CHARACTERIZATION OF A TYPE I TOXIN-ANTITOXIN SYSTEM IN *ESCHERICHIA COLI*

DECODING TOXICITY:
CHARACTERIZATION OF THE
IBSC/SIBC TYPE I TOXIN-ANTITOXIN SYSTEM OF
ESCHERICHIA COLI

By

WENDY W.K. MOK, BSc.

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

McMaster University Doctor of Philosophy (2012) Hamilton, Ontario

TITLE: Decoding toxicity: Characterization of the IbsC/SibC type I toxin-antitoxin system of *Escherichia coli*

AUTHOR: Wendy W.K. Mok, BSc.

SUPERVISOR: Professor Yingfu Li

NUMBER OF PAGES: xvi; 212

ABSTRACT

Small RNAs and small proteins encoded in diverse microbial genomes have been shown to play big parts in regulatory processes that are vital to the defence, adaptability, and survival of their hosts despite their small size. These elements also make up the two components of type I toxin-antitoxin (TA) systems that are present across many bacterial lineages. In these TA systems, the production of small toxic peptides can disrupt cellular processes and induce growth attenuation. However, toxin synthesis can be antagonized by its cognate RNA antitoxin, which binds toxin transcripts in order to inhibit their translation. As the functions of some of these TA pairs emerge, it has become evident that their presence is beneficial to the cell, taking on roles such as the guardians of genomic integrity or the coordinators of cellular responses. At present, the biological function and regulation of many type I TA pairs remain enigmatic. Among them is the IbsC/SibC pair found in *Escherichia coli*. In this dissertation, our efforts in identifying and characterizing IbsC/SibC are discussed. After coming across *ibsC/sibC* through a genetic screen and confirming that the production of the IbsC peptide is toxic to cells, we examined the sequence requirements for the toxicity of IbsC. We further probed into the effects of mutations on its structure and mechanism of action. In investigating the regulation of *ibsC* expression, we found that toxin production is tightly repressed under nutrient rich conditions at the transcriptional and post-transcriptional levels, the latter of which is mediated by the SibC RNA. Furthermore, we took advantage of the toxicity of IbsC by engineering an efficient molecular cloning system with its variants using the sequence information garnered from our prior studies. Taken together, our work has shed light on the biochemical properties, evolution, and potential applications of this TA pair.

ACKNOWLEDGEMENTS

First off, I must thank my supervisor, Dr. Yingfu Li, for giving me the opportunity to work in his lab as a fresh undergraduate all those years ago. Throughout this time, his support, motivation, and creative discussions have been most invaluable to me. I would also like to acknowledge my committee, Dr. Eric Brown and Dr. Marie Elliot, for their guidance and inspiration. I would especially like to thank Dr. Elliot for the experimental advice and help she had provided me when I first started on this project.

I have been fortunate to have many fantastic colleagues and friends along the way. Fellow "Li Labbers" past and present have provided an enjoyable and stimulating working environment. I am especially grateful to Casey Fowler, Naveen Kumar, and R. J. Sawchyn for their mentorship and friendship. I would also like to thank Lindsay Matthews for all those reassuring coffee breaks.

I am indebted to my parents, David and Mabel, for their love, advice, faith, and all those wonderful home-cooked takeaways. To Andrea, my "sister", thank you for your strength and for always believing in me ever since we were little kids. Finally, I turn to my fiancé Simon, thank you for sharing this adventure with me. Your patience, calmness, and ideas have kept me going. Thank you guys for being a part of this incredible journey!

Dedicated to my grandfather, Lo Tak Kwong, who taught me one of my first lessons in science with a light bulb, a battery, and two wires.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 5

CHAPTER 6

LIST OF TABLES

LIST OF ABBREVIATIONS

5'-UTR	5'-untranslated region
%N.G.	normalized growth (%)
%R.F.	relative fluorescence (%)
Atc	anhydrotetracycline
AI-2	autoinducer-2
CD	circular dichroism
c-di-GMP	3',5'-cyclic diguanylic acid
$DiBAC_4(3)$	Bis-(1,3-dibarbituric acid)-trimethine oxonal
DNA	deoxyribonucleic acid
FACS	fluorescence-assisted cell sorting
EDF	extracellular death factor
GFP	green fluorescent protein
h	hour
ICE	integrative and conjugative element
IGR	intergenic region
IPTG	isopropyl β-D-1-thiogalactopyranoside
LacR	lac repressor
LB	Luria-Bertani broth
MALDI	Matrix-assisted laser desorption/ ionization
MCS	multiple cloning site
mRNA	messenger RNA
MRSA	methicillin resistant *Staphylococcus aureus*

nt	nucleotides
ORF	open reading frame
OD_{600}	optical density measured at 600 nm
PAGE	polyacrylamide gel electrophoresis
PCD	programmed cell death
PCR	polymerase chain reaction
P_{BAD}	arabinose-inducible promoter
P_{ibsC}	*ibsC* promoter
P_{LlacO1}	lactose-inducible promoter
P_{LtetO1}	tetracycline-inducible promoter
P_{opt}	a synthetic strong and constitutive promoter
PBS	phosphate buffered saline
PMF	proton motive force
ppGpp	guanosine pentaphosphate
PSM	phenol-soluble modulins
RACE	rapid amplification of cDNA ends
RBS	ribosome binding site
RNA	ribonucleic acid
RNAP	RNA polymerase
RPS	randomized peptide-encoding sequence
rRygC	reverse complement of *rygC*
SASP	small acid soluble protein
SEC	Sec-dependent pathway signal peptide

sprotein	small protein
sRNA	small RNA
STOP	stop codon
TA	toxin-antitoxin
TetR	tetracycline repressor
TFE	2,2,2-trifluoroethanol
TIR	translation initiation region
TRD	target recognition domain
X-gal	5-bromo-4-chloro-3-indolyl-galactopyranoside
λ_{em}	emission wavelength
λ_{ex}	excitation wavelength
μ_n	average signal from negative control
μ_p	average signal from positive control

Chapter 1:
General Introduction

1.1. It's a small world

Bacterial genomes harbour a plethora of small proteins and RNA sequences with vital defensive and regulatory functions, modulating processes that range from stress response to pathogenesis (reviewed in (Gottesman et al., 2006) and (Toledo-Arana et al., 2007), respectively). Many of these elements were previously overlooked as they are found in what were once believed to be intergenic regions (IGRs), masked by larger flanking genes. Due to their small size, they have also evaded detection by mutational screens (Wassarman et al., 2001). With the advancements of bioinformatics and experimental techniques, such as deep sequencing and microarrays, higher resolution and more detailed maps of genomes emerged, and more of these small genetic entities were discovered. Herein, we will focus on two main classes of products encoded by these small genes: small proteins (sproteins) and small RNAs (sRNAs).

1.2. Small RNAs in bacteria

1.2.1. Characteristics and identification of sRNAs

Over the past decade, the number of small regulatory RNAs identified across bacterial genomes has vastly expanded. In the *E. coli* genome alone, approximately 100 sRNAs have been confirmed to exist. The majority of these RNA species are not protein coding. Rather, they modulate cellular physiology by directly interacting with target mRNAs or proteins in order to alter their fate and activity. sRNAs generally range from 50 to 500 nucleotides (nt). Due to their small size, identifying them through biochemical assays and mutational screens proves to be challenging (Altuvia, 2007; Vogel & Sharma, 2005). Some of the first sRNAs to be discovered were found serendipitously through studies on the transcriptional regulation of their neighbouring genes (Wassarman et al.,

1999). In the last 15 years, a number of techniques for the isolation of sRNAs were developed, leading to a dramatic increase in sRNA annotation (reviewed in Altuvia, 2007; Vogel & Sharma, 2005). For example, RNA synthesized by bacteria cultured under different growth conditions can be radioactively labelled before being fractionated by size. Specific bands corresponding to sRNA candidates can then be recovered and sequenced (Altuvia et al., 1997). Alternatively, using shotgun cloning, RNAs of specific sizes can be reverse transcribed and cloned following size fractionation, allowing potential sRNA sequences to be further characterized after excluding those that hybridize to rRNA, tRNA, and abundant sRNAs (Vogel et al., 2003). Total RNA isolated from bacteria grown under various conditions can also be subjected to microarray analyses in order for transcripts expressed from intergenic regions to be detected (Wassarman et al., 2001). As the interaction between many sRNAs and their mRNA targets is mediated by chaperone Hfq, co-purification of RNA sequences with this protein has also led to the identification of novel sRNAs (Zhang et al., 2003). The discovery of sRNAs is further accelerated by computational strategies, which allow candidate sRNAs to be identified based on homology with annotated sequences or based on the presence of flanking promoters and rho-independent transcription terminators around genes in intergenic regions (Wassarman et al., 2001; Argaman et al., 2001).

1.2.2. Protein-targeting sRNAs

sRNAs have been shown to act on both protein and mRNA targets in bacteria. Protein-targeting sRNAs often act on global regulatory proteins under stress conditions, thereby affecting multiple metabolic pathways. For instance, the CsrB sRNA in *E. coli* binds CsrA, a small polypeptide that acts as a regulator of carbon storage. During

stationary phase or under nutrient poor conditions, one CsrB sRNA can bind 18 copies of CsrA via its repetitive sequences that mimic CsrA binding sites in the leader regions of its target genes (Liu et al., 1997). This prevents CsrA from acting on its target mRNAs, resulting in an increase in glycogen levels in the cell. Another example is the 6S sRNA, whose structure resembles an open promoter (Barrick et al., 2005; Wassarman, 2007). 6S has been shown to bind RNA polymerase (RNAP) holoenzymes that are in complex with the housekeeping sigma factor, σ^{70}, during stationary phase growth to promote transcription of genes recognized by alternative sigma factor σ^{S}.

1.2.3. mRNA-targeting sRNAs

Regulation of gene expression at the post-transcriptional level using sRNAs has been suggested to offer several advantages over protein-based regulators (Mehta et al., 2008; Shimoni et al., 2007). sRNA regulation enables cells to more rapidly respond to environmental signals compared with regulation at the transcriptional level using protein transcription factors. Under instances where regulators have to be newly synthesized, sRNAs offer more rapid responses compared with protein regulators acting via protein-protein interactions. Once the external signal is removed, recovery of the target gene to its original level of expression is faster under sRNA-mediated post-transcriptional regulation compared with regulation at the transcriptional level depending on the ratio of sRNA-to-mRNA present in the cell and the degradation rate of the sRNA. In the repressed state where the number of sRNAs in the cell outnumbers their mRNA targets, sRNA provides tight regulation and is less prone to small fluctuations in signals or noise compared with protein regulators. Considering that this repression is dependent on the rate of sRNA production, the use of sRNAs as a regulator of gene expression allows for fine-tuning of

the expression of their target genes so that it is maintained at a desired steady-state level. Collectively, these factors allow cells to rapidly and reliably switch between gene expression states in response to changes in their environment.

sRNAs that target mRNAs can be encoded in *cis* or in *trans* relative to their target transcripts. Those that are encoded in *trans*, where the sRNA and mRNA encoding genes are found at divergent genetic loci, often share limited complementarity. It has been reported that while the region of potential base pairing between *trans*-acting sRNAs and their targets often encompasses 10 to 25 nt, only a few core nucleotides are critical to their interactions and are therefore intolerant of mutations (Kawamoto et al., 2006). As a consequence of this limited sequence complementarity, the interactions between *trans*-acting sRNAs and mRNAs often require mediation by chaperone Hfq. Hfq binds RNAs as a hexamer in order to remodel the RNAs, relieve inhibitory secondary structures, and promote binding (Brennan & Link, 2007; Beisel et al., 2012). The binding of Hfq to unpaired sRNAs often occurs at sites that are recognized by RNase E. Thus Hfq can also prevent the degradation of sRNAs that have not yet hybridized with their targets. On the other hand, Hfq can have negative effects on the stability of sRNAs by recruiting RNases to degrade sRNA::mRNA complexes.

The binding of sRNAs to their mRNA targets can alter the stability and affect translation initiation of the mRNA (reviewed in (Waters & Storz, 2009)). In most cases, such interactions are inhibitory, whereby sRNA binding targets the mRNA for degradation or occludes regions necessary for translation initiation. However, positive regulation has also been documented. It is also possible for sRNAs to alter gene

expression at the transcriptional level by promoting the formation of transcription terminators and antiterminators.

As *trans*-acting sRNAs often share limited complementarity with their targets and binding only requires a few nucleotides, they can target multiple mRNAs and regulate their expression at the post-transcriptional level. Many of them are induced under specific stress conditions and can coordinate multiple genes in a regulatory circuit, allowing bacteria to mount specific physiological responses and cope with such stresses (Gottesman et al., 2006). Some prominent examples of these include RyhB, which is expressed under iron-limiting conditions to regulate genes associated with iron metabolism (reviewed in (Masse et al., 2007)); OxyS, which is expressed under oxidative stress, and can regulate over 40 genes in order to reduce hydrogen peroxide production in the cell and protect it from oxidative damage (Altuvia et al., 1997); as well as RybB and MicA, which downregulate the expression of outer membrane porins following membrane stress (Johansen et al., 2006).

