin mRNA levels but also the levels of antitoxin. The data also shows that higher antitoxin levels are able to rescue growth without changing toxin RNA levels in these strains. This supports the notion that the antitoxin small RNA functions not only by promoting the degradation of the toxin mRNA (as previously supported by the reverse relationship between the toxin and antitoxin levels. i.e. higher levels of the toxin measured in the absence of the antitoxin in Figure 3.1 b and d) but also by inhibiting translation perhaps by blocking the ribosome binding site (RBS) or other currently unknown mechanisms.

As for the antitoxin levels in relation to growth, higher antitoxin levels generally led to higher levels of growth as previously described and illustrated in **Figure 3.2**d. However, there were examples where cells showed similar levels of the antitoxin but different levels of growth, for example, in strains highlighted by colors in Figure 3.4.Comparing toxin levels in these strains, the cells showing growth defects were measured to contain much higher levels of the toxin (middle panel below). The data summarized in Figure 3.3 and Figure 3.4 and collectively indicate that the interplay between the toxin and antitoxin plays an essential role in determining the growth fate of these cells.



Strain	DH5αz1PE	DH5αz1*ibsC*	DH5αz1*ibsC*
[AtC] ng/m	5	200	100
[IPTG] mM	1	1	1

Figure 3.3. An interesting exception to the general pattern observed for IbsC toxin RNA abundance vs. OD_{600}. RNA abundance levels are derived from RT-qPCR experiments. Top. The color-coded cell cultures all express similar levels of the toxin. However the one in burgundy experiences much more dramatic growth impairment relative to the other two. Middle. Comparing SibC antitoxin levels in these cells, indicate much lower antitoxin levels in these cells relative to the other ones. These results collectively suggest that a) the growth fate of a cell is determined not only by toxin levels but also the levels of antitoxin and b) the antitoxin small RNA functions not only by promoting the degradation of the toxin mRNA but also by inhibiting translation perhaps by blocking the ribosome binding site.

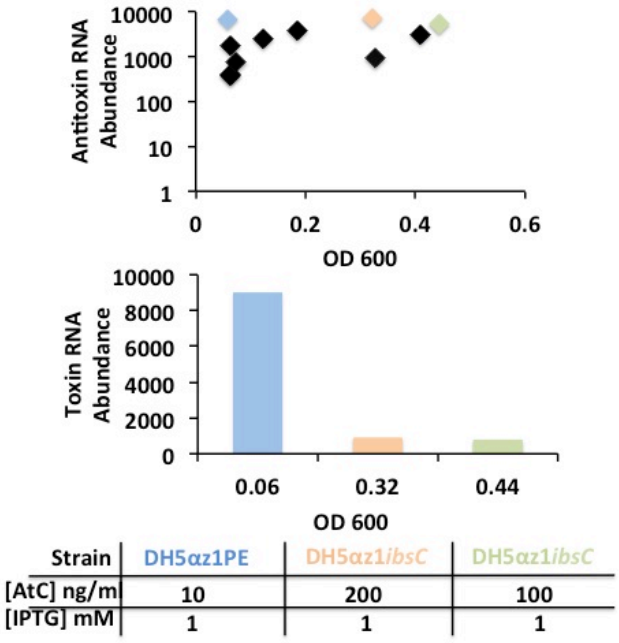


Figure 3.4. An interesting exception to the general pattern observed for antitoxin abundance vs. OD600. RNA abundance levels are derived from RT-qPCR experiments. Top. The color-coded cell cultures all express similar levels of the antitoxin. However, the one in blue experiences much more dramatic growth impairment relative to the other two. Middle. Comparing IbsC toxin levels in these cells indicates much higher toxin levels relative to the other ones. The data here and in **Figure 3.3** collectively indicate that the interplay between the toxin and antitoxin plays an essential role in determining the growth fate of the cells.

Another exception is summarized in **Figure 3.5** below. As previously mentioned, cells grew relatively normally when SibC/IbsC was greater than 1:1; below that, cell growth was severely inhibited. However, we observed that the sample colored burgundy in **Figure 3.5** below, in spite of showing a greater than 1 ratio of antitoxin to toxin, experienced impaired growth. We compared its toxin levels to that of another sample with similar [antitoxin]/[toxin] which grew relatively normally. We believe the higher toxin levels in this sample are the reason for its impaired growth. This further supports the notion above that toxin levels together

with antitoxin to toxin levels are important for determining the growth destiny of these cells. Maybe there is a threshold to toxin levels above which increasing amounts of antitoxin will not rescue growth.

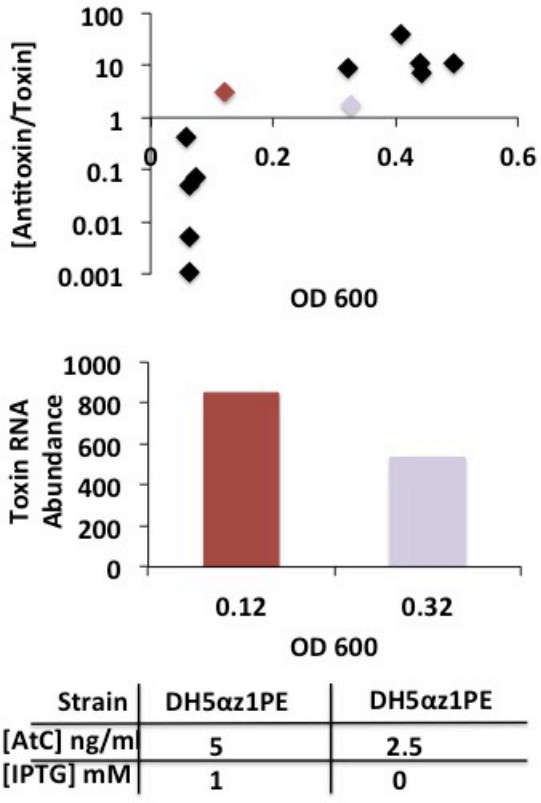


Figure 3.5. An interesting exception to the general pattern observed for [SibC]/[IbsC] vs OD600. Top. The sample in burgundy, in spite of having a greater than 1 ratio of [SibC] to [IbsC] which is generally associated with normal growth, shows impaired growth. Middle. Comparing toxin levels, we observe higher levels of the toxin in this sample relative to the other ones with similar [antitoxin]/[toxin] ratios but normal growth. This supports the notion above that toxin levels together with antitoxin to toxin levels are important for determining the growth destiny of the cells. Maybe there is a threshold to toxin levels above which increasing amounts of antitoxin will not rescue growth.

3.3.2 Characterization of the regulatory regions upstream of the ibsC and sibC genes

In an attempt to better understand the function of the IbsC/SibC pair in its native context and expression as driven by their natural promoters, we sought to identify possible regulatory elements, which governed the expression of the SibC and IbsC molecules, by truncating sequences upstream of their core promoter and assessing survival by culturing and counting colony forming units (CFUs).

To begin our characterization studies, we amplified the *ibsC/sibC* gene and its surrounding intergenic region (IGR) from the chromosome of *E. coli* DH5αZ1 by culture PCR, and used it as a template for PCR to truncate the sequences upstream of the *ibsC* and *sibC* core promoters (-35 to -10 region). These constructs were separately cloned into *E. coli* DH5αZ1. We subsequently monitored the colony forming abilities of the cells carrying the different constructs by culturing and colony-forming unit counting. This would assess activity by each of these constructs inside the cells and evaluate the role and significance of the regions under study in regulating IbsC, hence keeping its levels low enough to not cause growth impairments or death.

The *ibsC/sibC* encoding region and all of its surrounding IGR was named construct 1 (**Figure 3.6**). When cloned into a plasmid and expressed in *E. coli*, the construct was non- lethal, as expected. Cells transformed with this construct grew normally on solid culture plates. Next, we deleted the *sibC* promoter from construct 1 (to create construct 2) and found that the colony forming capability of the strains was almost intact. This indicated that in the absence of any SibC RNA, under normal physiological conditions, the toxin levels are still repressed to low enough levels to not cause any growth impairment. These results are in agreement with previous findings by others. [78,92] This was not unexpected, as our lab had

previously identified a negative regulatory region upstream of the core IbsC promoter (i.e. the -60 region of ibsC promoter) to tightly repress the *ibsC* gene expression.[87] Next, we deleted this negative regulatory region from construct 2 to generate construct 3 in **Figure 3.6**. This time, in the absence of the *sibC* promoter to drive its production and in the absence of the negative regulatory region upstream of the *ibsC* core promoter, the strains had lost their colony forming capability. No cells grew on the solid culture media. To assess whether SibC production in the presence of an intact *sibC* promoter could repress the IbsC to sub-toxic levels, in the absence of the negative regulatory region, we generated construct 4. Carrying this construct, very few cells could survive. This indicated to us that SibC production as driven by its native promoter is unable to repress and counteract the toxin effectively, in the absence of regulation from the negative regulatory region upstream of the *ibsC* core promoter.

Based on these observations, full toxicity by IbsC, as measured by survival rates in this study, is not achieved just by removing the *sibC* promoter, but in combination with the removal of the negative regulatory region upstream of the *ibsC* core promoter. This points to the existence of multiple levels of intracellular regulation for fine-tuning expression from these genes. In addition, our observations here are in agreement with previous observations in the literature about the toxin not causing any growth defects when the antitoxin promoter is deleted. However, the removal of the negative regulatory region seems to be sufficient for greatly reducing the survival rates of the cells. Therefore, the negative regulatory region upstream of *ibsC* plays the dominant role in inhibition of *ibsC* gene expression, compared to SibC.

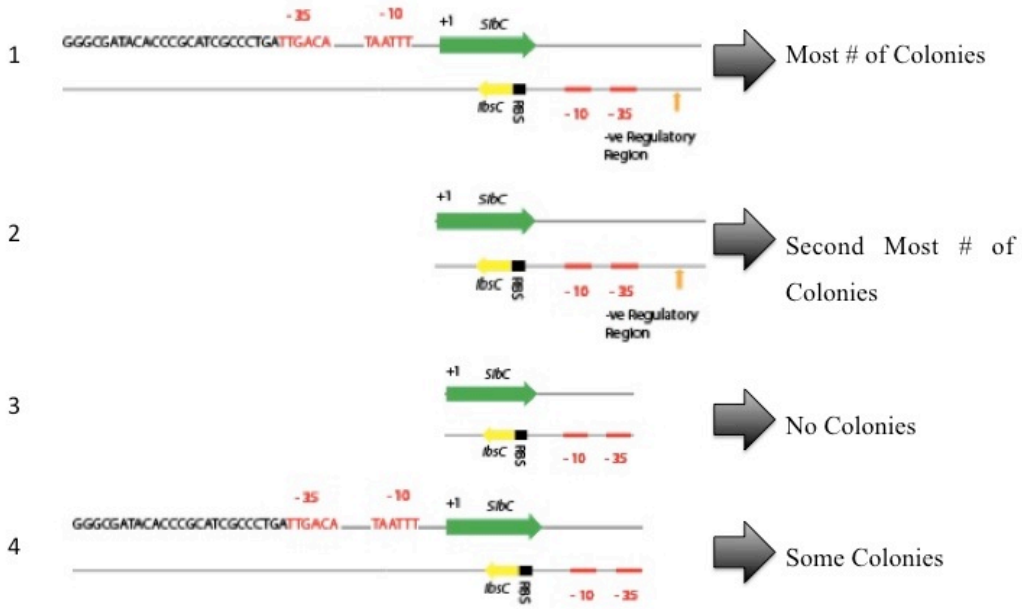


Figure 3.6. Constructs 1-4 synthesized and transformed into DH5αZ1 strains to assess the roles and significance of SibC production, (as driven by the *sibC* promoter) and a negative regulatory region (upstream of the *ibsC* core promoter) in regulating IbsC (so it is maintained at levels to not cause growth impairments or death). Colony forming capability of these cells was assessed. Construct activities inside cells were inferred by monitoring the colony forming abilities of the cells containing them. As shown by measured activity from construct 2, the negative regulatory region upstream of the *ibsC* gene can inhibit expression of the *ibsC* gene to rescue the cells from its detrimental effects to a great extent even in the absence of *sibC* expression, but the opposite is not true. As shown by construct 4, SibC expression driven by its full length promoter, in the absence of the negative regulatory region upstream of *ibsC,* cannot inhibit *ibsC* expression enough to rescue the cells from its detrimental effects to the same extent as construct 2. These results suggest that the negative regulatory region upstream of *ibsC* plays the dominant role in inhibition of *ibsC* gene expression, compared to SibC.

To examine the *sibC* promoter more closely, we created a series of truncation variants, 5 nucleotides at a time, starting from the -60 position of sibC (Figure 3.7) in the presence and absence of the negative regulatory region upstream of the *ibsC* gene. Sequences for these constructs and the oligonucleotides to synthesize them are presented in the **Supplementary Table 3.1**. These constructs were

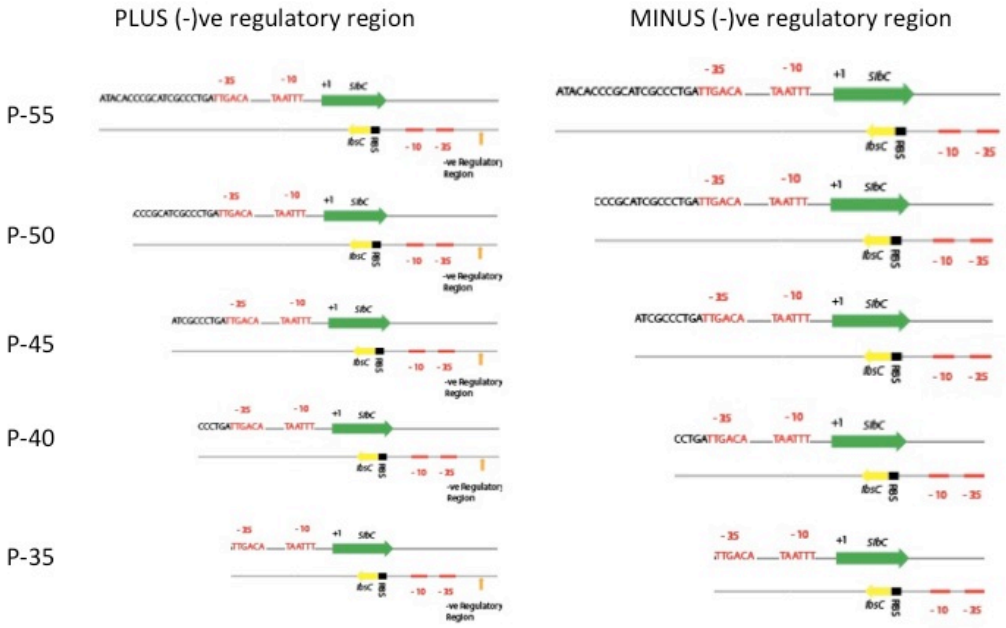
cloned into *E. coli* DH5αZ1 and assessed for their impact on survival.

Transformations were repeated two times and only the transformants carrying the following constructs were capable of colony formation: P-45 and P-35. This may indicate arrangements of alternating positive and negative regulatory regions, starting at the -60 position relative to *sibC*. When a positive regulatory region (e.g. -60 to -50 in P-50) is removed, we see no colony growth. However, when removed in a combination with a negative regulatory region (e.g. -50 to -45) as in P-45, we observe growth. Again, removing another positive regulatory region (-45 to -40 in P-40), we observe no growth but in combination with the removal of another negative regulatory region (-40 to -35) in P-35 we observe growth.

By scrambling and mutating these regions, future studies can further confirm their positive and negative regulatory roles. In addition, the regulatory elements discovered here can be fused to other promoters such as the strong, constitutive Popt, to see how promoter activities are affected. These constructs can add to the toolbox of promoters for manipulating gene expression. Fusing these promoter elements to GFP will enable us to monitor promoter activities by means of fluorescence instead of rescue assays.

It is also important to note that the colony-forming results presented here, indicate differences in survival rates on agar plates. It would also be interesting to examine how the surviving colonies, carrying different constructs, vary in their growth in liquid media as the result of their different levels of IbsC expression and

repression by SibC.



Construct PLUS (-)ve regulatory region	Colony forming ability	Construct MINUS (-)ve regulatory region	Colony forming ability
P-55	No	P-55	No
P-50	No	P-50	No
P-45	Yes	P-45	Yes
P-40	No	P-40	No
P-35	Yes	P-35	Yes

Figure 3.7 Constructs designed for the sibC promoter truncation studies. A series of truncation variants, 5 nucleotides at a time, starting from the -60 position of *sibC* were created and named as shown (Left constructs contain the *ibsC* negative regulatory region and Right constructs do not contain the *ibsC* negative regulatory region). The P_{sibC} truncation variants consist of different sequences upstream of the core promoter containing the -35 and -10 regions. They were cloned into *E. coli* DH5αZ1, and assessed for their impact on survival. The table summarizes the effect of each construct on the colony forming ability of the cells. Two independent xperiments were conducted to confirm the results.

3.4 Discussion

This chapter summarized our efforts in studying the regulation of the IbsC-SibC system both in artificial systems where we controlled their expression levels by utilizing inducible promoters and in their native context under regulation of their natural promoters.

Briefly, we engineered strains of *E. coli* with the capability to produce different and controlled levels of the toxin and antitoxin in these cells. These cells together with a method we have optimized (please see chapter 4 for details) and utilized here to reliably measure *ibsC* and *sibC* gene expression, present an opportunity to link expression levels from these genes to function. We believe these molecules can serve different roles to the cells depending on their abundance level. As such, the cells regulate expression from these genes carefully according to their cellular demands. As a start, we set to establish links between the RNA abundance of these molecules and growth impairment. We determined the RNA dosage at which the toxin caused growth defects and the optimal [SibC/IbsC] RNA ratio for rescue capabilities of the antitoxin. We found that both individual RNA levels of these molecules, as well as their ratios relative to one another are important for the fate of the cells. Studying the interplay between the toxin and antitoxin RNA abundance levels revealed clues about the regulatory features of these molecules that were stated in the results section and are worth investigating in depth in

future studies.

At the DNA level, we show different regulatory regions, both positive and negative, upstream of the core *ibsC* and *sibC* that contribute to fine regulation of expression from these genes. A previous study showed that deleting the *sibC* promoter altogether, from the chromosome, had no obvious impact on cell growth and survival, while consequences for the membrane potential and integrity were revealed.[78,92] Here, we also show the effect of *sibC* promoter deletion, but from a plasmid on survival. These results suggested that even in the absence of any (at least major) regulation by SibC, under normal physiological conditions, the cells are able to regulate toxin levels below levels that affect cell survival. Furthermore, we showed that the deletion of some intergenic regions at the 5' untranslated regions, upstream of the -35 hexamer core promoter of *sibC*, had major consequences for the survival of the cells, indicating they may have important regulatory roles. It is interesting to note that in the absence of the plasmid-encoded *sibC* promoter, the cells showed almost intact colony forming abilities (Figure 3.6, construct 2). However, in the absence of some regulatory regions upstream of the -35 hexamer of the plasmid-encoded *sibC* core promoter, the cells did not show any colony forming ability (Figure 3.7). We believe this might be due to the presence of the genomic copy of SibC. Repression by the genomic SibC sRNA might be contributing to the survival of the cells carrying construct 2 in Figure 3.6. However, in Figure 3.7, the constructs carrying more negative regulatory regions than positive might be “reserving” the RNA polymerase away from the chromosomal copy of *sibC*, hence inhibiting any contribution to survival by chromosomal expression of SibC sRNA. We can test this hypothesis by performing the experiments in *ibsC-sibC* deleted strains. In addition, for better comparison of the behavior by the cells in Figure 3.6 and those in Figure 3.7, we must conduct the experiments similarly (e.g. simultaneously, same batch of electrocompetent cells, similar DNA concentrations and growing media, etc). For

a quantitative analysis of how these regions affect growth, culture experiments in liquid media can be carried out. That noted, the presence of chromosomal copies of *ibsC-sibC* do not alter the conclusions we draw in this chapter.

Our data collectively suggest that cells use multiple and delicate strategies to regulate these molecules at the DNA, RNA and protein levels (both with respect to their existence and expression levels). We have previously made observations about cells carrying extra copies of the *ibsC* gene in addition to their chromosomal copy, which stopped showing growth defects potentially due to either having lost or mutated the toxin gene. However, since we never actively investigated these hypotheses, we take caution in making any claims. It would be interesting to perform evolution/time course experiments to study these hypotheses in detail. Based on our studies in this chapter, however, we suggest that upon overproduction of the toxin, the chromosomal copies of the antitoxin are expressed at higher levels. We consider this to be an attempt by cells to restore physiological balance. To test this hypothesis, we suggest performing the relevant experiments in *ibsC*/*sibC*-deleted strains. We believe that even the existence of multiple copies of these genes on the chromosome might be a strategy to fine-tune expression of these genes for different functions. We investigate this possibility in chapter 5 of this dissertation.

With the knowledge, and tools and techniques developed, we are now equipped to further optimize the design of the platforms we developed previously. In chapter 2, we created *ibsC-gfp* fusion constructs to assess the interaction between the toxin and wild type antitoxin and its derivatives by fluorescence. These platforms were developed to facilitate the study of the sequence and structural requirements for successful interaction and silencing by the antitoxin. Although functional, the Signal to Noise ratio (S/N) of the platform seemed to depend on the expression levels of the construct. However, the benefit of maximizing fluorescence by increasing expression levels seemed to be counteracted by toxicity from the

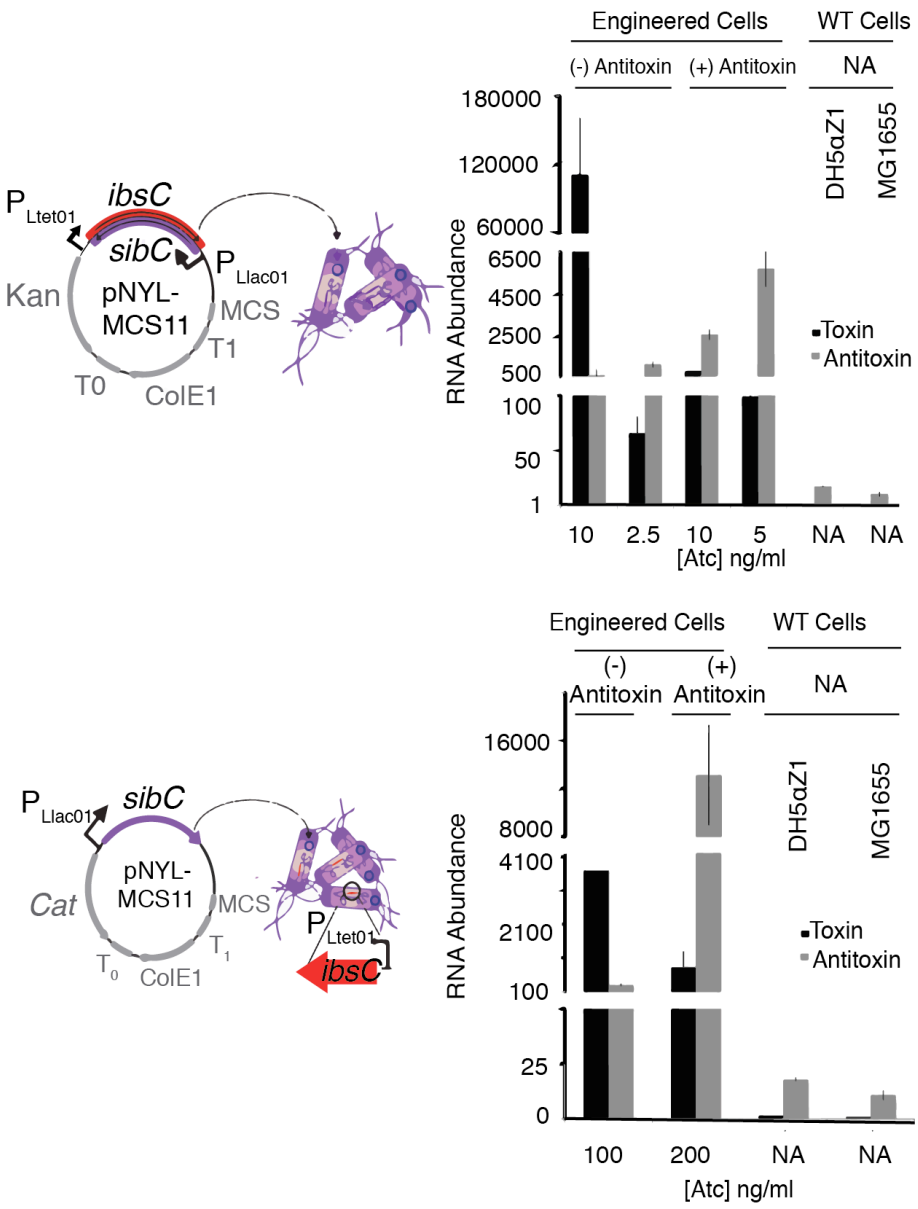
construct. The construct that seemed to have lost toxicity at the chromosomal levels of expression manifested its toxicity at higher levels of expression from plasmid. In addition, the rescue capabilities of the antitoxin seemed to be dose-dependent, but the effective antagonizing ratio was not known at the time. We believe the knowledge gained through the studies presented in this chapter can guide our decisions about the design parameters that affect expression levels (e.g. expression from plasmid or chromosome, different concentration of the inducing agents) of the fluorescence constructs in the aforementioned systems, so to maximize the desirable effects (e.g. fluorescence signals and rescue function by the antitoxin) and minimize undesirable side effects such as toxicity. With the tools and techniques developed here we can then achieve the desirable expression levels. After optimizing the design elements that involve expression, any remaining limitations can be resolved by other considerations such as the choice of fluorescence reporter for maximizing S/N, substituting normalization against OD_{600} in cell cultures with low and varying OD_{600} values to normalization against a second fluorescence protein [97] etc.

We believe that our determination of RNA levels here from wild type and engineered cells in the first part of the project, together with the characterization of the native regulatory regions driving expression of the toxin and antitoxin in the second part of the project, have revealed some key principles and elementary features of the IbsC toxin-SibC antitoxin RNA expression and regulation. In addition to the biological insights, having gained control over the production of these molecules in biological systems, and having developed a reliable method to track the changes in their expression are important. They will enable us to discover potentially different functions by these molecules at different concentrations. By establishing links between their expression levels and cellular behavior under different conditions (environmental, niche, stress, etc.), we can gain insights about the biological functions of these molecules. Moreover, the

future of many proposed toxin-antitoxin-based applications relies on the ability to create a certain balance in the amounts of the toxin relative to the antitoxin for a certain purpose and to monitor these amounts.

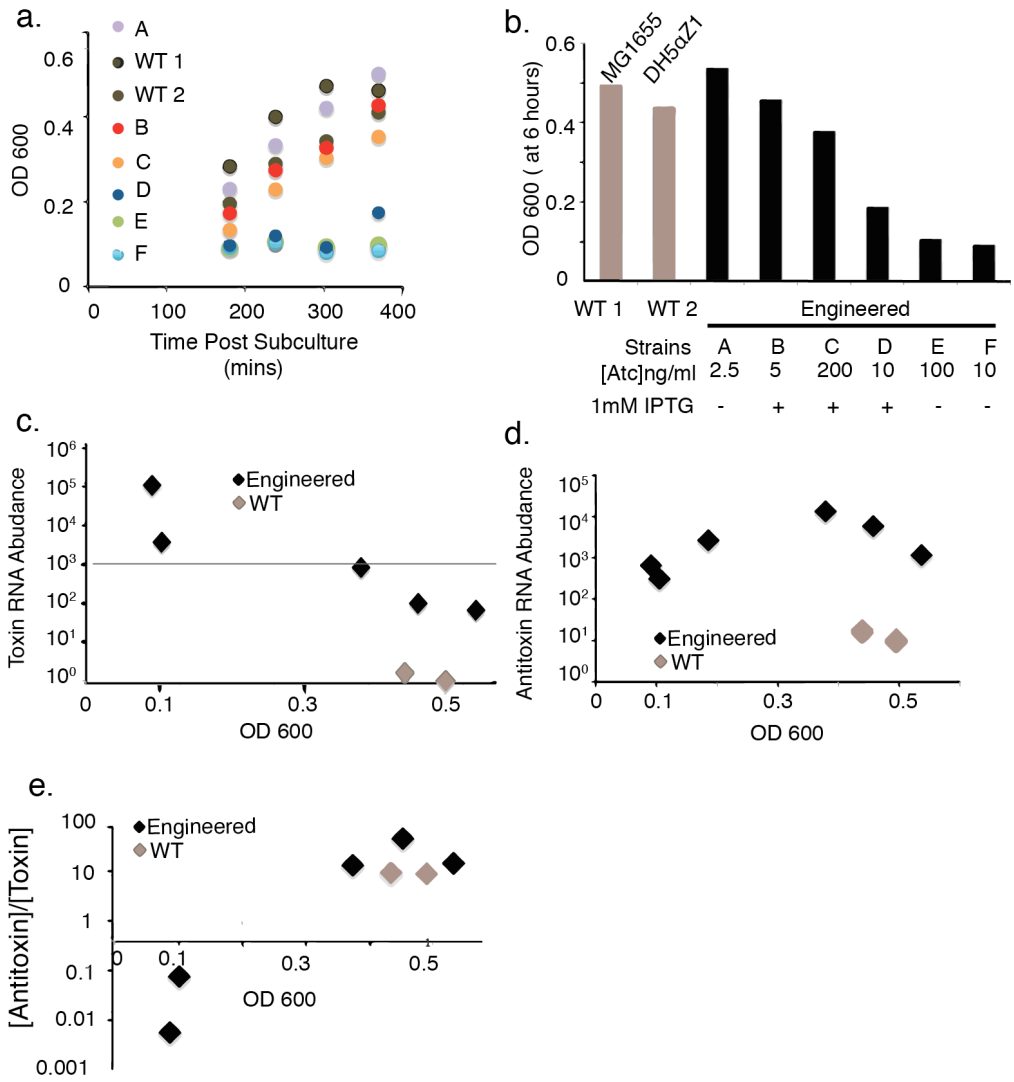
Fine regulation of the expression from these well-conserved genes by the cells, as shown in this chapter, implicates functions by them at endogenous levels. An impeding factor to fully elucidate the functions of endogenous levels of these TA molecules is lack of knowledge of their physiological absolute levels and how they change as the function of developmental stage or environmental conditions. Gaining this knowledge is the main next goal of our investigations presented in chapter 4 of this dissertation.

3.5 Supplementary figures and tables



Suplementary Figure 3.1. Controlled production and reliable measurement of IbsC and SibC RNAs in engineered E. coli stains. (a) Strain of DH5αZ1PE where ibsC and sibC are placed in the plasmid pNYL-MCSII. PLtetO1: tetracycline inducible promoter. PLlacO1: IPTG inducible promoter. (c) Strain of DH5αZ1ibsC containing a PLtetO1-regulated knocked-in ibsC in the chromosome of DH5αZ1 and a PLlacO1-controled sibC in pNYL-MCSII. (b) and (d) represent the abundance of IbsC and SibC RNAs in DH5αZ1PE and DH5αZ1ibsC

normalized to the level of IbsC RNA in MG1655, in the presence of the indicated concentrations of Atc with and without 1 mM IPTG. Schematics were borrowed from Wendy Mok. [87]



Suplementary Figure 3.2. Growth impairment vs. measured IbsC toxin and SibC antitoxin RNA levels in the engineered *E. coli* cells under various Atc and IPTG conditions. (a) Growth curves for DH5α (diamonds), DH5αZ1PE (filled circles) and DH5αZ1ibsC (filled triangles) under 10 induction conditions are shown as A-J on the right. (b) OD600 at the 6-hour time point. (c) Measured IbsC toxin RNA abundance vs. OD600. (d) Measured SibC RNA abundance vs. OD600. (E) SibC/IbsC ratios vs. OD600.

Oligonucleotide Name	Sequence 5'→3'	Purpose	Constructs to be Synthesized (5'→3')
Pfwd1	TCATCATGTTTAAACGGGCGATACACCCGC	To synthesize the 5' Inter Genic Region (IGR)+ *sibC* core promoter+ *sibC* encoding region+ 3' prime IGR (388 ntds)	GGGCGATACACCCGCATCGCCCTGATTGACATCGTTGATTCTTTCGACCTAATTTAGTGAGTAAGGGTAAGGGAGGATTGCTCCTCCCCTGAGACTGACTGTTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGATGACAAGTCGCATCATAACCCTTCTCCTTCAAGCCCTCGCTTCGGTGAGGGCTTTACCGTTACAGCCCCATGCTGCCCTGCCATCGTAAATCCCCATTAAATAAACACAACGCATTGATCTGACTTTGATTTATTTTCTGGAGCAGACTCGCAAAGTAGAATGCGCAACGCGGCAACGGTGTGGAGAAGGGATAAAAAAACGGGCAAGTCAGTGACCTGCCCGTTGATTTTCAGAGAAGGGGAA
Prev1	TCATCATG	Same as above	Same as above

	GATCCTTCC CCTTCTCTG AAAAT		
Pfwd 2	TCATCATGT TTAAACAA GGGTAAGG GAGGA	To synthesize the *sibC* encoding region+ 3' prime IGR: This construct will help evaluate the importance of the *sibC* promoter (consequently *sib* expression) in regulating IbsC levels as measured by cell survival	AAGGGTAAGGGAGG ATTGCTCCTCCCCTG AGACTGACTGTTAAT AAGCGCTGAAACTT ATGAGTAACAGTAC AATCAGTATGATGA CAAGTCGCATCATA ACCCTTCTCCTTCAA GCCCTCGCTTCGGTG AGGGCTTTACCGTTA CAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAACG CATTGATCTGACTTT GATTTATTTTCTGGA GCAGACTCGCAAAG TAGAATGCGCAACG CGGCAACGGTGTGG AGAAGGGATAAAAA AAC GGGCAAGTCAGTGA CCTGCCCGTTGATTT TCAGAGAAGGGGAA
Prev 2-1	Same as Prev1	Same as above	Same as above

Pfwd 3-2	TCATCATGTTTAAACAAGGGTAAGGGAGGA	To synthesize the *sibC* encoding region+ the *ibsC* core promoter only (i.e. The negative regulatory region previously discovered at the 60 ntd upstream region of the *ibsC* promoter is not included): This construct will help evaluate the combinatory importance of the *sibC* promoter (consequently the SibC RNA) and the *ibsC* (-)ve regulatory region in regulating IbsC levels, as measured by cell survival	**AAG**GGTAAGGGAGG ATTGCTCCTCCCCTG AGACTGACTGTTAAT AAGCGCTGAAACTT ATGAGTAACAGTAC AATCAGTATGATGA CAAGTCGCATCATA ACCCTTCTCCTTCAA GCCCTCGCTTCGGTG AGGGCTTTACC GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAA
Prev 3	TCATCATGGATCCTTGTGTTTATTTAATGG	Same as above	Same as above
Pfwd 4-1	TCATCATGTTTAAACGGGCGATACACCCGC	To synthesize the *sibC* promoter+ sibC encoding region + *ibsC* core promoter (i.e. without the negative regulatory region) This construct will help	GGGCGATACACCCG CATCGCCCTGATTGA CATCGTTGATTCTTT GACCTAATTTAGTGA GT**AAG**GGTAAGGGA GGATTGCTCCTCCCC TGAGACTGACTGTTA ATAAGCGCTGAAAC TTATGAGTAACAGTA

		evaluate the importance of the *ibsC* (-)ve regulatory region in regulating *ibsC* gene expression, as measured by cell survival	CAATCAGTATGATG ACAAGTCGCATCAT AACCCTTCTCCTTCA AGCCCTCGCTTCGGT GAGGGCTTTACC GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAA
Prev 4-3	TCATCATG GATCCTTGT GTTTATTTA ATGG	Same as above	Same as above
Pfwd -55	TCATCATGT TTAAACAT ACACCCGC ATCGC	For P*sibC* truncation studies, to remove the first 5 ntds of the 5' IGR upstream of the core promoter. This construct is denoted by P-55	ATACACCCGCATCGC CCTGATTGACATCGT TGATTCTTT GACCTAATTTAGTGA GTAAGGGTAAGGGA GGATTGCTCCTCCCC TGAGACTGACTGTTA ATAAGCGCTGAAAC TTATGAGTAACAGTA CAATCAGTATGATG ACAAGTCGCATCAT AACCCTTCTCCTTCA AGCCCTCGCTTCGGT GAGGGCTTTACC GTTACAGCCC CATGCTGCCCTGCCA

			TCGTAAATCCCCATT AAATAAACACAACG CATTGATCTGACTTT GATTTATTTTCTGGA GCAGACTCGCAAAG TAGAATGCGCAACG CGGCAACGGTGTGG AGAAGGGATAAAAA AAC GGGCAAGTCAGTGA CCTGCCCGTTGATTT TCAGAGAAGGGGAA
Prev -55	TCATCATG GATCCTTCC CCTTCTCTG AAAAT	To synthesize the P-55 construct in an otherwise native *ibsC-sibC* encoding context. The construct will help us evaluate the effect of the first 5 ntds in regulating *sibC* expression as measured by cell survival.	Same as above

Prev -55 MINUS(-)ve regulatory region)	TCATCATGGATCCTTGTGTTTATTTAATGG	To synthesize the P-55 construct in an *ibsC-sibC* encoding context, which also lacks the ibsC negative regulatory region). This construct is coined P-55MINUS(-)ve regulatory region The construct will help us evaluate the effect of the first 5 ntds in regulating *sibC* expression as measured by cell survival.	ATACACCCGCATCGCCCTGATTGACATCGTTGATTCTTT GACCTAATTTAGTGAGTAAGGGTAAGGGAGGATTGCTCCTCCCCTGAGACTGACTGTTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGATGACAAGTCGCATCATAACCCTTCTCCTTCAAGCCCTCGCTTCGGTGAGGGCTTTACCGTTACAGCCCCATGCTGCCCTGCCATCGTAAATCCCCATTAAATAAACACAA
Pfwd -50	TCATCATGTTTAAACCCCGCATCGCCCTGA	For P*sibC* truncation studies, to remove the first 10 ntds of the 5' IGR upstream of the core promoter. This construct is denoted by P-50	CCCGCATCGCCCTGATTGACATCGTTGATTCTTT GACCTAATTTAGTGAGTAAGGGTAAGGGAGGATTGCTCCTCCCCTGAGACTGACTGTTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGATGACAAGTCGCATCATAACCCTTCTCCTTCAAGCCCTCGCTTCGGTGAGGGCTTTACC

			GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAACG CATTGATCTGACTTT GATTTATTTTCTGGA GCAGACTCGCAAAG TAGAATGCGCAACG CGGCAACGGTGTGG AGAAGGGATAAAAA AAC GGGCAAGTCAGTGA CCTGCCCGTTGATTT TCAGAGAAGGGGAA
Prev -50	TCATCATG GATCCTTCC CCTTCTCTG AAAAT	To synthesize the P-50 construct in an otherwise native *ibsC-sibC* encoding context. The construct will help us evaluate the combinatory effect of the first 10 ntds in regulating *sibC* expression as measured by cell survival.	Same as above

Prev -50 MINUS(-)ve regulatory region)	TCATCATGGATCCTTGTGTTTATTTAATGG	To synthesize the P-50 construct in an *ibsC-sibC* encoding context, which also lacks the ibsC negative regulatory region). This construct is coined P-50MINUS(-)ve regulatory region The construct will help us evaluate the combinatory effect of the first 10 ntds in regulating *sibC* expression as measured by cell survival.	CCCGCATCGCCCTGATTGACATCGTTGATTCTTT GACCTAATTTAGTGAGTAAGGGTAAGGGAGGATTGCTCCTCCCCTGAGACTGACTGTTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGATGACAAGTCGCATCATAACCCTTCTCCTTCAAGCCCTCGCTTCGGTGAGGGCTTTACCGTTACAGCCCCATGCTGCCCTGCCATCGTAAATCCCCATTAAATAAACACAA

Pfwd -45	TCATCATGTTTAAACATCGCCCTGATTGAC	For P*sibC* truncation studies, to remove the first 15 ntds of the 5' IGR upstream of the core promoter. This construct is denoted by P-45	ATCGCCCTGATTGAC ATCGTTGATTCTTTG ACCTAATTTAGTGAG TAAGGGTAAGGGAG GATTGCTCCTCCCCT GAGACTGACTGTTA ATAAGCGCTGAAAC TTATGAGTAACAGTA CAATCAGTATGATG ACAAGTCGCATCAT AACCCTTCTCCTTCA AGCCCTCGCTTCGGT GAGGGCTTTACC GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAACG CATTGATCTGACTTT GATTTATTTTCTGGA GCAGACTCGCAAAG TAGAATGCGCAACG CGGCAACGGTGTGG AGAAGGGATAAAAA AAC GGGCAAGTCAGTGA CCTGCCCGTTGATTT TCAGAGAAGGGGAA

Prev -45	TCATCATG GATCCTTCC CCTTCTCTG AAAAT	To synthesize the P-45 construct in an otherwise native *ibsC-sibC* encoding context. The construct will help us evaluate the combinatory effect of the first 15 ntds in regulating *sibC* expression as measured by cell survival.	Same as above
Prev -45 MINUS(-)ve regulatory region)	TCATCATG GATCCTTGT GTTTATTTA ATGG	To synthesize the P-45 construct in an *ibsC-sibC* encoding context, which also lacks the ibsC negative regulatory region). This construct is coined P-45MINUS(-)ve regulatory region The construct will help us evaluate the combinatory effect of the first 15 ntds in	ATCGCCCTGATTGAG ATCGTTGATTCTTTG ACCTAATTTAGTGAG TAAGGGTAAGGGAG GATTGCTCCTCCCCT GAGACTGACTGTTA ATAAGCGCTGAAAC TTATGAGTAACAGTA CAATCAGTATGATG ACAAGTCGCATCAT AACCCTTCTCCTTCA AGCCCTCGCTTCGGT GAGGGCTTTACC

		regulating *sibC* expression as measured by cell survival.	GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAA
Pfwd -40	TCATCATGT TTAAACCC TGATTGAC ATCGT	For P*sibC* truncation studies, to remove the first 20 ntds of the 5' IGR upstream of the core promoter. This construct is denoted by P-40	CCTGATTGACATCGT TGATTCTTT GACCTAATTTAGTGA GTAAGGGTAAGGGA GGATTGCTCCTCCCC TGAGACTGACTGTTA ATAAGCGCTGAAAC TTATGAGTAACAGTA CAATCAGTATGATG ACAAGTCGCATCAT AACCCTTCTCCTTCA AGCCCTCGCTTCGGT GAGGGCTTTACC GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAACG CATTGATCTGACTTT GATTTATTTTCTGGA GCAGACTCGCAAAG

			TAGAATGCGCAACG CGGCAACGGTGTGG AGAAGGGATAAAAA AAC GGGCAAGTCAGTGA CCTGCCCGTTGATTT TCAGAGAAGGGGAA
Prev -40	TCATCATG GATCCTTCC CCTTCTCTG AAAAT	To synthesize the P-40 construct in an otherwise native *ibsC-sibC* encoding context. The construct will help us evaluate the combinatory effect of the first 20 ntds in regulating *sibC* expression as measured by cell survival.	Same as above

Prev -40 MINUS(-)ve regulatory region)	TCATCATG GATCCTTGT GTTTATTTA ATGG	To synthesize the P-40 construct in an *ibsC-sibC* encoding context, which also lacks the ibsC negative regulatory region). This construct is coined P-45MINUS(-)ve regulatory region The construct will help us evaluate the combinatory effect of the first 20 ntds in regulating *sibC* expression as measured by cell survival.	CCTGATTGACATCGT TGATTCTTT GACCTAATTTAGTGA GTAAGGGTAAGGGA GGATTGCTCCTCCCC TGAGACTGACTGTTA ATAAGCGCTGAAAC TTATGAGTAACAGTA CAATCAGTATGATG ACAAGTCGCATCAT AACCCTTCTCCTTCA AGCCCTCGCTTCGGT GAGGGCTTTACC GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAA
Pfwd -35	TCATCATGT TTAAACTT GACATCGT TGATT	For P*sibC* truncation studies, to remove the first 25 ntds of the 5' IGR upstream of the core promoter. This construct is denoted by P-35	TTGACATCGTTGATT CTTTGACCTAATTTA GTGAGTAAGGGTAA GGGAGGATTGCTCCT CCCCTGAGACTGACT GTTAATAAGCGCTG AAACTTATGAGTAA CAGTACAATCAGTAT GATGACAAGTCGCA TCATAACCCTTCTCC TTCAAGCCCTCGCTT CGGTGAGGGCTTTAC C GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAACG

			CATTGATCTGACTTT GATTTATTTTCTGGA GCAGACTCGCAAAG TAGAATGCGCAACG CGGCAACGGTGTGG AGAAGGGATAAAAA AAC GGGCAAGTCAGTGA CCTGCCCGTTGATTT TCAGAGAAGGGGAA
Prev -35	TCATCATG GATCCTTCC CCTTCTCTG AAAAT	To synthesize the P-35 construct in an otherwise native *ibsC-sibC* encoding context. The construct will help us evaluate the combinatory effect of the first 25 ntds in regulating *sibC* expression as measured by cell survival.	Same as above

Prev -35 MINUS(-)ve regulatory region)	TCATCATG GATCCTTGT GTTTATTTA ATGG	To synthesize the P-35 construct in an *ibsC-sibC* encoding context, which also lacks the ibsC negative regulatory region). This construct is coined P-45MINUS(-)ve regulatory region The construct will help us evaluate the combinatory effect of the first 25 ntds in regulating *sibC* expression as measured by cell survival.	TTGACATCGTTGATT CTTT GACCTAATTTAGTGA GTAAGGGTAAGGGA GGATTGCTCCTCCCC TGAGACTGACTGTTA ATAAGCGCTGAAAC TTATGAGTAACAGTA CAATCAGTATGATG ACAAGTCGCATCAT AACCCTTCTCCTTCA AGCCCTCGCTTCGGT GAGGGCTTTACC GTTACAGCCC CATGCTGCCCTGCCA TCGTAAATCCCCATT AAATAAACACAA

Supplementary Table 3.1. Primers to synthesize the entire *ibsC-sibC* encoding sequence with the 5' intergenic regions relative to their native promoters or truncations of these regions as specified. The complete native ibsC/sibC encoding region was synthesized by performing colony PCR on E. coli K12 or DH5αz1 genomic DNA. Using the native region as template, we performed PCR using the specified primers to generate subsequent truncations of this native context. Underlined regions are the restriction sites (PmeI on the forward primers and BamH1 on the reverse primers) to be incorporated into the constructs for their subsequent cloning into the pBSdeltaBAD plasmid. -35 and -10 regions of the core primer are highlighted in red. The transcription start site is highlighted in bold burgundy. Highlighted in green is the sibC core promoter and entire or parts of its 5'Inter Genic Regions (IGR)s. Highlighted in grey is the sibC/ibsC encoding region. Highlighted in yellow is the ibsC core promoter and entire or parts of its 5' IGRs.

4 Chapter 4: Quantitative Analysis Of RNA Expression Of The IbsC-SibC TA System In *E. coli*

4.1 Abstract

Toxin-Antitoxin (TA) systems are small genetic modules that consist of a toxin protein and a cognate antitoxin, either RNA or a protein. Although widespread and abundant in bacterial genomes, their physiological relevance is currently unknown. An impeding factor to fully elucidate the functions of endogenous levels of these TA molecules is lack of knowledge of their physiological absolute levels and how they change as the function of developmental stage or environmental conditions. Bacterial small RNAs are difficult to detect as they are only expressed under specific conditions. A TA pair that is of particular interest to us is the IbsC toxin-SibC antitoxin pair. Under many tested conditions, the toxin mRNA has never been detected, so the question has been raised: "Are they ever expressed"? Herein, we took an optimized RT-qPCR approach so that these challenging-to-detect small RNAs could be detected in wild type *E. coli* cells for the first time. We also determined the relative abundance of antitoxin to toxin RNAs in these cells. Previously, as presented in chapter 3, we engineered cells with the capability to express different RNA levels of the toxin and antitoxin and showed that our method could reliably quantify these levels. We further validated our approach by RNA-sequencing. We then used the refined approach to measure the relative abundance of the antitoxin to toxin RNAs in cells at different stages of growth. Furthermore, we estimated their average cellular copy numbers. This knowledge of expression from these TA genes, the engineered cells previously presented in chapter 2, and our optimized approach will now enable the study of TA functions at endogenous levels.

4.2 Introduction

Toxin-Antitoxin (TA) systems are genetic modules with two components, a stable protein toxin and its counteracting antitoxin. TAs have been grouped into six types so far based on the molecular nature of the antitoxin (protein or RNA) and its mode of action to neutralize the toxin.[1] They are widespread in bacterial genomes,[98,99] and in some bacteria, they account for ~3% of predicted open reading frames (ORFs).[100] Their surprising abundance raises many interesting but mostly unanswered questions about their biological roles and their evolution.[7,10,101] Proposed functions include: control over genetic material (i.e. bacterial DNA maintenance[8,9] or protection against DNA invasion[10,5]), coping with stress[12,5], persistence[20–27,30,31], biofilm formation[3–5,38–43,45,46], and virulence[51,53,55,102]. Related to these ideas, TAs have been viewed as worthy drug targets[103]. In addition, because of the cell killing ability of the toxin component and rescue by the antitoxin, TAs have been examined for anticancer[104], antiviral[105–107], antibacterial[108] applications, for containment of genetically modified organisms[109,110], and for other applications[56].

In spite of the aforementioned importance, the functions of many individual TA systems are still not fully understood. Particularly, the functions of chromosome-encoded type I TAs are less obvious.[111] They are speculated to have specialized functions under very specific conditions.[12,111] Because this particular type of TA system relies on RNA-RNA interactions to achieve regulation, the levels of the toxin mRNA and its antitoxin RNA need to be carefully regulated to ensure the most beneficial antitoxin/toxin mRNA ratio to serve these functions.[31] We show this to indeed be the case in chapter 3 for growth stasis and rescue functions by the toxin and antitoxin respectively and that multiple levels of control over their gene expression exist at the DNA and RNA levels. Therefore, it is important to

measure the levels of these RNA pairs as a function of developmental stage or environmental conditions in order to fully elucidate their functions.[73] However, it is an extremely difficult task because of their low copy numbers[74] and technical challenges for detecting low-abundant RNAs. Conventional techniques, such as cloning, northern blot, and microarray, are not sensitive enough to detect low-abundant RNAs[75,76,77]. Reverse transcription-polymerase chain reaction (RT-PCR) represents a better method [75–77,112].

The *ibs/sib* pairs in *E. coli* are of particular interest to us. They have also been in the genomes of other proteobacteria including the Enterobacteriaceae, Pasteurellaceae, and Helicobacteraceae families[113,114]. In *E. coli*, there are 5 pairs (*ibs-sib* A-E) with significant sequence similarity. Given the fact that no point mutations in either the coding sequences or the ribosome binding sites have been observed in all the *ibs* genes discovered to date, we believe at least some of them confer a selective advantage to cells.[111] However, their physiological roles are currently unknown. Some overexpression studies have shown that these toxin molecules cause cell death by localizing into the inner membrane, and causing membrane depolarization.[111] However, it is possible that the observed toxicity under these conditions could simply be an artificial outcome of overexpression, and functions of these proteins expressed at endogenous levels might be entirely independent of cell killing[111,115]. Unfortunately, although several groups have successfully detected all five Sib RNAs under various conditions examined[116,117], Ibs mRNAs expressed from the native chromosomal genes in *E. coli* have never been detected. It has been reported that Ibs mRNAs could be detected upon either deletion of the cognate Sib promoter (preventing expression of the antisense sRNA) or in a *E. coli* strain with the knockout of *rnc,* which encodes RNase III associated with cleavage of type I toxin-antitoxin complexes. These findings suggest that the cellular levels of Ibs mRNAs associated with chromosomal *ibs* genes are extremely low. But, how low is the expression of these toxin genes or

are they ever expressed? These questions have yet to be answered.[116]

We believe low abundance of the Ibs mRNAs has hampered their detection to date. We therefore made an effort to establish an effective RT-qPCR method that could reliably detect the IbsC mRNA produced by the native chromosomal *ibsC* gene. We further confirmed the reliability of the method using RNA-seq. We then used this method to measure the abundance of SibC and the IbsC mRNA in wild type *E. coli* cells at different stages of growth and showed that their transcription is growth phase dependent. Additionally, we estimated average copy numbers of SibC and the IbsC mRNA per cell. The last section of this chapter describes our efforts in measuring expression of these genes from a stressed population of cells that we came upon serendipitously and our attempts at mimicking these stressed conditions to reproduce the results. The effective RNA detection method, the new knowledge on the expression of the chromosomal *ibsC-sibC* genes acquired by this method in healthy and stressed population of cells, and the engineered cells containing additional, regulatable *ibsC-sibC* genes, will lend themselves as useful tools and information to finally interrupt the long pause in the study of this toxin/antitoxin pair and open the avenue to new discoveries pertaining to their biological relevance.

4.3 Materials and Methods

4.3.1 Oligonucleotides

To specifically detect SibC and the IbsC mRNA, highly specific primers were selected for RT-PCR, taking into consideration that there were another four homologous *ibs-sib* genes. The sequences of these primers, designed to bind

regions that showed the highest level of dissimilarity between these 5 genes, are provided in **Supplementary Table 4.1**. The sequences for all DNA oligonucleotides used in this study are listed in this table. They were chemically synthesized by Integrated DNA Technologies (Coralville, IA). Deep sequencing primers were synthesized using the same target binding sequences as RT-qPCR primers plus an index and an adaptor sequence for each primer. These primers were purified by 10% (8 M urea) denaturing PAGE prior to use according to a previously published protocol.[118]

4.3.2 E. coli strains

The *E. coli* strain used in this study, namely MG1655[119,120] was derived and named by Mark Guyer from strain W1485, which was derived in Joshua Lederberg's lab from a stab-culture descendant of the original K-12 isolate. [119,120] It is cured of the lambda phage by UV light and of the F plasmid by acridine orange so that it could no longer perform conjugation. It is considered to be the closest to the original K-12 strain, minus these two lost genes, mutations in the *ilvG* and *rfb* genes, uncertainty about the "true" wild-type allele of the *rpoS* gene, and a frameshift mutation in the *rph* gene encoding RNase PH.

4.3.3 Growth curves

To obtain cell growth curves, relevant cells were first cultured overnight in Lysogeny Broth (LB) medium and then subcultured in fresh medium (1:400 dilution) at 37 °C with shaking at 260 rpm. An aliquot was taken out hourly and its absorbance at 600 nm (A_{600}) was measured by the TECAN infinite M1000

microplate reader.

For glucose starvation experiments we created the M63 media based on an article by Karen Elbing and Roger Brent. [121] We grew the cells in LB medium for 5 hours, in M63 overnight, and sub cultured them in medium with glucose to a desired OD600 (0.3 in these experiments). We subsequently pelleted the cells for expression analysis in normally growing cells and for treatment in media with no glucose for starvation. As depicted in **Supplementary Figure 4.2**, no increase in growth was proof of starvation. Subsequently, these cells were pelleted for expression analysis in starved cells.

4.3.4 Colony counting by plating

Decimal dilutions of 10^{-2}, 10^{-3}, 10^{-4}, and others as appropriate, of a cell sample were prepared by diluting 100 microliters of the previous dilution into 1 ml of diluent by gently pipetting up and down. Hundred microliters of each dilution were evenly spread onto freshly prepared, duplicate LB agar plates, followed by overnight growth at 37 °C. The colony numbers were then counted only for the plates containing 30-300 CFUs (determined to be within linear dynamic range). Linear plots of CFU vs. dilutions were generated, from which the number of cells for a test sample was derived.

4.3.5 RNA isolation and quality test

The RNA isolation kit from Bio-Rad, small RNA purification kit from Omega as well Sureprep small RNA purification kit, microRNA kit and mirVana miRNA isolation kit from ThermoFisher Scientific were compared for their efficiencies of

IbsC and SibC RNA recovery. IbsC and SibC RNAs were transcribed *in vitro,* as described below, and purified using these kits to measure the relative loss of RNA (**Figure 4.1**a and b). In addition, we ran the total extracted RNA by these kits on agarose gels to compare the amount of tRNA isolated from the cells using each kit (**Figure 4.1**c). We ultimately chose the mirVana™ miRNA isolation kit for our experiments. The procedure was conducted using supplier-provided protocols. RNA samples obtained by the kit were treated with DNA-*free*™ DNase (to remove contaminating DNA) and subsequently with removal reagents (to remove divalent cations and DNase) from ThermoFisher Scientific, following supplier-provided protocol. To ensure the integrity of the isolated RNA, an aliquot of the RNA sample was run on a denaturing agarose gel. Sharp 23S and 16S rRNA bands signified intact total RNA (**Figure 4.3**a). For concentration and purity measurements, we used a NanoVue spectrophotometer (**Figure 4.3**b). RNA samples were ensured to display A260/280 values between 1.8-2.1 and slightly higher A260/A230 values, in accordance with MIQE guidelines.[122]

4.3.6 Synthesis of cDNA (Reverse Transcription)

SuperScript III Reverse Transcriptase from Invitrogen was used for the generation of cDNA of isolated IbsC and SibC RNAs according to the protocols provided by the supplier.

4.3.7 Real-time PCR

Real-time PCR was perfomed on cDNAs (produced as described) above to quantify the amounts of IbsC and SibC RNAs. This was carried out on the Bio-

Rad CFX Connect™realtime PCR detection system in the presence of Eva Green (Biotium, Burlington, Canada) and Phusion High-Fidelity DNA Polymerase (NEB, Pickering, Canada). Briefly, the reaction conditions consisted of 2 microliters of cDNA and 0.8 μM (final concentration) primers in a final volume of 50 μl of supermix. Each cycle consisted of denaturation at 98 °C for 10 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 30 seconds. No Reverse-Transcription Controls (NRTs) and No Template Controls (NTCs) were run together with the cDNA samples to ensure no genomic DNA contamination, no primer-dimer formation and no external DNA contamination. Calibrations against standards (prepared as described in the next subsections) were performed to quantify samples. Biological, technical and independent repeats of experiments were performed as indicated for each experiment separately in the Results section.

4.3.8 In-vitro transcription

To synthesize RNA, we used NEB's protocol for Standard RNA Synthesis. DNA template sequences containing the generic T7 promoter were generated by PCR. Primer sequences are provided in **Supplementary Table 4.1**. We used 500 mM EDTA to stop the RNA synthesis reactions. Subsequently, the samples were treated with DNase and purified using the mirVana™ miRNA isolation kit. Their quality was determined as previously described.

4.3.9 DNA standard preparation

DNA fragments encoding *ibsC* and *sibC* were generated by PCR and then gel purified. Their concentrations and purities were measured using NanoVue (to

ensure A260/280 values of ~1.8 and A260/A230 values of ~2.0-2.2). These samples were diluted over several orders of magnitude. Gene target copy numbers were determined in each sample using the concentration and dilution factor of the sample, size and nucleotide composition of the gene of interest.

4.3.10 Deep sequencing

The procedure is described in detail in the Results section.

4.4 Results

4.4.1 Developing and optimizing a detection protocol

RT-qPCR offers a sensitive method for RNA quantification. However, using this technique to specifically detect RNA species like the IbsC mRNA is not straightforward due to its low abundance and high-degree of homology to the mRNAs produced by the other four *ibs* genes in *E. coli*. Optimizations of protocols throughout the entire workflow are necessary, an effort we made to ensure the development of an effective method.

4.4.1.1 Comparison of RNA isolation methods

The IbsC mRNA was predicted to be present in scarce amounts inside cells, if at all[123], and therefore it was important to ensure its efficient isolation. As such, we compared the efficiency of different RNA isolation methods. Significant loss was

observed when the BioRad total RNA isolation kit was applied to the *in vitro* transcribed IbsC and SibC RNAs (Figure 4.1a). However, the SurePrep™ and mirVana miRNA isolation kits, optimized for small RNA purification, led to much better recovery of RNA (Figure 4.1b).

In particular, the mirVana microRNA isolation kit produced the best recovery, reflected by the intensive tRNA bands on agarose gel (Figure 4.1c) when applied to biological samples. Based on these findings, we chose this method for our RNA isolation.

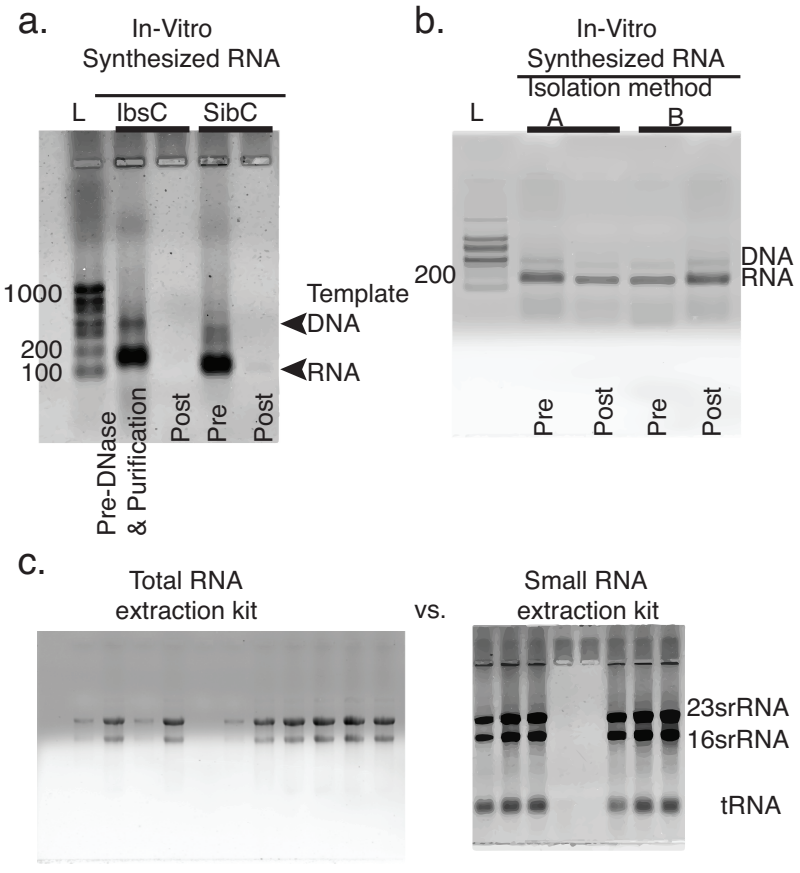
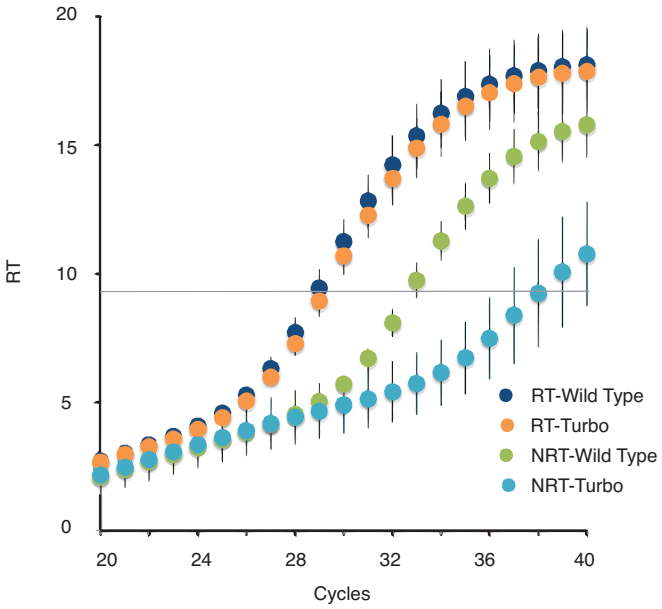


Figure 4.1. Comparing the efficiency of different RNA isolation methods for the purpose of yielding maximum amount of IbsC mRNA and SibC sRNAs from *E. coli* cells. a. Significant Loss of in vitro-transcribed IbsC and SibC small RNAs post-purification by the Bio-Rad total RNA isolation kit. b. Technologies optimized for small RNA purification (A and B denote the Sure prep and the microRNA kits respectively) result in significantly less loss of in vitro-transcribed RNA as compared to that in a. **c.** Small RNA purification technologies (Right) result in isolation of tRNA at high enough concentrations to be visualized on agarose gel whereas the BioRad total RNA extraction kit (Left) does not. This indicates improved efficiency of small RNA recovery by these small RNA purification kits.

4.4.1.2 Comparison of genomic DNA removing agents

For quantifying lowly expressed mRNA transcripts by RT-qPCR, it is of paramount importance to eliminate genomic DNA (gDNA), which is typically done by DNase digestion. Subjecting a DNase-treated RNA sample directly to PCR without revere transcription is an effective way to assess the digestion efficiency. These tests were denoted in our study as No Reverse Transcription controls (NRTs). We tested the digestion efficiency of two DNases: the regular DNase I and Turbo DNase engineered to have better activity. qPCR was performed on biological RNA samples purified by the mirVana kit using the primers designed to amplify either the *ibsC* or *sibC* gene (Figure 4.2). Turbo DNase was a better option as it produced a significantly right-shifted amplification curve (a larger Ct value for NRT-Turbo relative to NRT-Wild Type). It should be noted that the samples treated by either enzyme, when first reverse transcribed and then subjected to PCR, produced identical left-shifted amplification curves (Figure 4.2; RT-Turbo Ct = RT-Wild Type Ct < NRT-Turbo Ct & NRT-Wild Type Ct). This observation indicates that both enzymes are effective in digesting gDNA and capable of reducing it to levels that do not affect RT-qPCR results.

a.



b.

	Samples	OD600 0.2				OD600 0.7			
		Bioreplicate 1		Bioreplicate 2		Bioreplicate 1		Bioreplicate 2	
	Target / DNase	IbsC	SibC	IbsC	SibC	IbsC	SibC	IbsC	SibC
NRT	Turbo	36.42	31.46	37.63	32.92	37.56	30.14	37.56	34.77
NRT	Wild Type	32.31	28.61	31.87	30.1	30.37	29.15	32.54	28.96
NRT	ΔCt (Ct_{Turbo} - $Ct_{Wild\ type}$)	4.11	2.85	5.76	2.82	7.19	0.99	5.02	5.81
RT	Turbo	28.57	24.01	28.6	23.32	32.47	23.45	33.47	22.93
RT	Wild Type	28.32	25.7	28.3	24.75	30.51	24.21	31.87	24.35
RT	ΔCt (Ct_{Turbo} - $Ct_{Wild\ type}$)	0.25	-1.69	0.3	-1.43	1.96	-0.76	1.6	-1.42

Figure 4.2. Comparison of genomic DNA removing agents. a. Example amplification curves, comparing the efficiency of different gDNA removal reagents, namely the Turbo DNase and the Wild Type DNase. Same biological samples of RNA, were treated with either wild type DNase1 or turbo DNase in separate reactions. They were directly subjected to PCR to assess the effeciency of gDNA removal. Theses samples were labeled NRT for no reverse transcription. Amplification curves as the result of PCR with IbsC primers, on RNA samples isolated from *E. coli* cells grown to OD600 of about 0.2 are shown. The major shift to the right of the amplification curve corresponding to the Turbo DNase treated sample relative to the wild type DNase treated sample indicated enhanced enzymatic activity by the Turbo DNase. RT-qPCR for cDNA detection in these samples generated interesting results. The amplification curves corresponding to these two differentially treated samples of RNA, overlapped, showing exactly the same or similar Ct values. This indicated to us that in the presence of the target cDNA, presence of some gDNA would not affect the results of gene expression analyses . We performed these experiments on 4 different biological samples, for the detection of two different targets (IbsC and SibC cDNAs). The results are summarized in the Table. b. Average Ct values from two PCR repeats are presented. The higher Ct values, corresponding to the Turbo DNase treated NRT samples relative to the wild type DNase treated NRT samples (indicated by the positive ΔCt values) , confirmed enhanced enzymatic activity by the Turbo Dnase. Similar Ct values for the RT samples corresponding to these two differentially treated samples of RNA agreed closely (as indicated by small ΔCt values for RT samples relative to that for the corresponding NRT samples). This indicated to us that in the presence of the target cDNAs, presence of some residual gDNA would not affect the results of gene expression analyses. Therefore, both treatments were suitable for our purposes in spite of their fidderent efficiencies.

4.4.1.3 RNA integrity and purity.

RNA samples isolated with the mirVana kit and treated with DNase were analysed by gel electrophoresis and spectrophotometry, as required by MIQE standards for RT-qPCR experiments. [122] As shown with the example in the Figure 4.3, these procedures produced RNA molecules with excellent quality.

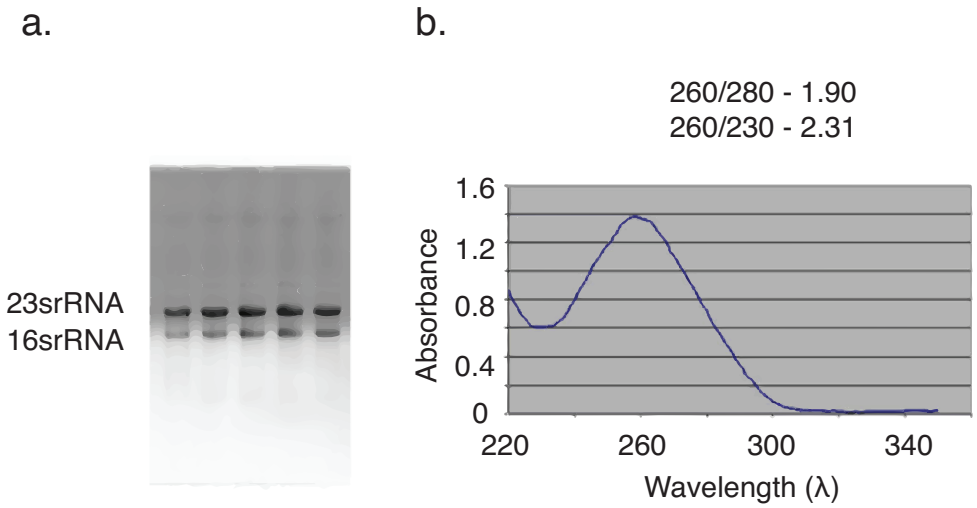


Figure 4.3. RNA integrity and purity tests. The quantity and quality of the total RNA were analyzed after isolation from each biological sample by the conventional methods, namely gel electrophoresis and spectrophotometry, as required by MIQE standards for RT-qPCR experiments. a. Typical electrophoresis gel visualization of total RNA. Distinct, intense bands corresponding to 23srRNA and 16srRNA indicate intactness of these large RNAs, hence a large likelihood for intactness of the smaller RNAs. b. Typical spectrophotometry readings. 260/280 ≈ 2 and slightly higher 260/230 values would indicate pure RNA.

4.4.1.4 Primer design and optimization of PCR conditions for the desired specificity and sensitivity

Despite the high sequence similarity among the five *sib* genes or five *ibs* genes, there are a few highly variable regions[124]. Our primers were designed to target the sequence regions of the least similarity, as shown in the Figure 4.4; the selected primer sequences are provided in Supplementary Table 4.1. We optimized the annealing temperatures (Figure 4.4b&c). Gel electrophoresis (Figure 4.5c), next-generation & Sanger sequencing (discussed in subsequent sections) and melt curve analyses (Supplementary Figure 4.3b) of the RT-qPCR amplicons, produced under the optimized conditions, altogether verified

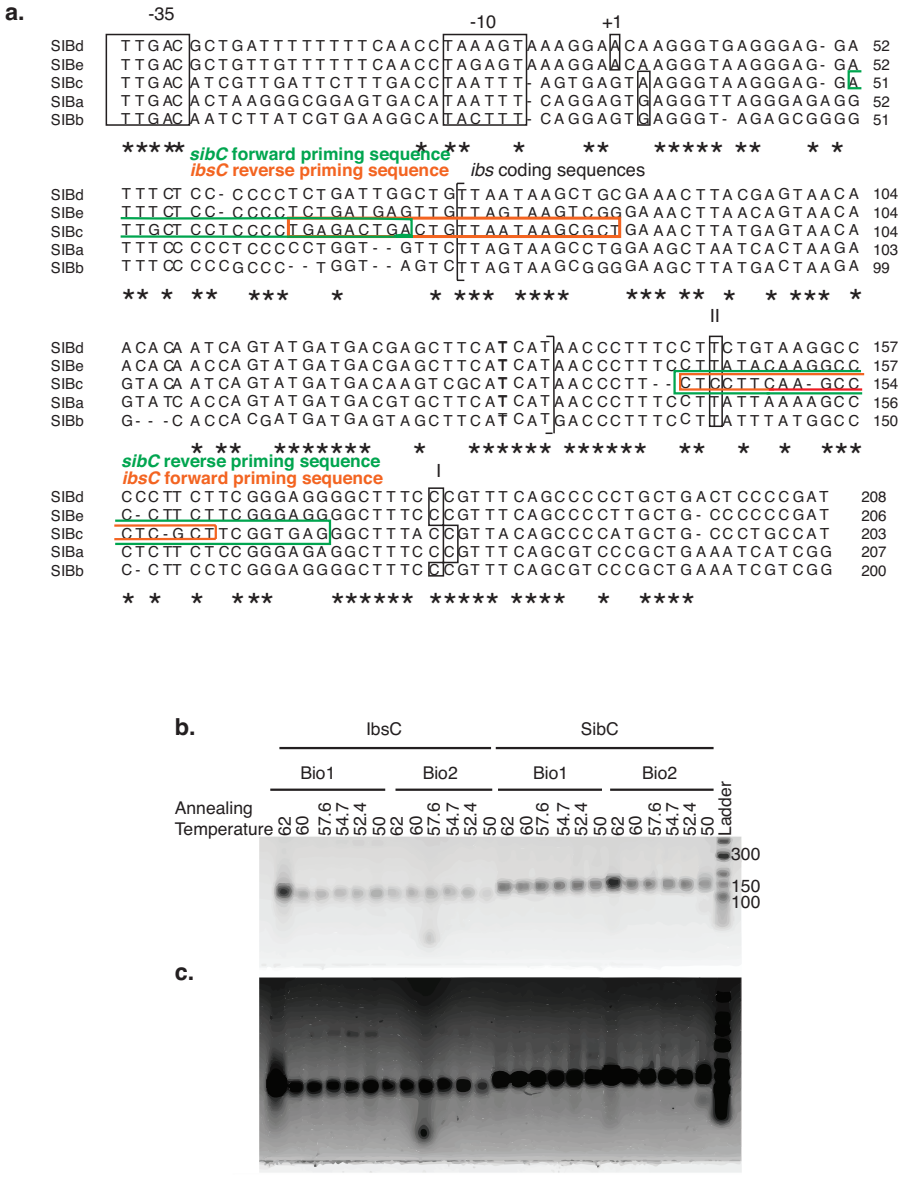


Figure 4.4. Primer design and annealing temperature optimizations. a. *sib* gene homology and organization. ClustalW alignment (with minimal manual alignment) of the five SIB repeats from E. coli MG1655. Figure is adopted and modified from Fozo. [117] The predicted −35 and −10 promoter elements for the *sib* genes are shown. The transcriptional start site as determined is indicated by +1. The two distinct 3′ ends are also indicated by I and II. Brackets are placed around the Ibs ORFs. Regions of least similarity shown by green and orange rectangles were chosen as priming regions for IbsC and SibC q-PCR primers respectively for specific amplification. The sequences for these primers are shown in the **Supplementary Table 4.1**. b. Agarose gel analyses from the annealing temperature optimization experiment. IbsC and SibC cDNAs generated from total RNA isolated from 2 biological replicates, were subjected to PCR under a range of temperatures above and below the calculated T_a of the primers. The resultant amplicons were run on an agarose gel. c. The intensified version of the same gel in b. The highest intensity of the band corresponding to the desired target (i.e. IbsC or SibC) and absence of non-specificity were achieved at 60 and 62 degrees.

the specificity of the IbsC and SibC RNA detection. Reaction sensitivities were confirmed by standard curves later in **Figure 4.5**d.