In contrast to *trans*-acting sRNAs, *cis*-acting sRNAs are encoded on the antisense strand of their targets and have the propensity to form more extensive base pairing interactions. Their initial interactions often only involve a few nucleotides and do not require the mediation of Hfq, though the duplex can be extended following initial binding (Waters & Storz, 2009). As the two RNA species are transcribed in close proximity, local concentration of sRNA can be easily enhanced to facilitate interactions (Georg & Hess, 2011). The sRNA can subsequently alter the stability of its target (Figure 1.1A) and modulate its translation (Figure 1.1B). Furthermore, these antisense RNAs can also regulate the expression of its target genes at the transcriptional level by promoting the

formation of transcription terminators or antiterminators (Figure 1.1C) or by occluding their promoters from RNAPs (Figure 1.1D) (reviewed in (Georg & Hess, 2011)).

Cis-acting sRNAs are found to be increasingly abundant across different microbial genomes in recent years (Dornenburg et al., 2010; Rasmussen et al., 2009; Tang et al., 2005). A number of them have been found to modulate the expression of genes in the same operon (reviewed in (Waters & Storz, 2009)). They can also inhibit the synthesis of replication primers from plasmid origins of replication (Tamm & Polisky, 1985) or regulate transposase translation from transposable elements in order to control the copy number of these mobile genetic elements (Tang et al., 2005). Many antisense sRNAs have been shown to be a component of toxin-antitoxin (TA) systems, functioning as an antitoxin to antagonize the expression of their cognate toxins. These sRNAs will be discussed in Section 1.6 of this chapter.

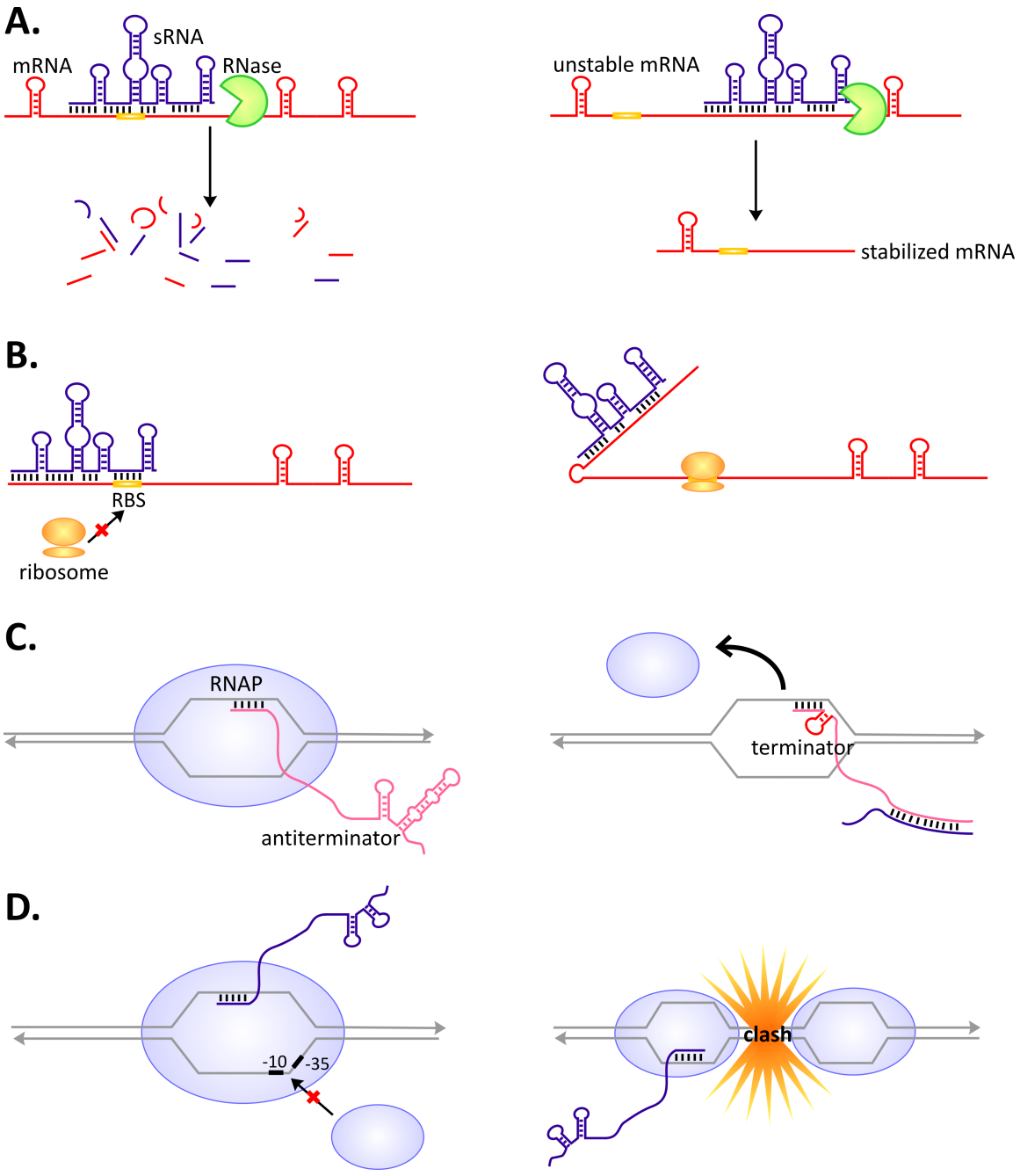


Figure 1.1. Effects of *cis*-acting sRNAs on target mRNA expression. A. Effects on mRNA stability. Interactions between sRNA (sequence in blue) and its target mRNA (sequence in red) can target the duplex for RNase (shown in green) degradation (left panel). Alternatively, binding of sRNA to an unstable mRNA can target it for processing in order to produce a stable product (right panel). B. Effects on mRNA translation. sRNA binding can occlude the ribosome binding site (RBS), the start codon, and/or other codons within the open reading frame (ORF) in order to prevent translation initiation (left panel). In other cases, the binding can alleviate structures near the translation initiation region

(TIR), thereby enabling the binding of ribosomes and translation initiation (right panel). C. Effects on transcription. Antisense sRNAs can interact with transcription antiterminators present on elongating mRNAs (sequence in pink) in order to promote the formation of transcription terminators (shown in red), dissociation of the RNA polymerase (RNAP; shown in blue), and the production of truncated transcripts. D. Effects at promoters. An elongating sRNA can prevent the binding of RNAPs to the promoter of its sense target (right panel). Transcription of the antisense sRNA can also interfere with RNAPs transcribing the sense mRNA, causing dissociation of the sense RNAP (left panel). If transcription has started from the sense promoter, the mRNA fragment that is produced is rapidly degraded.

1.3. Small proteins in bacteria

In addition to sRNAs, microbial genomes encode a diverse ensemble of sproteins (a term coined by the Storz laboratory) that mediate an array of important cellular processes. Defined as having 100 or fewer amino acids, their minute size has long impeded their annotation, isolation, and characterization. The initial mining of sproteins was as laborious and challenging as the search for sRNAs. Conventional biochemical techniques, such as two-dimensional gel electrophoresis and mass-spectrometry, tend to favour the identification of abundant proteins of standard size (between 30-200 kDa) from crude cell lysates (Link et al., 1997; Shevchenko et al., 1996). Their open reading frames (ORFs) are often too small to be disrupted by random mutagenesis, rendering them poor targets in genetic screens (Kastenmayer et al., 2006). Even when these ORFs are mutated, the lack of observable and robust phenotypic changes in many deletion strains further limited our understanding of their biological relevance. Consequently, the exact number of sproteins present in an organism, their identities, and their functions were challenging to elucidate.

In recent years, however, we have seen a remarkable surge in the number of sproteins found across various genomes (Kastenmayer et al., 2006; Peng et al., 2011;

Hemm et al., 2008). In the genome of *E. coli* K-12 alone, over 60 sproteins have been discovered (Hemm et al., 2008). Searches for their corresponding genes have been facilitated by bioinformatics approaches, including comparative genomics and ribosome binding site models. The production of these predicted sproteins was subsequently confirmed by epitope tagging followed by immunoblotting. Microarray analyses conducted with cells cultured under different stress-inducing conditions have led to the isolation of additional sproteins (Hemm et al., 2010; Thomassen et al., 2010). Microarrays along with competition assays and subcellular localization studies have also been shown to be useful in providing insight into the functions of a subset of sproteins (Hobbs et al., 2010; Fontaine et al., 2011).

Of the sproteins that have been characterized to date, many have been linked to functions that are important to the survival, adaptability, and competitiveness of microorganisms. Following their synthesis, these proteins can either be secreted into the extracellular milieu or retained in the cell. An overview of selected sproteins is provided in Figure 1.2. Their characteristics and functions are discussed in greater details in this chapter.

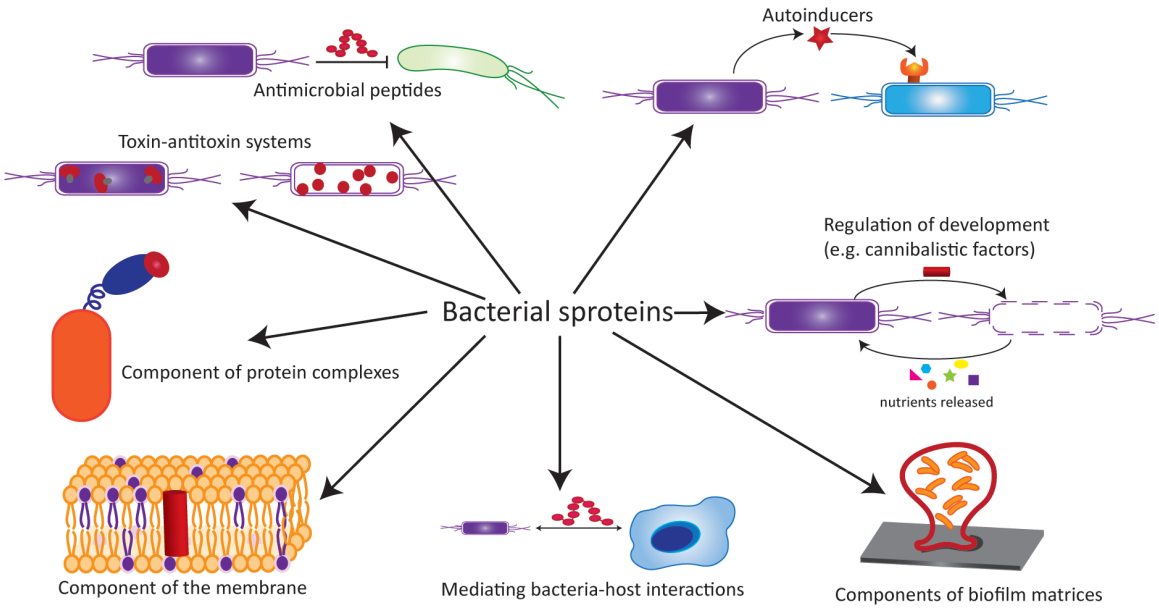


Figure 1.2. Overview of sprotein functions. The bacterial genome encodes a number of sproteins, which can function as (clockwise from top left) antimicrobial peptides, autoinducers, regulators of bacterial development, components of biofilm extracellular matrices, factors that mediate host-bacteria interactions, components of cellular membranes, components of larger protein complexes, and bacterial toxins.

1.4. Secreted small proteins

1.4.1. Signalling peptides

Once secreted outside the cell, peptides can mediate interactions between bacteria and their surroundings, allowing the cells to respond to and cope with changes in their environment. These peptides can stimulate signalling cascades and alter gene expression. Ultimately, these changes affect the proliferation, development, colonization, stress response, and pathogenicity of the organism. In gram-positive species, modified peptides or amino acids serve as signalling molecules for quorum sensing, a system that allow a community of cells to survey their population density and coordinate gene expression accordingly (Ng & Bassler, 2009). The peptides activate membrane-associated sensor

kinases, which subsequently phosphorylate a cognate response regulator, culminating in the stimulation of a specific set of genes. Examples of these peptide quorums include autoinducer peptides produced in *Staphylococcus aureus* (Novick & Geisinger, 2008), ComX (Okada et al., 2005) and CSF (Lazazzera et al., 1997) from *Bacillus subtilis*, and competence stimulating peptides from *Streptococcus pneumonia* (Havarstein et al., 1995). Moreover, cyclic dipeptides (also known as diketopiperazines) can promote interspecies cell-to-cell communication and cross talk between bacterial signalling pathways (Holden et al., 1999). These molecules can act as agonists of quorum sensing pathways or antagonists that compete with cognate autoinducers (Holden et al., 1999; Park et al., 2006; Li et al., 2011).

1.4.2. Secreted peptides with signalling and antimicrobial functions

Some peptides secreted by bacteria have dual functions, acting as signalling molecules and antibiotics (reviewed in (Romero et al., 2011)). At low concentrations typically observed under their natural settings, these peptides mediate cell-cell communication. However, these peptides are lethal to their recipients at high concentrations, allowing the producers to limit competition in their ecological niche. Many of these peptides, including ribosomally synthesized bacteriocins, are subjected to extensive post-translational modifications, contributing to diversity in their structures and properties. For example, lantibiotics, a class of bacteriocins, are modified with unusual amino acids, cyclized, and exported into the extracellular milieu following synthesis (Siezen et al., 1996). Some of these peptides, such as class A lantibiotics, are small, cationic, and amphiphilic, akin to antimicrobial or host-defence peptides secreted by higher organisms (O'Connor & Shand, 2002; Brotz & Sahl, 2000). These peptides have

been shown to induce pore formation in the cytoplasmic membranes of their target cells. Membrane permeabilization can be achieved via direct interactions with the membrane or through the binding of lipid II, a membrane-anchored precursor of the central building block of the bacterial cell wall (Breukink & de Kruijff, 2006). Alternatively, peptide antibiotics can target the cell wall, DNA replication machineries, and metabolic enzymes. These mechanisms of action are reviewed in (Brogden, 2005) and summarized in Figure 1.3.