4.4.2 IbsC and SibC detection using optimized protocols

The optimized protocols were then utilized for the IbsC mRNA and SibC detection in biological replicates. Briefly, purified RNA extracts from two cell cultures originated from two colonies of *E. coli* were reverse-transcribed to create the cDNAs for IbsC and SibC in separate reactions. The reaction products were then subjected to RT-qPCR. The workflow is presented in Figure 4.5a. The amplification plots are presented in Figure 4.5b. The very large differences of Ct values acquired for the biological samples relative to NRT and NTC controls (discussed in Materials and Methods) clearly indicate the presence of the IbsC mRNA and SibC in *E. coli* cells. To our knowledge, this is the first report of successful detection of IbsC transcripts in wild-type *E. coli*.

The PCR amplicons taken from cycles 21, 23, 28, 31 and 33 were analyzed on agarose gels along with the cycle-40 NTC and NRT PCR solutions (Figure 4.5c). Clear DNA bands corresponding to the expected DNA amplicons were observed only in biological samples and their intensities increased with cycle numbers. Equally important is the lack of amplicons in both controls, signifying that the detected amplicons did not originate from external DNA contamination or genomic DNA. The amplification specificity was further confirmed by melt curve analyses (Supplementary Figure 4.3b). Taken together, these results show that our optimized protocols can effectively detect both the IbsC mRNA and SibC present in wild-type *E. coli* cells.

We next established calibration curves of Ct value vs. the absolute copy number of IbsC and SibC, as shown in **Figure 4.5**d. Note that these calibration curves

were established using purified double-stranded IbsC and SibC DNA products because we can accurately measure their concentrations. Two points are worth commenting. First, the Ct value exhibits an excellent linear relationship with the copy number of IbsC and SibC over 4-5 orders of magnitude, indicating the reliability of qPCR for the detection of input IbsC and SibC templates. Second, the IbsC curve and the SibC curve exhibit a slope of 3.9 and 3.3 respectively, translating into an amplification efficiency of 1.80 and 2.01 (using the formula of PCR efficiency = $10^{-1/slope}$-1),[125] both of which are close to the ideal amplification efficiency of two (doubling of amplicons in each PCR cycle). These calculations show that the PCR primers and protocols can effectively amplify IbsC and SibC over a large concentration range.

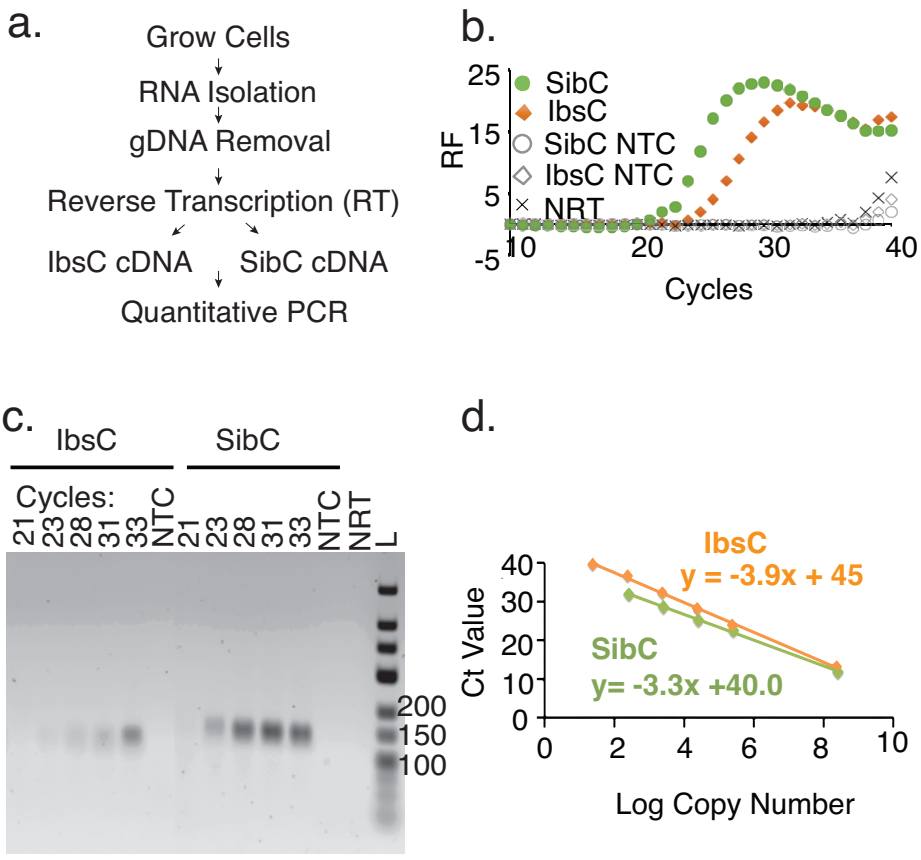


Figure 4.5. Establishing an effective RT-qPCR method for IbsC and SibC RNA detection. a. Overall workflow used to obtain the required materials for RT-qPCR. b. Amplification curves. NRT: No Reverse Transcription control; NTC: No Template Control. c. Agarose gel analysis of the IbsC and SibC PCR solutions taken at indicated cycles. d. Calibration curves of Ct value vs. logarithmic copy number of purified double-stranded IbsC and SibC DNA fragments.

4.4.3 Next-generation sequencing to validate RT-qPCR

Next-generation sequencing of cDNA (notably RNA-seq) provides both the complete sequence of all the genetic material in the sequenced sample and allows actual counting of RNA numbers, offering unbiased information on average cellular RNA concentrations in cell populations[73,126]. As such, we chose to perform an RNA-seq experiment for two purposes: 1) to confirm pure composition of our samples after RT-qPCR (i.e. specificity of RT-qPCR) and 2) to independently quantify RNA copy numbers to validate our q-PCR data.

We started by cross-validating q-PCR and deep sequencing-based quantifications. The approach is summarized in **Figure 4.6**a. Briefly, we prepared samples containing known concentrations of SibC and IbsC DNA (measured by OD_{260}) and subjected them to PCR with differently indexed sets of deep sequencing primers (**Supplementary Table 4.1**). These primers consisted of an adaptor region, an index sequence as well as a target binding sequence. Ten PCR reactions were performed on each sample, two of which were allowed to reach completion to generate an average Ct value for q-PCR, and the rest were used to generate the DNA amplicons for deep sequencing. These 8 reactions were stopped at a cycle below the Ct value (indicated by the arrow on the amplification plot) to avoid sequence quantification bias due to over-amplification. The PCR products from these reactions were combined and gel purified prior to sequencing. The sequence reads (SR) were counted from forward and reverse reads, averaged and plotted against the copy number obtained by the OD_{260} method (CN_{OD}) in **Figure 4.6**b.

The observed SR-CN_{OD} linearity between the sequence reads and copy numbers validates the quantitative value of the deep sequencing approach.

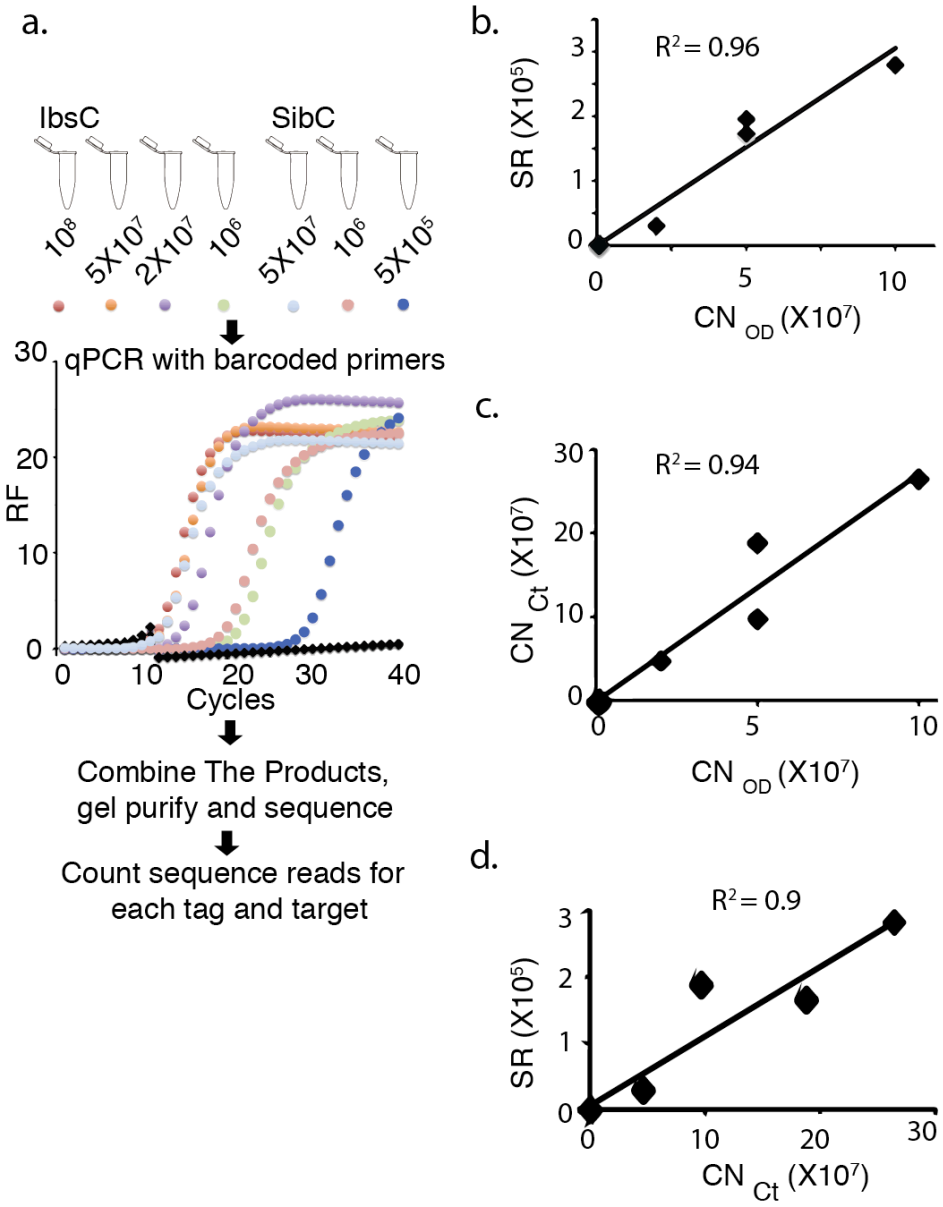


Figure 4.6. Next-generation sequencing with standard IbsC and SibC DNA samples. a. Workflow and list of samples tested. Each sample contained known copy numbers of SibC or IbsC DNA obtained by OD260. b. Copy numbers obtained by sequencing reads (SR) vs. copy numbers of the test samples obtained by OD260 (CN_{OD}). c. Copy numbers obtained by Ct values (CN_{Ct}) vs. copy numbers obtained by

OD260 (CN_{OD}). d. The sequence reads (SR) vs. copy numbers obtained by Ct values (CN_{Ct}).

The copy number of each sample was also calculated from the Ct value (CN_{Ct}), which was then plotted against CN_{OD} (**Figure 4.6**c). The CN_{Ct}-CN_{OD} linearity with a slope of ~1 indicates close agreement between the copy numbers obtained by the two methods. We also plotted SR versus CN_{Ct} (**Figure 4.6**d), which exhibits a linear relationship, as expected. These results cross-validated the quantitative value of all the 3 techniques.

We next applied the three techniques to measure the relative abundance of SibC to IbsC RNAs in biological samples. Of particular note is the usefulness of next-generation sequencing in confirming the IbsC and SibC detection specificity by RT-qPCR.

After RNA isolation, cDNA of IbsC and SibC were individually prepared for sequencing. This is because DNA sequencing is more mature than direct RNA sequencing. A schematic showing the library preparation is presented in Figure 4.7a.

To prepare the biological sample, the following steps common to RT-qPCR were performed: RNA isolation from *E. coli* cells grown to midlog phase, individual reverse transcription reactions to generate IbsC and SibC cDNAs separately, multiple individual real-time PCR reactions for each target using their corresponding cDNAs as template and deep sequencing primers that were uniquely indexed. Only one amplification curve of eight repeats is presented for each sample as an example. The primer sequences are presented in **Supplementary Table 4.1**.To validate the reliability of our deep sequencing run with respect to estimating the relative abundance of the target molecules, standard samples containing known concentrations of SibC relative to IbsC as measured by

spectrophotometry (Nanovue) were also prepared and PCR amplified similar to our biological sample but each with a differently indexed set of primers (Figure 4.7a).

Again, multiple PCR reactions were performed to amplify each target within each sample and only example amplification curves are shown for each one. It is worth mentioning that the amplification repeats for each sample produced very similar Ct values, pointing to the reproducibility of our technique. The two targets that were eventually combined to create a certain ratio for each sample were amplified with deep sequencing primers containing the same index sequence (color coded in Figure 4.7a).

The reactions for each target were allowed to complete 30 cycles of PCR and pooled together to generate enough substance for deep sequencing. We used ethanol precipitation and gel purification to purify the target sequences prior to combining them to create the intended ratios and sending them for sequencing.

The [SibC]/[IbsC] RNA quantification results are shown in Figure 4.7b). The estimates of [SibC]/[IbsC] in the standard samples A-C, by spectrophotometry, deep sequencing and qPCR agreed closely, cross validating all three techniques with respective to RNA quantification. In addition, the deep sequencing reads for the biological sample indicated a) pure composition, hence specific detection of the targets and b) A 5:1 ratio of SibC to IbsC concentration which closely agreed with the estimate by RT-qPCR.

In the subsequent sections we will use the optimized and validated RT-qPCR approach for further *ibsC* and *sibC* gene expression analyses.

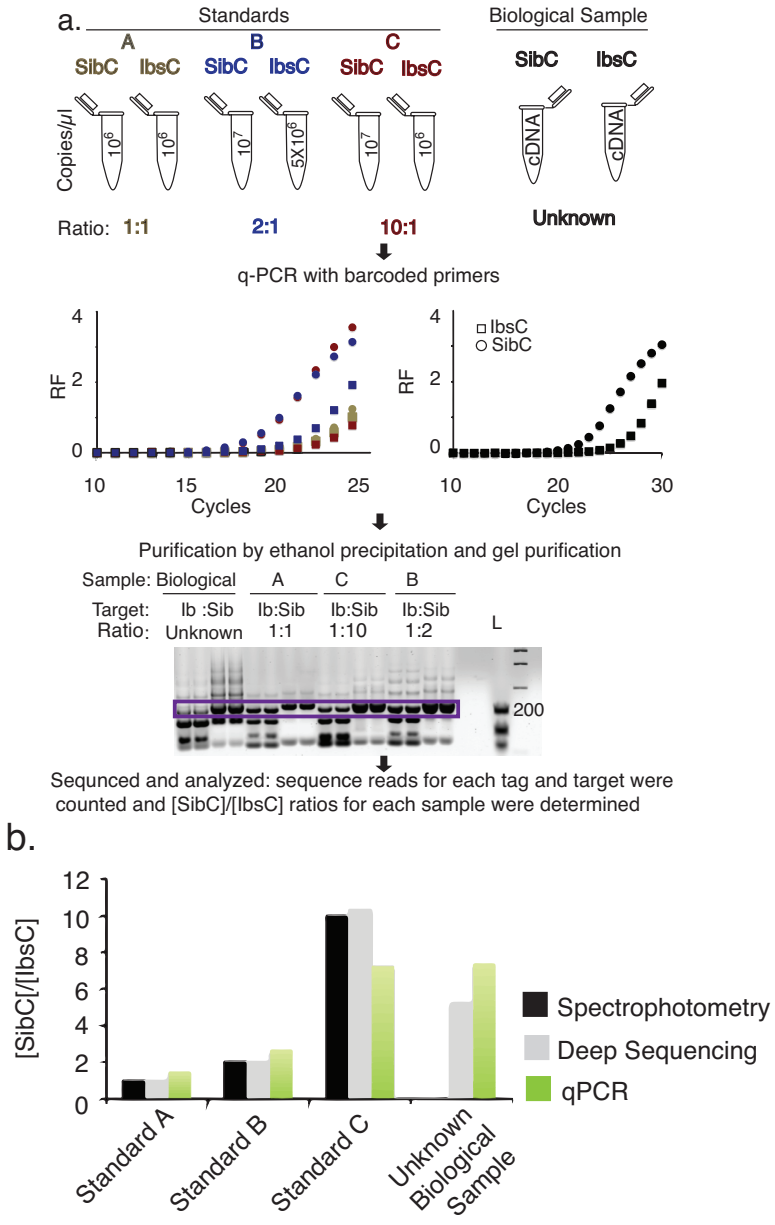


Figure 4.7. Deep sequencing to confirm the relative abundance of SibC sRNAs and IbsC mRNAs in a biological sample determined by RT-qPCR and to confirm specificity of RT-qPCR reactions with respect to distinguishing between the IbsC-sibC pair and the other four homologous copies (IbsA,B,D and E). a. Schematic showing DNA library generation for deep sequencing. The following steps common to RT-qPCR were performed to create the biological sample: RNA isolation from *E. coli* cells grown to midlog phase, individual

reverse transcription reactions to generate IbsC and SibC cDNAs separately, multiple individual real-time PCR reactions for each target using their corresponding cDNAs as template and deep sequencing primers that were uniquely indexed. To validate the reliability of our deep sequencing run with respect to estimating the relative abundance of the target molecules, standard samples containing known concentrations of SibC relative to IbsC as measured by spectrophotometry (Nanovue) were also prepared and PCR amplified similar to the biological sample but each with a differently indexed set of primers. Multiple PCR reactions were performed to amplify each target within each sample. Only example amplification curves are shown.The two targets, that were eventually combined to create a certain ratio for each sample, were amplified with deep sequencing primers containing the same index (color coded). The reactions for each target were allowed to complete 30 cycles of PCR, pooled together to generate enough substance for deep sequence. We used ethanol precipitation and gel purification to purify the target sequences prior to combining them to create the intended ratios and sending them for sequencing. The Ct values from the q-PCR experiments and sequencing reads from deep sequencing were analyzed as previously described. b. Comparing the relative abundance of antitoxin to toxin (i.e. [SibC]/[IbsC]) determined by spectrophotometry, deep sequencing and q-PCR in different samples. As shown, the estimates of [SibC]/[IbsC] in the standard samples, by deep sequencing in grey bars, spectrophotometry in black, and qPCR in geen matced closely. This cross-validated the three techniques. Therefore, an estimate of about 5fold higher SibC relative to IbsC in the biological sample, by deep sequencing was considered reliable. In addition, it closely agreed with the estimate by RT-qPCR.

4.4.4 Growth phase dependency of the IbsC toxin and SibC antitoxin RNA expressions

Gene expression changes over time. As such, the data presented so far, taken at a single OD_{600}, only create a snapshot. We performed time course experiments to observe changes in the expression of *ibsC* and *sibC* genes as a function of growth phase. Since RNA levels potentially reflect cellular requirements for different tasks[73], comparing the toxin-antitoxin gene expression during different phases of growth can provide hints about the function of these genes and whether their expression may be growth phase dependent.

We cultured two independent colonies for 48 hours in the LB media. Growth curves are shown in Figure 4.8a. At multiple time points representing different

growth phases, equal numbers of cells (as confirmed by plating and CFU counting, to normalize against cell numbers) were taken. Subsequently, RT-qPCR was used as previously described to quantify transcript levels at the different growth phases. The results were normalized against total RNA in these samples and reported in Figure 4.8b-d relative to one another.

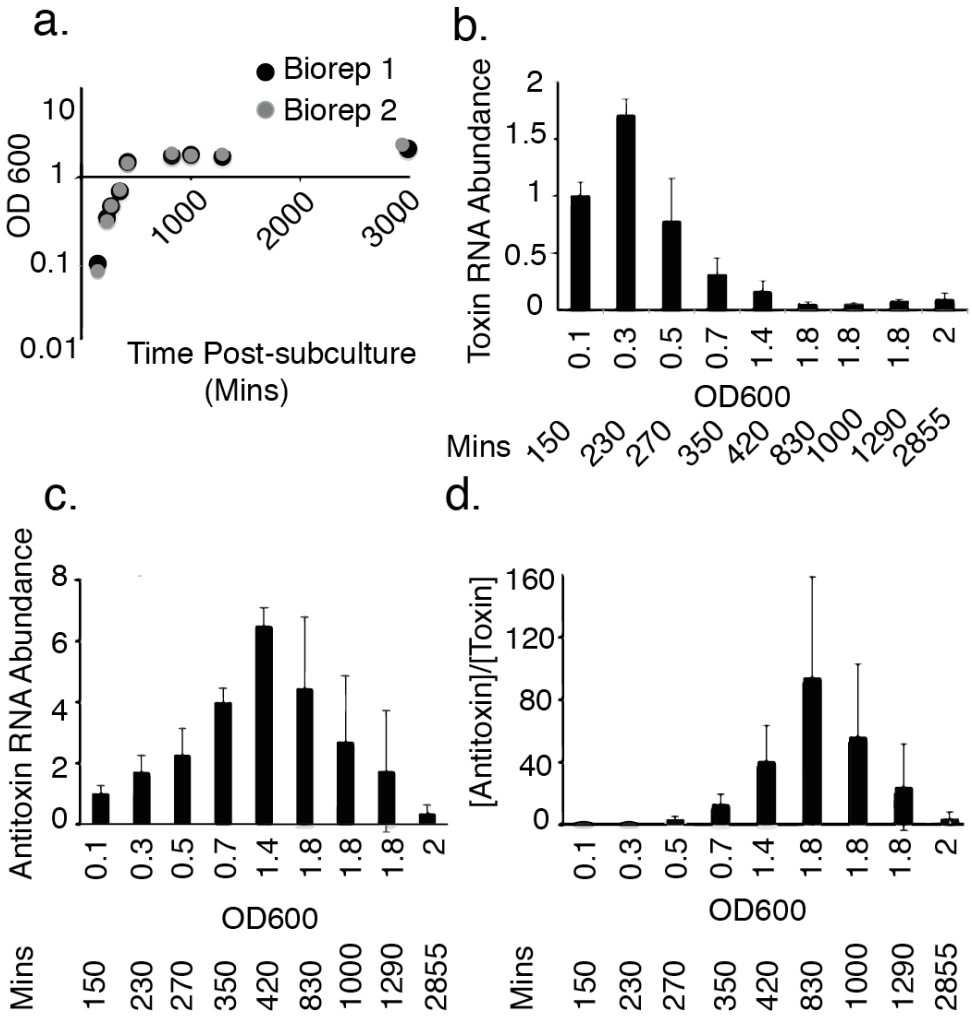


Figure 4.8. Growth phase dependency of IbsC toxin and SibC antitoxin RNA expressions. a. Growth curves for two independent colonies that were cultured for 48 hours in LB media. b. Toxin RNA abundance at multiple time points representative of different growth phases. Our optimized RT-qPCR based approach was used as previously described to analyze gene expression. The results were normalized against cell numbers experimentally by taking equal numbers of cells (confirmed by plating and

CFU counting) from different growth phases for analysis, and against total RNA mathematically and reported here relative to one another. c. Antitoxin RNA abundance at different growth phases. d. Antitoxin to toxin ratios during different growth phases. The data are averages of two biological replicates.

The toxin exhibited a switch like expression (Figure 4.8b). Its expression seemed to be restricted to the early to mid-log phase and repressed in the stationary phase. The peak of the toxin RNA levels during the mid-log phase, when cells are rapidly growing, is interesting considering the potential by these molecules to halt cellular growth. The expression pattern observed here suggests that the toxin might be important for physiological functions other than growth arrest.

The antitoxin showed an increase in transcript levels with growth, a peak at the early stationary phase and a gradual decrease into the late stationary phase (Figure 4.8c). This seems contrary to what is currently believed in the literature i.e. constitutive expression from these genes [123]. Due to the potential by the toxin to halt cellular growth, the balance between the levels of TA RNAs is critical for the cells' growth status. Antitoxin RNA levels should be sufficiently higher than toxin RNA levels to successfully inhibit the toxin. As such, one might expect the antagonizing ratio of the antitoxin to toxin to be at its peak during the mid-log phase when the cells are actively growing compared to the stationary phase where a lot of cells are in a dormant state. Using the toxin and antitoxin gene expression data, we calculated antitoxin to toxin ratios during different stages of growth and plotted them as a function of OD_{600} values in Figure 4.8d.

Contrary to our expectations, this ratio gradually increases throughout growth to reach its peak during the stationary phase. These physiological toxin-antitoxin ratios suggest physiological functions by these molecules other than growth arrest by the toxin and the antagonizing effect by the antitoxin (Please see discussion).

We have performed an independent time course study to reproduce these results.

The results are summarized in the **Supplementary Figure 4.4**.

4.4.5 Measurement of IbsC and SibC RNAs at different OD600s in a stressed, slowly growing population of cells we serendipitously arrived upon

During the course of our studies we accidentally stumbled upon some very slowly growing cell cultures originated from colonies on an old plate. Growth curves for these cells are shown in **Supplementary Figure 4.1**. Performing *ibsC* and *sibC* gene expression analyses on these cells, similar to that done on normally growing cells at different OD_{600} values, revealed interesting results. For the first time we observed that the IbsC toxin RNA was expressed at higher levels relative to the antitoxin RNA (**Figure 4.9**). In addition, the toxin RNA was expressed identically throughout the different phases of growth, suggesting that the expression of this gene in slowly growing cells was not tuned to specific requirements in different growth phases. Constitutive transcription from this gene in these stressed cells, also suggested some kind of functionality other than toxicity or death by IbsC, perhaps a role in protection against stress. This is one of the longest-lived hypotheses about the role of TA systems.

Gene expression analyses of *sibC* also revealed an interesting pattern. SibC RNA levels continued to gradually rise with OD_{600} levels such as that observed in healthy wildtype cells in Figure 4.8c. Therefore, it seems that irrespective of the toxin RNA expression patterns (i.e. lower amounts relative to the antitoxin in healthy or higher relative amounts in stressed cells, constant expression levels throughout growth in stressed cells or changing expression levels in healthy cells), the antitoxin expression pattern remains the same. This supports the notion that the antitoxin might moonlight to regulate additional targets as also later explained in the discussion.

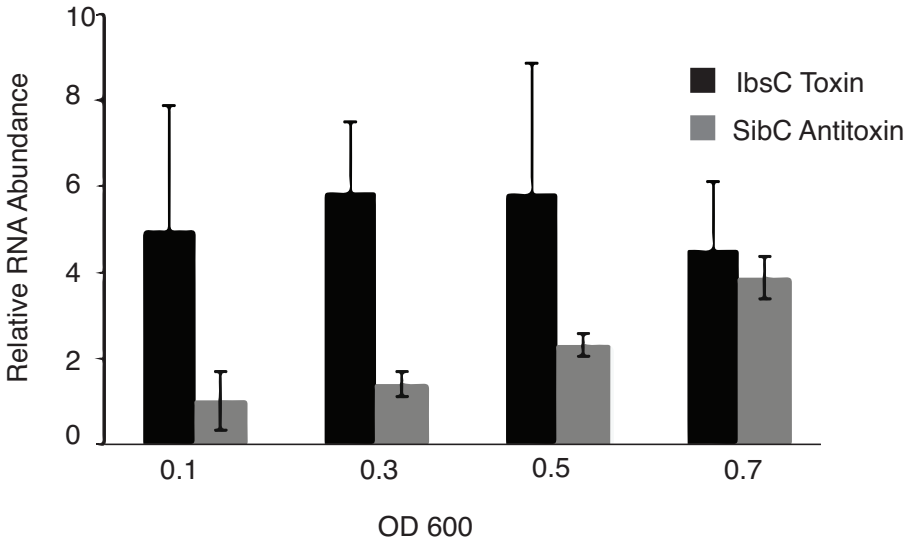


Figure 4.9. Gene expression analyses in a slowly growing, stressed population of cells. Data points are averages of two RT-qPCR reactions performed on two subcultures originated from three biological replicates (average of 12 repeats altogether)

4.4.6 Measurement of IbsC and SibC RNAs in a glucose-starved population of cells

As described in the previous section, slowly growing cells from an old plate showed a different pattern of *ibsC* and *sibC* expression than normally growing cells. To reproduce the results, we tried to mock the conditions that might have triggered the different expression patterns. We speculated that nutrient deprivation was probably the cause of slow growth and consequently the observed expression patterns. Therefore, we cultured *E. coli* cells in an M63 medium. M63 is a minimal, defined, low osmolarity media for *E. coli*, which results in slower growth rate of these cells. To create normally and slowly growing cells under regular and starvation conditions for comparison, we either included or removed glucose from the media. Details of the experiment are provided in the materials and methods section.

Gene expression analyses by RT-qPCR results showed no difference in the expression of the toxin and antitoxin between starved cells (grown in a no-glucose media) and normally growing cells (in media with Glucose) in spite of their very different growth behaviors (i.e. no growth by starved cells vs. normal growth by cells in complete media).

To uncover what exactly in the old plates might have triggered the different expression pattern may be a challenging task. Taking the glucose away from the media, to cause starvation was my first attempt at deducing this trigger. In the literature, lack of thymine and amino acids and some other nutrients have also been shown to cause elevation of certain toxins. [127–129] Studying expressions under these conditions (i.e. lack of the individual nutrients or a combination of them) may provide us with clues as to when these molecules are expressed.

4.4.7 Cellular copy numbers of IbsC-SibC toxin-antitoxin RNAs

Gene regulation has a central place in all domains of biology. Although transcriptomic approaches have generated ample data on relative expression changes between different conditions, little is known about actual cellular numbers of RNAs and how gene regulation affects them. These numbers will empower understanding of cellular functions and biological systems[73]. As such, we aimed to measure average cellular toxin and antitoxin RNA copy numbers in a population of cells at mid-log phase ($OD_{600} \approx 0.5$) in balanced growth at 37 degrees in LB media.

To estimate these numbers, RT-qPCR was used to measure target RNA copy numbers in total small-RNA isolated from different amounts of cells (i.e. Colony

Forming Units or CFUs). Briefly, *E. coli* cells were grown to mid-log phase. Total small-RNA was extracted from different amounts of cells (i.e. different volumes of cells that were grown to the same OD_{600}, mid-log). Plating and CFU counting were performed in parallel to estimate the cell numbers, from which RNA was isolated. Subsequently, RNA quality check, DNase treatment of RNA samples, reverse transcription, and q-PCR were carried out. As previously described, Ct values for known copy numbers were used to translate Ct values in these experiments into estimates of copy numbers in individual q-PCR reactions. Simple math was then used to arrive at target copy numbers per total RNA or certain CFUs. Calculated target RNA copy numbers were plotted versus CFUs in **Figure 4.10** a & b for the IbsC toxin and SibC antitoxin respectively. As expected, target RNA copy numbers increased linearly with cell numbers. The slopes of these graphs were used to estimate average copy numbers per cell (i.e. 0.31 for the toxin and 0.94 for the antitoxin).

To achieve the linearity observed in our figures (i.e. target RNA copy numbers increasing linearly with the increasing levels of CFUs), additional optimizations of our methodology were required. We had to find the linear dynamic range within which the efficiencies of lysing cells, RNA isolation, reverse transcription and qPCR were equal for the different amounts of starting material (i.e. cells, RNAs and cDNAs respectively). We experimented with different amounts of cells to find the optimal range presented in **Figure 4.10**. Panel c also shows the linearity between these cell amounts and the quantity of total RNA isolated from them, hence confirming that our lysing and RNA isolation techniques were equally efficient for these cell amounts. For any higher cell amounts we observed the curves would reach plateau (data not shown), due to either inefficient lysing of the cells or the limited capacity of RNA purification columns before they were clogged with cell debris and hence let through smaller amounts of RNA.

In addition, to accurately estimate average cellular RNA numbers, we had to

ensure that recovery of RNA by our RNA isolation method was not only efficient but also as close to complete as possible. To examine this, we created samples containing different amounts of RNA; in the range of RNA amounts collected from our cells, and purified them using our methodology. As shown in **Figure 4.10** d, the amounts of RNA recovered matched closely to the amounts of RNA pre-purification. The amounts were measured by spectrophotometry and confirmed by gel electrophoresis (**Supplementary Figure 4.5**). The results indicated close to 100% recovery for majority of the samples.

Average estimates of approximately 0.3 IbsC mRNA and 0.9 SibC sRNA copy numbers per individual cells of the population, put the IbsC toxin and SibC antitoxin in the category of tightly repressed genes (i.e. <1 copy per cell)[73] and suggested a 3:1 ratio of the antitoxin to toxin, consistent with our previous observations (Figure 4.8d and Supplementary Figure 4.4d). Our estimates of <1 RNA copy per cells are also consistent with the few attempts in the literature at determining cellular numbers for RNAs in yeast and bacteria [73,74]. Noncoding, regulatory RNAs have been extensively quantified in yeast.

The results from these studies suggest that most of these RNAs that do not produce detectable proteins are tightly repressed below 1 copy per cell and these low copy numbers are sufficient for roles in *cis*. Similar comprehensive quantifications of the regulatory RNAs are lacking in bacteria. However, the <1 copy/ cell copies of the SibC sRNA studied here, also expressed in *cis* relative to its known target IbsC might similarly be sufficient for roles by SibC. In addition, data from *E. coli* suggest that great majority of mRNAs in *E. coli* were expressed at <1 copy per cell, yet productive protein expression happened from such low mRNA concentrations. Therefore, it is not unusual for a toxin mRNA which could potentially have deleterious effects on the cell, to be expressed at <1 copy per cell.

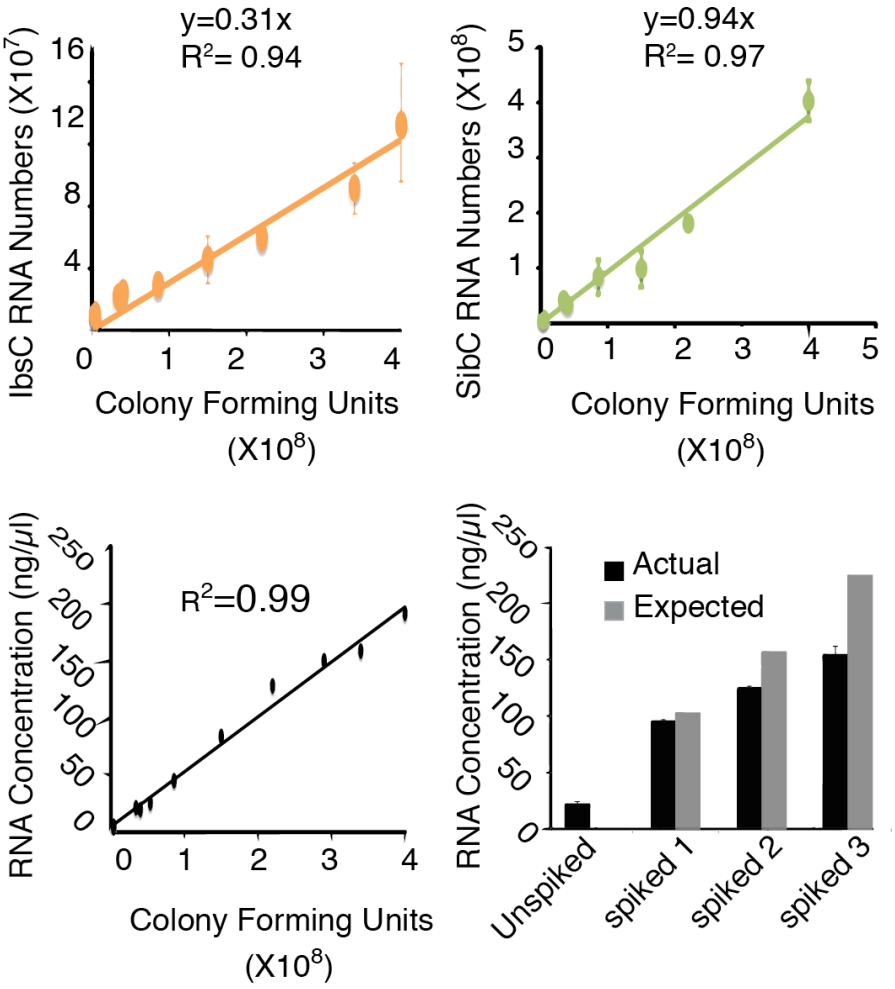


Figure 4.10. Average cellular IbsC toxin and SibC antitoxin RNA copy numbers in an average *E. coli* cell population, at mid-log phase (OD600≈0.5) in balanced growth at 37 degrees in LB media. RT-qPCR was used to measure target RNA copy numbers in total small-RNAs isolated from different amounts of cells (i.e. Colony Forming Units or CFUs). The data were pooled from two cell cultures, and are averages of at least two RT-qPCR repeats. a. Measured toxin RNA copy numbers are plotted against CFUs. Linearity is observed as expected. The slope of the graph estimates an average of approximately 0.31 RNA copy numbers per cell for the toxin. b. Measured antitoxin RNA copy numbers are plotted against CFUs. The slope of the graph estimates an approximate average of 0.94 RNA copy numbers per cell for the antitoxin c-d. Additional optimizations of our methodology to accurately estimate cellular RNA numbers. c. Efficient recovery of RNA. Linearity between cell amounts used in a and b and the quantity of total RNA isolated from them was necessary to achieve the linearity observed in a and b. We experimented with different amounts of cells to find the optimal range presented here. For any higher cell amounts we observed the curves would reach plateau (data not shown), due to either inefficient lysing of the cells or the limited capacity of RNA purification columns. The linearity observed in a, b and c indicated equally efficient lysing, RNA isolation, reverse transcription and qPCR for the different

amounts of starting material. d. Almost complete RNA recovery. The amounts of RNA recovered by our RNA isolation method matched closely to the amounts of RNA pre-purification. This indicated close to 100% recovery for majority of the samples. Thsee results from OD_{260} quantification of RNA were confirmed by running the samples on an agarose gel pre and post purification (**Supplementary Figure 4.5**).