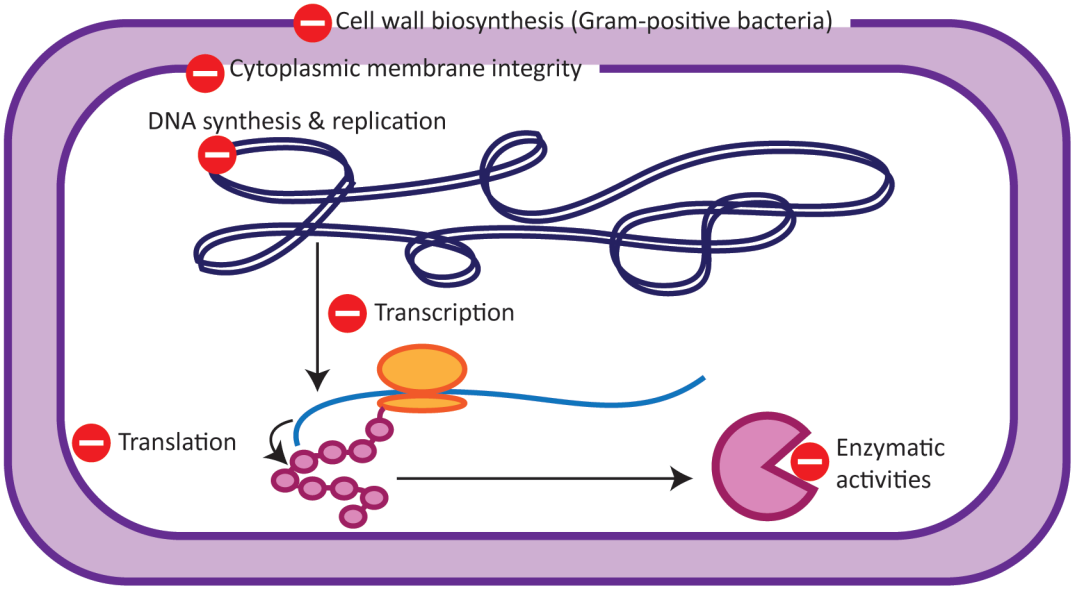


Figure 1.3. Mechanisms of peptide antibiotic action. The cellular processes and structures that can be inhibited by antimicrobial peptides are illustrated. In addition to disrupting the integrity of the cellular membrane, defence peptides are proposed to interfere with other key cellular processes, including DNA, RNA, protein, and cell wall biosynthesis.

Bacteria-derived antimicrobial peptides can be quite multifaceted. This quality is exemplified by nisin, a prominent member of class A lantibiotics that has been used in food preservation for over four decades (reviewed in (Peschel & Sahl, 2006)). Nisin is

synthesized in *Lactococcus lactis* as a 57-amino acid precursor that is processed to a 34-amino acid peptide. It displays broad-spectrum antimicrobial activity, inhibiting both gram-positive and gram-negative bacteria. Nisin can induce pore formation by directly targeting the cytoplasmic membrane and in a lipid II-dependent manner. By binding lipid II, nisin can also inhibit cell wall biosynthesis. Furthermore, it can displace cationic autolytic enzymes from anionic binding sites on bacterial cell walls and activate them, thereby inducing autolysis. Administration of nisin at subinhibitory concentrations has been demonstrated to stimulate the transcription of its own structural gene (Kuipers et al., 1995). Nisin has also been found to prevent spore outgrowth, induce extracellular matrix production, and promote biofilm formation in *Bacillus subtilis* (Lopez, Vlamakis, et al., 2009).

1.4.3. Secreted sproteins in development and differentiation

In response to quorum sensing molecules and antibiotics, bacteria can synthesize sproteins that kill susceptible members of its population and alter cellular differentiation. In the soil-dwelling bacterium *B. subtilis*, surfactin, a signalling lipopeptide, stimulates the production and secretion of toxins Skf (a 26-amino acid peptide) and Sdp (a 42-amino acid peptide) in a subpopulation of cannibalistic cells (Lopez, Fischbach, et al., 2009; Liu et al., 2010). These toxins selectively lyse and kill sensitive siblings (Gonzalez-Pastor et al., 2003). As these cells die, they release nutrients that are utilized by the toxin-resistant cannibals, allowing them to overcome nutrient limitation and delay the onset of sporulation, an energy demanding process that is difficult for cells to exit once they are committed. Cannibalistic toxins and antimicrobial peptides (including the aforementioned nisin) produced by *B. subtilis*, closely-related members of its own genus, and

microorganisms sharing the same ecological niche can also activate genes linked to extracellular matrix production in the cannibals (Lopez, Vlamakis, et al., 2009; Shank et al., 2011). The extracellular matrix, composed of exopolysaccharides and amyloid fibres, encases multicellular aggregates known as biofilms (Branda et al., 2001; Branda et al., 2006; Romero et al., 2010). This subsequently protects the encased cells from antibiotics secreted by other organisms in its surroundings and from harsh unfavourable growth conditions. Thus, the production of these cannibalistic toxins is proposed to eliminate cell types that are no longer needed in the community, while promoting the growth and proliferation of matrix-producing cells that can support biofilm development (Lopez, Vlamakis, et al., 2009).

Secreted sproteins can also participate in modulating the differentiation of bacterial populations without inducing cell death. In soil-dwelling gram-positive bacteria *Streptomyces coelicolor*, several sproteins are involved in the adaptation of cell surfaces as the bacteria transition from vegetative growth in aqueous environments to aerial hyphae and spore germination in the air. The cell surface of the vegetative hyphae is quite hydrophilic. During the vegetative phase, SapB, a 21-residue amphiphilic lantibiotic-like peptide, is produced (Kodani et al., 2004). It has been shown to act as a surfactant, coating the nascent aerial hyphae and the air-water interface in order to facilitate its emergence from the aqueous milieu into the air (Tillotson et al., 1998). In contrast, the cell surfaces of aerial hyphae are highly hydrophobic. Another class of eight proteins coined chaplins, of which five are short and hydrophobic, also act as strong surfactants (Di Berardo et al., 2008). The chaplins anchor to the cell wall and polymerize, forming a hydrophobic coat that encase the aerial filaments. Deletion of either the SapB and chaplin

biosynthetic genes have been found to impair aerial hypha formation (Capstick et al., 2007).

1.4.4. Modulation of host-bacteria interactions

Secreted peptides can also regulate host-bacteria interactions and govern bacterial pathogenesis. The α-type of phenol-soluble modulins (PSMα) produced in community-associated methicillin-resistant *S. aureus* (MRSA) was found to be a major virulence determinant. It was reported that elevated expression of PSMα, which have high cytolytic potential toward human neutrophils, contributed to enhanced virulence of this particular strain (Li et al., 2009). In contrast, the γ-type of PSM produced by *S. epidermis*, a commensal organism found in the skin microbiota, has been shown to act synergistically with host-derived antimicrobial peptides and modulate innate immunity of the host in order to offer defence against pathogens (Cogen et al., 2010).

The secreted peptides and sproteins identified and characterized thus far have been linked to a number of interesting and important intraspecies, interspecies, and host-pathogen interactions. The discovery of additional novel secreted proteins is expected to be facilitated by a technique referred to as imaging mass spectrometry, which was previously employed to identify *B. subtilis* cannibalistic factors Skf and Sdp (Liu et al., 2010; Yang et al., 2009). With this approach, bacteria are cultured on a thin-layer agar cast on top of a Matrix-assisted laser desorption/ ionization (MALDI) plate. Natural product MALDI-mass spectrometry is then conducted to uncover secreted metabolites. Bacteria can also be co-cultured in conjunction with other species of interest in order to eavesdrop on cell-cell communication and metabolic exchanges occurring between these organisms (D'Onofrio et al., 2010; Seyedsayamdost et al., 2011).

1.5. Functions of small proteins inside the cell

1.5.1. Membrane-targeting sproteins

Of the 60 sproteins identified in *E. coli* K-12 in a recent study by the Storz laboratory, over 65% were predicted to contain a single α-helical transmembrane domain (Hemm et al., 2008). In a subsequent study, the localization and orientation of a subset of these proteins were examined (Fontaine et al., 2011). The majority of these sproteins were found to localize to the inner membrane, with positively-charged residues near the transmembrane domain found close to the inner leaflet, thus complying with the "positive -inside" rule (von Heijne, 1992). While the biological functions of many of these membrane-bound sproteins remain elusive, they are suggested to modulate the architecture and character of the membrane in response to changes in their environment. In response to cold shock, which tends to decrease membrane fluidity, bacteria have to remodel their membrane composition in order to survive (Yamanaka, 1999). One method of achieving this is through the introduction of double bonds into the acyl chains of phospholipids (Cybulski et al., 2004; Garwin & Cronan, 1980). Altering membrane fluidity is also a mechanism by which bacteria resist the assault of antimicrobial peptides (Koprivnjak & Peschel, 2011). This can be done through the modification of phospholipids (Staubitz et al., 2004), adjusting phospholipid composition (Bayer et al., 2000), and incorporation of carotenoid compounds into the cytoplasmic membrane (Clauditz et al., 2006). However, these processes are often time-consuming, requiring multiple genes and gene products to synthesize, traffic, and insert new lipids. Thus, the production and insertion of sproteins is proposed to be a more time and energy efficient way of remodelling the cytoplasmic membrane (Hobbs et al., 2011).

Membrane-bound sproteins are also suggested to aggregate and localize to specific microdomains on the membrane, forming lipid raft-like structures (Hobbs et al., 2011). The sproteins associated with the lipid rafts then serve to recruit other larger proteins that can function together and orchestrate processes such as signal transduction and protein secretion (Lopez & Kolter, 2010). The ability of sproteins to dock at specific patches of the membrane is illustrated by SpoVM, a 26-amino acid amphipathic α-helical sprotein found in *B. subtilis*. During sporulation, SpoVM is produced in the mother cell. After the mother cell engulfs the forespore, thereby enveloping it with inner and outer membranes, SpoVM is deposited around the entire outer surface of the forespore and guides the assembly of the endospore protein coat. The localization of SpoVM is dictated by membrane curvature, as it specifically recognizes and binds the membrane of positive curvature (convex) surrounding the forespore rather than the negative curvature (concave) of the mother cell (Ramamurthi et al., 2009).

Small hydrophobic proteins are also components of many membrane-bound complexes found in eukaryotes and yeast (Beguin et al., 1997; Navarre et al., 1994; Simmerman et al., 1986). They are equally prevalent in bacteria. For instance, KdpF, a 29-amino acid peptide, stabilizes the KdpABC ATP-dependent potassium transporter in *E. coli* (Gassel et al., 1999). While phenotypic changes were not observed in the *kdpF* deletion strain, the absence of KdpF in purified KdpABC complexes resulted in a loss of ATPase activity. Another sprotein, the 49-amino acid AcrZ (formerly YbhT), was found to predominantly localize to the outer membrane and is proposed to be associated with the AcrAB-TolC multidrug efflux pump (Hobbs et al., 2011). Although the function of AcrZ in this complex is unknown, this peptide is suggested to stabilize the AcrB trimers in the

membrane. Membrane-associated sproteins are also suggested to be components of the electron transport chain. Two paralogs, YbgT and YccB, are encoded in the cytochrome oxidase operons. They have been shown to contain two conserved cysteine residues in their transmembrane domains that are suggested to coordinate heme groups in cytochrome oxidase (Hemm et al., 2008).

1.5.2. Sproteins in complexes

In addition to membrane-bound protein complexes, sproteins have been demonstrated to associate with cytoplasmic proteins and organelles, where they mediate fundamental processes such as protein synthesis and cell division, just to name a few (Hobbs et al., 2011). The ribosome, for instance, contains several sproteins of fewer than 50 amino acids (Hemm et al., 2008). Some of these ribosomal sproteins are primarily expressed under adverse conditions. The expression of Sra, a 45-residue sprotein associated with the 30S subunit of *E. coli* rapidly increases following entry into stationary phase. YkgO, a 46-amino acid sprotein paralogous to the L36 ribosomal protein found in the 50S subunit, is synthesized during stationary phase growth in minimal media (Panina et al., 2003; Hemm et al., 2008). Specifically, its expression is induced during zinc starvation and repressed by zinc in many bacteria (Panina et al., 2003). YkgO lacks the zinc-binding motif observed in L36 and is thought to mediate ribosomal RNA folding under zinc-limiting conditions.

Sproteins can also interact with larger proteins and/or protein complexes in order to impact their activities. During endospore formation in *B. subtilis*, MciZ, a 40-amino acid sprotein, binds FtsZ in order to inhibit its GTPase activity and prevent aberrant Z-ring formation at a stage when cytokinesis is normally ceased (Handler et al., 2008). In *E.*

coli, a 47-residue sprotein coined MgrB binds the PhoQ histidine kinase and inactivates the PhoPQ two-component system, which regulates the expression of over 40 genes. Many of these genes are linked to virulence, cell envelope modifications, and adaptation to magnesium limitations (Groisman, 2001).

1.5.3. Sproteins in regulation and stress response

Sproteins also participate in processes that are important for regulating development, maintaining homeostasis, and mediating responses to specific stresses. In *B. subtilis*, the expression of a number of sproteins is linked to endospore formation. Small acid soluble proteins (SASPs) of 60 to 80 amino acids found in gram-positive organisms, including *B. subtilis,* bind and repackage DNA during sporulation, thus protecting it from damages caused by UV radiation (Mason & Setlow, 1986; Dahl & Fordice, 2011). If the bacterium does encounter DNA damage, perturbations in replication initiation and elongation can induce the expression of another sprotein, Sda, which can prevent the phosphorylation of a histidine kinases required for the initiation of endospore development (Ruvolo et al., 2006). As sporulation progresses, two sproteins, Fin and Gin, act as anti-σ factors. Together, they control the activation and deactivation of specific alternative σ factors, thus enabling certain sets of genes to be expressed at different stages of development (Camp et al., 2011).

Sproteins are further involved in maintaining metal ion homeostasis in bacteria. Three small basic proteins in *B. subtilis*, coined FbpA (59-amino acids), FbpB (53-amino acids), FbpC (29-amino acids), have been shown to be a part of the iron-sparing response and act as RNA chaperones (Gaballa et al., 2008). They promote the interactions between the FsrA sRNA and its mRNA targets, allowing the production of iron-rich enzymes to be

repressed when iron is limited. In *E. coli*, the 42-amino acid MntS sprotein is required to maintain a delicate balance of manganese in the cell (Waters et al., 2011). It is suggested to act as a chaperone when the availability of manganese is limited, shuttling manganese to necessary cellular compartments. It is also implicated in affecting the activities of other proteins involved in manganese availability, including MntR (a transcription regulator) and MntH (a manganese importer).

Sproteins can also act as chaperones of mRNA and regulate gene expression in response to stresses and changes imposed by their environments. A prime example of this is the major cold shock protein CspA expressed in *E. coli*. When the bacteria experience a sudden temperature downshift from 37 ^{o}C to 15 ^{o}C, growth stalls as cells acclimatize to the cold shock. At this point, the production of the 70-amino acid sprotein is strongly induced, and it constitutes over 10% of the total proteins synthesized (Goldstein et al., 1990). As growth resumes following the acclimation phase, the expression of cold shock genes, including *cspA* is repressed. CspA acts by destabilizing the secondary structures of mRNAs, including in its own 5′-untranslated region (5′-UTR), in order to negatively regulate their expression (Bae et al., 1997). Sproteins can interact with other cellular targets and modulate their activities in response to stress. In *E. coli*, the 43-residue SgrT sprotein is induced during glucose-phosphate stress (Wadler & Vanderpool, 2007). It is translated from SgrS, a regulatory RNA that can bind and inhibit the expression of *ptsG*, which encodes a glucose transporter. SgrT is suggested to further limit glucose import by acting on the PtsG transporter. A subset of sproteins that has been linked to regulatory functions and stress response belong to type I toxin-antitoxin (TA) systems, which will be the subject of discussions in the subsequent sections of this chapter.

1.6. Overview of bacterial toxin-antitoxin systems

Since their discovery nearly three decades ago, a number of TA pairs have emerged in various bacterial genomes. The collection of these two-component systems, which consist of a stable toxin moiety and a labile antitoxin, is vast and is continuing to expand. To date, three classes of TA systems have been identified, and they are categorized based on the nature of the antitoxin and the mechanism by which the antitoxin represses the expression of function of their cognate toxin. The key characteristics of these TA systems are illustrated Figure 1.4.

In type I TA pairs, the toxin is a small, often hydrophobic sprotein encoded by a stable mRNA (Fozo, Hemm, et al., 2008). The antitoxin is a more labile regulatory sRNA. In the toxin repressed state, the binding of the antitoxin sRNA to the toxin mRNA can repress its translation or target the transcript for degradation, thereby antagonizing its toxicity. In the toxin active state, transcription of the antitoxin is suppressed, and the remaining antitoxin RNAs, which only tend to have half-lives of a few minutes, is rapidly degraded. Thus the toxin transcripts, which can have half-lives of over 30 minutes, can persist in the cell and become translated. The accumulation of toxins and their interactions with their cellular targets subsequently result in growth inhibition. Improvements in screening strategies for sRNAs and sproteins across genomes have contributed to the increase in the number of type I TA pairs discovered in recent years. In the genome of *E. coli* K-12, six families of type I TA systems have been found. Many of these TA systems have multiple homologs distributed throughout the genome.

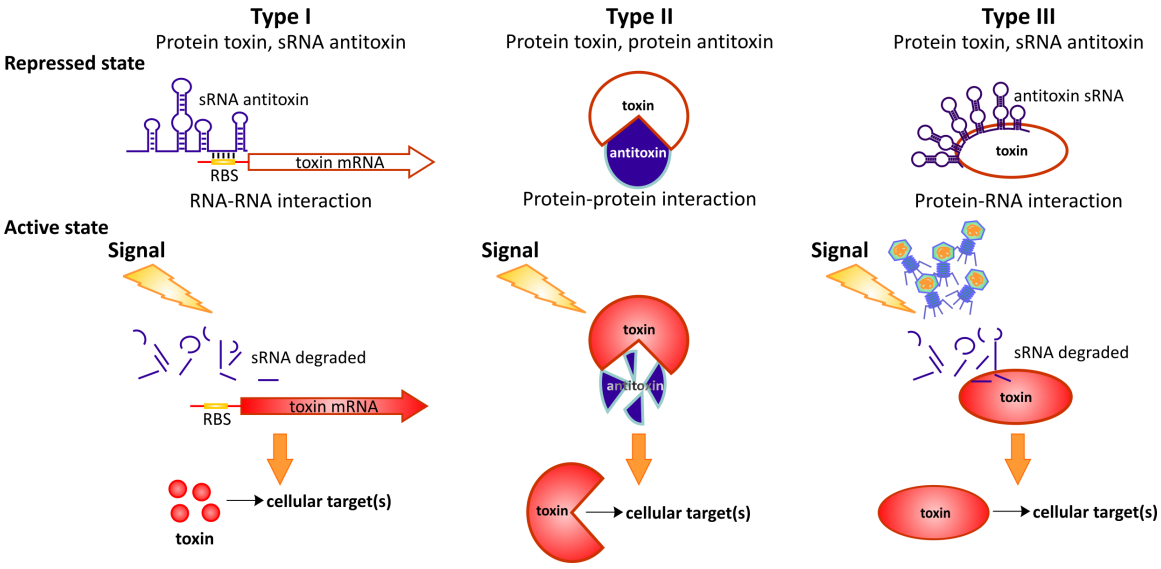


Figure 1.4. Characteristics of the three types of TA systems. In type I TA systems (left panel), the toxin is often a small hydrophobic peptide encoded by a stable mRNA (in red). The antitoxin is a labile sRNA (in blue). In the toxin repressed state, binding of the antitoxin to the toxin mRNA can repress its translation or target it for degradation (Fozo, Hemm, et al., 2008). This impedes the interaction between the toxin and its cellular target, thereby mitigating its toxicity. In the toxin active state, transcription of the antitoxin is suppressed, and remaining sRNAs are rapidly degraded. Therefore, the toxin can be translated and is free to act on its cellular target, resulting in growth suppression. In type II TA systems (centre panel), the toxin and antitoxin are both proteic. Binding of the toxin by the antitoxin in the repressed state prevents it from interacting with its cellular target (Van Melderen & Saavedra De Bast, 2009). Degradation of the antitoxin by proteases releases the toxin, allowing it to elicit its toxic effects on the cell. In type III TA systems (right panel), which may be regarded as a hybrid of type I and II TA pairs, the sRNA antitoxin directly binds the protein toxin, consequently sequestering it from its target until it is activated by bacteriophage infection (Fineran et al., 2009).

Type II TA systems are perhaps the most abundant and extensively studied of the three types of TA pairs annotated thus far. In these TA pairs, both the toxin and the antitoxin are small proteins, with the toxin ranging from 41 to 206 amino acids and the antitoxin ranging from 31 to 204 amino acids (Leplae et al., 2011). Similar to type I systems, the toxin is more stable and less prone to proteolytic degradation compared to the antitoxin. In the repressed state, the toxin is bound by the antitoxin, which sequesters

it away from its targets. The toxin-antitoxin complex can also act negatively to autoregulate the expression of the toxin. Under certain conditions, degradation of the antitoxin by proteases releases the toxin, enabling it to interfere with cellular processes and elicit its toxic effects on the cell.

Type III TA systems may be considered as a hybrid of type I and II TA systems. The toxin of this system is proteic and its antitoxin is a regulatory RNA akin to type I TAs. However, the RNA directly interacts with the toxin in order to inhibit its toxicity. The first type III TA pair was identified on a plasmid found in plant pathogen *Erwinia carotovora* subspecies *atrosepticum* and is linked to the phage abortive infection system (Fineran et al., 2009). In the absence of infection, ToxN is inhibited by the 36 nt ToxI RNA. Through x-ray crystallography, it was revealed that three ToxI RNAs bind three ToxN monomers, forming a trimeric ToxIN complex (Blower et al., 2011). Moreover, ToxN exhibits endoribonuclease activity, processing ToxI from a repetitive precursor. When the bacteria are challenged by certain bacteriophages, the ToxN toxin is activated and promotes the "altruistic suicide" of infected members of the population in order to limit phage replication within the community (Fineran et al., 2009). The exact mechanism by which ToxIN is activated by certain phages and the mechanism of ToxN-mediated cell death remain to be elucidated.

1.7 Distribution of toxin-antitoxin systems

TA systems are widely distributed in bacterial genomes. The decoding of genomes through large-scale sequencing endeavours has led to the discovery of many TA pairs, including ones that belong to novel families. In a recent survey of 2181 prokaryotic genomes, which encompassed their chromosomes, phages, and plasmids, over 7000 type

II toxins and 10000 antitoxins were predicted (Leplae et al., 2011). Eighteen of these antitoxins and 23 of these toxins were validated experimentally in *E. coli*, and novel toxin and antitoxin families were revealed. TA mining efforts demonstrated that the abundance and copy number of these systems could vary drastically between organisms. In some genomes, over 100 type II TAs have been predicted, representing over 1% of their total open reading frames. On the other hand, TA loci may be completely absent in other genomes. The number of TA pairs present in an organism is independent of the size of its genome. Dramatic differences in the TA abundance are also observed between organisms of the same genus and even the same species. For instance, the laboratory strain of *E. coli*, *E. coli* K-12 MG1655, carries 16 type I and 12 type II TA pairs. The enterohemorragic strain *E. coli* O157:H7 and the uropathogenic strain CFT1073 contain only 10 and three type II TA pairs respectively, even though these three strains share approximately 40% of the total coding sequences. Some TA families also appear to be more prevalent across different genomes than others. The RelBE (ParE) TA family is perhaps one of the most common systems. It is highly represented in Proteobacteria and is found in other bacterial phyla, including Actinobacteria, Cyanobacteria, and Firmicutes (Leplae et al., 2011).

For many years, screening efforts have been skewed toward identifying type II TA families, while type I TA loci appeared to be under-represented. The small size and hydrophobicity of type I toxins and the sRNA nature of their corresponding antitoxins have imposed challenges for their search. However, type I TA pairs have been successfully isolated in recent searches and validated in subsequent assays, suggesting that they may be as abundant as their type II counterparts (Fozo et al., 2010). In this particular study, exhaustive BLAST searches and additional computational searches for

small ORFs and tandem repeats were carried out to identify putative type I toxins. Some of the common features used to define potential toxins were the presence of ORFs encoding peptides of <70 amino acids that contained bulky polar residues in the C-terminus and a possible transmembrane domain. The propensity for type I TA pairs to occur in tandem repeats in an intergenic region was also considered in one of their searches. The locations of antisense sRNA antitoxins were predicted using local free-energy minima of RNA folding. From this study, several type I toxin candidates in enterohemorrhagic *E. coli*, *Enterococcus faecalis*, and *B. subtilis* were validated. Moreover, type I toxins in Enteroproteobacteria and Firmicutes lineages outside those that were previously found to contain type I TA pairs were unveiled. As for type III TA systems, a recent study by the Salmond laboratory has identified three additional independent families based on structure-based homology and iterative protein sequence comparisons (Blower et al., 2012). Representatives from all three families were confirmed to exhibit toxin and antitoxin functions in *E. coli*, and the anti-phage activity of two was validated. These TA pairs are abundant and widespread in bacteria of the Firmicutes, Fusobacteria, and Proteobacteria phyla. Similarly, ToxIN has been detected in other gram-positive and gram-negative bacteria in addition to the aforementioned *E. carotovora* subspecies (Fineran et al., 2009).

An examination of the arrangement of type II TA systems in bacterial genomes suggests that they were likely acquired through horizontal gene transfer. They are frequently found in cryptic prophages, genomic islands, conjugative transposons, superintegrons, and intergenic regions. In *E. coli* K-12 MG1655, many TA pairs are found in proximity of repetitive palindromic repeats, which are often conserved in

inverted repeats associated with mobile genetic elements. In contrast, type I TA pairs are suggested to be inherited in a lineage-specific manner. Tandem repeats of type I TA pairs are thought to have arisen from duplication of a single locus.