We also estimated average cellular RNA numbers for the toxin and antitoxin in cells from different phases of growth (**Supplementary Figure 4.6**). The numbers were consistently measured below 1 copy per cell and the patterns of expression agreed with our relative RNA abundance data presented above, for *ibsC* and *sibC* transcription during different phases of growth (Figure 4.8 and Supplementary Figure 4.4).

4.5 Discussion

Here, we optimized a real-time quantitative reverse transcription-PCR based approach to detect IbsC toxin mRNAs in *E. coli* wild type cells for the first time. Our methodology proved to overcome challenges associated with the detection of these molecules such as their low abundance and the great homology between them and the other four copies of the *ibs* and *sib* genes in *E. coli.* In addition to their detection, we measured the relative abundance of antitoxin to toxin in these cells. The results suggested higher concentrations of SibC relative to IbsC in these cells consistent with the role of antisense transcription in repressing the corresponding sense transcription. We also showed that in *E. coli* cells capable of producing different levels of the IbsC and SibC RNAs (described in chapter 3), our optimized methodology could reliably track the changes that we artificially induced. This confirmed the accuracy and reliability of our method in detecting these molecules and measuring their relative abundance. We further validated our RT-qPCR based approach by next generation sequencing. By performing

simultaneous q-PCR and deep sequencing-based quantification of pure DNA samples and biological cDNA samples and achieving similar results, we validated our q-PCR based approach with respect to absolute quantification of nucleic acid copy numbers and relative quantification of intracellular levels of sibC and IbsC RNAs. The sequencing data also confirmed pure composition of the amplicons generated in the RT-qPCR reactions. This importantly showed that our RT-qPCR technique was able to discriminate between IbsC-SibC RNA transcripts and their highly similar homologs (Ibs-Sib A, B, D and E RNAs) in *E. coli*.

Having developed an optimized RT-qPCR based for quantification of low abundant IbsC mRNAs and SibC sRNAs, and having it validated by multiple strategies discussed above, we were next motived to analyze expression from *ibsC* and *sibC* genes more extensively. Gene expression changes over time. As such, we performed time course experiments to observe changes in the expression of *ibsC* and *sibC* genes as a function of growth phase. Since cells fine-tune their RNA levels according to their requirements for different tasks, comparing the toxin-antitoxin gene expression during different phases of growth can provide hints about the function of these genes and whether their expression may be growth phase dependent.

We observed that in wild type *E. coli* cells, the toxin gene peaked in its expression during the exponential phase and was repressed during the stationary phase. These observations suggested to us some kind of functionality other than toxicity or death by IbsC, during growth. Similar pattern of expression has been observed for other type I TA systems such as the ShoB-OhsC TA pair where the toxin mRNA has been reported to be highest in the exponential phase[111,117,130]. Also, recently, new type 1 TA systems have been discovered adjacent to CRISPR arrays in *Clostridium difficile*, to confer immunity to these bacteria against stress. Their gene expression analyses also revealed higher expression during exponential phase than stationary[5].

Interestingly, during the course of our studies we accidentally stumbled upon some very slowly growing cell cultures originated from colonies on an old plate. We Performed *ibsC* and *sibC* gene expression analyses on these cells, similar to that done on normally growing cells. For the first time, we observed that the IbsC toxin was expressed at higher levels relative to antitoxin (**Figure 4.9**). In addition, the toxin was expressed identically throughout the different phases of growth, suggesting that the expression of this gene in slowly growing cells was not tuned to specific requirements in different growth phases. Constitutive transcription from this gene in these stressed cells also suggested some kind of functionality other than toxicity or death by IbsC, perhaps a role in protection against stress.

Previously, people have linked expression patterns with cellular functions. It has been concluded that mRNAs functioning in the same or related biological processes, usually share similar expression levels and these levels reflect cellular requirements for different tasks or complexes. [73] Such studies in yeast cells have compared copy numbers for mRNAs from selected functional categories in proliferation and quiescence stages. According to these studies, most functional categories were substantially down-regulated in quiescence, whereas a few retained similar numbers. Messenger RNAs and proteins involved in cell maintenance, such as adaptation to stress and nutrient limitation, DNA repair and cell wall, become relatively more abundant in quiescence, whereas those involved in translation and growth show decreased abundance at the mRNA and protein levels in quiescence. These findings highlight how cells change in their physiology from a growth program in proliferation to a maintenance program in quiescence to protect against stress and for long-term endurance.[73]

Similar quantitative studies in bacteria remain to be performed. However, assuming that bacteria have also implemented these two fundamental programs by balancing the expression of stress versus growth related genes, during stationary versus exponential phases of growth, and based on the expression pattern of the

toxin in slowly growing cells (i.e. retaining similar levels of expression during stationary phase as in other growth phases, whereas majority of other mRNAs shrink in expression levels), a role in cellular maintenance could be hypothesized for the toxin, such as adaptation to stress and nutrient limitation. This would be relevant as these slowly growing cells were nutrient-deprived and stressed. Also interestingly, stress response has previously been suggested as a hypothetical role for these toxins and has been proven to be the case for certain ones[99,129,131–134]. In normally growing cells, however, higher levels of the toxin during the mid-log phase relative to the stationary phase suggest a role for the toxin in growth rather than maintenance. These findings are not contradictory to one another as these toxins might serve more than one role inside cells. SibC RNA expression analyses also revealed an interesting pattern. As shown in Figure 4.8, RNA levels continued to gradually increase in wild type *E. coli* cells (and in slowly growing cells, shown in the **Figure 4.9**), peaked at early stationary and gradually decreased into late stationary. Whether the expression of the antitoxin is independently regulated or controlled together with a set of critical regulators by a regulatory switch during different growth phases is not known yet. It is tempting to speculate that the SibC antitoxin, as it is usually more highly expressed than its toxin mRNA and varies in its expression throughout growth phases even when the toxin is fully repressed, moonlights to regulate additional target RNAs.[135] This is not unprecedented, as many *trans*-encoded sRNAs have more than one target (e.g. *S. aureus* RNAIII or *E. coli* RybB) [136] and frequently use different mechanisms to regulate them. In some cases, only translation is affected, in others both translation and RNA degradation. [135,136] Same sRNA can also have activating or inhibitory effects on different targets (e.g. *S. aureus* RNAIII).

Normally, toxins are expected to be tightly repressed by their cognate antitoxin inside cells to allow for normal growth. Therefore, one might expect the antagonizing ratio of antitoxin to toxin to peak during the mid-log phase when

cells are actively growing. However, our results show that this ratio gradually increases throughout growth to reach its peak during the early stationary phase and gradually decreases as the cells proceed into late stationary phase. Perhaps such physiological toxin-antitoxin ratios are important for some physiological functions by these molecules (other than growth arrest by the toxin and the antagonizing effect by the antitoxin). For example, consider the hypothesized function of causing bacterial persistence for the toxin.[102,134,137–140] The smaller antitoxin levels relative to toxin levels in the mid-log phase, as compared to that in the stationary phase could be a great advantage upon rapid changes of environmental conditions during active growth, in terms of a greater percentage of bacterial populations becoming persisters in the face of challenge. Perhaps, cells could afford even lower toxin levels relative to the antitoxin during the stationary phase, as compared to the growth phase, because they are already in a dormant state and the role of the toxin to induce dormancy may not be as critical for them in creating persistence.[28]

The role of the toxin might become more prominent again in the later stationary phase; perhaps in protection against stress and long-term endurance, hence the lower observed abundance levels of the antitoxin.

These hypotheses by others and us still await experimental evidence. We believe that collection of expression data under different physiological and environmental conditions would contribute to unravelling the function(s) of these molecules.

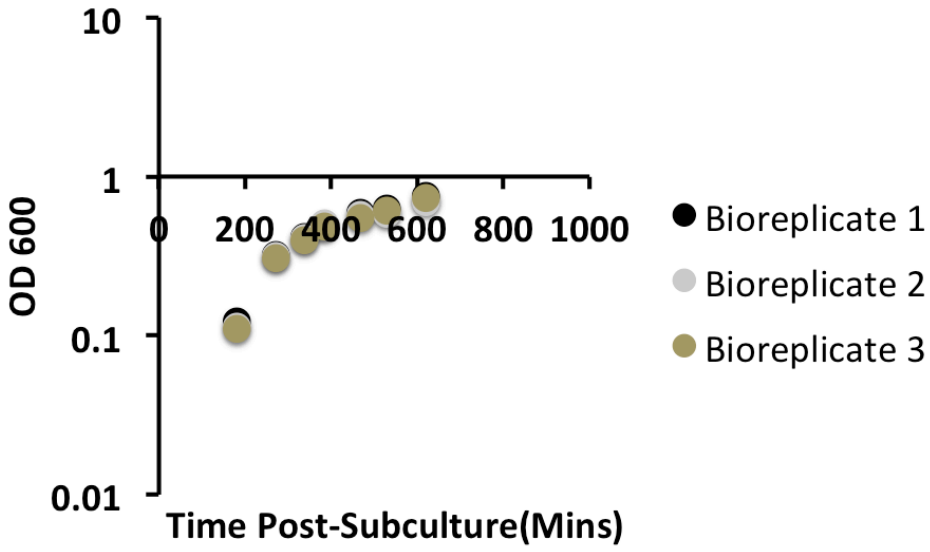
In general, most data in biology are qualitative or relatively quantitative. Similarly, transcriptomics has focused on relative expression changes between different conditions. Therefore, little is known about actual cellular numbers of RNAs and how gene regulation affects these numbers, but ultimately many biological processes will only be understood if investigated with absolute quantitative data. In addition quantitative data allow for comparison of data sets,

current and future[73]. As such, we aimed to measure average cellular toxin and antitoxin RNA copy numbers in a population of cells at mid-log phase ($OD_{600} \approx 0.5$) in balanced growth at 37 degrees in LB media. These absolute quantifications required additional levels of optimization as discussed in the results section. Our studies estimated an average of 0.3 IbsC mRNA and 0.9 SibC sRNA copy numbers per individual cells of the population. These calculations (i.e. <1 copy per cell) suggested tight repression of the IbsC toxin and SibC antitoxin genes and a 3:1 ratio of the antitoxin to toxin, consistent with our previous observations (Figure 4.8 and Supplementary Figure 4.4d at OD_{600} of 0.5). Our estimates of <1 RNA copy per cells are also consistent with the few attempts in the literature, as explained in the results section, at determining cellular numbers for RNAs in yeast and bacteria[73,74]. We also estimated average cellular RNA numbers for the toxin and antitoxin in cells at different phases of growth (**Supplementary Figure 4.6**). The numbers were all measured below 1 copy per cell and the pattern of expression agreed with the relative abundance data during the different growth phases presented in Figure 4.8.

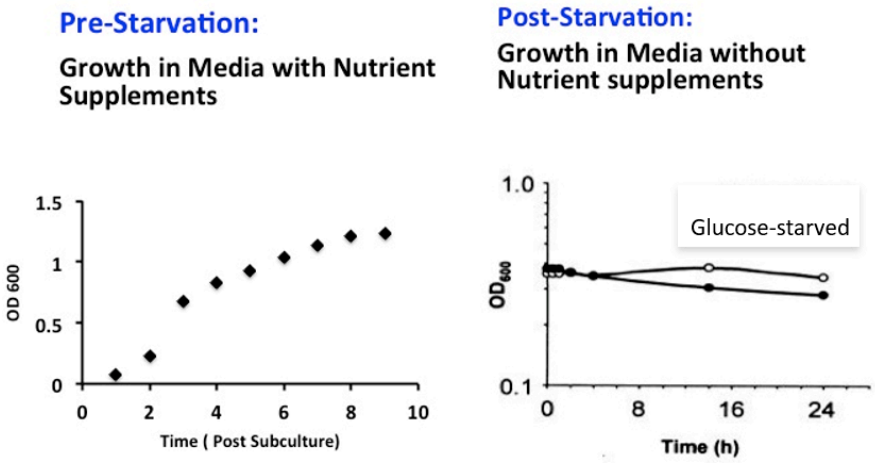
In conclusion, we believe that the knowledge of the physiological levels of these molecules in *E. coli* gained throughout our studies presented here, control over their production inside cells and the reliable method we developed to track the changes in their expression levels, together will empower us to establish links between their expression levels and cellular behavior under different conditions (environmental, niche, stress, etc.). Such links would provide hints about the biological functions of these molecules. This is important from the basic research point of view. In addition, the future of many proposed toxin-antitoxin based applications (e.g. antiviral, anticancer, antibacterial therapies, containment of GMOs previously described in chapter 1) also, would require controlled induction of these toxin-antitoxin molecules for controlled cytotoxic or cytoprotective action by them in undesired or desired cells respectively, and knowledge of their

accurate dosages. Our studies here pave the way for future applications.

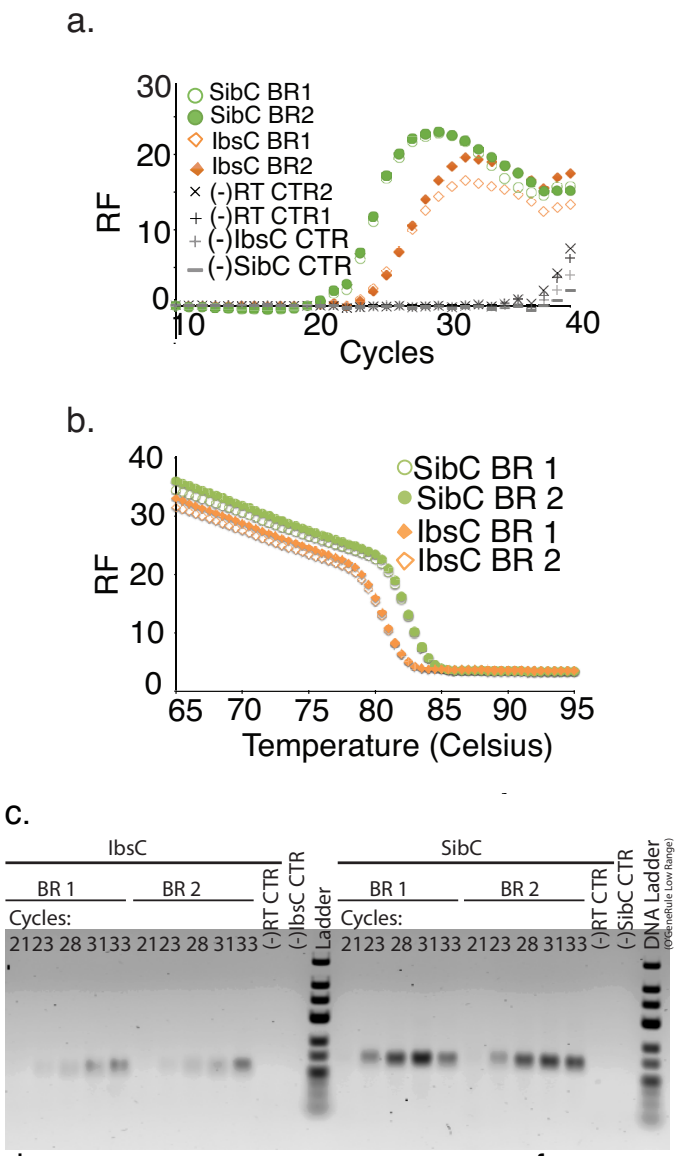
4.6 Supporting figures



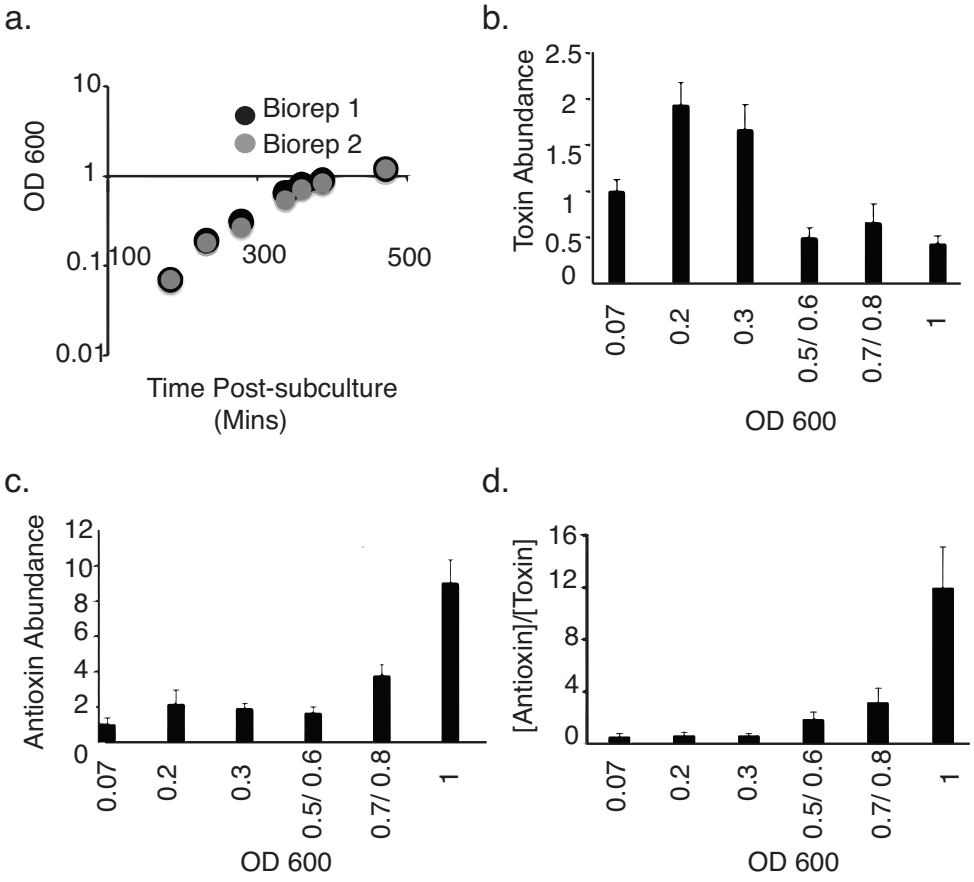
Supplementary Figure 4.1. Growth curves for a slowly-growing population of *E. coli* MG1655 cells from an old culture plate. We arrived upon these cells serendipitously and performed IbsC and SibC RNA expression analyses on them.



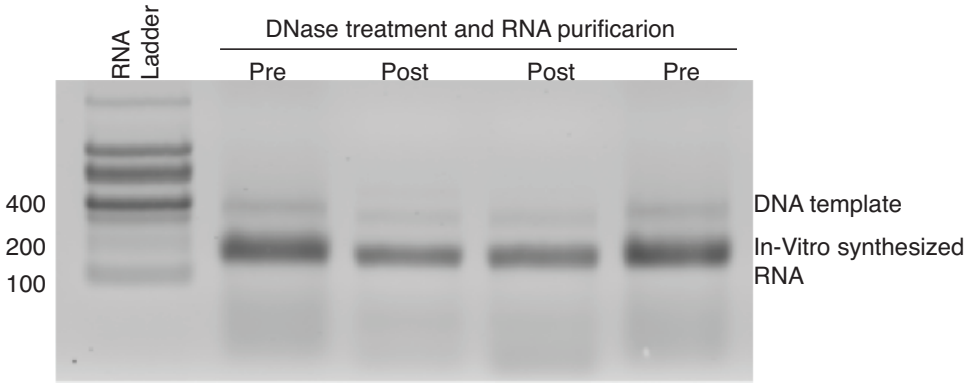
Supplementary Figure 4.2. Creating glucose-starved cells. The cells were grown in M63 media to OD_{600} of about 0.3. They were subsequently divided into two batches. L. One bactch was left in the same media as before and continued normal growth in the presence of nutrient supplements**.** Right. The second bactch was moved to an M63 media without glucose and showed halted growth post treatment. No growth in these cells indicated starvation.



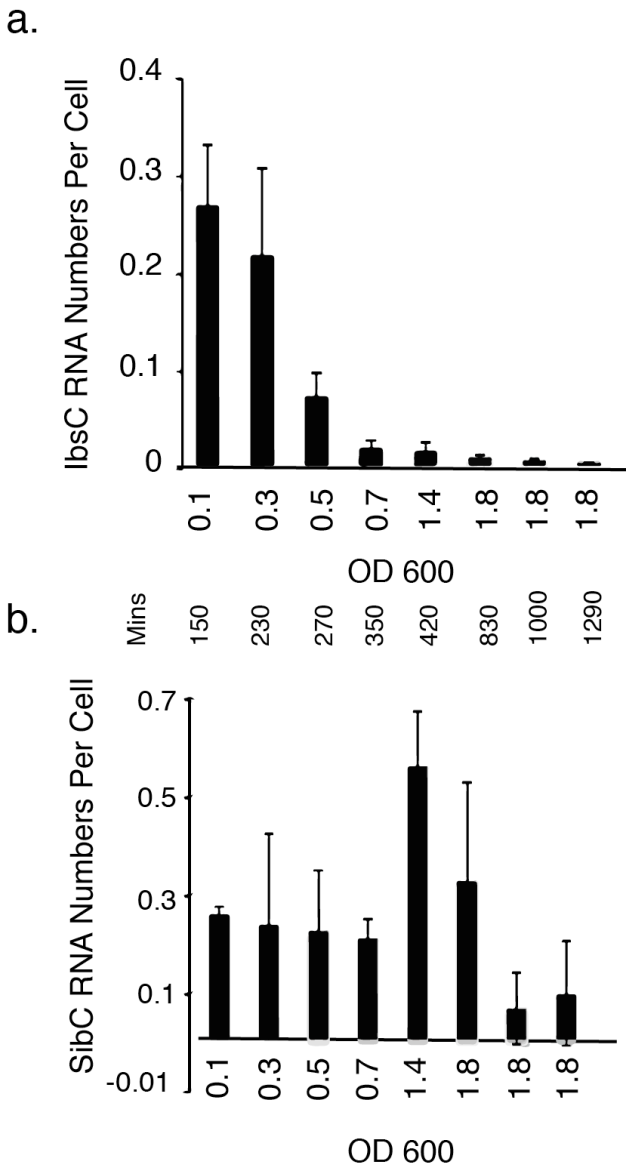
Supplementary Figure 4.3. Specific detection of IbsC and SibC RNAs in wild type *E. coli* cells. a. The optimized RT-qPCR approach was applied to RNA extracts from two cell cultures originated from two colonies of *E. coli*. Amplification curves are shown. The threshold cycle (Ct) values acquired in these biological samples relative to the much higher values in the control samples (i.e. No Reverse Transcription controls (NRTs) to control for genomic DNA and No Template Controls (NTCs) to control for primer dimers and external DNA contamination) pointed to the presence of target RNAs in the biological samples. **b.** Melt curve analyses of end PCR products (IbsC and SibC targets) also confirmed the specificity of the amplification reactions. A single dip in fluorescence at a single melting temperature confirms homogeneity of the reaction products. **c.** Gel visualization of IbsC and SibC amplification products from independent reactions from the 2 biological replicates together with the negative controls also showed specificity of the approach.



Supplementary Figure 4.4. Growth phase dependency of IbsC toxin and SibC antitoxin RNA expressions. a. Growth curves for two independent colonies that were cultured for 8 hours in LB media. b. Toxin RNA abundance at multiple time points representative of different growth phases. Our optimized RT-qPCR based approach was used as previously described to analyze gene expression. The results were normalized against cell number experimentally by taking equal numbers of cells (confirmed by plating and CFU counting) from different growth phases for analysis, and against total RNA mathematically and reported here relative to one another. c. Antitoxin RNA abundance at different growth phases. d. Antitoxin to toxin ratios during different growth phases. The data are averages of two biological replicates and two RT-qPCR technical replicates.



Supplementary Figure 4.5. RNA recovery as measured by gel electrophoresis. The sample containing 200 ng/microliter of RNA in **Figure 4.10** d was run on an agarose gel pre and post purification using our RNA isolation method. The difference in band intensities measured loss of about 30% RNA, which matched that in **Figure 4.10**d, measured by spectrophotometry.



Supplementary Figure 4.6. Average cellular (a) IbsC toxin and (b) SibC antitoxin RNA copy numbers in an average *E. coli* cell population, at different phases of growth, in balanced growth at 37 degrees in LB media. These numbers are averages of two biological replicates, each subjected to RT-qPCR twice. The numbers were consistently measured below 1 copy per cell and the patterns of expression agreed with our relative RNA abundance data presented in Figure 4.8 and Supplementary Figure 4.4 for *ibsC* and *sibC* transcription during different phases of growth.

Primer	Sequence
IbsC RT-q-PCR forward primer	AGCGAGGGCTTGAAGGAG
IbsC RT-q-PCR reverse primer	TGAGACTGACTGTTAATAAGCGCT
SibC RT-q-PCR forward primer	ATTGCTCCTCCCCTGAGACTGA
SibC RT-q-PCR reverse primer	CTCACCGAAGCGAGGGCTTGAAGGAG
IbsC Deep Sequencing forward primer (IbsC-F-MS)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAG CGA GGG CTT GAA GGA G
IbsC deep Sequencing Reverse primer (IbsC-R-61)	CAA GCA GAA GAC GGC ATA CGA GAT CAT CTT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
IbsC-R-62	CAA GCA GAA GAC GGC ATA CGA GAT CAC ATA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
IbsC-R-63	CAA GCA GAA GAC GGC ATA CGA GAT CCA ATT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
IbsC-R-64	CAA GCA GAA GAC GGC ATA CGA GAT CGA TTA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
IbsC-R-65	CAA GCA GAA GAC GGC ATA CGA GAT GTT AGT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
IbsC-R-75	CAA GCA GAA GAC GGC ATA CGA GAT ACA CGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
IbsC-R-76	CAA GCA GAA GAC GGC ATA CGA GAT AGG TTC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
IbsC-R-77	CAA GCA GAA GAC GGC ATA CGA GAT CAT GAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
IbsC-R-78	CAA GCA GAA GAC GGC ATA CGA GAT GCT ATC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG AGA CTG ACT GTT AAT AAG CGC T
SibC Deep Sequencing forward primer (SibC-F-MS)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAT TGC TCC TCC CCT GAG ACT GA
SibC deep Sequencing Reverse primer (SibC-R-61)	CAA GCA GAA GAC GGC ATA CGA GAT CAT CTT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT CAC CGA AGC GAG GGC TTG AAG GAG
SibC-R-62	CAA GCA GAA GAC GGC ATA CGA GAT CAC ATA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT CAC CGA AGC GAG GGC TTG AAG GAG
SibC-R-63	CAA GCA GAA GAC GGC ATA CGA GAT CCA ATT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT CAC CGA AGC GAG GGC TTG AAG GAG
SibC-R-75	CAA GCA GAA GAC GGC ATA CGA GAT ACA CGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT CAC CGA AGC GAG GGC TTG AAG GAG
SibC-R-76	CAA GCA GAA GAC GGC ATA CGA GAT AGG TTC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT CAC CGA AGC GAG GGC TTG AAG GAG
SibC-R-77	CAA GCA GAA GAC GGC ATA CGA GAT CAT GAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT CAC CGA AGC GAG GGC TTG AAG GAG
SibC-R-78	CAA GCA GAA GAC GGC ATA CGA GAT GCT ATC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT CAC CGA AGC GAG GGC TTG AAG GAG
IbsC-fwd-RS (RNA Synthesis in vitro)	**GAAAT***TAATACGACTCACTATAGGG*AGCGAGGGCTTGAAGGAG
SibCfwd-RS (RNA Synthesis in vitro)	**GAAAT***TAATACGACTCACTATAGGG*ATTGCTCCTCCCCTGAGACTG A

Supplementary Table 4.1. Primers used in this study. RT-qPCR primers, deep sequencing primers consisting of an adaptor region, an index sequence (coloured here) as well as the target binding sequence (underlined sequences), and primers used to synthesize the DNA templates for in vitro transcription of RNA, consisting of a five nucleotide sequence (in bold) upstream of the T7 polymerase promoter sequence (in Italics), and the target binding sequence (underlined).

Sample	Target	Forward Deep Sequencing Primer	Reverse Deep Sequencing Primer	Concentration
1	IbsC	IbsC-F-MS	IbsC-R-61	50000000
2	IbsC	IbsC-F-MS	IbsC-R-62	1000000
3	IbsC	IbsC-F-MS	IbsC-R-63	1000000
4	IbsC	IbsC-F-MS	IbsC-R-64	100000000
5	IbsC	IbsC-F-MS	IbsC-R-65	20000000
6	SibC	SibC-F-MS	SibC-R-61	500000
7	SibC	SibC-F-MS	SibC-R-62	50000000
8	SibC	SibC-F-MS	SibC-R-63	1000000

Supplementary Table 4.2. Standards to cross-validate qPCR and deep sequencing approaches forquantification of nucleic acids.

5 Chapter 5: Examining Expression From Homologous *ibs/sib* Genes

5.1 Introduction

In chapter 4, we extensively analyzed expression from the *ibsC/sibC* locus. While the antitoxin expression was previously detected, toxin RNAs were never detected in wild type cells. There were ambiguities in literature as to whether this toxin gene was ever expressed and if yes, under what circumstances. For the first time, we reported expression from the toxin locus under natural physiological conditions and normal growth in wild type *E. coli* cells. Although transcribed, our experimental evidence suggested that expressions from these genes were under tight and multiple layers of regulation at the DNA, RNA and perhaps protein levels. We estimated absolute cellular copies of the toxin and antitoxin RNA to be <1, which further supported the notion about their tight regulation.

By analyzing expression under different phases of natural growth and stressed conditions, we were able to probe the genes' expression profiles. Nutrient-deprivation was the only tested condition during the course of our studies where we detected the toxin RNA levels to be higher relative to the antitoxin (data are presented in chapter 4). This supported the long-lived and most prevalent hypothesized role of toxins in combating stress.[5,12,20,41,54,131] Interestingly, we found that under normal physiological conditions, the toxin was highly expressed during the log phase of growth when the cells were rapidly growing and repressed during the stationary phase. In addition, the antitoxin showed an opposite trend of expression, in other words, less expression in the log phase than in stationary, hence causing less repression of the toxin during the log phase. These expression patterns support a yet-unknown role of the toxin during growth. These findings are not contradictory as the toxin may play multiple roles according to cellular

demands.

One of the means by which the cells might achieve different roles of the toxin, is by fine-tuning their expression level. Through our own work and that of others [78,117], we have learned that the currently known effects of the toxin ranging from membrane damage to reversible pause in growth (cytostatic) to irreversible growth stasis (cytotoxic, i.e. death) are dependent on its RNA and protein levels. Cellular concentrations of these molecules are regulated at multiple levels through promoter strength, RNA stability, rates of protein synthesis and degradability. Another layer of regulation might be through the existence of multiple homologous genomic copies to fine tune expression of these molecules, assuming all homologous copies function redundantly. [78]

Multiple copies of the *ibs* and *sib* genes exist within the genome, at multiple locations of multiple bacterial species. [72,78] The genetic organization of the *ibs-sib* loci in *E. coli* MG1655 might have implications for their gene expressions, and consequently functions. The *ibs-sib* gene pairs are oriented such that the *sib* antitoxin gene is directly antisense to the coding region of the *ibs* toxin gene. All toxins are extremely hydrophobic and their over productions have been shown to cause cell stasis and at higher levels death. The Sib RNAs have also been shown to repress the toxicity of the Ibs proteins. Furthermore the repression has been shown to be specific. [78,117] This has confirmed that despite the high degree of similarity between the multiple *sib* genes, there is specificity in regulation. Han et al. later showed that the specificity was achieved through two highly variable regions. [124]

"What Is the Benefit to *Escherichia coli* of Having Multiple Toxin-Antitoxin Systems in Its Genome?" [141] What counteracts the cost of maintaining and specifically regulating each gene individually? In the case of Ibs-Sib, it is still unknown. As previously mentioned, it has been difficult to detect Ibs mRNA

expression, and the combined data suggest that the toxin is expressed at very low levels if at all. [78] However, their ribosome binding sites are nearly canonical with little, if any, degeneration. The coding sequences are also very well conserved across species. This suggests that the sequences are maintained for a purpose; but their expression is low. There are several hypotheses in the literature as to why multiple copies of these genes exist inside bacteria.[78]

One hypothesis suggests that each Ibs protein has a different function. However, due to their extensive similarity, small size and tolerance to multiple amino acid mutations, they likely have the same target, but may interact with the target differently. [78,90]

Another hypothesis suggests that the multiple *ibs* genes are maintained to allow the bacteria to fine tune Ibs production by varying levels of expression from each gene. Examining expression from each *ibs* gene (i.e. amount and rate of production under different conditions) and the parameters about each gene that influence its production (e.g. promoters, mRNA stability, translational control, etc.) will validate this hypothesis. [78]

Lastly, the multiple *ibs* genes may exist to ensure sufficient total Ibs protein production while keeping expression from each gene very low to avoid toxicity. This hypothesis assumes redundant function by these proteins and demand by the cells to have the total Ibs protein levels at a certain level. [78]

To begin validating some of these hypotheses, endogenous levels of expression from the chromosomal copies of these genes were measured and examined for how they compared with one another. Are they all expressed? Are they expressed equivalently or non- equivalently?

Most studies of Ibs protein function have relied upon overexpression from multicopy plasmids. [79,117,124] To better understand how endogenous levels of Ibs

affect *E. coli*, it is important to measure expression from these genes in their native biological context. As previously shown for *ibsC* and *sibC*, knowledge of endogenous expression levels and how they change under different physiological and environmental conditions, reveal insights about the true function of these genes.

As with previous *sibC* and *ibsC*, at the time we began our studies in this chapter; knowledge of expression from the other homologous loci was limited. The antitoxins had been detected under all tested conditions and shown to have very strong promoter elements. These observations had led to the notion that they were likely transcribed constitutively.[78,117] However, Ibs mRNAs had not been detected from the chromosome of a wild type *E. coli* strain under natural physiological conditions and normal growth. They had only been detected upon either deletion of the cognate Sib promoter (i.e. in the absence of the cognate antisense sRNA to cause repression) or upon deletion of the *rnc* gene (i.e. in the absence of the ribonuclease, RNase III to cause toxin-antitoxin complex cleavage) (Fozo E, unpublished observations) [78,79]. These observations collectively suggested that in a wild type strain, Sib RNA would be expressed to form an RNA duplex with the Ibs toxin mRNA, which would be subsequently cleaved by RNase III, to keep Ibs mRNA levels low.

All the experimental evidence points to the tight regulation of the *ibs* expression. But, are these genes ever expressed? In this project, we utilized the optimized Reverse Transcription-quantitative PCR (RT-qPCR) approach we previously developed (presented in chapter 4) to examine endogenous expression from these loci in their native chromosomal context and under different growth stages. In addition, we have designed probes and developed protocols to facilitate the use of high throughput techniques such as NanoString and next-generation sequencing for more in-depth and quantitative analyses of RNA expression from these genes in the future under a multitude of different conditions.

5.2 Materials and Methods

Materials and protocols for the RT-qPCR approach utilized in this chapter have been previously described in Chapter 4. Any modifications will be specified in the results section.

5.3 Results

5.3.1 Design of RT-qPCR primers

The design of primers for *ibs* and *sib* genes was not straightforward given the short length of these transcripts (about 150 nucleotides) and high similarity among them (up to >80%).

RT-qPCR primers were designed to target regions of most dissimilarity between the target and the rest of the four homologous copies. These regions were determined for each target by performing pairwise alignments. The primer sequences are shown in **Supplementary Table 5.1**. For some targets multiple primer sets were designed and validated.

5.3.2 Validating specific detection and amplification by the primers through gel electrophoresis, sequencing of PCR reaction products, and melt curve analysis

Considering the great homology between the different Ibs and Sib targets, we had to validate that each primer set only amplified its respective target and did not show cross reactivity against the other homologous targets. To test this, we performed PCR reactions using the designed primer sets and the genomic DNA (gDNA) as template. The reaction products were run on an agarose gel, shown in Figure 5.1, purified and sent for Sanger sequencing. Later, the end products of RT-qPCR reactions on biological cDNA were subjected to melt curve analyses to further confirm specificity of the reactions.

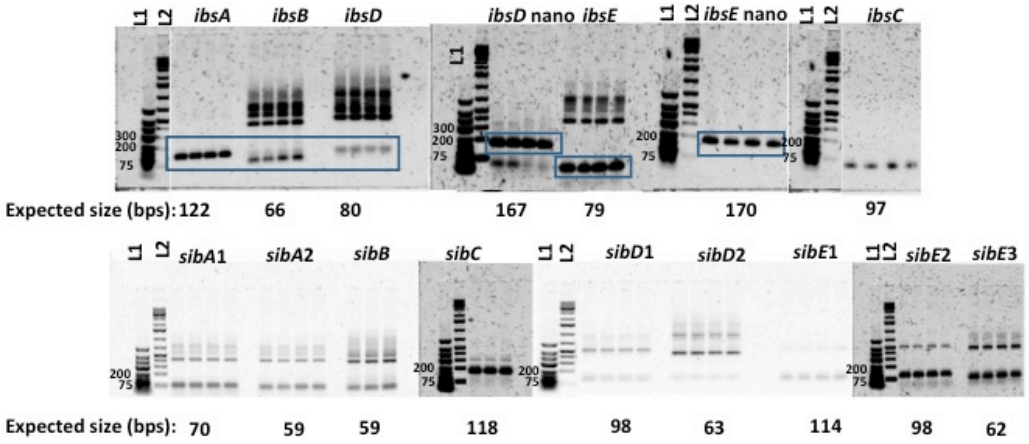


Figure 5.1. Gel electrophoresis visualization and purification of end reaction products of PCR with individual Ibs or Sib primer sets using the gDNA as template. Bands corresponding to the right size for the desired targets existed in all cases, pointing to a) successful amplification by these individual primer sets b) specific amplification by the designed primers to discern among homologous copies of slightly different lengths. For further confirmation, these bands were excised and sent for Sanger sequencing. L1= Gene ruler LowRange DNA L2=Thermo 1kb plus ladder. Numbers refer to multiple forward primers designed for some targets. ibsD and E nano refer to amplifications with forward 1 primers for ibsD and E respectively, that target regions farther upstream of what we know is transcribed. These primer sets were validated to later be used for assessment of whether these regions were transcribed or not for the purpose of designing Nano string probes.

It is worth mentioning that the appearance of multiple bands on some of the gels above was expected and not a concern. The gDNA is a complex template, and any primer set may target multiple regions within the gDNA. However, these primers

will eventually be used on specific cDNA templates from biological samples. Therefore, we were only interested in the band corresponding to the right size for each expected target, to purely contain that target as opposed to any of the homologous copies. As shown in the gel electrophoresis Figure 5.1 the differences in size between the homologous copies appeared on the gel, which pointed to their specific amplification. For further confirmation of specificity, the reaction products were excised and sent for Sanger sequencing. The results for reaction products that were long enough to give a sequencing read (about 100 base pairs (bps), namely for IbsA: 122bps, IbsC: 97, IbsD nano: 167, IbsE: 79, IbsE nano: 170, sibC 118, sibD1: 98bps, SibE1: 114, SibE2: 98) are summarized in Supplementary Figure 5.1-5.7.

The Sanger reads for purified reaction products from each PCR reaction with unique primer sets were aligned against the sequence for the expected target and also against the sequences for other homologous copies. In all cases, we found 100% alignment with the expected target sequence and misalignments with the other homologous copies. The results from alignment of IbsC and SibC Sanger reads against all the other copies are shown in Figure 5.2 for IbsC and Figure 5.3 for SibC as examples. The rest can be found in Supplementary Figure 5.1-5.7.

The results indicated that the primers could successfully distinguish between the targets of interest and other homologous copies.

Against *ibsC* (100.0%)

```
100.0% identity in 85 residues overlap; Score: 476.0; Gap frequency: 0.0%

UserSeq1        1 AGCGAGGGCTTGAAGGAGAAGGGTTATGATGCGACTTGTCATCATACTGATTGTACTGTT
UserSeq2       26 AGCGAGGGCTTGAAGGAGAAGGGTTATGATGCGACTTGTCATCATACTGATTGTACTGTT
                  ************************************************************

UserSeq1       61 ACTCATAAGTTTCAGCGCTTATTAA
UserSeq2       86 ACTCATAAGTTTCAGCGCTTATTAA
                  *************************
```

Against *ibsA* (69.0%)

```
69.0% identity in 87 residues overlap; Score: 265.0; Gap frequency: 4.6%

UserSeq1        4 GAGGGCTT---GAAGGAGAAGGGTTATGATGCGACTTGTCATCATACTGATTGTACTGTT
UserSeq2       96 GAGGGCTTTTAATAAGGAAAGGGTTATGATGAAGCACGTCATCATACTGGTGATACTCTT
                  ********     * *  *************   *  ************ *  **** **

UserSeq1       61 ACTCATAAGTTTAGCG-CTTATTAACA
UserSeq2      156 AGTGATTAGCTTCCAGGCTTACTAAGA
                  * * ** ** **   * **** *** *
```

Against *ibsB* (65.6%)

```
65.6% identity in 93 residues overlap; Score: 238.0; Gap frequency: 8.6%

UserSeq1        4 GAGGGCTTGAAGGAG--AAGGGTTATGATGCGACTTGTCATCATACTGATTGTACTGTTA
UserSeq2       95 GGGGCCATAAATAAGGAAAGGGTCATGATGAAGCTACTCATCATCGTG---GTGCTCTTA
                  * ** * * **  **  ****** ******   **  *******  **   ** ** ***

UserSeq1       62 CTCATAAGTTTAGC-GCTTATTAA--CAGTCAG
UserSeq2      152 GTCATAAGCTTCCCCGCTTACTAAGACTACCAG
                   ******* **  * ***** ***  *   ***
```

Against *ibsD* (79.2%)

```
79.2% identity in 96 residues overlap; Score: 382.0; Gap frequency: 3.1%

UserSeq1        1 AGCGAGGGCTTGAAGGAG--AAGGGTTATGATGCGACTTGTCATCATACTGATTGTACTG
UserSeq2       34 AAGGGGGCCTTACAGAAGGAAAGGGTTATGATGAAGCTCGTCATCATACTGATTGTGTTG
                  *  * ** ***  ** **  *************   ** *****************  **

UserSeq1       59 TTACTCATAAGTTTAGC-GCTTATTAACAGTCAGTC
UserSeq2       94 TTACTCGTAAGTTTCGCAGCTTATTAACAGCCAATC
                  ****** ******* ** ************ ** **
```

Against *ibsE* (73.1%)

```
73.1% identity in 93 residues overlap; Score: 328.0; Gap frequency: 3.2%

UserSeq1        4 GAGGGCTTGAAGGAG--AAGGGTTATGATGCGACTTGTCATCATACTGATTGTACTGTTA
UserSeq2       37 GGGGCCTTGTATAAGGAAAGGGTTATGATGAAGCTCGTCATCATACTGGTTGTGTTGTTA
                  * ** **** *  **  **************   ** ************ ****  *****
```

Figure 5.2. Alignment of Sanger sequencing reads for the end reaction products of PCR on genomic DNA with IbsC primers. Results indicate pure composition of the sample and specific amplification by the designed primers.

Against *sibC* (100.0% and 100.0% for paired end reads)

```
100.0% identity in 51 residues overlap; Score: 321.0; Gap frequency: 0.0%

UserSeq1        1 GTATGATGACAAGTCGCATCATAACCCTTCTCCTTCAAGCCCTCGCTTCGG
UserSeq2       78 GTATGATGACAAGTCGCATCATAACCCTTCTCCTTCAAGCCCTCGCTTCGG
                  ***************************************************

100.0% identity in 37 residues overlap; Score: 228.0; Gap frequency: 0.0%

UserSeq1        1 ATTGCTCCTCCCCTGAGACTGACTGTTAATAAGCGCT
UserSeq2       15 ATTGCTCCTCCCCTGAGACTGACTGTTAATAAGCGCT
                  *************************************
```

Against *sibA* (75% and 63.6% for paired end reads)

```
75.0% identity in 56 residues overlap; Score: 197.0; Gap frequency: 10.7%

UserSeq1        1 GTATGATGACAAGTCGCATCATAACCCTTCTCCTTC----AAGCCCTCG-CTTCGG
UserSeq2       77 GTATGATGACGTGCTTCATCATAACCCTT-TCCTTATTAAAAGCCCTCTTCTCCGG
                  **********  *   ************ *****     ********  ** ***

63.6% identity in 33 residues overlap; Score: 99.0; Gap frequency: 9.1%

UserSeq1        2 TTGCTCCTCCCCTGAGACTGACTGTTAATAAGC
UserSeq2       18 TTCCCCCTCCCCCTGGTGT---TCTTAGTAAGC
                  ** * *******   *  *   * *** *****
```

Against *sibB* (63.5% and 61.5% for paired end reads)

```
63.5% identity in 52 residues overlap; Score: 169.0; Gap frequency: 5.8%

UserSeq1        3 ATGATGACAAGTCGCATCATAACCCTTCTCCT---TCAAGCCCTCGCTTCGG
UserSeq2       73 ATGATGAGTAGCTTCATCATGACCCTTTCCTTATTTATGGCCCCTTCCTCGG
                  *******  **   ****** ******  * *   *   ****   * ****


61.5% identity in 26 residues overlap; Score: 93.0; Gap frequency: 0.0%

UserSeq1        1 ATTGCTCCTCCCCTGAGACTGACTGT
UserSeq2      109 ATGGCCCCTTCCTCGGGAGGGGCTTT
                  ** ** *** **  * **  * ** *
```

Against *sibD* (76.8% and 82.4% for paired end reads)

```
76.8% identity in 56 residues overlap; Score: 206.0; Gap frequency: 10.7%

UserSeq1        1 GTATGATGACAAGTCGCATCATAACCCTTCTCCTTC----AAGCCCTCG-CTTCGG
UserSeq2       79 GTATGATGACGAGCTTCATCATAACCCTT-TCCTTCTGTAAGGCCCCCTTCTTCGG
                  ********** **   ************ ******    * **** *  ******

82.4% identity in 34 residues overlap; Score: 152.0; Gap frequency: 2.9%

UserSeq1        1 ATTGCTCCTCCCCTGAGACTGACTGTTAATAAGC
UserSeq2       17 ATTTCTCC-CCCCTCTGATTGGCTGTTAATAAGC
                  *** **** *****  ** ** ************
```

Against *sibE* (75.9% and 69.7% for paired end reads)

```
75.9% identity in 54 residues overlap; Score: 209.0; Gap frequency: 5.6%

UserSeq1        1 GTATGATGACAAGTCGCATCATAACCCTT--CTCCTTCAAG-CCCTCGCTTCGG
UserSeq2       79 GTATGATGACGAGCTTCATCATAACCCTTTCCTTATACAAGGCCCCTTCTTCGG
                  ********** **   *************  **  * **** ***   ******

69.7% identity in 33 residues overlap; Score: 116.0; Gap frequency: 3.0%

UserSeq1        1 ATTGCTCCTCCCCTGAGACTGACTGTTAATAAG
UserSeq2       17 ATTTCTCC-CCCCTCTGATGAGTTGTTAGTAAG
                  *** **** *****  **     ***** ****
```

Figure 5.3. Alignment of Sanger sequencing reads for the end reaction products of PCR on genomic DNA with SibC primers. Since the forward and reverse reads did not overlap, we aligned them individually against all the other copies. Results indicate pure composition of the sample and specific amplification by the designed primers.

5.3.3 ibsA-E gene expression analysis under two different growth stages (i.e. early log (OD_{600} ~ 0.2) and late log (OD_{600} ~ 0.8))

In chapter 4, we extensively analyzed expression from the *ibsC* toxin gene and for the first time reported detection of IbsC RNA transcripts under different conditions. We were next interested to know whether the other homologous copies were also expressed in wild type *E. coli* cells under physiological conditions, and how their RNA expressions compared with that of *ibsC.*

To analyze expression from the *ibsA, B, D* and *E* toxin genes, similar to that described for *ibsC* in Chapter 4, we performed RT-qPCR experiments on biological samples using the validated PCR primers for each target. Through our studies presented in the previous chapter, we learned that the IbsC-toxin RNA expression was the highest at the early to the mid-log phase (up to OD_{600} of about 0.5) and was switched off during the mid-log into the stationary phase. Therefore, to begin studying expression from the other homologous copies, we chose OD_{600} of 0.2 and 0.8 during which the *ibsC* gene was differentially expressed. Studying expression of the other 4 copies during these two growth stages (early and late log) would allow us to compare the expression behavior during the different growth phases by these homologous copies.

The amplification curves for the different *ibs* copies are presented in Figure 5.4 (i.e. *ibsA, B, D* and *E* from top to bottom). The left panels present the data from cells grown to OD_{600} of 0.2 and the right panels present the data from the same cell cultures grown to OD_{600} of 0.8.

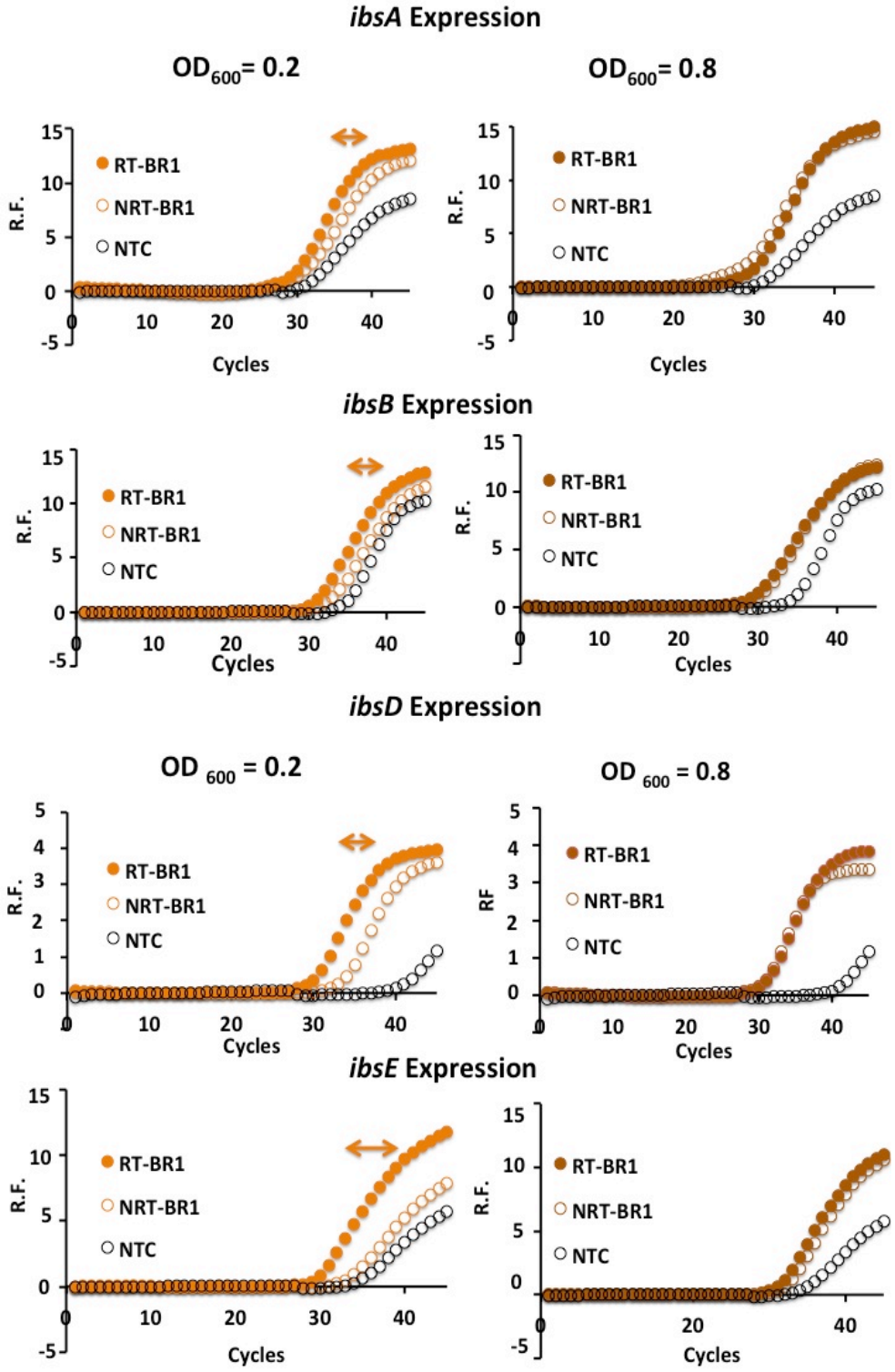

ibsA Expression
OD600= 0.2
OD600= 0.8
RT-BR1
NRT-BR1
NTC
R.F.
Cycles
ibsB Expression
ibsD Expression
OD 600 = 0.2
OD 600 = 0.8
RF
ibsE Expression

Figure 5.4. Gene expression analyses from the *ibsA*, *B*, *D* and *E* genes. (top to bottom). Left: cultures grown to OD_{600} of 0.2 Right: cultures grown to OD_{600} of 0.8. BR1: Biological Replicate 1; RT: Short for reverse transcription and signifies amplification curves generated by PCR on target-cDNA samples obtained by RT; NRT: No Reverse Transcription control for genomic DNA contamination; NTC: No Template Control for primer-dimer and external DNA contamination. The left shift of the RT curves relative to the controls in early log (OD_{600} = 0.2) samples, for all *ibs* genes indicates expression from these genes. However, the small magnitude of the shift and high Ct values (~30) indicate small expression. Overlap between RT and negative controls in late log (OD_{600} = 0.8) smaples indicates no expression during this phase. RNA expression behaviors for all *ibs* genes are similar.

The expression data presented in Figure 5.4 show that similar to *ibsC* (previously investigated in chapter 4), the other four homologous *ibs* copies were indeed expressed, but only at OD_{600} of 0.2. This is indicated by the left shift of the amplification curves for the RT samples (that contain the target cDNA obtained by reverse transcription) relative to those for the NRT (no reverse transcription to control for genomic DNA contamination) and NTC (to control for primer dimer and external DNA contamination) samples. However the small magnitude of the shift and high Ct values (~30) suggested that RNA levels from each individual locus were maintained at low levels. Expression seemed to turn off at OD_{600} of 0.8 as indicated by the overlap between the RT and the negative controls. Therefore, all 5 *ibs* gene copies seemed to be expressed equivalently, at least under the natural physiological conditions tested here.

These results were reproduced for another biological replicate on a different run. They are summarized in the Supplementary Figure 5.8.

Melt curves for these reactions are presented in **Supplementary Figure 5.9** and confirm specificity of detection.

5.3.4 sibA-E gene expression analysis under two different growth stages (i.e. early log (OD_{600} ~ 0.2) and late log (OD_{600} ~ 0.8))

We investigated expression from the *sib A, B, D* and *E* genes under the aforementioned conditions (i.e. OD_{600} of 0.2 and 0.8). Please see Figure 5.5 , Figure 5.6 and Figure 5.7 below. We detected expression from all copies except for *sibE*. This was indicated by the left shift of the amplifications curves for the test samples relative to the negative controls. For the detected RNA transcripts, our results demonstrated much higher levels of Sib RNAs in the test samples relative to the negative controls, as compared with that observed for the toxins. Together with much lower Ct values observed for Sib targets relative to Ibs targets, these results suggested higher levels of the antitoxin sRNAs relative to the toxin mRNAs. These relative abundance levels of antitoxin to toxin were similar to what we previously reported for IbsC and SibC in chapter 4. Otherwise, and interestingly, *sibA*, *B*, and *D* showed expression patterns that were different from that of *sibC*.

In chapter 4, we showed that *sibC* was more highly transcribed at the late log (e.g. OD_{600} ~ 0.8) phase of growth than in the early log (e.g. OD_{600} ~ 0.2). Here, we observed the opposite trend for both *sibA* in Figure 5.5 and *sibD* in Figure 5.7. In other words, higher RNA transcript levels were observed at OD_{600} of 0.2 relative to 0.8. This was indicated by the shift to the left of the amplification curves for RT-OD 0.2 (light green) relative to RT-OD 0.8 samples (dark green) in all panels. Different panels in each figure represent data from different biological replicates and amplifications with different (previously validated) primer sets for the target under investigation. As shown the results were reproducible among the different biological replicates, amplified with different primer sets for each target.

sibB, presented in Figure 5.6, showed equal expression during the early log and

late log stages of growth, as indicated by the overlap between the RT-OD 0.2 (light green) and RT-OD 0.8 (dark green) curves. Each panel represents results from a different biological replicate. As shown the results were reproducible among the different biological replicates.

Melt curve analyses for the reactions performed here are demonstrated in **Supplementary Figure 5.10** and confirm specific detection of the individual targets.

We believe different expression patterns by different *sib* genes during the different phases of growth may indicate different functions by them. For more explanation please see discussion.

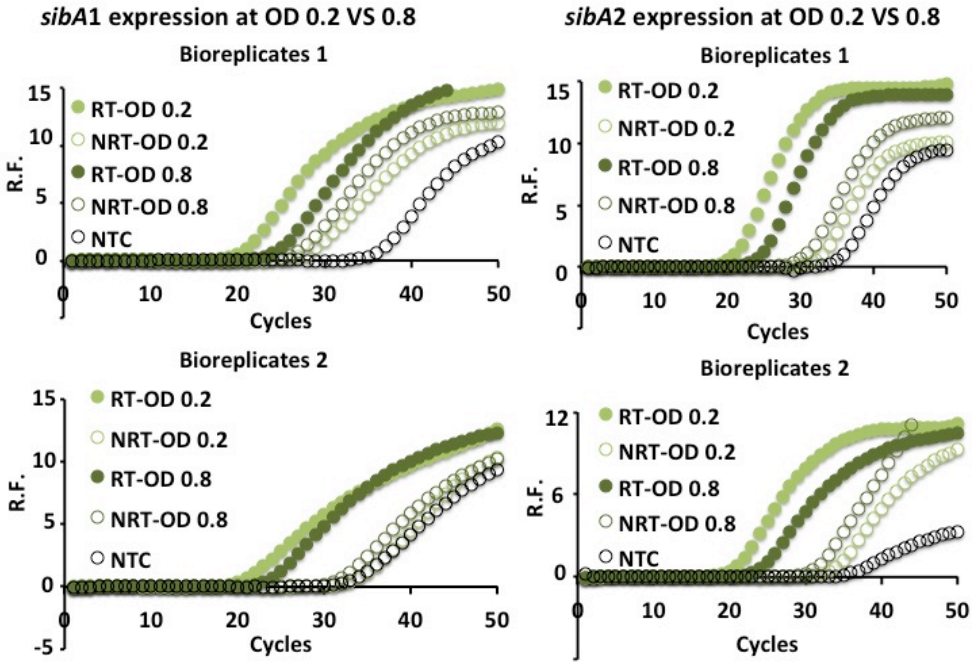


Figure 5.5. Gene expression analyses of the *sibA* gene during the early log (OD_{600} = 0.2) and late log (OD_{600} =0.8) phases of growth. Results for two biological replicates (top and bottom), each amplified with two different sets of primes against *sibA* (Left and Right) are shown. RT: Short for reverse transcription signifies amplification curves generated by PCR on target-cDNA samples obtained by RT; NRT: No Reverse Transcription Control for genomic DNA contamination; NTC: No Template Control for primer-dimer and external DNA contamination. The left shift of RT samples relative to controls indicate presence of the target RNAs. The left shift of the RT-OD 0.2 curves relative to the RT-OD 0.8 curves indicates higher SibA RNA levels at the early log phase relative to the late log phase.

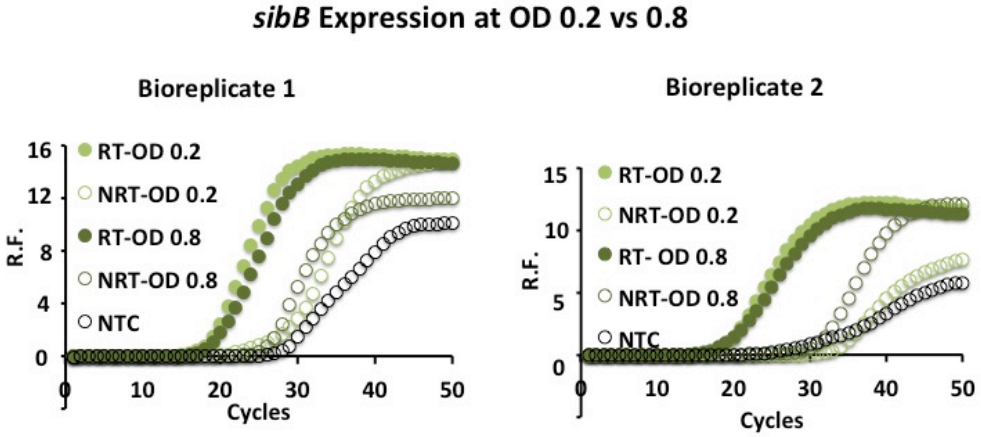


Figure 5.6. Gene expression analyses from the *sibB* gene. during the early log (OD_{600} = 0.2) and late log (OD_{600} =0.8) phases of growth. Results for two biological replicates in Left and Right are shown. RT: Short for reverse transcription signifies amplification curves generated by PCR on target-cDNA samples obtained by RT; NRT: No Reverse Transcription Control for genomic DNA contamination; NTC: No Template Control for primer-dimer and external DNA contamination. The left shift of RT samples relative to negative controls indicate presence of the target RNAs. The overlap between the RT-OD 0.2 and RT-OD 0.8 curves indicate similar expression levels between the two growth stages.

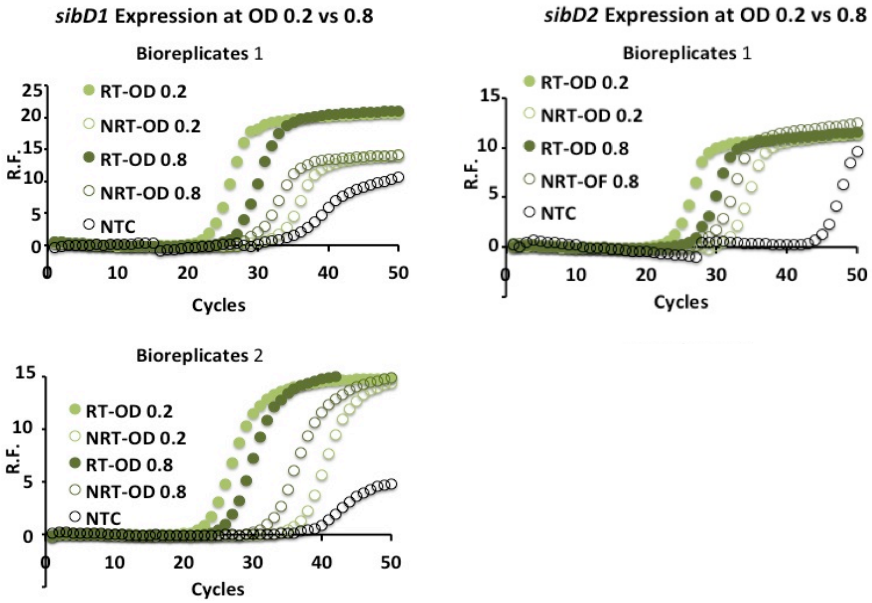


Figure 5.7. Gene expression analyses of the *sibD* gene during the early log (OD_{600} = 0.2) and late log (OD_{600} =0.8) phases of growth. Results for two biological replicates (Top and bottom), are shown. Right panel shows amplification with a different (previously validated) primer set against *sibD* in bioreplicate 1. RT: Short for reverse transcription signifies amplification curves generated by PCR on target-cDNA samples obtained by RT; NRT: No Reverese Transcription Control for genomic DNA contamination; NTC: No Template Control for primer-dimer and external DNA contamination. The left shift of RT samples relative to controls indicate presence of the target RNAs. The left shift of the RT-OD 0.2 curves relative to the RT-OD 0.8 curves indicates higher SibA RNA levels at the early log phase relative to the late log phase.

5.3.5 Design and theoretical validation of Nanostring probes for specific and highthrouput quantification of ibs/sib gene pairs

To truly understand function by the *ibs/sib* gene pairs, we need to study their RNA expression levels under a multitude of conditions. For these investigations, we recommend techniques with high multiplexing power such as nanostring.

We consulted with the team of bioinformaticians at NanoString to design probes to distinguish among the different *ibs* and *sib* transcripts. These are pairs of (A, reporter and B capture) probes for each target, and about 50 base pairs in length that would sit right next to each other against their targets as shown for the Ibs targets in Figure 5.8 and for the Sib targets in Figure 5.9. The highlighted hanging sequences are the capture and reporter portions of the probes. The design of the probes is based on GC content to ensure all probes have similar melt temperatures. Therefore occasionally, we need to slide the probes upstream or downstream by a few nucleotides to find a suitable location. As such, instead of sitting immediately next to one another, probes A and B are within just a few bases between them, in case certain bases need to be avoided. A stretch of 200-300 bps would provide enough space for probe design, within which one can maneuver. However the shorter lengths of our targets were just enough to fit the probes.

Fortunately, these probes met the percent similarity cut offs to avoid cross hybridization. For the Sib set, each probe pair was under the threshold for specificity (~85-86%) against the other sequences. The Ibs probes, although more similar, particularly in the probe B region, had enough mismatches in the probe A region to not cross hybridize. IbsD and IbsE were problematic, since they were more similar than the other three homologous genes (i.e. IbsA, B and C). Therefore, the probes for these genes were designed to target regions within the RNA transcript further upstream at the 5' end relative to the region targeted for other Ibs copies. To confirm that these regions were transcribed, we performed PCR with primers targeting these region (i.e. IbsE and SibE Forward 1 primers) on their respective cDNA pools. In Supplementary Figure 5.8, the results show that this region is indeed transcribed for *ibsD*. More experiments need to be done to determine this for *ibsE* too. If not transcribed, the probes can be designed to target the same region as the other Ibs probes, however, they will be more likely to cross hybridize. All in all, most, if not all of the Ibs and Sib transcripts could be experimentally targeted.

It is important to note that due to the complementarity between the *ibs* and *sib* gene pairs, we need to avoid including probes to both the Ibs and Sib transcripts in the same assay since the probe pairs to each Ibs/Sib pair would be complementary to one another and cause background signal issues. To measure the full set of genes, we would need to run separate assays for the toxins and antitoxins. In addition, for absolute quantification of the RNA transcripts, standards of known quantity for each target transcript need to be included in the assay to factor in differences in hybridization efficiency.

The probe designs are finalized and theoretically validated. Their sequences are presented in **Table 5.1**.

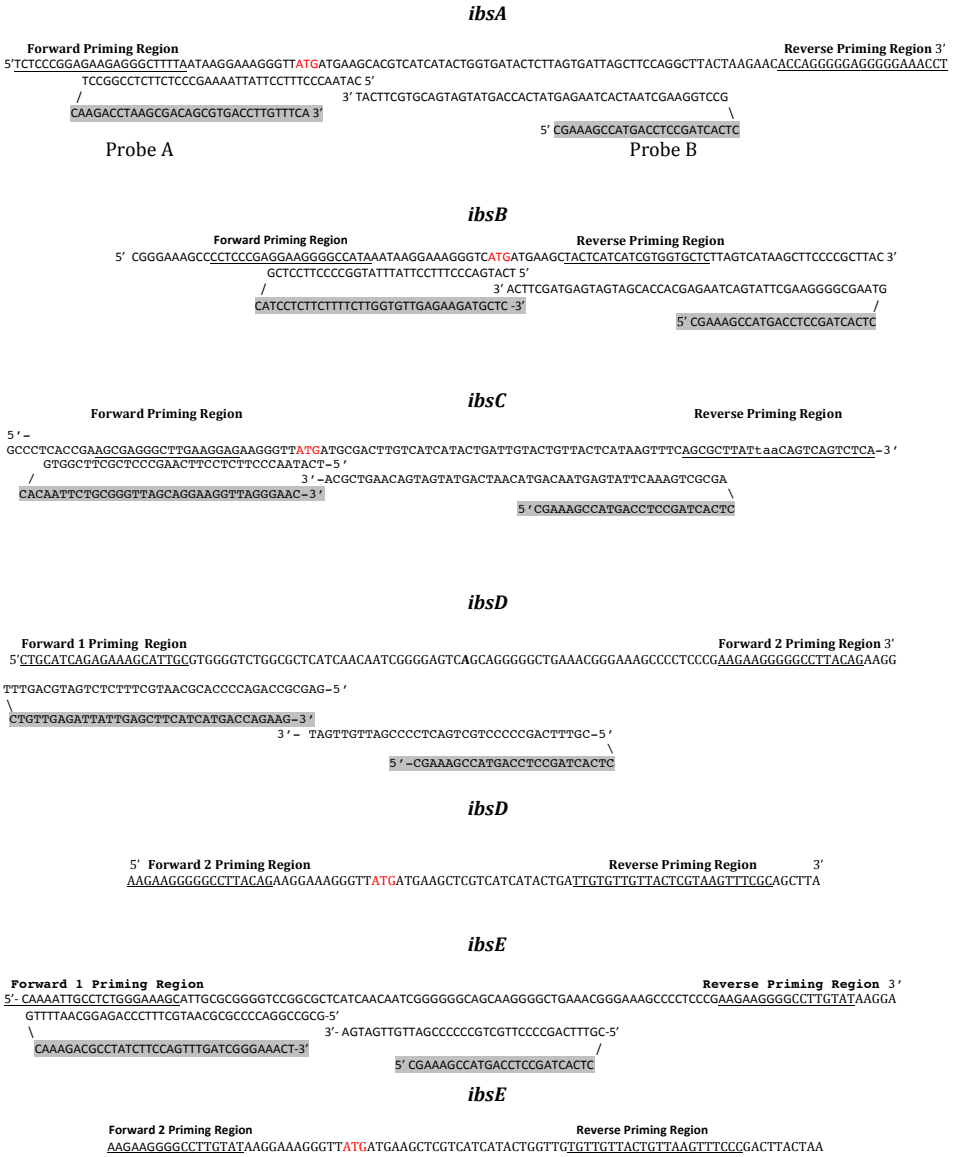


Figure 5.8. Binding of nanostring probes to their respective targets (*ibsA-E*). Probes A (reporter probes) and B (Capture probes) are labeled for *ibsA* as an example. These are a pair of probes for each target, and about 50 base pairs in length that would sit right next to each other or within a few nucleotides from one another against their targets as shown. The highlighted hanging sequences are the reporter and capture portions of the probes A and B respectively. Underlined regions within the target sequences are priming sites for the qPCR primers. These probes meet the percent similarity cut offs to avoid cross hybridization. Both nanostring probes and qPCR primers were designed within regions of least similarity among all five homologous copies.

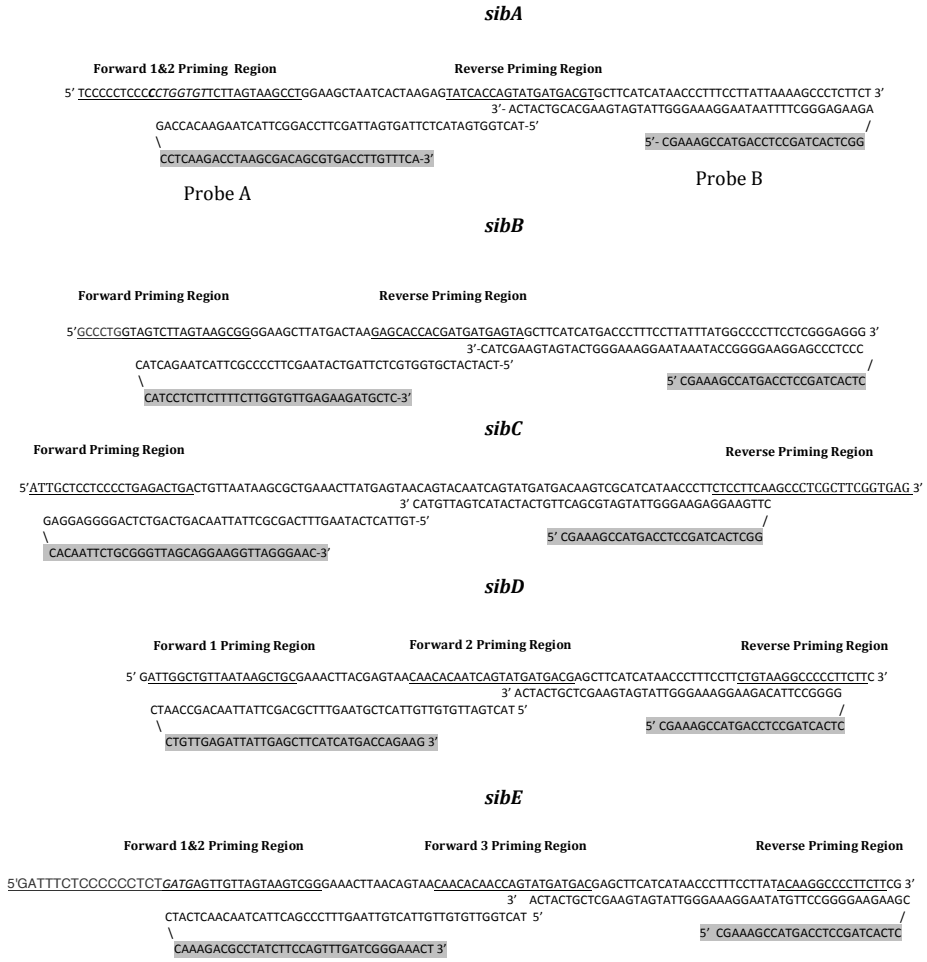


Figure 5.9 . Binding of nanostring probes to their respective targets (*sibA-E*). Probes A (reporter probes) and B (Capture probes) are labeled for *sbsA* as an example. Highlighted in grey are reporter portions for Probe A and capture portions for probe B. Underlined regions within the target sequences are priming sites for the qPCR primers. These probes meet the percent similarity cut offs to avoid cross hybridization. each probe pair is under the threshold for specificity (~85-86%) against the other sequences. Both nanostring probes and qPCR primers were designed within regions of least similarity among all five homologous copies.