1.8. Biological functions of TA systems

1.8.1. Plasmid addiction

TA systems were first identified on plasmids in the early 1980s. Coined CcdAB for coupled cell division, this type II TA pair was found on the mini-F plasmid in *E. coli* and was shown to couple the proliferation of the plasmid to cell division (Figure 1.5A) (Ogura & Hiraga, 1983). Loss of the TA-encoding plasmid in a segregant followed by the rapid degradation of unstable CcdA antitoxins would allow the CcdB toxin to accumulate in the cell. CcdB is then free to act on DNA gyrase and inhibit its ability to catalyse DNA supercoiling (Bernard & Couturier, 1992). This ultimately inhibits replication, induces the SOS response by introducing double-strand breaks, and results in post-segregational cell death. Since the discovery of CcdB, additional TA systems have been identified on plasmids and have been linked to plasmid addiction. These include type I systems such as FlmAB on the F plasmid (Loh et al., 1988), Hok/Sok on plasmid R1 in *E. coli* (Gerdes et al., 1986), and RNAI-RNAII on plasmid pAD in *E. faecalis* (Greenfield et al., 2000).

1.8.2. Stabilization of genomes

In recent years, TA systems have been found to be prevalent in chromosomes as well as on plasmids. It is suggested that some of these chromosomally encoded TA pairs are merely selfish genomic parasites, which have been acquired from plasmids or other sources, integrated into the genome, and maintained due to their addictive properties. Nevertheless, many TA systems have been implicated in functions that are important to

the physiology of the bacteria. Akin to plasmid encoded TA pairs, some chromosomal TA systems can promote the maintenance of self-transmissible mobile genetic elements in bacteria by eliciting post-segregational cell death (Figure 1.5B). For example, the MosAT TA pair found in the SXT integrative and conjugative elements (ICE) of many clinical isolates of *Vibrio cholera* can stabilize SXT in their genomes (Wozniak & Waldor, 2009). Overexpression of factors that promote SXT excision and transfer can induce *mosAT* expression, thereby preventing the loss of SXT in daughter cells should cell division occur before reintegration of the ICE into the genome. Many chromosomal TA systems are also associated with cryptic prophages. These include the *ypjF-yfjZ* pair found in the CP4-57 cyrptic prophage and *relBE* found in the cryptic Qin prophage of *E. coli* K-12 (Wang et al., 2010). These elements may have been instrumental in the integration of the cryptic prophages in the genome and might have acquired additional functions once they are stabilized.

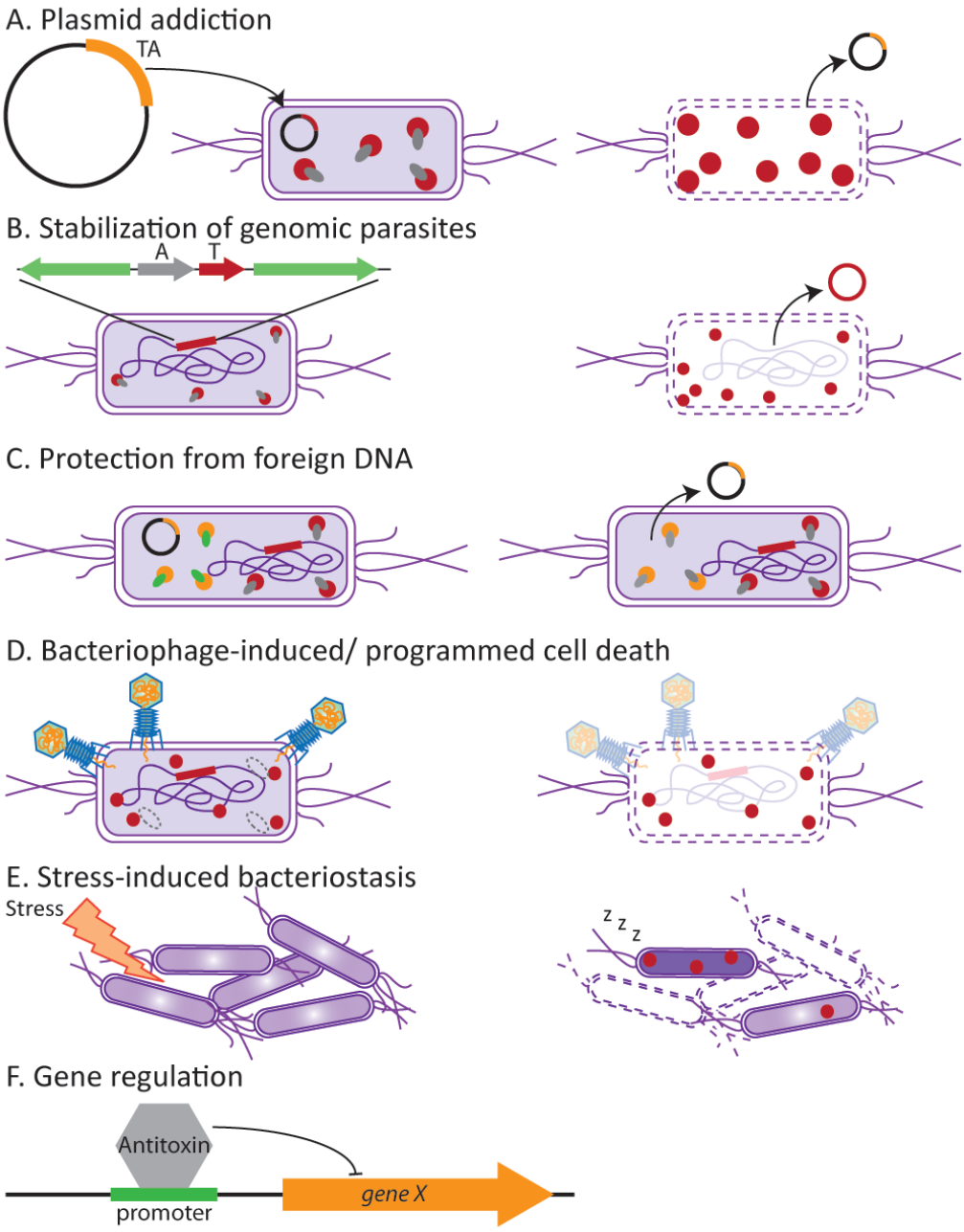


Figure 1.5. Potential functions of plasmid and chromosomally encoded TA systems. In these diagrams, toxin moieties are depicted by circles in red or in orange (panel C), while antitoxins are depicted in grey or in green (panel C).

1.8.3. Protecting the genome from foreign DNA

Chromosomally encoded TA systems can also safeguard the integrity of bacterial genomes by protecting them against foreign mobile genetic elements, such as those from bacteriophages and plasmids (Figure 1.5C) (Van Melderen, 2010). Saavedra De Bast and colleagues previously demonstrated that integration of the *ccdAB* TA pair derived from *Erwinia chrysanthemi* (ccd_{Ech}) into the genome of *E. coli* MG1655 prevented the postsegregational killing mediated by an F-plasmid derived *ccdAB* homolog expressed on a replication-thermosensitive vector (Saavedra De Bast et al., 2008). Similarly, the ChpBI antitoxin, which is part of the ChpBK-ChpBI TA system encoded on the chromosome of *E. coli* K-12, could neutralize the plasmid-borne ParD toxin (Santos Sierra et al., 1998). Sequence alignment of the ChpBK and ParD toxins revealed that the two share 18% identical and 19% conserved residues. These observations further suggest that there may be cross talk between homologous TA systems encoded on chromosomes and on plasmids, and they further support the role of certain chromosomal TA pairs as anti-addiction modules. Alternatively, the presence of chromosomal TA pairs may allow bacteria to take up pieces of DNA that only contain toxin alleles in the absence of their corresponding antitoxins.

1.8.4. TA systems for anti-phage defence

As bacteriophages outnumber bacteria by a ratio of approximately 10 to 1, it is important for bacteria to establish a viable defence mechanism against their viral parasites. Some TA systems have been implicated as part of that defence. It has been proposed that bacteriophage infections can perturb transcription and translation in bacterial hosts, thereby activating their endogenous TA systems (Otsuka & Yonesaki,

2012). ToxIN, the type III TA system previously discussed, has been demonstrated to be a part of the phage abortive infection system (Fineran et al., 2009). Upon bacteriophage infections, ToxN is activated. Acting as an endoribonuclease, ToxN cleaves cellular mRNA, inducing reversible growth stasis or cell death (Blower et al., 2011). Type II toxin MazF and type I toxin Hok in *E. coli* are protective against phages P1 and T4, respectively (Hazan & Engelberg-Kulka, 2004; Pecota & Wood, 1996). The production of toxins delayed the maturation of phage particles, reduced cell lysis, and minimized the release of phage particles (Figure 1.5D). These toxins further induced altruistic suicide of phage-infected cells. While this act of altruism may not be favourable for the individual cell, this mechanism can benefit the fitness of cells growing in a community, such as in biofilms. As cells in the outer layer of the biofilm become infected and die, they can mitigate the spread of viral particles and form a protective layer to further exclude phages from siblings residing in the interior of the biofilm (Pecota & Wood, 1996).

Interestingly, it was recently reported that bacteriophages have countermeasures to defend against bacterial toxins. In *E. coli*, toxins RnlA and LsoA, part of the RnlAB and LsoAB TA pairs, are activated during T4 infections, where they act as endoribonucleases to degrade host and phage mRNAs (Koga et al., 2011; Otsuka & Yonesaki, 2012). On the other hand, phage T4 was demonstrated to produce a 60-amino acid protein referred to as Dmb, which can act as an antitoxin against both RnlA and LsoA (Otsuka & Yonesaki, 2012). This study suggests that T4 may encode additional antitoxins that can counteract other toxins expressed in *E. coli*.

1.8.5. Programmed cell death and toxin expression

In addition to eliciting the altruistic killing of phage-infected cells, TA systems have also been implicated in the programmed cell death (PCD) of a large fraction of the population in response to a number of different stressful conditions. These conditions include amino acid starvation, high temperature, oxidative stress, thymine-less death, and antibiotic treatments. In a study by the Engelberg-Kulka group, who has been one of the major advocates for the PCD hypothesis, five type II TA pairs were implicated in PCD in *E. coli* (Kolodkin-Gal et al., 2009). They showed that the TA pairs had differential effects on *E. coli* depending on whether the cells were growing in the free-swimming planktonic phase or in surface-associated biofilms. Two TA pairs in particular, *dinJ-yafQ* and *mazEF*, are important for biofilm formation, as their deletion resulted in defective biofilms. It is suggested that these two TA pairs are responsible for killing a large proportion of the population, and the threshold of dead cells in a biofilm is important for its structural integrity. It is likely that the death of these cells provides the surviving population with nutrients, proteins, signalling molecules, and extracellular matrix molecules that can sustain the remaining cells and support the architecture of the biofilm.

MazEF, the paradigm of TA systems associated with PCD, has been extensively studied by the Engelberg-Kulka group over the last two decades. It has been demonstrated that when cells are challenged by certain stresses, the expression of *mazEF* is downregulated in response to alarmone guanosine pentaphosphate ((p)ppGpp) (Aizenman et al., 1996; Hazan et al., 2004). Under these circumstances, the labile MazE antitoxin is rapidly degraded by ATP-dependent ClpAP serine protease. MazF is then free to cleave specific mRNAs at their translation initiation regions and remove the anti-Shine-Dalgarno

sequence of the 16S rRNA in the 30S ribosomal subunit, thereby further interrupting translation initiation of canonical mRNAs. *mazEF*-mediated PCD depends on the presence of a quorum sensing pentapeptide coined extracellular death factor (EDF), which mediates the liberation of MazF from MazE (Belitsky et al., 2011). Under severe DNA damage, *mazEF*-EDF prevents the "apoptotic-like *E. coli* death pathway" mediated by *recA* and *lexA* and directs cells toward a non-apoptotic programmed cell death pathway that does not involve membrane depolarization and DNA fragmentation (Erental et al., 2012). This phenomenon can result in the death of 99% of the population, and the remaining 1% is believed to seed the future population once the stress is alleviated. It has also been reported that MazF production induces the synthesis of a group of proteins of <20 kDa, some of which are required for the death of the majority of the population (Amitai et al., 2009). The other subset of MazF-induced proteins, however, promotes the survival of a small subpopulation. These include SoxS and SoxR, which have been shown to detoxify reactive oxygen species. They further suggest that the activation of *mazEF*-EDF following the treatment with some antibiotics can alter the downstream effects of these drugs, rendering some traditionally bacteriostatic drugs bactericidal (Kolodkin-Gal et al., 2008; Engelberg-Kulka et al., 2009).

At present, the factors that determine which subpopulation is destined to survive and which subpopulation is destined to commit altruistic suicide in response to specific stresses and subsequent toxin action are unknown. The selection may be an entirely stochastic process. While a number of studies supporting the role of TA in bacterial PCD have been published, it is still a topic of active debate, if not controversy, in the field. Several groups have reported that they were not able to reproduce *mazEF*-mediated PCD

under similar conditions (Van Melderen, 2010; Christensen et al., 2003; Christensen & Gerdes, 2003). Perhaps cell death caused by some of these adverse conditions may partially depend on TA systems, which act in concert with other components to inhibit the affected cell (Yamaguchi & Inouye, 2011). A MazF homolog from *Myxococcus xanthus* is essential for fruiting body formation, where it is responsible for the lysis and obligatory altruistic cell death of 80% of the cell population (Nariya & Inouye, 2008). As obligatory cell death is not required in *E. coli* under stresses such as amino acid starvation, it has been proposed that the cell death observed upon MazF overproduction in *E. coli* may only occur under artificial conditions (Yamaguchi & Inouye, 2011). It is possible that MazF causes cells to enter a state of dormancy from which they can recover once the stress is alleviated and sufficient MazE is produced to neutralize MazF (Yamaguchi & Inouye, 2011).