Target	Nanostring Probe A	Probe B
IbsA	5'CATAACCCTTTCCTTATTAAAAGCCCTCTTCTCCGGCCTCAAGAC CTAAGCGACAGCGTGACCTTGTTTCA-3'	5'CGAAAGCCATGACCTCCGATCACTCGCCTGGAAGCTAATCACTA AGAGTATCACCAGTATGATGACGTGCTTCAT-3'
IbsB	5'TCATGACCCTTTCCTTATTTATGGCCCCTTCCTCGCATCCTCTTCTT TTCTTGGTGTTGAGAAGATGCTC-3'	5'CGAAAGCCATGACCTCCGATCACTCGTAAGCGGGGAAGCTTAT GACTAAGAGCACCACGATGATGAGTAGCTTCA-3'
IbsC	5'TCATAACCCTTCTCCTTCAAGCCCTCGCTTCGGTGCACAATTCTG CGGGTTAGCAGGAAGGTTAGGGAAC -3'	5'CGAAAGCCATGACCTCCGATCACTCAGCGCTGAAACTTATGAGT AACAGTACAATCAGTATGATGACAAGTCGCA-3'
IbsD	5'GAGCGCCAGACCCCACGCAATGCTTTCTCTGATGCAGTTTCTGT TGAGATTATTGAGCTTCATCATGACCAGAAG-3'	5'CGAAAGCCATGACCTCCGATCACTCCGTTTCAGCCCCCTGCTGA CTCCCCGATTGTTGAT-3'
IbsE	5'GCGCCGGACCCCGCGCAATGCTTTCCCAGAGGCAATTTTGCAAA GACGCCTATCTTCCAGTTTGATCGGGAAACT-3'	5'CGAAAGCCATGACCTCCGATCACTCCGTTTCAGCCCCTTGCTGCC CCCCGATTGTTGATGA-3'
SibA	5'TACTGGTGATACTCTTAGTGATTAGCTTCCAGGCTTACTAAGAA CACCAGCCTCAAGACCTAAGCGACAGCGTGACCTTGTTTCA-3'	5'CGAAAGCCATGACCTCCGATCACTCGGAGAAGAGGGCTTTTAA TAAGGAAAGGGTTATGATGAAGCACGTCATCA-3'
SibB	5'TCATCATCGTGGTGCTCTTAGTCATAAGCTTCCCCGCTTACTAAG ACTACCATCCTCTTCTTTTCTTGGTGTTGAGAAGATGCTC-3'	5'CGAAAGCCATGACCTCCGATCACTCCCCTCCCGAGGAAGGGGC CATAAATAAGGAAAGGGTCATGATGAAGCTAC-3'
SibC	5'TGTTACTCATAAGTTTCAGCGCTTATTAACAGTCAGTCTCAGGG GAGGAGCACAATTCTGCGGGTTAGCAGGAAGGTTAGGGAAC-3'	5'CGAAAGCCATGACCTCCGATCACTCGGCTTGAAGGAGAAGGGT TATGATGCGACTTGTCATCATACTGATTGTAC-3'
SibD	5'TACTGATTGTGTTGTTACTCGTAAGTTTCGCAGCTTATTAACAGC CAATCCTGTTGAGATTATTGAGCTTCATCATGACCAGAAG-3'	5'CGAAAGCCATGACCTCCGATCACTCGGGGCCTTACAGAAGGAA AGGGTTATGATGAAGCTCGTCATCA-3'
SibE	5'TACTGGTTGTGTTGTTACTGTTAAGTTTCCCGACTTACTAACAAC TCATCCAAAGACGCCTATCTTCCAGTTTGATCGGGAAACT-3'	5'CGAAAGCCATGACCTCCGATCACTCCGAAGAAGGGGCCTTGTA TAAGGAAAGGGTTATGATGAAGCTCGTCATCA-3'

Table 5.1. Nano string Probe sequences. The black are the binding sequences and the grey are the reporter portion for Probe A and the capture sequence for Probe B.

5.3.6 A cost efficient and specific next-generation sequencing protocol for high throughput quantification of ibs/sib gene pairs

Another powerful technique for specific detection and quantification of transcripts is next generation sequencing as previously shown for IbsC and SibC transcripts in chapter 4. The high throughput capacity of the technique is advantageous when performing specific detection and quantification of the 10 targets (IbsA-E and SibA-E) under a multitude of conditions, and in different biological replicates. However, with the many number of targets, biological replicates, test conditions, and standards of known quantity for each target for absolute quantification, the

number of samples that require unique indexing can add up and become expensive. To minimize cost, without sacrificing specificity of the reactions, we adopted a protocol that involved 2 rounds of PCR (as opposed to the single PCR protocol we developed in Chapter 4 for detection and quantification of IbsC and SibC RNAs). In the first round of PCR, the previously validated specific primers could be used to amplify each of the 10 targets (Ibs and Sib A-E; cDNA templates in biological samples and double stranded DNA in standard samples) individually and specifically. These primers are short and inexpensive. In addition to the priming sequence, they include a short part of the adaptor sequence that will serve as the binding site for "indexing primers" in a second round of PCR. Indexing primers are the more expensive primers, however, they are A) reusable for different projects, as long as they are not used on a shared run B) can be used in combinations to reduce their numbers for unique indexing of samples (e.g. 20 samples can be indexed by using only 9 primers: 4X5 → to create 20 unique indices). This approach is specific and cost effective. These primers are shown **Supplementary Table 5.2**.

The second round of PCR solely serves to uniquely index different samples, therefore only few cycles of PCR are required. Different targets (i.e. Ibs and SibA-E) and their technical replicates for each sample (either a biological replicate under a specific test sample or a standard sample of known template quantity) which have been amplified in round 1 of q-PCR, can be pooled, purified and subsequently indexed altogether in one PCR reaction with round 2, indexing primers to minimize bias and reagents. These sequences and barcoding plans are confirmed by next gen sequencing technicians, and are presented in **Supplementary Table 5.3**. There are other experimental considerations such as purifying PCR reactions both prior to round II PCR and before submission for sequencing, primer purification and concentrations to use. For these considerations, please refer to the illumina protocols.

5.4 Discussion

In this project, we measured transcript levels for the 4 pairs of *ibs/sib* genes (A, B, D &E) homologous to the *ibsC*/*sibC* gene pair whose expression was extensively analyzed in chapter 4. As described in chapter 4, the *ibsC* toxin expression was the highest at the early to mid log phase (up to OD_{600} of about 0.5) and switched off during the mid-log into the stationary phase. The antitoxin *sibC* was low during the early log phase, continued to rise with OD_{600} and peaked in the stationary phase. As such, both the toxin and antitoxin were shown to be distinctly expressed between OD_{600} of 0.2 and 0.8. Therefore, in an attempt to investigate whether expression could be detected from the other 4 homologous gene pairs and how their RNA expression behavior compared with that of the *ibsC* and *sibC* genes, we studied expression under OD_{600} of 0.2 and 0.8. Similar to *ibsC,* the other four copies were indeed expressed, but only at OD_{600} of 0.2 and our expression data relative to the negative controls suggested that RNA levels from each individual locus were maintained at low levels. Expression seemed to turn off at OD_{600} of 0.8. Therefore, all 5 *ibs* gene copies seemed to be expressed equivalently, at least under natural physiological conditions. As Fozo suggests, similar expression behavior between the 5 homologous copies may indicate redundant function by all copies, which is also in accordance with their incredible homology, small size and great mutability. [78] In addition, our observations of little expression from all 5 copies of *ibs*, further support Fozo's hypothesis about why there are multiple copies of the *ibs* locus within a bacteria's chromosome: to ensure adequate production of "total" Ibs proteins while keeping individual Ibs RNA levels low, to not cause toxicity.

It should be noted that we only confirmed low and equivalent expression from all the 5 copies under normal physiological conditions and two different OD_{600}

levels. Whether diverse environmental conditions might induce different expression behaviors among these genes remains to be investigated. Different expression levels from these genes under different conditions would allow the bacterium to fine tune Ibs production for different cellular demands. Therefore, we need to study expression from these genes, together with the parameters that influence expression (i.e. promoter strength, RNA stability, translation control, etc.) under a multitude of different conditions.

In this chapter we also investigated expression from the *sib A, B, D* and *E* genes under the aforementioned conditions (i.e. OD_{600} of 0.2 and 0.8). We detected expression from all copies except for *E*. For the transcribed RNAs, our results suggested much higher levels of Sib RNAs relative to their cognate toxin RNAs, consistent with that reported for the *ibsC/sibC* pair in chapter 4. Otherwise, *sibA, B, D* interestingly showed expression patterns that were different from that of *sibC*.

In chapter 4, we showed that *sibC* was more highly transcribed at the OD_{600} of 0.8 relative to 0.2. Here, we observed the opposite trend for *sibA* and *D* and equal expression during the two stages for *sibB*. Different expression behaviors by these genes suggest that the different *sib* genes may have different functions. In chapter 4, we discussed multiple lines of evidence suggesting that *sibC* might have targets other than *ibsC,* which is not unprecedented for trans acting small RNAs. To add to the complexity of functions by these genes, here our data suggest that the different *sib* genes may also have different functions from one another. Pull-down experiments and modified CLASH (crosslinking, ligation and sequencing hybrids) or similar technologies may be employed to elucidate protein and RNA targets for these small non-coding RNAs.

To gain a better understanding of transcription and function of these genes, more expression analyses under specific physiological and environmental conditions

(e.g. various OD_{600} for gaining a better resolution of expression behavior throughout growth such as that performed in Chapter 4 for *ibsC* and *sibC*, persistence, stressed conditions, etc.) will be helpful. To conduct these experiments, we suggest techniques that have multiplexing power such as next-generation sequencing and nanostring. We have designed probes and protocols to facilitate the use of these techniques in future studies. They can also be used as independent approaches to verify the RT-qPCR results presented in this chapter.

We believe the fundamental knowledge of expression from the *ibsA-E* and *sib A-E* genes presented in this chapter resolves a lot of ambiguities that currently exist in the literature about the expression of these genes. Also, together with the techniques and protocols developed in chapter 4 and here, we believe we have paved the path to begin investigating expression from these genes under different physiological and environmental conditions to gain an understanding of when they are expressed and therefore their functions.

5.5 Supplementary Figures

Against *ibsA* (100%)

```
100.0% identity in 77 residues overlap; Score: 421.0; Gap frequency: 0.0%

UserSeq1        1 AGAAGAGGGCTTTTAATAAGGAAAGGGTTATGATGAAGCACGTCATCATACTGGTGATAC
UserSeq2       19 AGAAGAGGGCTTTTAATAAGGAAAGGGTTATGATGAAGCACGTCATCATACTGGTGATAC
                  ************************************************************

UserSeq1       61 TCTTAGTGATTAGCTTC
UserSeq2       79 TCTTAGTGATTAGCTTC
                  *****************
```

Against *ibsB* (77.9%)

```
77.9% identity in 77 residues overlap; Score: 310.0; Gap frequency: 0.0%

UserSeq1        1 AGAAGAGGGCTTTTAATAAGGAAAGGGTTATGATGAAGCACGTCATCATACTGGTGATAC
UserSeq2       89 AGAAGGGGGCCTTACAGAAGGAAAGGGTTATGATGAAGCTCGTCATCATACTGATTGTGT
                  ***** **** **  * ********************** ************* *  *

UserSeq1       61 TCTTAGTGATTAGCTTC
UserSeq2      149 TGTTACTCGTAAGTTTC
                  * *** *  * ** ***
```

Against *ibsC* (68.8%)

```
68.8% identity in 77 residues overlap; Score: 255.0; Gap frequency: 3.9%

UserSeq1        1 AGAAGAGGGCTTTTAATAAGGAAAGGGTTATGATGAAGCACGTCATCATACTGGTGATAC
UserSeq2       21 AAGCGAGGGCTT---GAAGGAGAAGGGTTATGATGCGACTTGTCATCATACTGATTGTAC
                  *   ********     * *  *************   *  ************ *  ***

UserSeq1       61 TCTTAGTGATTAGCTTC
UserSeq2       78 TGTTACTCATAAGTTTC
                  * *** * ** ** ***
```

Against *ibsD* (77.9%)

```
77.9% identity in 77 residues overlap; Score: 296.0; Gap frequency: 3.9%

UserSeq1        1 AGAAGAGGGCTTTTAATAAGGAAAGGGTTATGATGAAGCACGTCATCATACTGGTGATAC
UserSeq2       18 AGGAAGGGGCCATAAATAAGGAAAGGGTCATGATGAAGCTACTCATCATCGTGGTG---C
                  ** *  ****  * ************** **********   *******  *****   *

UserSeq1       61 TCTTAGTGATTAGCTTC
UserSeq2       75 TCTTAGTCATAAGCTTC
                  ******* ** ******
```

Against *ibsE* (79.2%)

```
 79.2% identity in 48 residues overlap; Score: 203.0; Gap frequency: 0.0%

 UserSeq1       30 ATGATGAAGCACGTCATCATACTGGTGATACTCTTAGTGATTAGCTTC
 UserSeq2        1 ATGATGAAGCTCGTCATCATACTGGTTGTGTTGTTACTGTTAAGTTTC
                   ********** ***************  *  * *** ** * ** ***
```

Supplementary Figure 5.1. Alignment of Sanger sequencing reads for the end reaction products of PCR on gDNA with IbsA primers. Results indicate pure composition of the sample and specification amplification by the designed primers.

Against *ibsD* (100.0%)

```
100.0% identity in 100 residues overlap; Score: 594.0; Gap frequency: 0.0%

UserSeq1        1 GCGCTCATCAACAATCGGGGAGTCAGCAGGGGGCTGAAACGGGAAAGCCCCTCCCGAAGA
UserSeq2       32 GCGCTCATCAACAATCGGGGAGTCAGCAGGGGGCTGAAACGGGAAAGCCCCTCCCGAAGA
                  ************************************************************

UserSeq1       61 AGGGGGCCTTACAGAAGGAAAGGGTTATGATGAAGCTCGT
UserSeq2       92 AGGGGGCCTTACAGAAGGAAAGGGTTATGATGAAGCTCGT
                  *****************************************
```

Against *ibsA* (86.9%)

```
86.9% identity in 61 residues overlap; Score: 304.0; Gap frequency: 0.0%

UserSeq1       40 CGGGAAAGCCCCTCCCGAAGAAGGGGGCCTTACAGAAGGAAAGGGTTATGATGAAGCTCG
UserSeq2        1 CGGGAAAGCCTCTCCCGGAGAAGAGGGCTTTTAATAAGGAAAGGGTTATGATGAAGCACG
                  ********** ****** ***** **** **  * ********************** **

UserSeq1      100 T
UserSeq2       61 T
                  *
```

Against *ibsB* (86.9%)

```
86.9% identity in 61 residues overlap; Score: 288.0; Gap frequency: 1.6%

UserSeq1       40 CGGGAAAGCCCCTCCCGAAGAAGGGGGCCTTACAGAAGGAAAGGGTTATGATGAAGCTCG
UserSeq2        1 CGGGAAAGCCCCTCCCGAGGAAGGGG-CCATAAATAAGGAAAGGGTCATGATGAAGCTAC
                  ****************** ******* ** ** * *********** ***********

UserSeq1      100 T
UserSeq2       60 T
                  *
```

Against *ibsC* (71.4%)

```
71.4% identity in 63 residues overlap; Score: 215.0; Gap frequency: 6.3%

UserSeq1       38 AACGGGAAAGCCCCTCCCGAAGAAGGGGGCCTTACAGAAGGAAAGGGTTATGATGAAGCT
UserSeq2        2 AACGGTAAAGCCCTCACCGAAGCGAGGG--CTTGAAGGAG--AAGGGTTATGATGCGACT
                  ***** *******   ******   ***  ***  ** **  *************   **

UserSeq1       98 CGT
UserSeq2       58 TGT
                   **
```

Against *ibsE* (93.0%)

```
93.0% identity in 100 residues overlap; Score: 519.0; Gap frequency: 2.0%

UserSeq1        1 GCGCTCATCAACAATCGGGGAGTCAGCAGGGGGCTGAAACGGGAAAGCCCCTCCCGAAGA
UserSeq2       37 GCGCTCATCAACAATCGGGGGG-CAGCAAGGGGCTGAAACGGGAAAGCCCCTCCCGAAGA
                  ******************** * ***** ********************************

UserSeq1       61 AGGGGGCCTTACAGAAGGAAAGGGTTATGATGAAGCTCGT
UserSeq2       96 AGGGG-CCTTGTATAAGGAAAGGGTTATGATGAAGCTCGT
                  ***** ****  * *************************
```

Supplementary Figure 5.2. Alignment of Sanger sequencing reads for the end reaction products of PCR on gDNA with IbsD (forward 1/ nano) primers to amplify beyond regions that are known to be transcribed. Results indicate pure composition of the sample and specific amplification by the designed primers.

Against *ibsE* (100.0%)

```
100.0% identity in 170 residues overlap; Score: 1009.0; Gap frequency: 0.0%

UserSeq1        1 CAAAATTGCCTCTGGGAAAGCATTGCGCGGGGTCCGGCGCTCATCAACAATCGGGGGGCA
UserSeq2        1 CAAAATTGCCTCTGGGAAAGCATTGCGCGGGGTCCGGCGCTCATCAACAATCGGGGGGCA
                  ************************************************************

UserSeq1       61 GCAAGGGGCTGAAACGGGAAAGCCCCTCCCGAAGAAGGGGCCTTGTATAAGGAAAGGGTT
UserSeq2       61 GCAAGGGGCTGAAACGGGAAAGCCCCTCCCGAAGAAGGGGCCTTGTATAAGGAAAGGGTT
                  ************************************************************

UserSeq1      121 ATGATGAAGCTCGTCATCATACTGGTTGTGTTGTTACTGTTAAGTTTCCC
UserSeq2      121 ATGATGAAGCTCGTCATCATACTGGTTGTGTTGTTACTGTTAAGTTTCCC
                  ***************************************************
```

Against *ibsA* (83.3%)

```
83.3% identity in 96 residues overlap; Score: 432.0; Gap frequency: 1.0%

UserSeq1       75 CGGGAAAGCCCCTCCCGAAGAAG-GGGCCTTGTATAAGGAAAGGGTTATGATGAAGCTCG
UserSeq2        1 CGGGAAAGCCTCTCCCGGAGAAGAGGGCTTTTAATAAGGAAAGGGTTATGATGAAGCACG
                  ********** ****** ***** **** **  ********************** **

UserSeq1      134 TCATCATACTGGTTGTGTTGTTACTGTTAAGTTTCC
UserSeq2       61 TCATCATACTGGTGATACTCTTAGTGATTAGCTTCC
                  *************  *  * *** ** * ** ****
```

Against *ibsB* (81.2%)

```
81.2% identity in 96 residues overlap; Score: 414.0; Gap frequency: 3.1%

UserSeq1       75 CGGGAAAGCCCCTCCCGAAGAAGGGGCCTTGTATAAGGAAAGGGTTATGATGAAGCTCGT
UserSeq2        1 CGGGAAAGCCCCTCCCGAGGAAGGGGCCATAAATAAGGAAAGGGTCATGATGAAGCTACT
                  ***************** ********* *  ************* ***********  *

UserSeq1      135 CATCATACTGGTTGTGTTGTTACTGTTAAGTTTCCC
UserSeq2       61 CATCATC---GTGGTGCTCTTAGTCATAAGCTTCCC
                  ******    ** *** * *** *  **** *****
```

Against *ibsC* (77.1%)

```
77.1% identity in 96 residues overlap; Score: 376.0; Gap frequency: 3.1%

UserSeq1       73 AACGGGAAAGCCCCTCCCGAAGAAGGGGCCTTGTATAAGGAAAGGGTTATGATGAAGCTC
UserSeq2        2 AACGGTAAAGCCCTCACCGAAGCGAGGGC-TTGAAGGAG--AAGGGTTATGATGCGACTT
                  ***** *******   ******   **** *** *  **  *************   **

UserSeq1      133 GTCATCATACTGGTTGTGTTGTTACTGTTAAGTTTC
UserSeq2       59 GTCATCATACTGATTGTACTGTTACTCATAAGTTTC
                  ************ ****  *******  ********
```

Against *ibsD* (90.9%)

```
90.9% identity in 99 residues overlap; Score: 529.0; Gap frequency: 1.0%

UserSeq1        3 AAATTGCCTCTGGGAAAGCATTGCGCGGGGTCCGGCGCTCATCAACAATCGGGGGG-CAG
UserSeq2        1 AAACTGCATCAGAGAAAGCATTGCGTGGGGTCTGGCGCTCATCAACAATCGGGGAGTCAG
                  *** *** ** * ************ ****** ******************** * ***

UserSeq1       62 CAAGGGGCTGAAACGGGAAAGCCCCTCCCGAAGAAGGGG
UserSeq2       61 CAGGGGGCTGAAACGGGAAAGCCCCTCCCGAAGAAGGGG
                  ** ***********************************
```

Supplementary Figure 5.3. Alignment of Sanger sequencing reads for the end reaction products of PCR on gDNA with IbsE (forward 1/nano) primers, to amplify beyond regions that are known to be transcribed. Results indicate pure composition of the sample and specific amplification by the designed primers.

Against *ibsE* (100%)

```
100.0% identity in 26 residues overlap; Score: 140.0; Gap frequency: 0.0%

UserSeq1        1 AGAAGGGGCCTTGTATAAGGAAAGGG
UserSeq2       93 AGAAGGGGCCTTGTATAAGGAAAGGG
                  **************************
```

Against *ibsA* (85.2%)

```
85.2% identity in 27 residues overlap; Score: 101.0; Gap frequency: 3.7%

UserSeq1        1 AGAAG-GGGCCTTGTATAAGGAAAGGG
UserSeq2       19 AGAAGAGGGCTTTTAATAAGGAAAGGG
                  ***** **** **  ************
```

Against *ibsB* (88.0%)

```
88.0% identity in 25 residues overlap; Score: 120.0; Gap frequency: 0.0%

UserSeq1        2 GAAGGGGCCTTGTATAAGGAAAGGG
UserSeq2       20 GAAGGGGCCATAAATAAGGAAAGGG
                  ********* *  ************
```

Against *ibsC* (54.2)

```
54.2% identity in 24 residues overlap; Score: 66.0; Gap frequency: 0.0%

UserSeq1        1 AGAAGGGGCCTTGTATAAGGAAAG
UserSeq2       21 AAGCGAGGGCTTGAAGGAGAAGGG
                  *   * ** **** *  ** *  *
```

Against *ibsD* (70.8%)

```
70.8% identity in 24 residues overlap; Score: 73.0; Gap frequency: 4.2%

UserSeq1        1 AGAAGGGGCCTTGTATAAGGAAAG
UserSeq2       59 AGCAGGGGGCT-GAAACGGGAAAG
                  ** ***** ** * *   ******
```

Supplementary Figure 5.4. Alignment of Sanger sequencing reads for the end reaction products of PCR on gDNAwith IbsE (forward 2) primers. Results indicate pure composition of the sample and specific amplification by the designed primers.

Against *sibD* (100.0%)

```
100.0% identity in 92 residues overlap; Score: 540.0; Gap frequency: 0.0%

UserSeq1        1 ATTGGCTGTTAATAAGCTGCGAAACTTACGAGTAACAACACAATCAGTATGATGACGAGC
UserSeq2        1 ATTGGCTGTTAATAAGCTGCGAAACTTACGAGTAACAACACAATCAGTATGATGACGAGC
                  ************************************************************

UserSeq1       61 TTCATCATAACCCTTTCCTTCTGTAAGGCCCC
UserSeq2       61 TTCATCATAACCCTTTCCTTCTGTAAGGCCCC
                  ********************************
```

Against *sibA* (74.7%)

```
74.7% identity in 87 residues overlap; Score: 347.0; Gap frequency: 0.0%

UserSeq1        5 GCTGTTAATAAGCTGCGAAACTTACGAGTAACAACACAATCAGTATGATGACGAGCTTCA
UserSeq2       17 GTTCTTAGTAAGCCTGGAAGCTAATCACTAAGAGTATCACCAGTATGATGACGTGCTTCA
                  * * *** *****   *** ** *  * *** *  *  * ************* ******

UserSeq1       65 TCATAACCCTTTCCTTCTGTAAGGCCC
UserSeq2       77 TCATAACCCTTTCCTTATTAAAAGCCC
                  *************** *  ** ****
```

Against *sibB* (76.7%)

```
76.7% identity in 86 residues overlap; Score: 336.0; Gap frequency: 3.5%

UserSeq1        7 TGTTAATAAGCTGCGAAACTTACGAGTAACAACACAATCAGTATGATGACGAGCTTCATC
UserSeq2       11 TCTTAGTAAGCGGGGAAGCTTATGACTAAGAGCAC---CACGATGATGAGTAGCTTCATC
                  * *** ***** * *** **** ** *** * ***   **  *******  *********

UserSeq1       67 ATAACCCTTTCCTTCTGTAAGGCCCC
UserSeq2       68 ATGACCCTTTCCTTATTTATGGCCCC
                  ** *********** * ** ******
```

Against *sibC* (79.3%)

```
79.3% identity in 92 residues overlap; Score: 376.0; Gap frequency: 2.2%

UserSeq1        1 ATTGGCTGTTAATAAGCTGCGAAACTTACGAGTAACAACACAATCAGTATGATGACGAGC
UserSeq2       18 ACTGACTGTTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGATGACAAGT
                  * ** ************   ******** ********  ***************** **

UserSeq1       61 TTCATCATAACCCTTTCCTTCTGTAAGGCCCC
UserSeq2       78 CGCATCATAACCCTT--CTCCTTCAAGCCCTC
                    *************  ** **  *** ** *
```

Against *sibE* (85.9%)

```
85.9% identity in 92 residues overlap; Score: 440.0; Gap frequency: 0.0%

UserSeq1        1 ATTGGCTGTTAATAAGCTGCGAAACTTACGAGTAACAACACAATCAGTATGATGACGAGC
UserSeq2       18 ATGAGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGACGAGC
                  **  * ***** ****  * ********  ************* ****************

UserSeq1       61 TTCATCATAACCCTTTCCTTCTGTAAGGCCCC
UserSeq2       78 TTCATCATAACCCTTTCCTTATACAAGGCCCC
                  ******************* *  ********
```

Supplementary Figure 5.5. Alignment of Sanger sequencing reads for the end reaction products of PCR on gDNA with SibD Nano (forward 1) primers. Results indicate pure composition of the sample and specific amplification by the designed primers.

Against *sibE* (100%)

```
100.0% identity in 109 residues overlap; Score: 651.0; Gap frequency: 0.0%

UserSeq1        1 GATTTCTCCCCCCTCTGATGAGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAA
UserSeq2        1 GATTTCTCCCCCCTCTGATGAGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAA
                  ************************************************************

UserSeq1       61 CCAGTATGATGACGAGCTTCATCATAACCCTTTCCTTATACAAGGCCCC
UserSeq2       61 CCAGTATGATGACGAGCTTCATCATAACCCTTTCCTTATACAAGGCCCC
                  **************************************************
```

Against *sibA* (78.1%)

```
78.1% identity in 105 residues overlap; Score: 458.0; Gap frequency: 1.9%

UserSeq1        4 TTCTCCCCCCTCTGATGAGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCA
UserSeq2        1 TCCCCCTCCCCCTG--GTGTTCTTAGTAAGCCTGGAAGCTAATCACTAAGAGTATCACCA
                  * * ** *** ***  * *** ******** * **** ** * ** *** *  *  ****

UserSeq1       64 GTATGATGACGAGCTTCATCATAACCCTTTCCTTATACAAGGCCC
UserSeq2       59 GTATGATGACGTGCTTCATCATAACCCTTTCCTTATTAAAAGCCC
                  *********** ************************  ** ****
```

Against *sibB* (75.3%)

```
75.3% identity in 97 residues overlap; Score: 348.0; Gap frequency: 5.2%

UserSeq1       13 CTCTGATGAGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGA
UserSeq2        2 CCCTGGT-AGTC-TTAGTAAGCGGGGAAGCTTATGACTAAGAGCACCAC---GATGATGA
                  * *** * ***  ********  ***** ****  * *** * *** **    *******

UserSeq1       73 CGAGCTTCATCATAACCCTTTCCTTATACAAGGCCCC
UserSeq2       57 GTAGCTTCATCATGACCCTTTCCTTATTTATGGCCCC
                    *********** *************  * ******
```

Against *sibC* (72.5%)

```
72.5% identity in 109 residues overlap; Score: 398.0; Gap frequency: 2.8%

UserSeq1        2 ATTTCTCC-CCCCTCTGATGAGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAA
UserSeq2        1 ATTGCTCCTCCCCTGAGACTGACTGTTAATAAGCGCTGAAACTTATGAGTAACAGTACAA
                  *** **** *****  **     ***** ****    ********  *******  ****

UserSeq1       61 CCAGTATGATGACGAGCTTCATCATAACCCTTTCCTTATACAAGGCCCC
UserSeq2       61 TCAGTATGATGACAAGTCGCATCATAACCCTT--CTCCTTCAAGCCCTC
                   ************ **   *************  **  * **** ** *
```

Against *sibD* (86.0%)

```
86.0% identity in 93 residues overlap; Score: 446.0; Gap frequency: 0.0%

UserSeq1       17 GATGAGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGACGAG
UserSeq2        1 GATTGGCTGTTAATAAGCTGCGAAACTTACGAGTAACAACACAATCAGTATGATGACGAG
                  ***  * ***** ****  * ********  ************* ***************

UserSeq1       77 CTTCATCATAACCCTTTCCTTATACAAGGCCCC
UserSeq2       61 CTTCATCATAACCCTTTCCTTCTGTAAGGCCCC
                  ******************** *  ********
```

Supplementary Figure 5.6. Alignment of Sanger sequencing reads for the end reaction products of PCR on gDNA with SibE Nano (forward 1) primers. Results indicate pure composition of the sample and specific amplification by the designed primers.

Against *sibE* (100.0%)

```
100.0% identity in 93 residues overlap; Score: 542.0; Gap frequency: 0.0%

UserSeq1        1 AGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGACGAGCTTC
UserSeq2       21 AGTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGACGAGCTTC
                  ************************************************************

UserSeq1       61 ATCATAACCCTTTCCTTATACAAGGCCCCTTCT
UserSeq2       81 ATCATAACCCTTTCCTTATACAAGGCCCCTTCT
                  ********************************
```

Against *sibA* (81.7%)

```
81.7% identity in 93 residues overlap; Score: 414.0; Gap frequency: 1.1%

UserSeq1        2 GTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGACGAGCTTCA
UserSeq2       17 GTTCTTAGTAAGCCTGGAAGCTAATCACTAAGAGTATCACCAGTATGATGACGTGCTTCA
                  *** ******** * **** ** * ** *** *  *  *************** ******

UserSeq1       62 TCATAACCCTTTCCTTATACAAGGCCC-CTTCT
UserSeq2       77 TCATAACCCTTTCCTTATTAAAAGCCCTCTTCT
                  ******************  ** **** *****
```

Against *sibB* (77.5%)

```
77.5% identity in 89 residues overlap; Score: 354.0; Gap frequency: 3.4%

UserSeq1        4 TGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGACGAGCTTCATC
UserSeq2       11 TCTTAGTAAGCGGGGAAGCTTATGACTAAGAGCACCAC---GATGATGAGTAGCTTCATC
                  * ********  ***** ****  * *** * *** **    *******  *********

UserSeq1       64 ATAACCCTTTCCTTATACAAGGCCCCTTC
UserSeq2       68 ATGACCCTTTCCTTATTTATGGCCCCTTC
                  ** *************  * *********
```

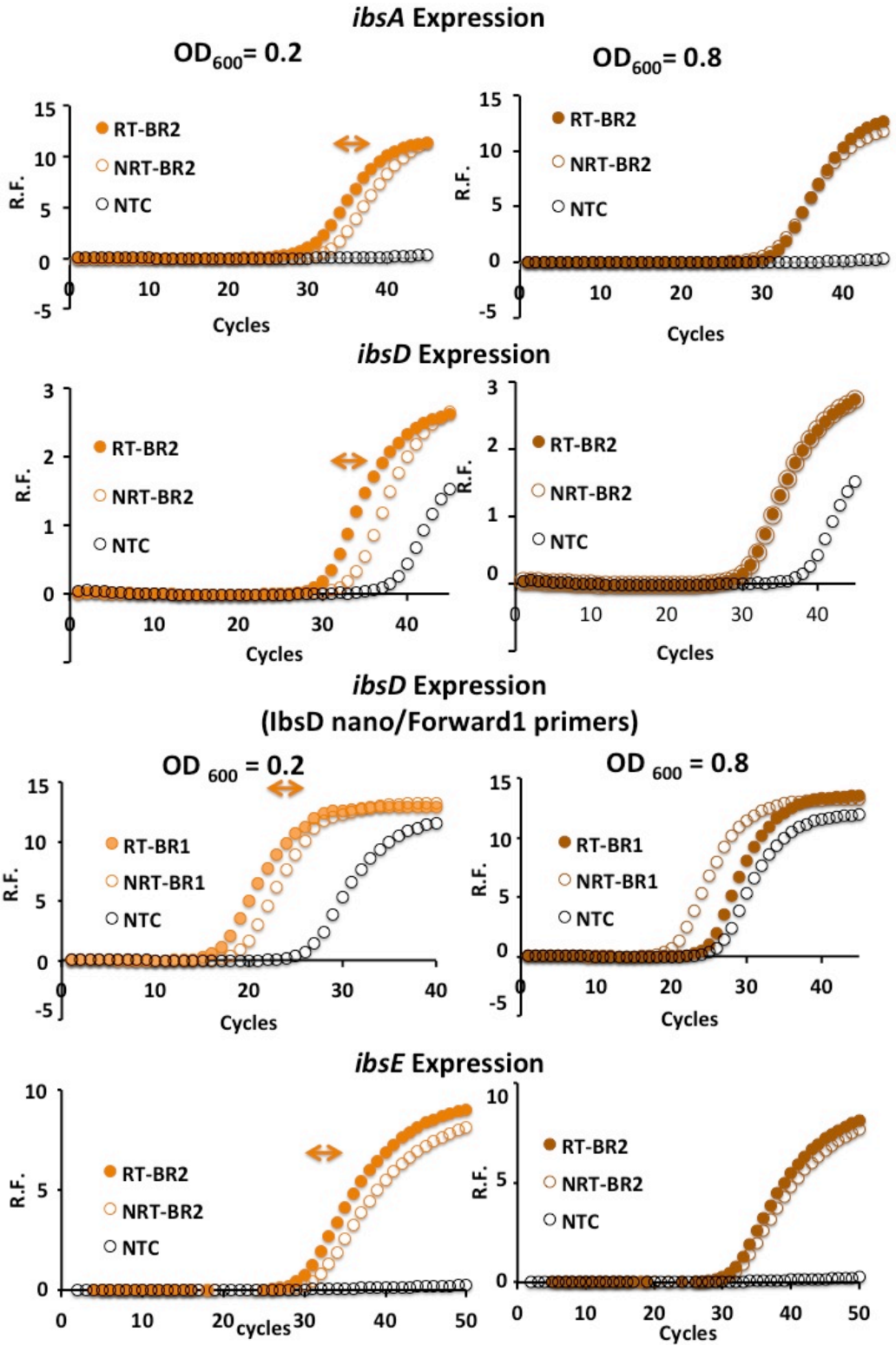
Against *sibC* (73.3%)

```
73.3% identity in 90 residues overlap; Score: 327.0; Gap frequency: 2.2%

UserSeq1        4 TGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGACGAGCTTCATC
UserSeq2       24 TGTTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGATGACAAGTCGCATC
                  ***** ****    ********  *******  **** ************ **   ****

UserSeq1       64 ATAACCCTTTCCTTATACAAGGCCCCTTCT
UserSeq2       84 ATAACCCTT--CTCCTTCAAGCCCTCGCTT
                  *********  **  * **** ** *   *
```

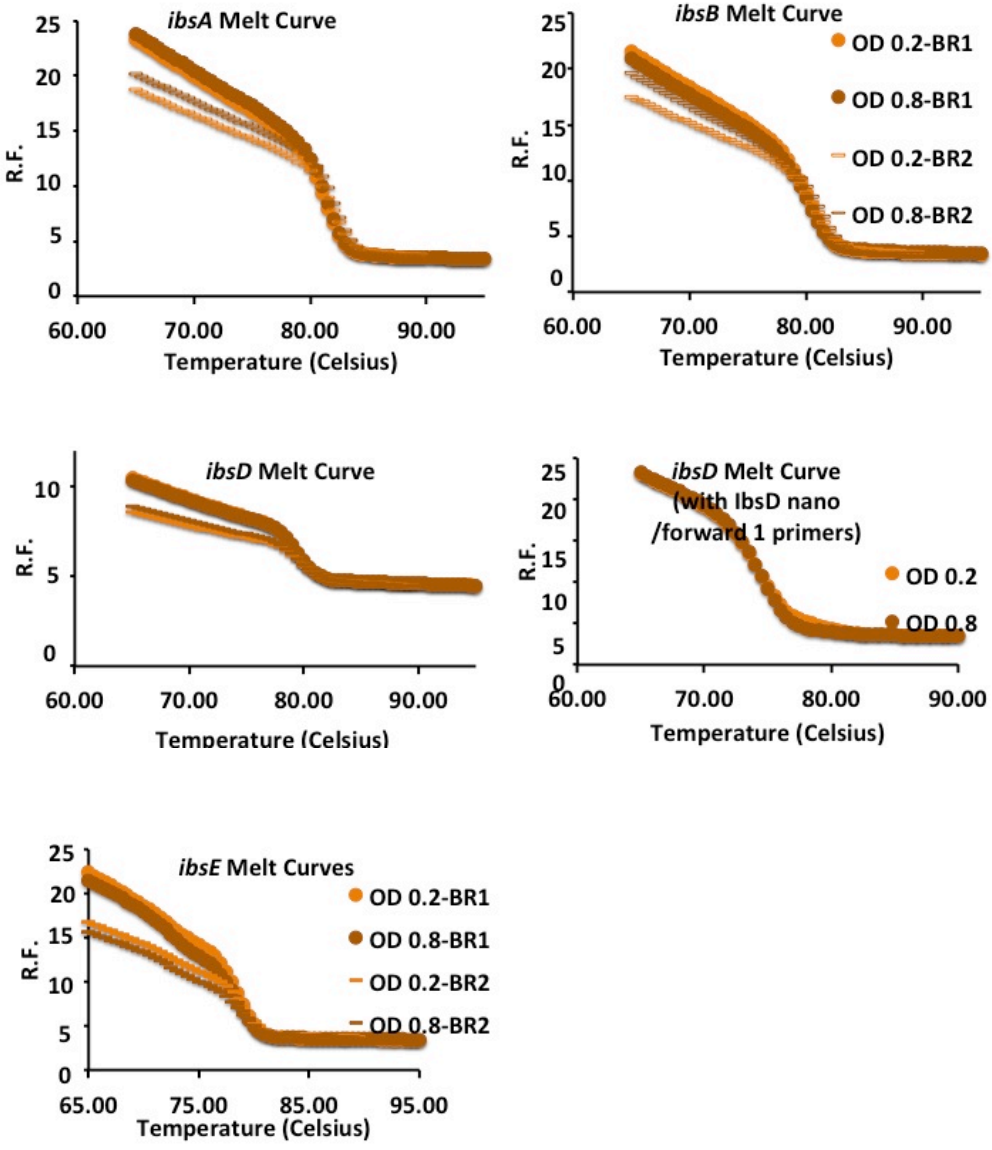
Against *sibD* (87.1%)

```
87.1% identity in 93 residues overlap; Score: 441.0; Gap frequency: 1.1%

UserSeq1        2 GTTGTTAGTAAGTCGGGAAACTTAACAGTAACAACACAACCAGTATGATGACGAGCTTCA
UserSeq2        6 GCTGTTAATAAGCTGCGAAACTTACGAGTAACAACACAATCAGTATGATGACGAGCTTCA
                  * ***** ****  * ********  ************* ********************

UserSeq1       62 TCATAACCCTTTCCTTATACAAGGCCCC-TTCT
UserSeq2       66 TCATAACCCTTTCCTTCTGTAAGGCCCCCTTCT
                  **************** *  ******** ****
```

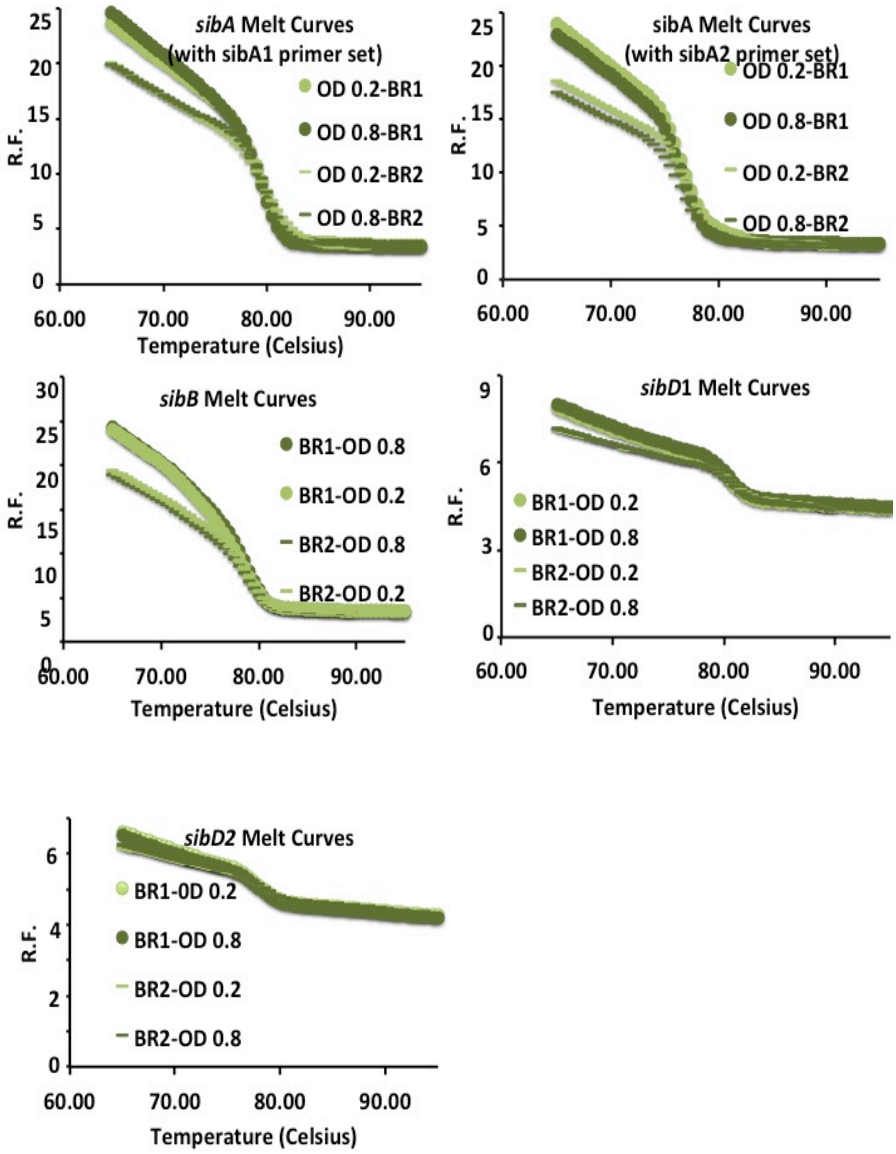
Figure 5.7. Alignment of Sanger sequencing reads for the end reaction products of PCR on gDNA with SibE (forward 2) primers. Results indicate pure composition of the sample and specific amplification by the designed primers.


ibsA Expression
OD_{600}= 0.2
OD_{600}= 0.8
RT-BR2
NRT-BR2
NTC
R.F.
Cycles
ibsD Expression
ibsD Expression
(IbsD nano/Forward1 primers)
OD 600 = 0.2
OD 600 = 0.8
RT-BR1
NRT-BR1
ibsE Expression

Supplementary Figure 5.8. Gene expression analyses from the *ibsA*, *B*, *D* and *E* genes. (Top to Bottom) Left: cultures grown to OD_{600} of 0.2 Right: cultures grown to OD_{600} of 0.8. BR2: Biological Replicate 2; RT: Short for reverse transcription signifies amplification curves generated by PCR on target-cDNA samples obtained by RT; NRT: No Reverese Transcription Control for genomic DNA contamination; NTC: No Template Control for primer-dimer and external DNA contamination. The Left shift of The RT curves relative to the controls in early log (OD_{600} = 0.2) samples, for all *ibs* genes indicates expression from these genes. However, the small magnitude of the shift and high Ct values (~30) indicate small expression. Overlap between RT and negative controls in late log (OD_{600} = 0.8) smaples indicates no expression during this phase. RNA expression behaviors for all *ibs* genes are similar.



Supplementary Figure 5.9. *ibs A, B, D& E* melt curve analyses. A single melting point (as indicated by a single dip) point to the homogeneity/purity of the reaction products, hence specific amplification/detection of the targets.



Supplementary Figure 5.10. *sibA, B, D &E* melt curve analyses. A single melting point (as indicated by a single dip) point to the homogeneity/purity of the reaction products, hence specific amplification/detection of the targets.

Target	RT-qPCR Forward Primer	RT-qPCR Reverse Primer
IbsA	TCTCCCGGAGAAGAGGGCTTTTA	AGGTTTCCCCCTCCCCCTGGT
IbsB	CCTCCCGAGGAAGGGGCCATA	GAGCACCACGATGATGAGTA
IbsC	AGCGAGGGCTTGAAGGAG	TGAGACTGACTGTTAATAAGCGCT
IbsD	Forward 1:CTGCATCAGAGAAAGCATTGC Forward 2:AAGAAGGGGGCCTTACAG	GCGAAACTTACGAGTAACAACACAA
IbsE	Forward1:CAAAATTGCCTCTGGGAAAGC Forward2: AAGAAGGGGCCTTGTAT	GGGAAACTTAACAGTAACAACA
SibA	Forward 1:TCCCCCTCCCCCTGGTGT Forward 2:CCTGGTGTTCTTAGTAAGCCT	ACGTCATCATACTGGTGATA
SibB	GCCCTGGTAGTCTTAGTAAGCGG	TACTCATCATCGTGGTGCTC
SibC	ATTGCTCCTCCCCTGAGACTGA	CTCACCGAAGCGAGGGCTTGAAGGAG
SibD	Forward 1: ATTGGCTGTTAATAAGCTGC Forward2: CAACACAATCAGTATGATGACG	AAGAAGGGGGCCTTACAG
SibE	Forward 1: GATTTCTCCCCCCTCTGATG Forward2: GATGAGTTGTTAGTAAGTCGG Forward 3: CAACACAACCAGTATGATGAC	AAGAAGGGGCCTTGT

Supplementary Table 5.1. RT-qPCR Primers used in this study to specifically amplify IbsA-E and SibA-E RNAs in wild type E. coli MG1655 under natural physiological conditions.

Target	Next-gen-Round I PCR Forward Primers	Next-gen-Round I PCR Reverse Primers
IbsA	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT TCTCCCGGAGAAGAGGGCTTTTA 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGTTTCCCCCTCCCCCTGGT 3'
IbsB	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-CCTCCCGAGGAAGGGGCCATA 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAGCACCACGATGATGAGTA 3'
IbsC	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-AGCGAGGGCTTGAAGGAG 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGAGACTGACTGTTAATAAGCGCT 3'
IbsD	Forward 1: 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT-CTGCATCAGAGAAAGCATTGC 3' Forward 2: 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT-AAGAAGGGGGCCTTACAG 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCGAAACTTACGAGTAACAACACAAT 3'
IbsE	Forward1: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCT-CAAAATTGCCTCTGGGAAAGC 3' Forward2: 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT-AAGAAGGGGCCTTGTAT 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGAAACTTAACAGTAACAACA 3'
SibA	Forward 1: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-TCCCCCTCCCCCTGGTGT 3' Forwards 2: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-CCTGGTGTTCTTAGTAAGCCT 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT ACGTCATCATACTGGTGATA 3'
SibB	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-GCCCTGGTAGTCTTAGTAAGCGG 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTACTCATCATCGTGGTGCTC 3'
SibC	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-ATTGCTCCTCCCCTGAGACTGA 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCACCGAAGCGAGGGCTTGAAGGAG
SibD	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-ATTGGCTGTTAATAAGCTGC 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAGAAGGGGGCCTTACAG 3'
SibE	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-GATTTCTCCCCCCTCTGATG 3'	5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAGAAGGGGCCTTGT 3'

Supplementary Table 5.2. Round I PCR primers to create Next-gen sequencing libraries. The underlined sequences are specific binding sequences against the target of interest. These are previously validated to specifically amplify their respective targets. The rest of the sequence for each forward and reverse primer is part of the "adaptor" sequence that will later serve as binding sites for the indexing primers in round II PCR. Forward and reverse adaptor sequences are different for forward and reverse primers but common among all targets. Indexing primers are shown in **Supplementary Table 5.3** .

Illumina I5 Indices for Forward Primers	Forward Primer Sequences for Round II PCR	Illumina I7 Indices For Reverse Primers	Reverse Primer Sequences for Round II PCR
S517	5'AATGATACGGCGACCACCGAGATCT ACACGCGTAAGAACACTCTTTCCCTACA CGACGCTCTTCCGATCT-3'	N726	5'CAAGCAGAAGACGGCATACGAGATGTC TTAGGGTGACTGGAGTTCAGACGTGTGCT CTTCCGATCT-3'
S518	5'AATGATACGGCGACCACCGAGATCT ACACCTATTAAG ACACTCTTTCCCTACACGACGCTCTTCC GATCT-3'	N727	5'CAAGCAGAAGACGGCATACGAGATACT GATCGGTGACTGGAGTTCAGACGTGTGCT CTTCCGATCT-3'
S520	5'AATGATACGGCGACCACCGAGATCT ACACAAGGCTAT ACACTCTTTCCCTACACGACGCTCTTCC GATCT-3'	N728	5'CAAGCAGAAGACGGCATACGAGATTAG CTGCAGTGACTGGAGTTCAGACGTGTGCT CTTCCGATCT-3'
S521	5'AATGATACGGCGACCACCGAGATCT ACACGAGCCTTA ACACTCTTTCCCTACACGACGCTCTTCC GATCT3'	N729	5'CAAGCAGAAGACGGCATACGAGATGAC GTCGAGTGACTGGAGTTCAGACGTGTGCT CTTCCGATCT3'
S522	5'- AATGATACGGCGACCACCGAGATCTAC ACTTATGCGA ACACTCTTTCCCTACACGACGCTCTTCC GATCT-3'		

Supplementary Table 5.3. Indexing primers for round II PCR. The 8ntd indices are color-coded. These are chosen from Illumina's Nextra XT v2 index set. They can be used in combinations to tag 20 different samples uniquely. They are reusable by multiple projects as long as they do not share the same run.

6 Chapter 6: Discussion And Future Directions

6.1 Summary of key findings

TA systems are widespread in bacterial genomes. [98,99] They might represent up to 3% of the total predicted open reading frames (ORFs) in some genomes, with some genomes containing more than 90 TA systems.[100] The Ibs/Sib family of type I TA systems has been the focus of our studies. Its members are abundantly present in *E. coli K-12* and are highly conserved in related Proteobacterial species. They exist in 1 to 7 copies across different species. [72,79]

The surprising abundance of these entities in bacterial genomes raises interesting questions regarding their possible biological roles and evolution. Why would bacteria continue to harbor so many genetic elements that can potentially kill them, if not tightly regulated? These questions are mostly unanswered, and the role of chromosomally encoded TA systems in bacterial physiology is highly debated in the field. [7,10,101]

In spite of the debate about the specific roles of these entities, there is growing evidence that TA systems are important for bacterial physiology and pathogenicity, hence hold great capacity to be turned into antibacterial agents. Yet, they are still only poorly understood, especially in the case of type I toxins encoded on bacterial chromosomes. We believe the reason for our poor understanding of these systems to lie in a) the complexity of their underlying mechanisms, and b) their low expression. TA systems have been proposed to have very specialized functions and be activated under very specific conditions. In addition, they are expected to be expressed at low endogenous levels. Current approaches are not high throughput and sensitive enough to test a multitude of conditions under some of which a specific toxin may become activated and to

detect these low abundant molecules. We believe technical challenges are major factors delaying the study of these molecules.

Our journey toward characterizing the Ibs/Sib type I family of TA, as described in this dissertation, began with developing fluorescence-based platforms to facilitate the study of these molecules, with a focus on studies that could potentially exploit the antibacterial capacities of both the toxins and antitoxins.

We were particularly interested in the functions of the endogenous levels of these proteins. One of the prevalent hypothesized roles for these toxins is in persistence where the production of toxins in a sub population of cells causes them to enter a temporary dormant, multidrug tolerant state and to become culprits of recalcitrant chronic infections [20–23,25–27,30,31,142]. Based on recent studies that have associated persister formation with dormancy and TA systems and our own observations of IbsC causing reversible growth stasis, we considered studying the role of the toxin in persistence. However, persister cells make up an extremely low percentage of a bacterial population and are only transient. For the purpose of investigating the potential role of the IbsC toxin in persistence, we needed to facilitate isolation and identification of these rare cells. In the first component of this project, we developed constructs and engineered *E. coli* strains to link persistence to fluorescence. Subsequently, high throughput, sensitive, accurate and high-resolution (single-cell fluorescent) techniques such as Fluorescence Assisted Cell Sorting (FACS) and Time-lapse microscopy can be used to substitute antibiotic sensitivity assays in identification and isolation of these persister cells. We believe our efforts to facilitate the study of persistence will take us closer to understanding the mechanisms of bacterial persistence and ways to combat them. If the toxin is indeed involved in causing persistence, it can be potentially used as an antibiotic target.

In the second component of this project, we were interested in studying the

interactions between the toxin and antitoxin. Through RNA-RNA interactions, SibC silences the IbsC toxin. Therefore, as a natural RNA molecule with gene silencing capabilities, it has an exciting potential for generation of new functional RNA-based genetic parts, including antimicrobial agents. To exploit this potential, however, further characterization of the molecule and its interactions with IbsC (i.e. minimum sequence and structural requirements) are required. Traditionally, these studies are carried out by performing death/rescue assays which are time consuming and laborious. Herein, we developed constructs and strains to facilitate the study of the interactions by replacing the death/rescue phenotypes with simple fluorescence. Fluorescence-based assays simplify the analytical techniques (e.g. eliminate the need for protein purification) and measurements. They also improve accuracy, sensitivity, throughput and resolution of the measurements by allowing us to employ screening strategies such as FACS. Furthermore, use of fluorescence will allow for studies in which IbsC is produced at sub-lethal levels. This will take us closer to understanding their functions at endogenous levels.

While developing the fluorescence-based platform to study the interactions between the toxin and antitoxin, some aspects about the toxicity of the IbsC molecule and its regulation by the antitoxin revealed themselves to us. Hence, we embarked on a journey to follow these clues to expand our knowledge of the biology and regulation of these systems. We believe this knowledge will feed directly into improving the design of the platforms and protocols developed, by others and ourselves, to study these molecules and into advancing TA-based applications.

Our lab has previously deduced the sequence requirements of IbsC for its toxicity.[90] However, in chapter 2 we learnt that the toxicity of IbsC is not solely dependent on its sequence and structure but also is dosage dependent. Therefore, we set out to determine the RNA expression levels that could cause growth stasis.

We engineered *E. coli* strains with the capability to express different levels of the toxin and then established links between toxin RNA expression and growth to determine the toxic dosage. This work is important because IbsC seems to exert different functions in the cells, ranging from membrane damage, to growth stasis to death (and perhaps other functions yet to be discovered), according to its abundance levels. [78,117] The majority of studies have thus far emphasized on studying the effects of the IbsC molecule at overexpressed levels.[79,117,124] These engineered cells will allow us to study the function of these molecules at different abundance levels.

Evidence from our own work and that of others suggests that cells utilize multiple levels of regulation to fine-tune levels of toxin expression according to cellular requirements. Therefore, we decided to study regulation at the transcriptional (DNA level) and post-transcriptional (RNA) levels.

Since the interaction between the toxin and antitoxin is regulated by RNA-RNA interactions, cells must tightly control the ratio between the antitoxin and toxin RNAs to allow for a desired function by these systems. In chapter 2, we observed for the first time that the rescue capability of the antitoxin was concentration dependent. We utilized the engineered strains mentioned above, and performed simultaneous gene expression analyses and growth assays and determined the effective ratio between the two molecules for rescue. This is especially important because in bacteria, transcription and translation happen simultaneously in the same location. Consequently, for successful gene silencing by natural or artificial sRNAs and perhaps a desired phenotype to appear, the expressed sRNA must bind the target mRNA before the ribosome, which is present in the cells abundantly. Therefore, the sRNAs have to be present at sufficiently higher levels relative to the mRNA. We believe the knowledge of natural [sRNA]/[mRNA] ratios, will provide insights to guide decisions about these design parameters.

To study the regulation of the toxin at transcriptional levels, we decided to study the natural promoters. Our lab had previously studied the *ibsC* promoter [87]; therefore we characterized the *sibC* promoter. Interestingly, we elucidated multiple positive and negative regulatory regions upstream of the *sibC* core promoter. In addition, we discovered that a negative regulatory region that existed upstream of the *ibsC* core promoter had a more dominant role in keeping the toxin levels low, rather than repression by the antitoxin at the RNA level. Our observations supported our hypothesis that multiple levels of regulation exist to fine-tune the toxin levels. In addition, we believe that the positive and negative regulatory elements discovered could serve as elements to add to the current available repertoire of promoters for fine-tuning their power to drive expression of their downstream genes.

Having successfully measured expression of the toxin and antitoxin in engineered cells, we decided to study their physiological expression levels next. This was one of the biggest pieces of the IbsC/SibC puzzle and took us steps closer to understanding the functions of these molecules. As previously mentioned, because this TA pair relies on RNA-RNA interactions to achieve regulation, the levels of the toxin mRNA and its antitoxin RNA need to be carefully regulated to ensure the most beneficial antitoxin RNA/toxin mRNA ratio for a desired function. Therefore, it is important to measure the levels of these RNA pairs as a function of developmental stage or environmental conditions in order to fully elucidate their functions.[73] However, it is an extremely difficult task because of their low copy numbers[74] and technical challenges for detecting low-abundant RNAs. [75] Here, we made an effort to establish an effective RT-qPCR method that could reliably detect the IbsC mRNA produced by the native chromosomal *ibsC* gene. To our knowledge this is the first report of *ibsC* toxin expression in WT *E. coli* cells. We further confirmed the reliability of the method using RNA-seq. We then used this method to measure the abundance of SibC and the IbsC mRNA in wild

type *E. coli* cells at different stages of growth and showed that their transcription was growth phase dependent. Contrary to our expectations, we observed that in wild type *E. coli* cells, the toxin gene peaked in its expression during the exponential phase and was repressed during the stationary phase. These observations suggested to us some kind of functionality other than toxicity or death by IbsC, during growth. Our *sibC* gene expression analyses also revealed an interesting pattern. RNA levels continued to gradually rise in wild type *E. coli* cells, peaking at early stationary and gradually decreasing into late stationary. Again, this was contrary to our expectation for the [antitoxin]/[toxin] ratio to be at its peak during the exponential phase to ensure effective repression of the toxin, hence allowing for growth. These physiological levels of antitoxin sRNA levels relative to the toxin mRNA may have implications for functions by these molecules other than growth stasis by the toxin and rescue by the antitoxin.

Additionally, we estimated average copy numbers of SibC and the IbsC mRNA per cell. In general, most data in biology are qualitative or relatively quantitative. Similarly, transcriptomics has focused on relative expression changes between different conditions. Therefore, little is known about actual cellular numbers of RNAs and how gene regulation affects these numbers, but ultimately many biological processes will only be understood if investigated with absolute quantitative data. In addition, quantitative data allow for comparison of data sets, current and future [73]. We found that in a population of cells at mid-log phase ($OD_{600} \approx 0.5$) in balanced growth at 37 degrees in LB media, these RNAs existed at < 1copy per cell, suggesting that these genes are tightly repressed.

The last section of chapter 4 describes our efforts in measuring expression of the *ibsC/sibC* genes from a stressed population of cells that we came upon serendipitously and our attempts at mimicking these stressed conditions to reproduce the results. For the first time, we observed higher levels of the toxin relative to the antitoxin, suggesting a role for these toxins in protection against

stress. Our *sibC* gene expression analyses revealed a similar pattern as that observed in wild type *E. coli* cells during different phases of growth, except at lower levels relative to the toxin. Whether the expression of the antitoxin is independently regulated or together, as a part of a set of regulators is not yet known. It is tempting to speculate that the SibC antitoxin, as it is more highly expressed than its toxin mRNA in WT cells under natural physiological conditions, and varies in its expression throughout growth phases even when the toxin is fully repressed (or when the toxin is maintained at the same level throughout different phases of growth in the case of stressed cells), moonlights to regulate additional target RNAs.[135]

The effective RNA detection method we optimized; the new knowledge on the expression of the chromosomal *ibsC-sibC* genes acquired by this method in healthy and stressed population of cells; and the engineered cells containing additional, regulatable *ibsC-sibC* genes, will lend themselves as useful tools to finally interrupt the long pause in the study of this toxin/antitoxin pair. They will open the avenue to new discoveries pertaining to the biological relevance of TA systems. In addition, the future of many proposed toxin-antitoxin based applications (i.e. antiviral, anticancer, antibacterial therapies, containment of GMOs), would require controlled induction of these toxin-antitoxin molecules for controlled cytotoxic or cytoprotective action by them in undesired or desired cells respectively, and knowledge of their accurate dosages. Our studies here have paved the way for future applications.

Data chapters 2-4 mainly focus on the IbsC-SibC TA pair. In chapter 5, we examined the other Ibs and Sib homologous copies. We had been wondering for a long time about why five copies of these homologous genes existed in *E. coli*. To prove functionality by these genes, we performed expression analyses.

In spite of the great similarity between all five copies, we showed there was

enough dissimilarity in some regions to allow for specific amplification, hence detection by RT-qPCR. We confirmed the reaction specificities by gel electrophoresis, Sanger sequencing and melt curve analyses. We measured expression from all five *ibs/sib* gene pairs during early log and late log phases of growth. Interestingly, we detected expression from all five toxin genes but only during the early log phase. Expression levels for all genes, seemed to be switched off during the late log phase, suggesting a role for these toxin genes during early phases of active growth by the cells. Also, expression levels seemed to be maintained at low levels. Among Fozo's hypotheses for the existence of multiple copies of these genes was to ensure sufficient levels of combined production from all the different copies while keeping expression levels from each individual gene at low levels. Our observations support this hypothesis. The hypothesis assumes that the genes have redundant functions and the cell requires a certain level of these molecules. [78]

Intriguingly, the antitoxins seemed to display different expression behaviors from one another. *sibA* and *D* seemed to be expressed at higher levels during the early log phase than late log. This was in contrast to what we previously reported for *sibC*. *sibB* did not seem to differ in its expression levels between the two growth stages and no expression was detected from *sibE* from either of the growth stages. These observations point to different functions by the different sRNAs. Also, their different expression patterns from one another, in spite of their cognate toxins showing similar patterns further supports our earlier hypothesis about these sRNA having targets other than their cognate Ibs toxins.

We are now at an exciting phase to gain an understanding of function of these genes by performing expression analyses under different physiological and environmental conditions. We believe we have laid the groundwork by taking the

initiative to establish the fundamental knowledge of expression from these genes and by optimizing approaches for their specific detection. In addition to RT-qPCR, we developed protocols in this chapter that enable us to take advantage of sensitive, and high throughput techniques such as NanoString and next generation sequencing to further facilitate the study of expression from these genes under a multitude of conditions.

6.2 SibC as a potentially useful tool for synthetic biology

Since ancient times, humans have enjoyed the benefits of microorganisms such as their use for the production of fermented food (e.g. cheese, sourdough, beer, wine, and vinegar) and other products such as enzymes, amino acids and fine chemicals. Continuously, efforts have been made to identify, breed and select for better performing strains. In recent years, this "forward" genetics approach has given its place to an opposite "reverse" genetics approach. In the latter, genetic perturbations are introduced to a particular gene or genes of interest to accumulate genetic information and engineer the strains for higher performance. [143]

The discovery of important and diverse functional RNAs in cells which act by virtue of their interactions with other nucleic acids, proteins and small molecules, has inspired scientists to design synthetic RNA-based switches of gene expression which are well-suited to high throughput and large-scale analyses. [81,82] These switches allow for the fine modulation of chromosomal gene expression without the modification of chromosomal sequences (as inevitable by conventional, homologous recombination-based techniques of knockout, knock-in and allelic exchange). This makes them widely applicable, especially to essential genes whose complete deletion cannot be tolerated by cells. [143] In

addition, there are unique properties of RNAs that make them more attractive design substrates relative to proteins. [82,144]

Great applications from environment and agriculture to health and medicine have been offered by artificial sRNAs (afsRNAs) and are well documented in the literature, but mostly in eukaryotes. RNA interference has become a powerful gene-silencing tool in eukaryotes, while such effective and convenient RNA silencing methods are still lacking for prokaryotic organisms and urgently needed. Non-coding small RNAs (nc sRNA) that post transcriptionally modulate bacterial gene expression were first discovered in 1984. [144]However, owing to a lack of understanding about the RNA silencing mechanisms in bacteria, only a few synthetic sRNAs have been designed for proof of concept studies ever since. Moreover, most of these synthetic sRNAs are synthetic riboswitches that are *cis*-acting and alter the downstream secondary structure of mRNAs in a particular, context dependent fashion. This makes them tricky to apply to genome-wide studies. [145] Even those that were not riboswitches remained ineffective, especially in *E. coli* due to their low and variant silencing efficiencies, until recently. In 2006, a group discovered that by just adding a hairpin structure to previously designed afsRNAs, their silencing efficiencies and life times in cells could be significantly improved. [143] Ever since, afsRNA-mediated gene silencing has been many times proven to be effective in *E. coli* and other bacteria.

Recently, a group used a library of 130 synthetic sRNAs, to identify gene targets whose silencing enabled substantial increase in the production of valuable compounds in *E. coli*. [145] It is also anticipated that afsRNAs will be useful for our battles against bacterial resistance by designing them to target resistance, virulence or essential genes.

Evidently, much progress has been made in synthetic biology by the design of

RNA-based devices, which offer great tools for diverse applications in bacteria (from their metabolic engineering for improved functions to killing them) and eukaryotes. However, a main limitation in the field is the availability of new and diverse functional RNA components, which form the basic building blocks for the construction of more complex genetic circuits.

Functional RNA components can be simply derived from naturally occurring elements. The RNA component can be "harvested" from its native surroundings and used in a synthetic genetic context to output a desired activity. Also, the native function of the naturally occurring RNA element can be changed to new functions.[82] This is the future that I envision for the Sib family of antitoxins.

The Sibs, as natural, antisense, gene-silencing sRNAs, have exciting potential for generation of new functional RNA-based genetic parts that can be expanded into more complex genetic circuits.

We believe that by studying the interaction between the Sib sRNAs and their target mRNAs, Ibs, we can gain insight into the mechanistic action of the Sib antitoxins and deduce their sequence and structural requirements. This knowledge will subsequently enable us to design efficient artificial scaffolds by substituting the original target recognition sequence for the cognate Ibs with one to target any gene of interest. Eventually, transferrable capabilities into other organisms and extendable architecture into more complex genetic devices are some of the desired features that we envision for these molecules. Such engineered riboregulators may lend further insight into mechanistic actions of the endogenous Sib molecules.

In chapter 2, we presented a fluorescence-based platform to monitor activity of WT or mutant SibC against ibsC. It would be useful to utilize this system and determine the minimal requirements for their interactions. Based on this information, it should be possible to design a generic gene knockdown mechanism for *E. coli* or perhaps more clinically relevant pathogens, by

engineering the smallest and most efficient scaffold that shares common features with SibC and has an exchangeable target recognition domain.

In a similar attempt, a group generated a series of SibC derivatives by performing 5′ and 3′ truncations, and constructing SibC/D chimeras. Subsequently, they assayed the ability of these molecules to repress IbsC expression, thereby revealing the regions of SibC that are essential for IbsC mRNA recognition/interaction. Their data indicated at least two target recognition domains, TRD1 and TRD2 that function independently.[124] To complement these studies, one may investigate if there are more of these domains, by performing more systematic truncations studies.

In a subsequent study by the same group, an effective artificial small RNA (afsRNA) was designed to target lacZ. They adopted one of the effective SibC RNA derivatives from their previous study, including TRD2 in its intact surrounding specialized structure, and replaced TRD2 with a new sequence recognizing the LacZ mRNA while maintaining the original structure and stability.[91] This study confirmed the effectiveness of this approach for specific gene silencing in *E. coli*. One may address whether by replacing the target recognition domains of the scaffolds developed by this group while maintaining their structures, they could be used to target other genes; perhaps encoding physiologically or clinically relevant proteins. Antibacterial resistance, virulence and essential genes would make interesting targets.

In addition to SibC-derived scaffolds, I believe that scaffolds derived from some well-known trans-RNAs in *E. coli* can be designed against the same selected targets to provide additional insights. Although the *sib/ibs* pairs are operated in *cis* in the cell, it has been shown by others and us that SibC and IbsC expressed from different plasmids inside a cell are able to interact with each other in *trans*. This indicates that the mechanisms of *cis*-encoded and *trans*-encoded sRNAs do not

significantly differ from one another.[124]It would be interesting if effective sRNAs against IbsC could be isolated among the non-cognate scaffolds and if any similarities between them could be found. By comparing and contrasting effective and non-effective afsRNA-target pairs, useful insights can be gained into the underlying mechanisms of these sRNAs, especially those of the less well known SibC, and how these mechanisms depend on the sRNA scaffolds and/or target mRNAs. The knowledge from these studies will subsequently enable the rational design of effective afsRNAs [144]. In addition, afsRNAs and targets in these studies can all be combined to systematically study their specificities. [146] These efforts will be useful as specificities of a lot of the designed and discovered sRNAs to date, including SibC, have not been examined thoroughly.

Based on the modular architecture of natural sRNAs that have inspired the design of their artificial counterparts, two major elements of any afsRNAs to be designed are a scaffold (containing a consensus secondary structure for recruiting cellular proteins involved in gene silencing such as hfq) and an mRNA basepairing region. [147] Principles, processes and different approaches for their designs can be found in the cited literature here. The designs can be of a rational or randomized nature or a combination of the two. Predictable and rational design of afsRNAs relies on the bits of information about variables that can enhance the silencing efficiency of these molecules and is a desirable long-term goal. However, numerous variables are likely to influence the overall performance of these sRNAs, many of which are still unknown. Unlike the mammalian microRNAs, bacterial sRNAs are highly diverse in their sequences and target interactions (e.g. number of base pairings, location of the target mRNA, etc.) and our current incomplete knowledge about them still limits our ability to rationally design effective afsRNAs. To avoid unforeseen limitations of a rational design, complete randomization of the target-binding region seems like the better alternative at the moment. It potentially allows for the identification of potent sRNAs without

knowing their underlying mechanisms.[144]

To design randomized afsRNAs, selected well-known bacterial sRNAs and SibC derivatives, whose scaffolds are desired for the design of new afsRNAs, can be cloned into different plasmids. By running an inverse PCR on these plasmids with primers containing randomized nucleotide sequences of 20-30 base pairs and re-ligating the mutagenized plasmids, plasmid libraries based on the different scaffolds can be prepared and later mixed in equal amounts to generate a combined library. The library is then transformed to cells that harbor target-*gfp* fusions (for fluorescence based screening) in their chromosome. [144] Having the *gfp* fused to the target genes, Fluorescence Activated Cell Sorter (FACS) can be used to sort out promising silencing afsRNAs in a vast library based on lowered detected levels of fluorescence in cells expressing them. This method of screening well suits to high throughput analyses. Also, some functions on the FACS machine will enable differentiation between the cells with highly repressed levels of fluorescence and those with no fluorescence due to, for example, mutations developed in the *gfp* gene. As a result, the number of false positives will be dramatically reduced.

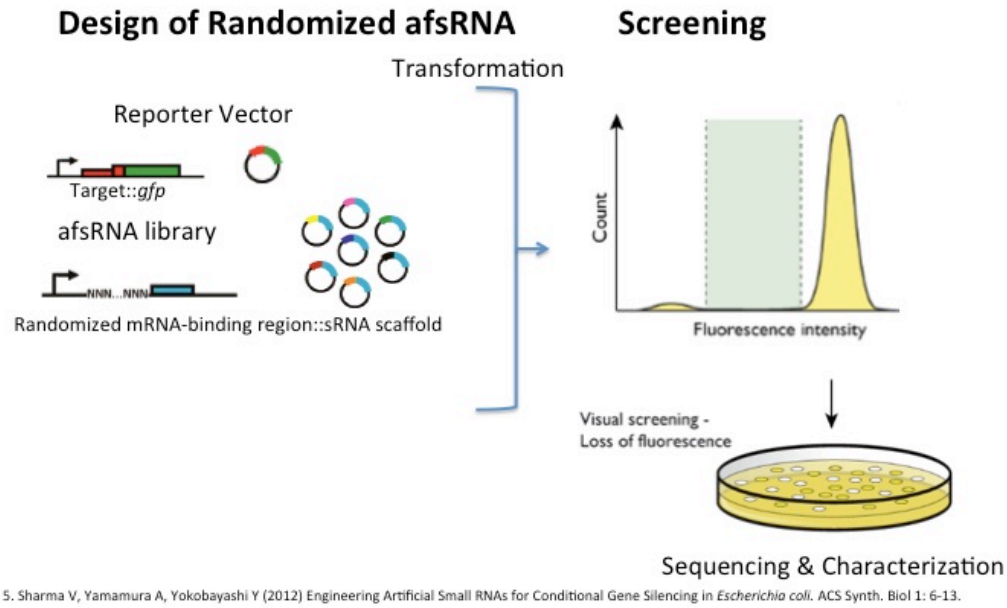


Figure 6.1. Design of Randomized afsRNAs, selection of the functional one with respect to gene-knockdown activities and downstream sequencing and characterization of the successful clones. Figure is Adapted from Sharma et al. [144], with permission from American Chemical Society, copyright (2012).

Furthermore, the cells sorted based on lowered fluorescence can be plated on selective agar plates to be potentially screened visually using a UV transilluminator. They can be restreaked on fresh plates to confirm their activity and clonal isolation. By purifying the plasmids and retransforming them into a fresh host, one can eliminate the possibility of host/plasmid mutations. Promising afsRNAs can subsequently be sequenced and further characterized in terms of their secondary structures, binding kinetics, stability, effective cellular ratio of [afsRNA]/[target mRNA], primary mechanism of action (i.e. Translation inhibition or mRNA degradation), specificity, interaction with other regulatory proteins, common features among afsRNAs and also among targets). Eventually we would like to broadly extend and deliver these afsRNA-base regulatory devices to systems other than *E. coli.*

Researchers are continuously generating afsRNA expression libraries to gradually cover whole bacterial genomes. These libraries will eventually make valuable and permanent research resources. These riboregulators will lend insight into mechanistic actions of endogenous RNA-based processes. Reciprocally, research progress on natural sRNAs will contribute to the improvement of afsRNA technology. Hence, we find it necessary to study the natural sRNA such as Sib both in artificial as well as their native biological contexts.