1.8.6. To resist and persist: TA systems induce reversible bacteriostasis

In response to adverse conditions, toxin production can cause cells to enter a "quasi-dormancy" state in which they are more resistant to stressors (Figure 1.5E). MazF-induced cells were fully active in a number of metabolic pathways and retained transcription and translation capabilities, suggesting that while these cells were unable to proliferate they were still viable (Suzuki et al., 2005). In another study, the production of the RelE toxin, whose expression is induced under nutritional stress, in *E. coli* severely impaired translation, consequently suppressing colony formation (Pedersen et al., 2002). However, the induction of its cognate antitoxin, *relB*, was found to alleviate its inhibitory effect on protein synthesis and reverse growth stasis. Similar observations were made following the induction of *mazF* and its corresponding antitoxin, *mazE*. Together, these

results suggest that these toxins modulate global macromolecular synthesis during nutritional stress and prompt the cells to enter a reversibly viable but nonculturable state.

In a search for TA loci in 126 sequenced prokaryotic genomes, it was found that the genomes of free-living and/or slow-growing prokaryotes have abundant TA pairs (Pandey & Gerdes, 2005). For example, *M. tuberculosis* carried 38 pairs, while *Nitrosomonas europeae* harboured 43 pairs. On the other hand, they are absent in many obligate intracellular organisms, such as *Rickettsia prowazekii* and *Mycobacterium leprae*. As free-living organisms often need to thrive under variable and limiting growth conditions, the presence of TA pairs in their genomes may be important for stress response and quality control of gene expression. In fact, bacteria cultured under different stress conditions can augment their expression of specific TA genes. For example, the aforementioned *mazEF* and *relBE* systems are induced in cells grown under amino acid and glucose starvation (Christensen et al., 2003; Christensen et al., 2001). In *M. tuberculosis*, two of its 30 confirmed TA pairs are induced by hypoxia, while two others are upregulated during macrophage infection (Ramage et al., 2009).

Although gene expression analyses implicate TA systems in stress responses, many gene deletion studies failed to produce phenotypic changes in cells cultured under stress-inducing conditions. For example, deleting *relBE* in *M. tuberculosis* did not increase the sensitivity of the organism to hypoxia, nitrosative, or oxidative stress (Singh et al., 2010). Deletion of five type II TA pairs from *E. coli*, including *mazEF* and its homolog *chpB* as well as *relBE* and two of its homologs, did not reduce the competitiveness of the deletion mutant relative to the wild-type strain in the presence of a number of stresses (Tsilibaris et al., 2007). These findings suggest that more than one TA

pair may act together, perhaps in conjunction with other factors, to respond to specific stresses. Individual TA pairs may each play a small role in stress response, thus their deletion may not produce a prominent phenotype nor would it alter competitiveness.

Further studies conducted using the five TA deletion strain implicated the involvement of these TA pairs in biofilm formation (Kim et al., 2009). Unlike the model describing the involvement of MazF and YafQ in the development of these multicellular structures, these five toxins are not thought to elicit the lysis of cells and the release of macromolecules to support biofilm development. Rather, they cause differential expression of *tabA*, whose product represses type I fimbriae. Consequently, the loss of these five TA genes decreases biofilm formation at 8 h of growth. On the other hand, as these TA pairs are involved in the dispersal of cells, biofilm formation increased after 24 h of growth in the deletion strain. In addition to *mazEF, relBE*, and their homologs, other TA pairs have been linked to biofilm formation. These include *mqsRA*, which exhibits differential expression in *E. coli* biofilms (Ren et al., 2004). The production of the toxin MqsR and the deletion of the antitoxin MqsA have been shown to enhance the production of cellulose (an exopolysaccharide) and curli (a matrix protein), which in turn promote cell adhesion and biofilm development (Wang et al., 2011). As MqsA constitutes one of the components that regulate RpoS, the stationary phase sigma factor that controls over 70 genes conferring resistance to oxidative stress, UV radiation, heat shock, hyperosmolarity, low pH, and ethanol stress, this TA system is implicated in diverse stress response pathways.

Bacteria in biofilms tend to be phenotypically heterogeneous: some of these cells are extracellular matrix producers, some are metabolically inactive, some are slow

growing, and others can survive as "persisters" (Vlamakis, 2011). While antibiotic treatments can target cellular processes of growing cells and kill a large fraction of the population, they have little effect on non-growing persister cells whose transient decrease in metabolic activity offers them protection from antibiotics (Lewis, 2010). These persister cells are thought to form stochastically in the population as a bet-hedging strategy to enable a subset of the population to swiftly cope with specific stress conditions without needing to take the time to activate stress response pathways (Balaban, 2011). The presence of persisters in biofilms is a major culprit for the recalcitrance of chronic infections. Unlike antibiotic resistant mutants, persister cells have not undergone genetic changes and cannot grow in the presence of the drugs. Instead, they arise from differential gene expression. Transcriptome analysis has revealed that the expression of several TA pairs, including *relBE*, *mazEF*, *dinJ-yafQ*, and *mqsRA* are elevated in persister cells. In fact, one of the first genes linked to persister cell formation was *hipA*, a toxin-encoding gene. Ectopic overexpression of *hipA* resulted in multidrug tolerance (Vazquez-Laslop et al., 2006). HipA, which has a eukaryotic serine/threonine kinase-like fold, phosphorylates elongation factor EF-Tu, rendering it inactive (Schumacher et al., 2009). This event subsequently blocks translation and induces dormancy. A similar mechanism of EF-Tu phosphorylation and deactivation leading to dormancy has been proposed for sporulation in *B. subtilis*. A recent study by the Gerdes group demonstrated that persister cell formation in *E. coli* may be induced by the Lon protease-mediated degradation of ten antitoxins that repress mRNase toxins (Maisonneuve et al., 2011). The activation of these mRNases, which include RelE, MazF, HigB, MqsR, and YafO, inhibits global translation, consequently resulting in dormancy and persistence.

Another factor that has been linked to bacterial persistence is type I toxin TisB. Although overexpression of this toxin inhibits growth of the whole population, mild ectopic expression of *tisB* increases persistence, protecting cells from the deleterious effects of antibiotics (Unoson & Wagner, 2008). Deletion of the *tisAB-istR1* locus reduces persister cell formation by 10 to 100-fold in the presence of RecA. TisB production may be induced by the SOS response, which can be triggered by treating cells with fluoroquinolones, including ciprofloxacin (Dorr et al., 2010). DNA damage induced by fluoroquinolones activates RecA, which then leads to LexA cleavage and *tisB* induction. It has been proposed that the production of TisB, a 29-amino acid membrane-targeting hydrophobic peptide, can result in the formation of ion-conductive pores, dissipation of the proton motive force (PMF) and reduction of cellular ATP (Unoson & Wagner, 2008; Gurnev et al., 2012). This consequently causes cells to enter a multidrug-tolerant and dormant state. The level of *tisB* expression during the SOS response has been suggested to be heterogeneous between cells and may fluctuate around a mean level (Lewis, 2010). Cells that stochastically achieve a higher level of *tisB* expression go on to become persisters. As the expression of *tisB* is stimulated by the SOS response, it is possible that other stress response pathways may also be linked to persister cell formation (Lewis, 2010). Pathogens in the host environment are constantly exposed to stresses, such as reactive oxygen species, high temperature, and low pH. These conditions may also induce the production of additional toxins and promote the formation of persister cells.

1.8.7. TA systems as global regulators

Toxins and antitoxins can act independently or in complex to regulate gene expression on a global level (Figure 1.5F). As described earlier, a number of type II

toxins act as endoribonucleases and their activity can lead to the degradation of a large proportion of mRNAs while favouring the translation of a subset that are resistant to their cleavage. For example, MazF has been proposed to cleave most of the cellular mRNAs and promote the synthesis of proteins necessary for both cell death and survival in response to oxidative stress (Amitai et al., 2009). MqsR, a sequence-specific RNase that has been linked to biofilm and persister cell formation, is a global regulator that enriches specific mRNAs encoding stress-related proteins (Yamaguchi et al., 2009; Wang & Wood, 2011). mRNAs spared by this toxin typically lack its cognate cleavage site or harbour secondary structures that occlude the recognition sequence. The action of MqsR provides a means for some cells to rapidly express new proteins in order to cope with stress by forming biofilms or becoming dormant persister cells in the biofilms (Hong et al., 2012).

The regulation by MqsRA is quite multifaceted. MqsR can repress *rpoS*, which encodes the general stress response regulator σ^S. This in turn decreases the ability of the cell to respond to antibiotic and acid stress, consequently causing the damaged cell to enter a dormant state and increase persistence (Hong et al., 2012). Alternatively, the MqsA antitoxin represses *rpoS* by directly binding to its promoter under normal growth conditions (Wang et al., 2011). Following oxidative stress, MqsA is rapidly degraded and its repressive effect on *rpoS* is alleviated. This subsequently stimulates the production of second messenger 3′,5′-cyclic diguanylic acid (c-di-GMP), which prompts the cells to adopt a more sessile lifestyle. The motility of the bacteria is further impaired following MqsA degradation. Since MqsA also suppresses the expression of *csgD,* the degradation of the antitoxin along with an increase in σ^S activity stimulate *csgD*, resulting in increased

curli and cellulose production. Together, these responses promote biofilm formation upon oxidative stress. Furthermore, MqsR acts as a conduit between autoinducer-2 (AI-2) and motility genes (Gonzalez Barrios et al., 2006). It promotes motility in *E. coli* strains devoid of conjugative plasmid Rldrd19 in response to AI-2, allowing these cells to attach to surfaces and form biofilms. It was also reported that 14% of the genes downregulated following *mqsR* deletion are controlled by AI-2. The MqsRA complex and MqsA alone can further regulate the transcription of other toxins. This TA complex represses the expression of *cspD*, a toxin that can act as a DNA replication inhibitor, by binding to its promoter (Kim et al., 2010). When MqsA is degraded under stress, *cspD* is derepressed and the activity of this toxin is hypothesized to act together with MqsR to regulate biofilm and persister production.

In addition to gene expression, toxins can act on other proteins to govern other cellular processes. Like many other toxins, YeeV (also known as CbtA) inhibits growth and decreases cell viability following ectopic expression (Brown & Shaw, 2003). These growth defects can be suppressed by the expression of its cognate antitoxin, YeeU. Interestingly, the expression of this toxin also results in morphological changes in *E. coli*. Its production causes the formation of filamentous, round, and "lemon-shaped" cells as opposed to cells with the typical rod-shaped morphology. This change in morphology is attributed to the effect of CbtA on the activities of two cytoskeleton proteins, the tubulin-like FtsZ and the actin-like MreB (Tan et al., 2011). CbtA can inhibit the GTPase activity of FtsZ and prevent the GTP-dependent polymerization of FtsZ, which can give rise to filamentous cells. It can also inhibit the ATP-dependent polymerization of MreB, resulting in round cells. The simultaneous inhibition of both FtsZ and MreB contribute to

the "lemon-shaped" appearance of the affected cells. It can also impede cell division, chromosomal segregation, and cell polarity. At present, the conditions triggering YeeV activation and cell division inhibition are unknown.

While the functions of some TA systems, namely type II and type III pairs, have been elucidated, the biological relevance of many is still obscure. The proposed functions for a few of these systems, such as MazEF, remain a matter of debate in the field. With the exception of *tisAB-istR1*, the biological roles of many type I TA pairs are less well defined. Their functions are hypothesized based on those determined from their better characterized type II counterparts (Fozo, Hemm, et al., 2008). Many of these toxins are suggested to be membrane active like TisB, because they share similar small and hydrophobic characteristics as well as the propensity to adopt transmembrane α-helices. Therefore, they too, may act by depolarizing the cytoplasmic membrane and reducing ATP production in order to induce transient growth stasis in response to specific stress conditions. The toxicity observed with these peptides may solely be an artefact of their ectopic overexpression. At endogenous levels, these toxins may have additional cellular targets and regulatory functions, and they may not exert deleterious effects on the bacteria. Thus, in order to determine the physiological roles of TA systems, it is imperative for their gene expression, localization, binding partners, and activities to be examined under various growth conditions.

1.9. The IbsC/SibC TA System

One type I TA pair whose function remains to be elucidated is IbsC/SibC belonging to the Ibs/Sib family of TA systems. Encoded in between the *fau* (previously *ygfA*) and *serA* genes in *E. coli* K-12, the 141 nt SibC antitoxin was first discovered by

the Storz and Gottesman laboratories through a screen for novel noncoding sRNAs using comparative genomics and high density microarrays (Wassarman et al., 2001). Later, the Storz group observed that when an overexpression construct harbouring an inducible copy of *sibC* was transformed into a *sibC* deletion strain, viable transformants did not arise unless the plasmid-borne copy of *sibC* was induced (Fozo, Kawano, et al., 2008). This was similar to the plasmid addiction phenotype observed in daughter cells that failed to inherit a TA-encoding plasmid.

Our group first came across this TA pair through two genetic screens: one aimed at the discovery of novel sRNAs in *E. coli* K-12 (*Navani, Barker, Li*, *et al., unpublished results*) and a second directed at the identification of phenotypic changes following the overexpression of the antisense sequences of predicted sRNAs. In the first screen, Navani, Barker, and colleagues utilized a shotgun cloning approach and cloned fragments of DNA of less than 400 bp derived from *E. coli* K-12 (Figure 1.6A). They subsequently screened for growth inhibition upon inducing the expression of these sequences. After examining approximately 3,500 constructs, they isolated one that caused the desired phenotypic changes. This sequence mapped to the reverse complement of *sibC*. In the second screen, we examined growth inhibition following the overexpression of the reverse complement of a subset of sRNAs (Figure 1.6B). The rationale, methods, and findings of this screen are discussed in detail in Chapter 2 of this thesis. Likewise, the antisense sequence of *sibC* was found to be deleterious to the bacteria in this screen. A closer examination of the reverse complements of *sibC* by Fozo and colleagues and by our group revealed an open reading frame encoding a peptide of 19 amino acids in its antisense strand. The gene encoded on the opposite strand of *sibC* was coined *ibsC* for

induction brings stasis (Fozo, Kawano, et al., 2008). SibC and IbsC were further confirmed to be a type I TA pair.

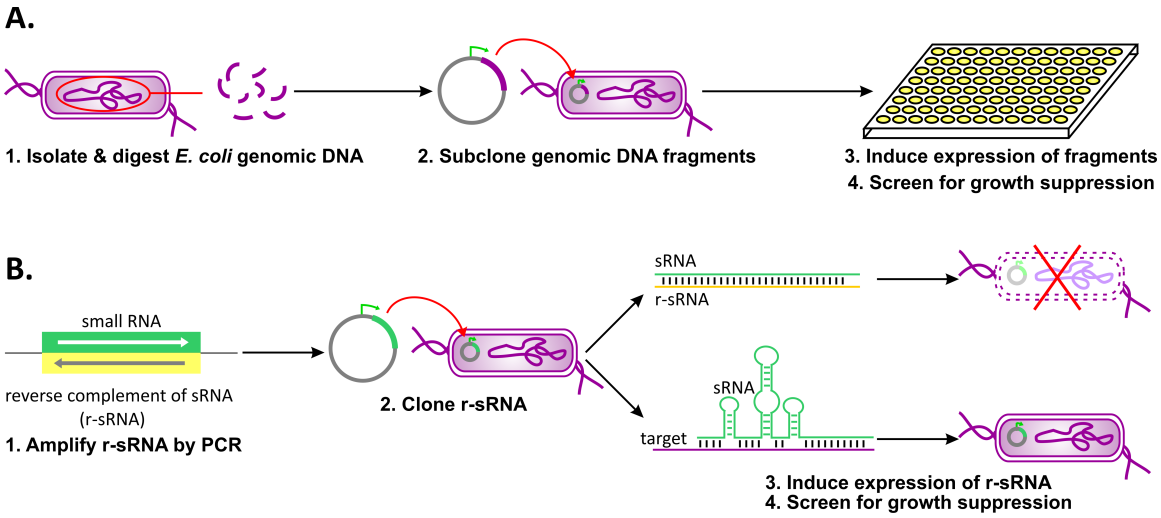


Figure 1.6. Screens for novel sRNAs and for sRNAs with essential cellular functions. A. Screening for novel sRNAs in *E. coli* K-12. A shotgun cloning approach was used to clone gene fragments of less than 400 bp from the *E. coli* K-12 genome into a tetracycline-inducible system. The expression of these sequences was then induced, and those that were capable of eliciting growth suppression were isolated. B. Screening for sRNAs with critical cellular function. The reverse complement sequences of a subset of sRNAs from *E. coli* K-12 were amplified by PCR and cloned into a tetracycline-inducible system. Sequences that could elicit growth suppression, presumably by sequestering their complementary sRNAs from their cognate targets, were isolated.

The genome of *E. coli* K-12 contains five homologs of the *ibs*/*sib* TA pairs, coined *ibs/sibA-E* (Figure 1.7A). *ibsC/sibC* is the sole pair that is found between *fau* and *serA*. The other four pairs are found in tandem at two other loci: *ibsA*/*sibA* and *ibsB*/*sibB* are found between *ygeL* and *mdtA*, while *ibsD*/*sibD* and *ibsE*/*sibE* are found between *yqiK* and *rfaE*. In a search for type I TA pairs across 774 bacterial genomes, *ibs/sib* homologs were identified in the genomes of other γ-Proteobacteria (Figure 1.7B) (Fozo et al., 2010). They are found in other *E. coli* strains, including *E. coli* O157:H7, in addition

to species that are closely related to *E. coli (*such as *Salmonella typhimurium* and *Shigella flexneri* strains) and some that are more distantly related (such as *Helicobacter pylori*). Similar to *E. coli* K-12, many of these bacteria carry more than one copy of this TA pair. For instance, the genomes of *Shigella boydii* and *Haemophilus somnus* harbour seven *ibs/sib* homologs. An examination of the neighbouring genes flanking the *ibs/sib* homologs from various *E. coli*, *S. flexneri*, and *S. typhimurium* strains indicate that the identities of the adjacent genes are not necessarily conserved (Hershberg et al., 2003). This suggests that a functional relationship between the TA pairs and their flanking genes may not exist.

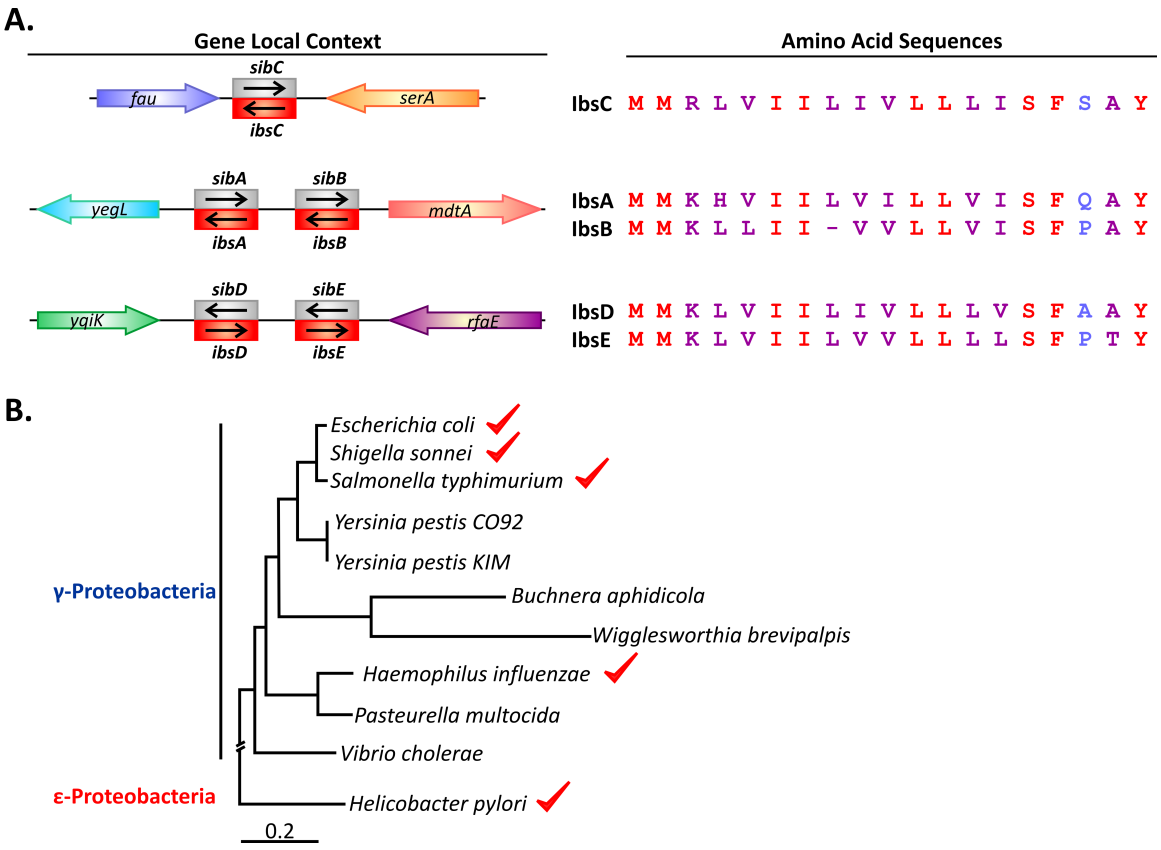


Figure 1.7. Location and orientation of *ibs/sib* TA pairs in the *E. coli* genome and their distribution across the Proteobacteria lineage. A. The location of *ibsC/sibC* and its four homologs in the *E. coli* K-12 genome are presented along with their flanking genes. The amino acid sequences of the Ibs toxins are also shown. B. Distribution of the *ibs/sib* TA pairs in the genomes of Proteobacterial species. This TA system is found in a number of γ-Proteobacteria and in ε-Proteobacteria *Helicobacter pylori*. Species in which *ibs/sib* homologs are present are denoted with a red checkmark.

1.10. Properties and regulation of the IbsC toxin and the SibC antitoxin

IbsC and its homologs in *E. coli* K-12, with only 18 to 19 amino acids, are amongst some of the smallest type I toxins identified thus far. Similar to other type I toxins, the Ibs peptides are highly hydrophobic. IbsC, for example, has nearly 75% hydrophobic residues. All five Ibs homologs carry an overall charge of +1 due to the presence of a lysine residue (in the case of IbsA, B, D, and E) or an arginine residue (in

the case of IbsC) at position three. These peptides were hypothesized to adopt α-helical structures. Our preliminary immunoblotting experiments suggest that IbsC is not secreted into the extracellular milieu (*Mok and Li, unpublished results*). A study by Fozo and colleagues demonstrated that the overexpression of IbsC compromised the integrity of the inner membrane of *E. coli*, resulting in reduction in membrane potential (Fozo, Kawano, et al., 2008). This suggests that this toxin likely localizes to and accumulates in the inner membrane when produced at high levels. Whether it inserts in the membrane as a transmembrane helix or alters membrane architecture by aggregating on the cytoplasmic face of the membrane as a multimer remains to be elucidated.

Another missing piece of the IbsC puzzle concerns its expression. It is not yet known if and when the expression of the toxin is induced. Fozo and colleagues reported that they were unable to detect the *ibsC* mRNA via northern analysis in wildtype *E. coli* K-12. However, when *sibC* transcription is impaired in a *sibC* promoter deletion strain, the *ibsC* transcript was detected. The levels of *ibsC* expression were comparable in cells cultured in rich media (Luria-Burtani broth) and in minimal media, regardless of whether the cells were in exponential phase or stationary phase. This suggests that SibC is expressed in excess of *ibsC* under normal growth conditions, and it exerts negative effects on the transcript levels of *ibsC*. The SibC-mediated *ibsC* repression was confirmed by Han and colleagues, who demonstrated that SibC and *ibsC* first recognize each other via two stem loop regions, coined target recognition domains (TRD) 1 and 2, forming a transient kissing complex before a more stable, though incomplete, duplex is formed (Han et al., 2010). The binding of the two RNAs is hypothesized to prevent the translation initiation of *ibsC* and may target the duplex for degradation. The precise mechanism of

ibsC and SibC interaction will be discussed in Chapter 4 of this thesis. With the exception of *ibsA*, cross interactions between *ibs* and Sib homologs in K-12 were not observed, despite their high sequence homology. It is likely that the sequence variability at the two TRDs ensures that recognition and binding only occurs between cognate pairs of toxin transcripts and antitoxin sRNAs.

In previous northern analyses, two transcripts have been observed for each of the *sib* genes. In the case of *sibC*, the larger of the two species was reported to be 141 nt and the smaller species was approximately 110 nt (Fozo, Kawano, et al., 2008; Wassarman et al., 2001). Through 5′ and 3′ rapid amplification of cDNA ends (RACE) experiments, it was found that the shorter transcript resulted from the 3′ post-transcriptional processing of the longer transcript (Fozo, Kawano, et al., 2008). It was observed that the 141 nt SibC variant dominated in cells in stationary phase when cultured in rich or minimal media. Considering that one of the TRDs resides between nucleotides 99 to 115 of SibC, the longer transcript is likely the active species of the two, because 3′ cleavage of the sRNA may diminish its activity. These observations indicate that *sibC* is constitutively expressed under normal and nutrient-limiting growth conditions. They further suggest that *ibsC* expression is repressed under these circumstances. Deletion of one or all five *ibs/sib* TA pairs did not affect the growth of *E. coli* K-12 under these conditions (Fozo, Kawano, et al., 2008). However, the precise conditions and events that may trigger the production of IbsC remain unknown.

1.11. Outstanding questions pertaining to IbsC/SibC

Upon confirming that *ibsC* encodes a small peptide toxin that is deleterious to bacteria when expressed ectopically, many aspects pertaining to its properties, mechanism

of action, regulation, and function remained obscure. At the beginning of this study, the physical properties of this toxin were predicted based on its sequence, but they had not yet been validated. The biological function of this TA pair and its homologs was also a mystery. It was 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 questions pertaining to the significance of IbsC/SibC, details related to its biochemical properties and expression must be elucidated. This present work lays the groundwork necessary to reveal the true role of the enigmatic IbsC/SibC and its related homologues.

1.12 Strains and plasmids used to study ibsC/sibC

The experiments detailed in this thesis were primarily conducted in *E. coli* DH5αZ1, which was kindly provided by the laboratory of Hermann Bujard (Lutz & Bujard, 1997). This strain is a variant of the *E. coli* K-12-derived cloning strain *E. coli* DH5αZ. It has been engineered to constitutively express a tetracycline repressor (TetR) and a lac repressor (LacR). This strain was used in conjunction with a plasmid modified from pZE21-MCS1, which was also obtained from the Bujard group (Figure 1.8). This plasmid contains a tetracycline-inducible promoter (P_{LtetO1}), which was synthesized using the strong phage lambda P_L promoter with operator sequences from the promoter of Tn10 tetracycline resistance operon inserted within. It also carries the ColE1 origin of replication, which is associated with high plasmid copy numbers. This promoter is tightly

repressed by TetR, but it can be induced over 2500-fold in the presence of anhydrotetracycline (Atc). Our group further removed the consensus ribosome binding site (RBS) present on pZE21-MCS1 and restored its multiple cloning site (*Swanson and Li, unpublished data*). This vector was subsequently coined pNYL-MCSII. Using this system, expression of desired DNA sequences can be tightly controlled using P_{LtetO1}.