6.3 Drug-ability of the toxin

6.3.1 Potential involvement in persistence

The functions of endogenous levels of the proteins are still obscure. Studying the hypothesized role of persistence is of a particular interest to us because in spite of the fact that persistence poses a major challenge for the treatment of infectious diseases, as most common bacteria kill only the growing cells, it has been underappreciated for some time as a mechanism for bacteria to evade antibiotics. [80]Based on recent studies that have associated persister formation with dormancy and TA systems and our own observations of IbsC causing reversible growth stasis (discussed in chapter 4), we hypothesize that induction of *ibsC* leads to persister cell formation and therefore survival under times of stress such as antibiotic exposure. Sufficient evidence for our hypothesis will present IbsC with the potential to be an antibiotic combination target for *E. coli* and perhaps other more clinically relevant pathogens. I believe that answers to two smaller questions can raise sufficient evidence for our hypothesis: 1) Do persisters exhibit higher toxin expression relative to wild type cells? 2) Are the cells with higher levels of

the toxin expression, more likely to form persisters?

We have laid the groundwork to answer these questions. In chapter 2, we reported fluorescence constructs that we have created to report on the transient metabolic state of the cells (i.e persistence vs. no persistence). These intracellular reporters provide an opportunity to utilize high sensitivity techniques such as FACS in the study of persistence, to facilitate identification and isolation of dormant persisters among a population of cells. Additionally, we have elucidated some conditions under which persistence happens in an *E. coli* population of cells. To answer the first question above, we need to measure the *ibsC* toxin expression in these cells relative to wild type cells. Having optimized independent approaches for sensitive quantification of expression from these genes, we are now well equipped for expression analyses in these cells. To answer the second part of the question, we can utilize the cells we engineered to allow for fine control over production of these toxin genes (as described chapter 3). Transforming these cells simultaneously with our fluorescence persistence-reporter constructs would facilitate isolation and count of persister cells by FACS to investigate whether the over production of the toxin does indeed contribute to persister cell formation under the persistence-causing conditions established by others or us. These engineered cells would allow us to test the effects of varying levels of toxin on persister formation to establish a quantitative link between the two. The isolated persisters would have to be confirmed by subsequent growth (under antibiotic and no antibiotic conditions to confirm reversibility of the dormant state, hence persistence) and expression analyses of genes that are known to be affected by persistence.

It should be noted that IbsC might play a physiological role in persistence only under certain stress conditions and/or at certain levels of expression. Choice of the antibiotic and its mode of administration can also be important to cause persistence. Therefore, *ibsC* induction may need to be accompanied with different

temperature, nutrient, oxidative, acidity, osmotic and antibiotic stress conditions to (elucidate the right conditions for the toxin to) cause persistence. With the fluorescence-based platform we have developed, we can now take advantage of high throughput techniques to elucidate these specific conditions.

6.3.2 Artificial activation of toxins

Due to cytotoxic and cytostatic (bactericidal) capabilities of toxins, artificial activation of these genes for specific targeting and killing of pathogens has already been explored in the literature. [148,149] These strategies rely on discovery of compounds that either degrade the antitoxin or disrupt the interaction with the toxin (antisense mediated or peptide mediated disruptions) to liberate the toxin to exert its effects inside the cells that carry it. These approaches alter the ways in which nature has intended for these molecules to promote bacterial fitness. They either promote persistence, so the infecting bacteria can be treated with anti-persistence drugs or alternatively, they 'wake up' cells from persistence, so they are susceptible to existing antibiotics. Lastly, they activate toxins to exert their death effects.

6.3.3 Potential role of the ibs toxins in bacterial host fitness inside its host

It would be interesting to examine whether these toxin genes exist and are expressed in host-associated (aka intracellular) bacteria. These studies will complement the studies of

TA systems in free-living cells. If they are, their potential role in promoting the fitness of its carrier inside the infected cells should be examined. Precedents for this phenomenon exist in the literature. As an example, *Mycobacterium tuberculosis* has been shown to up-regulate the type II TA modules inside macrophages. [150] Similarly, *Salmonella enterica serovar Typhimurium* (*S. typhimurium*) has been shown to depend on a type II TA module named sehAB for survival in mice. [55] Production of the TA modules can be confirmed by highly sensitive gel-free proteomics applied to bacteria directly isolated from its host cells. High-resolution mass spectroscopy can identify proteins as small in size as about 100 AAs, in intracellular bacteria. For smaller sized peptides (eg. 30-50 AA), however, quantitative RT-PCR (qRT-PCR) can be confidently used to confirm their production and measure their amounts. Next, it would be useful to examine whether there is a link between toxin production and the pathogen's proliferation inside the infected cell. Given the anti proliferative activity of the toxin in free living cells, it would be interesting to determine whether the toxin assists the pathogen to adapt to distinct intracellular environments (e.g. by slowing down proliferation in a hostile environment). To this aim, varied host cell types in which bacteria display different growth rates should be utilized and TA production should be measured for the bacteria infecting each of the different hosts. Results will indicate whether there is a link between bacterial proliferation and the level of TA production inside the host cells. Finally, to examine whether there is any advantage to the bacteria by the differences observed in toxin production in the different host cells, a cell line with enhanced toxin production can be compared with the same host line with an isogenic mutant of the bacteria lacking the toxin for survival of the intracellular bacteria. These investigations will collectively throw light on whether or not the TA pair has an impact on the fitness of its carrying pathogen inside infected cells and will be important for the purpose of drug discovery.

6.4 Other Biological questions pertaining to toxin and antitoxin

6.4.1 Other biological functions

The functions of endogenous levels of the proteins are still obscure. The greatest impeding factor has been the lack of knowledge of expression. We have now gained knowledge of physiological levels of expression from these genes. In addition, we have optimized a strategy to quantitatively measure and track expression from genes that are expressed at low quantities. Furthermore, we have engineered strains that allow fine control over production of these molecules. As such, we are now at an exciting phase, well equipped to establish links between their expression levels and cellular behavior under different conditions (environmental, niche, stress, etc.), to gain insights about the biological functions of these molecules. To elucidate when these genes are expressed, it is important to study a variety of different conditions as all evidence so far suggests that these molecules have very specialized functions and are expressed under very specific conditions. The high-throughput techniques such as next-generation sequencing and Nanostring will allow us to screen a wide range of conditions.

6.4.2 Potential regulators of ibsC/sibC gene expression and targets for their RNAs

We have performed characterization of the toxin and antitoxin promoters and identified regions with positive and negative regulatory roles. To identify

potential regulators that can bind and act on these regions, pull-down assays can be carried out. Their identities can be subsequently determined using mass spectrometry. Also, the interactions between these regions and the isolated regulators can be studied by electromobility shift assays. The effects of such interactions on the expression of downstream genes can be evaluated by reporter assays. It should also be noted that autoregulation has been demonstrated in some type II TA systems. [151] It may be possible for IbsC to be involved in its own regulation.

We could also employ CLASH (crosslinking, ligation and sequencing hybrids) or similar technologies to detect RNA-RNA interactions of the antitoxin, and quantify the potential targets during different phases of growth in *E. coli*. These experiments would be valuable as our data suggest there might be other targets for Sib RNAs than just the cognate Ibs RNAs, hence different function than just repressing the toxin.

6.4.3 Absolute quantification of the toxin protein under different conditions

We have so far emphasized on quantification of RNAs as the antitoxin components of type I TA systems are sRNA and the interactions between type I toxin-antitoxin systems happen through RNA-RNA interactions. Therefore, their levels must be tightly controlled to ensure a desired function according to cellular demands. Measuring RNA levels can provide us with clues about the function by these molecules. However, to complement these studies, it would also be useful to detect and measure the toxin protein levels. It would be interesting to measure the relative abundance of the toxin proteins throughout different growth stages of *E. coli.* We could employ an absolute quantification method named AQUA.

We first need to examine whether the protein is detectable in biological samples. To do so, the whole cell lysate from the biological sample is prepared, proteolytically digested and assessed for the presence of digested peptides from the target protein. If the peptides are detectable, quantitatively measurements of concentration can be pursued. To create an internal standard, heavy-isotope labeled peptides (with the same sequence as target) using 13C- and/or 15N-labeled amino acids (typically arginine or lysine if trypsin is used to digest the sample) are ordered and used. The internal standard will be spiked into the biological samples. Knowing the exact amount of the internal standard peptide, we can measure the peak area ratios between the standard peptide and the target peptide to calculate the abundance of the target protein in the biological sample.

We believe that the suggested future studies above, will tremendously contribute to a comprehensive understanding of the TA biological systems.

6.5 References

1. Page, R. & Peti, W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* (2016). doi:10.1038/nchembio.2044

2. Ramisetty, B. C. M., Natarajan, B. & Santhosh, R. S. MazEF-mediated programmed cell death in bacteria: 'What is this?' *Critical Reviews in Microbiology* (2015). doi:10.3109/1040841X.2013.804030

3. Zhao, J. *et al.* Escherichia coli toxin gene hipA affects biofilm formation and DNA release. *Microbiol. (United Kingdom)* (2013). doi:10.1099/mic.0.063784-0

4. Wang, Y. *et al.* Functional RelBE-family toxin-antitoxin pairs affect

biofilm maturation and intestine colonization in Vibrio cholerae. *PLoS One* (2015). doi:10.1371/journal.pone.0135696

5. Maikova, A. *et al.* Discovery of new type I toxin–antitoxin systems adjacent to CRISPR arrays in Clostridium difficile. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky124

6. Fozo, E. M. New type i toxin-antitoxin families from 'wild' and laboratory strains of E. coli: Ibs-Sib, ShoB-OhsC and Zor-Orz. *RNA Biology* (2012). doi:10.4161/rna.22568

7. Van Melderen, L. & De Bast, M. S. Bacterial toxin-Antitoxin systems: More than selfish entities? *PLoS Genet.* **5**, (2009).

8. Jaffe, A., Ogura, T. & Hiraga, S. Effects of the ccd function of the F plasmid on bacterial growth. *J. Bacteriol.* (1985).

9. Gerdes, K., Rasmussen, P. B. & Molin, S. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc. Natl. Acad. Sci.* (1986). doi:10.1073/pnas.83.10.3116

10. Van Melderen, L. Toxin-antitoxin systems: Why so many, what for? *Curr. Opin. Microbiol.* **13**, 781–785 (2010).

11. Wen, Y., Behiels, E. & Devreese, B. Toxin-Antitoxin systems: Their role in persistence, biofilm formation, and pathogenicity. *Pathogens and Disease* (2014). doi:10.1111/2049-632X.12145

12. Ramage, H. R., Connolly, L. E. & Cox, J. S. Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: Implications for pathogenesis, stress responses, and evolution. *PLoS Genet.* (2009). doi:10.1371/journal.pgen.1000767

13. Aizenman, E., Engelberg-Kulka, H. & Glaser, G. An Escherichia coli

chromosomal ‘addiction module’ regulated by guanosine [corrected] 3’,5’-bispyrophosphate: a model for programmed bacterial cell death. *Proc. Natl. Acad. Sci.* (1996). doi:10.1073/pnas.93.12.6059

14. Zhang, Y. *et al.* MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Mol. Cell* (2003). doi:10.1016/S1097-2765(03)00402-7

15. Hazan, R. & Engelberg-Kulka, H. Escherichia coli mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Mol. Genet. Genomics* (2004). doi:10.1007/s00438-004-1048-y

16. Engelberg-Kulka, H. mazEF: a chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria. *J. Cell Sci.* (2005). doi:10.1242/jcs.02619

17. Christensen, S. K., Mikkelsen, M., Pedersen, K. & Gerdes, K. RelE, a global inhibitor of translation, is activated during nutritional stress. *Proc. Natl. Acad. Sci.* **98**, 14328–14333 (2001).

18. Bigger, J. W. Treatment of staphylococcal infections with penicillin by intermittent sterilisation. *Lancet* (1944). doi:10.1016/S0140-6736(00)74210-3

19. Hobby, G. L., Meyer, K. & Chaffee, E. Observations on the Mechanism of Action of Penicillin. *Exp. Biol. Med.* (1942). doi:10.3181/00379727-50-13773

20. Wang, X. & Wood, T. K. Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Appl. Environ. Microbiol.* (2011). doi:10.1128/AEM.05068-11

21. Yamaguchi, Y. & Inouye, M. Regulation of growth and death in

Escherichia coli by toxin-antitoxin systems. *Nature Reviews Microbiology* (2011). doi:10.1038/nrmicro2651

22. Lewis, K. Persister Cells. *Annu. Rev. Microbiol.* (2010). doi:10.1146/annurev.micro.112408.134306

23. Gil, F., Pizarro-Guajardo, M., Álvarez, R., Garavaglia, M. & Paredes-Sabja, D. Clostridium difficile recurrent infection: Possible implication of TA systems. *Future Microbiology* (2015). doi:10.2217/fmb.15.94

24. Dörr, T., Vulić, M. & Lewis, K. Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli. *PLoS Biol.* (2010). doi:10.1371/journal.pbio.1000317

25. Moyed, H. S. & Broderick, S. H. Molecular cloning and expression of hipA, a gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. *J. Bacteriol.* (1986). doi:10.1128/jb.166.2.399-403.1986

26. Slattery, A., Victorsen, A. H., Brown, A., Hillman, K. & Phillips, G. J. Isolation of highly persistent mutants of Salmonella enterica serovar typhimurium reveals a new toxin-antitoxin module. *J. Bacteriol.* (2013). doi:10.1128/JB.01397-12

27. Van Acker, H., Sass, A., Dhondt, I., Nelis, H. J. & Coenye, T. Involvement of toxin-antitoxin modules in Burkholderia cenocepacia biofilm persistence. *Pathog. Dis.* (2014). doi:10.1111/2049-632X.12177

28. Peeters, C. *et al.* Burkholderia pseudomultivorans sp. nov., a novel Burkholderia cepacia complex species from human respiratory samples and the rhizosphere. *Syst. Appl. Microbiol.* (2013). doi:10.1016/j.syapm.2013.06.003

29. Voloudakis, A. E., Reignier, T. M. & Cooksey, D. A. Regulation of resistance to copper in Xanthomonas axonopodis pv. vesicatoria. *Appl. Environ. Microbiol.* (2005). doi:10.1128/AEM.71.2.782-789.2005

30. Muranaka, L. S., Takita, M. A., Olivato, J. C., Kishi, L. T. & de Souza, A. A. Global expression profile of biofilm resistance to antimicrobial compounds in the plant-pathogenic bacterium Xylella fastidiosa reveals evidence of persister cells. *J. Bacteriol.* (2012). doi:10.1128/JB.00436-12

31. Merfa, M. V., Niza, B., Takita, M. A. & De Souza, A. A. The MqsRA toxin-antitoxin system from Xylella fastidiosa plays a key role in bacterial fitness, pathogenicity, and persister cell formation. *Front. Microbiol.* (2016). doi:10.3389/fmicb.2016.00904

32. Lewis, K. Persister cells, dormancy and infectious disease. *Nature Reviews Microbiology* (2007). doi:10.1038/nrmicro1557

33. T. Rybtke, M. *et al.* The Implication of Pseudomonas aeruginosa Biofilms in Infections. *Inflamm. Allergy - Drug Targets* (2011). doi:10.2174/187152811794776222

34. Stewart, P. S. & Franklin, M. J. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology* (2008). doi:10.1038/nrmicro1838

35. Mah, T. F. C. & O'Toole, G. A. Mechanisms of biofilm resistance to antimicrobial agents. *Trends in Microbiology* (2001). doi:10.1016/S0966-842X(00)01913-2

36. Bjarnsholt, T. The role of bacterial biofilms in chronic infections. *APMIS* (2013). doi:10.1111/apm.12099

37. Kim, Y. & Wood, T. K. Toxins Hha and CspD and small RNA regulator Hfq are involved in persister cell formation through MqsR in Escherichia

coli. *Biochem. Biophys. Res. Commun.* (2010). doi:10.1016/j.bbrc.2009.11.033

38. González Barrios, A. F. *et al.* Autoinducer 2 controls biofilm formation in Escherichia coli through a novel motility quorum-sensing regulator (MqsR, B3022). *J. Bacteriol.* (2006). doi:10.1128/JB.188.1.305-316.2006

39. Kasari, V., Kurg, K., Margus, T., Tenson, T. & Kaldalu, N. The Escherichia coli mqsR and ygiT genes encode a new toxin-antitoxin pair. *J. Bacteriol.* (2010). doi:10.1128/JB.01266-09

40. Ren, D., Bedzyk, L. A., Thomas, S. M., Ye, R. W. & Wood, T. K. Gene expression in Escherichia coli biofilms. *Appl. Microbiol. Biotechnol.* (2004). doi:10.1007/s00253-003-1517-y

41. Wang, X. *et al.* Antitoxin MqsA helps mediate the bacterial general stress response. *Nature Chemical Biology* (2011). doi:10.1038/nchembio.560

42. Soo, V. W. C. & Wood, T. K. Antitoxin MqsA represses curli formation through the master biofilm regulator CsgD. *Sci. Rep.* (2013). doi:10.1038/srep03186

43. Brown, B. L., Lord, D. M., Grigorius, S., Peti, W. & Pages, R. The Escherichia coli toxin MqsR destabilizes the transcriptional repression complex formed between the antitoxin MqsA and the mqsRA operon promoter. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M112.421008

44. García-Contreras, R., Zhang, X. S., Kim, Y. & Wood, T. K. Protein translation and cell death: The role of rare tRNAs in biofilm formation and in activating dormant phage killer genes. *PLoS One* (2008). doi:10.1371/journal.pone.0002394

45. Kim, Y., Wang, X., Ma, Q., Zhang, X. S. & Wood, T. K. Toxin-antitoxin

systems in Escherichia coli influence biofilm formation through YjgK (TabA) and fimbriae. *J. Bacteriol.* (2009). doi:10.1128/JB.01465-08

46. Kolodkin-Gal, I., Verdiger, R., Shlosberg-Fedida, A. & Engelberg-Kulka, H. A differential effect of E. coli toxin-antitoxin systems on cell death in liquid media and biofilm formation. *PLoS One* (2009). doi:10.1371/journal.pone.0006785

47. Chatterjee, S., Wistrom, C. & Lindow, S. E. A cell-cell signaling sensor is required for virulence and insect transmission of [i]Xylella fastidiosa[/i] . *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0712236105

48. Provenzano, D. & Klose, K. E. Altered expression of the ToxR-regulated porins OmpU and OmpT diminishes Vibrio cholerae bile resistance, virulence factor expression, and intestinal colonization. *Proc. Natl. Acad. Sci.* (2000). doi:10.1073/pnas.170219997

49. Wang, H. *et al.* Catalases Promote Resistance of Oxidative Stress in Vibrio cholerae. *PLoS One* (2012). doi:10.1371/journal.pone.0053383

50. Yildiz, F. H. & Visick, K. L. Vibrio biofilms: so much the same yet so different. *Trends in Microbiology* (2009). doi:10.1016/j.tim.2008.12.004

51. Iqbal, N., Guérout, A. M., Krin, E., Le Roux, F. & Mazel, D. Comprehensive functional analysis of the 18 Vibrio cholerae N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron. *J. Bacteriol.* (2015). doi:10.1128/JB.00108-15

52. Christensen-Dalsgaard, M., Jørgensen, M. G. & Gerdes, K. Three new RelE-homologous mRNA interferases of Escherichia coli differentially induced by environmental stresses. *Mol. Microbiol.* (2010). doi:10.1111/j.1365-2958.2009.06969.x

53. Georgiades, K. & Raoult, D. Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. *PLoS One* (2011). doi:10.1371/journal.pone.0017962

54. Norton, J. P. & Mulvey, M. A. Toxin-Antitoxin Systems Are Important for Niche-Specific Colonization and Stress Resistance of Uropathogenic Escherichia coli. *PLoS Pathog.* (2012). doi:10.1371/journal.ppat.1002954

55. De la Cruz, M. A. *et al.* A Toxin-Antitoxin Module of Salmonella Promotes Virulence in Mice. *PLoS Pathog.* (2013). doi:10.1371/journal.ppat.1003827

56. Yeo, C., Abu Bakar, F., Chan, W., Espinosa, M. & Harikrishna, J. Heterologous Expression of Toxins from Bacterial Toxin-Antitoxin Systems in Eukaryotic Cells: Strategies and Applications. *Toxins (Basel).* **8**, 49 (2016).

57. Kristoffersen, P., Jensen, G. B., Gerdes, K. & Piškur, J. Bacterial toxin-antitoxin gene system as containment control in yeast cells. *Appl. Environ. Microbiol.* **66**, 5524–5526 (2000).

58. De la Cueva-Méndez, G., Mills, A. D., Clay-Farrace, L., Diaz-Orejas, R. & Laskey, R. A. Regulatable killing of eukaryotic cells by the prokaryotic proteins Kid and Kis. *EMBO J.* **22**, 246–251 (2003).

59. Zielenkiewicz, U., Kowalewska, M., Kaczor, C. & Cegłowski, P. In vivo interactions between TA proteins of streptococcal plasmid pSM19035, Epsilon and Zeta, in yeast. *J Bacteriol* **191**, 3677–3684 (2009).

60. Kobayashi, K., Munemura, I., Hinata, K. & Yamamura, S. Bisexual sterility conferred by the differential expression of barnase and barstar: A simple and efficient method of transgene containment. *Plant Cell Rep.* **25**,

1347–1354 (2006).

61. Bisht, N. C. *et al.* Effective restoration of male-sterile (barnase) lines requires overlapping and higher levels of barstar expression: A multi-generation field analysis in Brassica juncea. *J. Plant Biochem. Biotechnol.* **24**, 393–399 (2015).

62. Slanchev, K., Stebler, J., de la Cueva-Mendez, G. & Raz, E. Development without germ cells: The role of the germ line in zebrafish sex differentiation. *Proc. Natl. Acad. Sci.* **102**, 4074–4079 (2005).

63. Chen, Z. *et al.* Enhancement of the Gene Targeting Efficiency of Non-Conventional Yeasts by Increasing Genetic Redundancy. *PLoS One* **8**, (2013).

64. Nehlsen, K., Herrmann, S., Zauers, J., Hauser, H. & Wirth, D. Toxin-antitoxin based transgene expression in mammalian cells. *Nucleic Acids Res.* **38**, (2009).

65. Chono, H. *et al.* Acquisition of HIV-1 Resistance in T Lymphocytes Using an ACA-Specific E. coli mRNA Interferase . *Hum. Gene Ther.* **22**, 35–43 (2010).

66. Unterholzner, S. J., Poppenberger, B. & Rozhon, W. Biology , identification , and application Toxin – antitoxin systems. *Mob. Genet. Elements* **3**, e26219: 1-13 (2013).

67. Chono, H. *et al.* In vivo safety and persistence of endoribonuclease gene-transduced CD4+ t cells in cynomolgus macaques for HIV-1 gene therapy model. *PLoS One* **6**, (2011).

68. Park, J.-H., Yamaguchi, Y. & Inouye, M. Intramolecular Regulation of the Sequence-Specific mRNA Interferase Activity of MazF Fused to a MazE

Fragment with a Linker Cleavable by Specific Proteases. *Appl. Environ. Microbiol.* **78**, 3794–3799 (2012).

69. Preston, M. A. *et al.* Repurposing a Prokaryotic Toxin-Antitoxin System for the Selective Killing of Oncogenically Stressed Human Cells. *ACS Synth. Biol.* **5**, 540–546 (2016).

70. Yamamoto, T. A., Gerdes, K. & Tunnacliffe, A. Bacterial toxin RelE induces apoptosis in human cells. *FEBS Lett.* **519**, 191–194 (2002).

71. Shimazu, T. *et al.* NBK/BIK antagonizes MCL-1 and BCL-XL and activates BAK-mediated apoptosis in response to protein synthesis inhibition. *Genes Dev.* **21**, 929–941 (2007).

72. Fozo, E. M. *et al.* Abundance of type I toxin-antitoxin systems in bacteria: Searches for new candidates and discovery of novel families. *Nucleic Acids Res.* **38**, 3743–3759 (2010).

73. Marguerat, S. *et al.* Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151**, 671–683 (2012).

74. Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science (80-.).* **329**, 533–538 (2010).

75. Walkowiak, M., Mleczko, A. M. & Bąkowska-Żywicka, K. Evaluation of methods for the detection of low-abundant snoRNA-derived small RNAs in Saccharomyces cerevisiae. *Biotechnologia* **97**, 19–26 (2016).

76. Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B. & Bartel, D. P. Vertebrate microRNA genes. *Science* **299**, 1540 (2003).

77. Krichevsky, A. M., King, K. S., Donahue, C. P., Khrapko, K. & Kosik, K.

S. A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA* **9**, 1274–1281 (2003).

78. Fozo, E. M. New type I toxin-antitoxin families from "wild" and laboratory strains of E. coli . *RNA Biol.* **9**, 1504–1512 (2012).

79. Fozo, E. M., Hemm, M. R. & Storz, G. Small Toxic Proteins and the Antisense RNAs That Repress Them. *Microbiol. Mol. Biol. Rev.* **72**, 579–589 (2008).

80. Dawson, C. C., Intapa, C. & Jabra-Rizk, M. A. "Persisters": Survival at the Cellular Level. *PLoS Pathog.* **7**, e1002121 (2011).

81. Isaacs, F. J., Dwyer, D. J. & Collins, J. J. RNA synthetic biology. *Nature Biotechnology* **24**, 545–554 (2006).

82. Liang, J. C., Bloom, R. J. & Smolke, C. D. Engineering Biological Systems with Synthetic RNA Molecules. *Molecular Cell* **43**, 915–926 (2011).

83. Lutz, R. & Bujard, H. Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).

84. Mok, W. W. K. *et al.* Identification of a toxic peptide through bidirectional expression of small RNAs. *ChemBioChem* **10**, 238–241 (2009).

85. Campbell, T. L. & Brown, E. D. Characterization of the depletion of 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase in Escherichia coli and Bacillus subtilis. *J. Bacteriol.* **184**, 5609–5618 (2002).

86. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–5 (2000).

87. Mok, W. Characterization of a type I toxin-antitoxin system in Escherichia

coli. (2012).

88. Bartlett, M. S. & Gourse, R. L. Growth rate-dependent control of the rrnB P1 core promoter in Escherichia coli. *J. Bacteriol.* **176**, 5560–4 (1994).

89. Han, K., Kim, K. S., Bak, G., Park, H. & Lee, Y. Recognition and discrimination of target mRNAs by Sib RNAs, a cis-encoded sRNA family. *Nucleic Acids Res.* **38**, 5851–5866 (2010).

90. Mok, W. W. K., Patel, N. H. & Li, Y. Decoding toxicity: Deducing the sequence requirements of IbsC, a type i toxin in Escherichia coli. *J. Biol. Chem.* **285**, 41627–41636 (2010).

91. Park, H., Bak, G., Kim, S. C. & Lee, Y. Exploring sRNA-mediated gene silencing mechanisms using artificial small RNAs derived from a natural RNA scaffold in Escherichia coli. *Nucleic Acids Res.* **41**, 3787–3804 (2013).

92. Fozo, E. M. *et al.* Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.* **70**, 1076–1093 (2008).

93. Nakashima, N. & Miyazaki, K. Bacterial cellular engineering by genome editing and gene silencing. *International Journal of Molecular Sciences* **15**, 2773–2793 (2014).

94. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).

95. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97**, 6640–6645 (2000).

96. Mok, W. W. K. *et al.* Identification of a toxic peptide through bidirectional

expression of small RNAs. *ChemBioChem* **10**, 238–241 (2009).

97. Kuryllo, K., Jahanshahi, S., Zhu, W., Brown, E. D. & Li, Y. A dual reporter system for detecting RNA interactions in bacterial cells. *ChemBioChem* **15**, 2703–2709 (2014).

98. Leplae, R. *et al.* Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* **39**, 5513–5525 (2011).

99. Pandey, D. P. & Gerdes, K. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* **33**, 966–976 (2005).

100. Goormaghtigh, F. *et al.* Reassessing the role of type II toxin-antitoxin systems in formation of escherichia coli type II persister cells. *MBio* **9**, (2018).

101. Magnuson, R. D. Hypothetical functions of toxin-antitoxin systems. *Journal of Bacteriology* **189**, 6089–6092 (2007).

102. Norton, J. P. & Mulvey, M. A. Toxin-antitoxin systems are important for Niche-Specific Colonization and Stress Resistance of Uropathogenic Escherichia coli. *PLOS Pathog.* (2012).

103. Chan, W. T., Balsa, D. & Espinosa, M. One cannot rule them all: Are bacterial toxins-antitoxins druggable? *FEMS Microbiology Reviews* **39**, 522–540 (2015).

104. Preston, M. A. *et al.* Repurposing a Prokaryotic Toxin-Antitoxin System for the Selective Killing of Oncogenically Stressed Human Cells. *ACS Synth. Biol.* (2016). doi:10.1021/acssynbio.5b00096

105. Chono, H. *et al.* Acquisition of HIV-1 Resistance in T Lymphocytes Using

an ACA-Specific *E. coli* mRNA Interferase. *Hum. Gene Ther.* **22**, 35–43 (2011).

106. Chono, H. *et al.* In vivo safety and persistence of endoribonuclease gene-transduced CD4+ T cells in cynomolgus macaques for HIV-1 gene therapy model. *PLoS One* **6**, e23585 (2011).

107. Shapira, A. *et al.* Removal of hepatitis C virus-infected cells by a zymogenized bacterial toxin. *PLoS One* **7**, (2012).

108. Williams, J. J. & Hergenrother, P. J. Artificial activation of toxin-antitoxin systems as an antibacterial strategy. *Trends in Microbiology* (2012). doi:10.1016/j.tim.2012.02.005

109. Bisht, N. C., Jagannath, A., Gupta, V., Kumar Burma, P. & Pental, D. A two gene - Two promoter system for enhanced expression of a restorer gene (barstar) and development of improved fertility restorer lines for hybrid seed production in crop plants. *Mol. Breed.* (2004). doi:10.1023/B:MOLB.0000038002.45312.08

110. Kobayashi, K., Munemura, I., Hinata, K. & Yamamura, S. Bisexual sterility conferred by the differential expression of barnase and barstar: A simple and efficient method of transgene containment. *Plant Cell Rep.* **25**, 1347–1354 (2006).

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

112. Stankovic, K. M. & Corfas, G. Real-time quantitative RT-PCR for low-abundance transcripts in the inner ear: Analysis of neurotrophic factor expression. *Hear. Res.* **185**, 97–108 (2003).

113. Fozo, E. M. *et al.* Abundance of type I toxin-antitoxin systems in bacteria: Searches for new candidates and discovery of novel families. *Nucleic Acids Res.* **38**, 3743–3759 (2010).

114. Hershberg, R., Altuvia, S. & Margalit, H. A survey of small RNA-encoding genes in Escherichia coli. *Nucleic Acids Research* **31**, 1813–1820 (2003).

115. Mok, W. W. K., Patel, N. H. & Li, Y. Decoding toxicity: Deducing the sequence requirements of IbsC, a type i toxin in Escherichia coli. *J. Biol. Chem.* **285**, 41627–41636 (2010).

116. Fozo, E. M. New type I toxin-antitoxin families from "wild" and laboratory strains of *E. coli*. *RNA Biol.* **9**, 1504–1512 (2012).

117. Fozo, E. M. *et al.* Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.* **70**, 1076–1093 (2008).

118. Lopez-Gomollon, S. & Nicolas, F. E. Purification of DNA oligos by denaturing polyacrylamide gel electrophoresis (PAGE). in *Methods in Enzymology* **529**, 65–83 (2013).

119. Bachmann, B. J. Theodoro agostinho peters filho. **36**, 525–557 (1972).

120. Hayashi, K. *et al.* Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. *Mol. Syst. Biol.* **2**, (2006).

121. Karen, E. & Roger, B. Media Preparation and Bacteriological Tools. *Curr. Protoc. Mol. Biol.* **59**, 1.1.1-1.1.7 (2002).

122. Bustin, S. A. *et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).

123. Fozo, E. M. New type I toxin-antitoxin families from 'wild' and laboratory strains of E. coli: Ibs-Sib, ShoB-OhsC and Zor-Orz. *RNA Biol.* **9**, 1504–

1512 (2012).

124. Han, K., Kim, K. S., Bak, G., Park, H. & Lee, Y. Recognition and discrimination of target mRNAs by Sib RNAs, a cis-encoded sRNA family. *Nucleic Acids Res.* **38**, 5851–5866 (2010).

125. Bustin, S. A. *et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).

126. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98 (2011).

127. Shimizu, K. Regulation Systems of Bacteria such as Escherichia coli in Response to Nutrient Limitation and Environmental Stresses. *Metabolites* (2013). doi:10.3390/metabo4010001

128. Mandel, M. J. & Silhavy, T. J. Starvation for different nutrients in Escherichia coli results in differential modulation of RpoS levels and stability. *J. Bacteriol.* (2005). doi:10.1128/JB.187.2.434-442.2005

129. Aizenman E, Engelberg-Kulka H, G. G. An Escherichia coli chromosomal 'addiction module' regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci U S A.* **93**, 6059–63 (1996).

130. Kawano, M., Reynolds, A. A., Miranda-Rios, J. & Storz, G. Detection of 5???- and 3???-UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli. *Nucleic Acids Res.* **33**, 1040–1050 (2005).

131. Christensen SK, Pedersen K, Hansen FG, G. K. Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. *J. Mol. Biol.* **332**, 809–19 (2003).

132. Ghafourian, S., Raftari, M., Sadeghifard, N. & Sekawi, Z. Toxin-antitoxin systems: Classification, biological function and application in biotechnology. *Curr. Issues Mol. Biol.* **16**, 9–14 (2014).

133. Pedersen, K., Christensen, S. K. & Gerdes, K. Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Mol. Microbiol.* **45**, 501–510 (2002).

134. Wang, X. & Wood, T. K. Toxin-Antitoxin Systems Influence Biofilm and Persister Cell Formation and the General Stress Response Toxin-Antitoxin Systems Influence Biofilm and Persister Cell Formation and the General Stress Response. *Appl. Environ. Microbiol.* **77**, 5577–83 (2011).

135. Brantl, S. & Jahn, N. SRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiology Reviews* **39**, 413–427 (2015).

136. Brantl, S. Bacterial chromosome-encoded small regulatory RNAs. *Futur. Microbiol* **4**, 85–103 (2009).

137. Ren, Dabin, Walker, Anna N, Daines, D. A. Toxin-antitoxin loci vapBC-1 and vapXD contribute to survival and virulence in nontypeable Haemophilus influenzae. *BMC Microbiol.* **12**, (2012).

138. Lewis, K. Multidrug tolerance of biofilms and persister cells. *Current Topics in Microbiology and Immunology* **322**, 107–131 (2008).

139. Wagner, E. G. H. & Unoson, C. The toxin-antitoxin system tisB-istR1: Expression, regulation, and biological role in persister phenotypes. *RNA biology* **9**, 1513–9 (2012).

140. Maisonneuve, E., Shakespeare, L. J., Jørgensen, M. G. & Gerdes, K. Bacterial persistence by RNA endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13206–13211 (2011).

141. Tsilibaris, V., Maenhaut-Michel, G., Mine, N. & Van Melderen, L. What is the benefit to Escherichia coli of having multiple toxin-antitoxin systems in its genome? *J. Bacteriol.* **189**, 6101–6108 (2007).

142. Dörr, T., Vulić, M. & Lewis, K. Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli. *PLoS Biol.* **8**, (2010).

143. Nakashima, N. & Miyazaki, K. Bacterial Cellular Engineering by Genome Editing and Gene Silencing. *Int. J. Mol. Sci.* **15**, 2773–2793 (2014).

144. Sharma, V., Yamamura, A. & Yokobayashi, Y. Engineering Artificial Small RNAs for Conditional Gene Silencing in Escherichia coli. 6–13 (2012). doi:10.1021/sb200001q

145. Na, D. *et al.* Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31**, 170–174 (2013).

146. Urban, J. H. Translational control and target recognition by Escherichia coli small RNAs in vivo. **35**, 1018–1037 (2007).

147. Man, S. *et al.* Artificial trans -encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. **39**, (2011).

148. Nieto, C. *et al.* The chromosomal relBE2 toxin – antitoxin locus of Streptococcus pneumoniae : characterization and use of a bioluminescence resonance energy transfer assay to detect toxin – antitoxin interaction. **59**, 1280–1296 (2006).

149. Nieto, C., Sadowy, E., Campa, A. G. De, Hryniewicz, W. & Espinosa, M. The relBE2Spn Toxin-Antitoxin System of Streptococcus pneumoniae : Role in Antibiotic Tolerance and Functional Conservation in Clinical Isolates. **5**, (2010).

150. Ramage, H. R., Connolly, L. E. & Cox, J. S. Comprehensive functional

analysis of Mycobacterium tuberculosis toxin-antitoxin systems: Implications for pathogenesis, stress responses, and evolution. *PLoS Genet.* **5**, (2009).

151. Cataudella, I., Trusina, A., Sneppen, K., Gerdes, K. & Mitarai, N. Conditional cooperativity in toxin-antitoxin regulation prevents random toxin activation and promotes fast translational recovery. *Nucleic Acids Res.* **40**, 6424–6434 (2012).