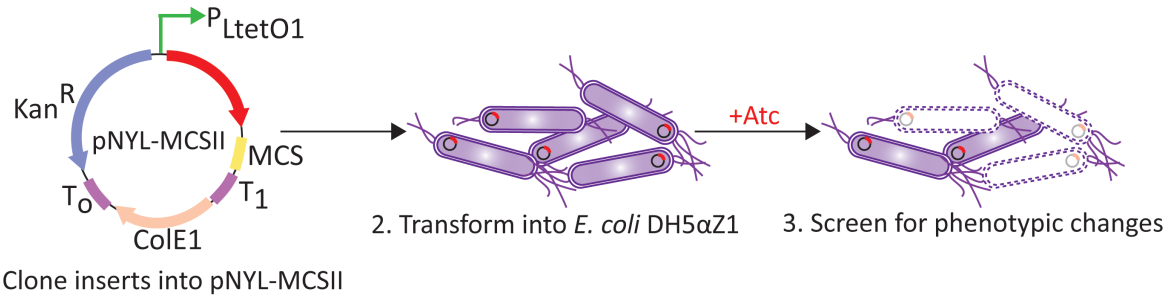


Figure 1.8. Design of pNYL-MCSII and assays carried out in *E. coli* DH5αZ1. pNYL-MCS11, derived from pZE21-MCS1, is depicted. Sequences of interest are cloned into this vector downstream of the tetracycline-inducible promoter (P_{LtetO1}). These constructs were then transformed into *E. coli* DH5αZ1, which constitutively expresses a tetracycline repressor (TetR). Using this system, the expression of the genes of interest can be induced following the addition of anhydrotetracycline (Atc). Phenotypic changes, such as growth suppression or cell death, caused by the expression of these sequences can be assessed thereafter.

1.13. Objectives of present study and overview of individual chapters

In Chapter 2 of this thesis, our efforts in screening for phenotypic changes following the overexpression of the antisense counterparts of sRNAs are described. In the first component of this project, we cloned the reverse complement of a subset of confirmed and predicted sRNAs from *E. coli* K-12 into pNYL-MCS11. The expression of these sequences was induced in *E. coli* DH5αZ1 using Atc, and we screened for growth suppression following their induction. Our initial objective was to isolate antisense sequences that could dimerize with sRNAs in order to sequester them and inhibit their

activities. Should the sRNAs carry out regulatory functions that are critical for cell survival, their inhibition would be deleterious to the cell. From this screen, one antisense sequence, corresponding to the reverse complement of *sibC* (referred to as *rygC* at the time of the study), was found to be growth inhibitory. This sequence did not inhibit growth via the antisense knock-down of SibC. Rather, it was determined to encode a peptide that was toxic to cells when produced. Our work toward confirming that the antisense sequence of *sibC* encoded a peptide and that the over-production of this peptide was the culprit for the growth inhibitory phenotype observed in our screen is also discussed in this chapter. The work presented in this chapter was published in an article in Chembiochem.

The second component of this project, presented in Chapter 3, involved the sequence and biochemical characterization of IbsC. As the five Ibs homologs in *E. coli* K-12 and homologs annotated in other Gram-negative organisms are well conserved, it is unknown whether this toxin is tolerant of mutations. Active mutants isolated from these experiments may be utilized as templates to search for novel toxins and Ibs homologs in additional microbial genomes. In this work, over 200 point mutants and a number of truncation variants of IbsC were generated. Through growth experiments carried out in *E. coli* expressing these *ibsC* derivatives, it was observed that this toxin could withstand extensive sequence changes despite being conserved in nature. The solution structures of IbsC and some of its mutants were investigated in this study. Furthermore, when the mechanism of toxicity of the IbsC derivatives was examined, it was determined that toxicity could be correlated with their ability to act on the inner membrane of *E. coli*. This work has since been published in the Journal of Biological Chemistry.

In Chapter 4 of this dissertation, the regulation of *ibsC* expression at the DNA and RNA levels was investigated. To examine the regulation of *ibsC* at the transcriptional level, sequences upstream of the *ibsC* ORF were fused to the gene encoding the green fluorescent protein (GFP). Fluorescence assays were carried out with these promoter-reporter constructs to delineate the promoter region of *ibsC* and to isolate potential regulatory elements. We further probed into the regulation of *ibsC* at the post-transcriptional level, which is mediated by SibC. Mutations were introduced in the two TRD sequences in SibC that were previously identified by Han and colleagues (Han et al., 2010). Growth assays carried out with cells expressing *ibsC* and mutant *sibC* indicate that TRD1 acts as the dominant sequence element for mediating SibC-*ibsC* interactions. Collectively, our observations indicate that the expression of *ibsC* is tightly repressed under the growth conditions examined.

The results described in Chapters 2 to 4 of this thesis have advanced our understanding of the sequence requirements and mechanism of IbsC and SibC function. Taking advantage of the toxicity and membrane active properties of IbsC, potential therapeutics and molecular tools based on this TA pair can be designed. The sequence and mechanistic information gathered from our studies can be applied to ameliorate and optimize such designs. As an example, we investigated the use of *ibsC* and toxic derivatives as positive selection markers in molecular cloning, as described in Chapter 5. The incorporation of *ibsC* and *ibsC* mutants into cloning vectors greatly improved the efficiency of ligase-dependent molecular cloning strategies, while the background was reduced. As a proof-of-concept, the cloning of *gfp* into these vectors consistently produced over 95% positive clones. The results presented in Chapters 4 and 5 have been

prepared for publication. Taken together, this thesis details our efforts from the identification of the IbsC/SibC TA system to the characterization of their action and the application of this enigmatic TA pair as a molecular tool.

1.14. References

Aizenman, E., H. Engelberg-Kulka, and G. Glaser (1996). 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: 6059-6063.

Altuvia, S. (2007). Identification of bacterial small non-coding RNAs: experimental approaches. Curr Opin Microbiol 10: 257-261.

Altuvia, S., D. Weinstein-Fischer, A. Zhang, L. Postow, and G. Storz (1997). A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. Cell 90: 43-53.

Amitai, S., I. Kolodkin-Gal, M. Hananya-Meltabashi, A. Sacher, and H. Engelberg-Kulka (2009). *Escherichia coli* MazF leads to the simultaneous selective synthesis of both "death proteins" and "survival proteins". PLoS Genet 5: e1000390.

Argaman, L., R. Hershberg, J. Vogel, G. Bejerano, E. G. Wagner, H. Margalit, and S. Altuvia (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr Biol 11: 941-950.

Bae, W., P. G. Jones, and M. Inouye (1997). CspA, the major cold shock protein of *Escherichia coli*, negatively regulates its own gene expression. J Bacteriol 179: 7081-7088.

Balaban, NQ. "Persister Bacteria." *Bacterial stress responses*. Ed. Storz, G., Hengge, R. 2 ed. Washington: ASM Press, 2011. 375-382. Print.

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

Bayer, A. S., R. Prasad, J. Chandra, A. Koul, M. Smriti, A. Varma, R. A. Skurray, N. Firth, M. H. Brown, S. P. Koo, and M. R. Yeaman (2000). In vitro resistance of *Staphylococcus aureus* to thrombin-induced platelet microbicidal protein is associated with alterations in cytoplasmic membrane fluidity. Infect Immun 68: 3548-3553.

Beguin, P., X. Wang, D. Firsov, A. Puoti, D. Claeys, J. D. Horisberger, and K. Geering (1997). The gamma subunit is a specific component of the Na,K-ATPase and modulates its transport function. EMBO J 16: 4250-4260.

Beisel, C. L., T. B. Updegrove, B. J. Janson, and G. Storz (2012). Multiple factors dictate target selection by Hfq-binding small RNAs. EMBO J 31: 1961-1974.

Belitsky, M., H. Avshalom, A. Erental, I. Yelin, S. Kumar, N. London, M. Sperber, O. Schueler-Furman, and H. Engelberg-Kulka (2011). The *Escherichia coli* extracellular death factor EDF induces the endoribonucleolytic activities of the toxins MazF and ChpBK. Mol Cell 41: 625-635.

Bernard, P., and M. Couturier (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J Mol Biol 226: 735-745.

Blower, T. R., X. Y. Pei, F. L. Short, P. C. Fineran, D. P. Humphreys, B. F. Luisi, and G. P. Salmond (2011). A processed noncoding RNA regulates an altruistic bacterial antiviral system. Nat Struct Mol Biol 18: 185-190.

Blower, T. R., F. L. Short, F. Rao, K. Mizuguchi, X. Y. Pei, P. C. Fineran, B. F. Luisi, and G. P. Salmond (2012). Identification and classification of bacterial Type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes. Nucleic Acids Res

Branda, S. S., F. Chu, D. B. Kearns, R. Losick, and R. Kolter (2006). A major protein component of the *Bacillus subtilis* biofilm matrix. Mol Microbiol 59: 1229-1238.

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

Brennan, R. G., and T. M. Link (2007). Hfq structure, function and ligand binding. Curr Opin Microbiol 10: 125-133.

Breukink, E., and B. de Kruijff (2006). Lipid II as a target for antibiotics. Nat Rev Drug Discov 5: 321-332.

Brogden, K. A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3: 238-250.

Brotz, H., and H. G. Sahl (2000). New insights into the mechanism of action of lantibiotics--diverse biological effects by binding to the same molecular target. J Antimicrob Chemother 46: 1-6.

Brown, J. M., and K. J. Shaw (2003). A novel family of *Escherichia coli* toxin-antitoxin gene pairs. J Bacteriol 185: 6600-6608.

Camp, A. H., A. F. Wang, and R. Losick (2011). A small protein required for the switch from {sigma}F to {sigma}G during sporulation in *Bacillus subtilis*. J Bacteriol 193: 116-124.

Capstick, D. S., J. M. Willey, M. J. Buttner, and M. A. Elliot (2007). SapB and the chaplins: connections between morphogenetic proteins in *Streptomyces coelicolor*. Mol Microbiol 64: 602-613.

Christensen, S. K., and K. Gerdes (2003). RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. Mol Microbiol 48: 1389-1400.

Christensen, S. K., M. Mikkelsen, K. Pedersen, and K. Gerdes (2001). RelE, a global inhibitor of translation, is activated during nutritional stress. Proc Natl Acad Sci U S A 98: 14328-14333.

Christensen, S. K., K. Pedersen, F. G. Hansen, and K. Gerdes (2003). Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. J Mol Biol 332: 809-819.

Clauditz, A., A. Resch, K. P. Wieland, A. Peschel, and F. Gotz (2006). Staphyloxanthin plays a role in the fitness of *Staphylococcus aureus* and its ability to cope with oxidative stress. Infect Immun 74: 4950-4953.

Cogen, A. L., K. Yamasaki, J. Muto, K. M. Sanchez, L. Crotty Alexander, J. Tanios, Y. Lai, J. E. Kim, V. Nizet, and R. L. Gallo (2010). *Staphylococcus epidermidis* antimicrobial delta-toxin (phenol-soluble modulin-gamma) cooperates with host antimicrobial peptides to kill group A Streptococcus. PLoS One 5: e8557.

Cybulski, L. E., G. del Solar, P. O. Craig, M. Espinosa, and D. de Mendoza (2004). *Bacillus subtilis* DesR functions as a phosphorylation-activated switch to control membrane lipid fluidity. J Biol Chem 279: 39340-39347.

D'Onofrio, A., J. M. Crawford, E. J. Stewart, K. Witt, E. Gavrish, S. Epstein, J. Clardy, and K. Lewis (2010). Siderophores from neighboring organisms promote the growth of uncultured bacteria. Chem Biol 17: 254-264.

Dahl, J. L., and D. Fordice (2011). Small acid-soluble proteins with intrinsic disorder are required for UV resistance in *Myxococcus xanthus* spores. J Bacteriol 193: 3042-3048.

Di Berardo, C., D. S. Capstick, M. J. Bibb, K. C. Findlay, M. J. Buttner, and M. A. Elliot (2008). Function and redundancy of the chaplin cell surface proteins in aerial hypha formation, rodlet assembly, and viability in *Streptomyces coelicolor*. J Bacteriol 190: 5879-5889.

Dornenburg, J. E., A. M. Devita, M. J. Palumbo, and J. T. Wade (2010). Widespread antisense transcription in *Escherichia coli*. MBio 1: e00024-10.

Dorr, T., M. Vulic, and K. Lewis (2010). Ciprofloxacin causes persister formation by inducing the TisB toxin in *Escherichia coli*. PLoS Biol 8: e1000317.

Engelberg-Kulka, H., I. Yelin, and I. Kolodkin-Gal (2009). Activation of a built-in bacterial programmed cell death system as a novel mechanism of action of some antibiotics. Commun Integr Biol 2: 211-212.

Erental, A., I. Sharon, and H. Engelberg-Kulka (2012). Two Programmed Cell Death Systems in *Escherichia coli*: An Apoptotic-Like Death Is Inhibited by the mazEF-Mediated Death Pathway. PLoS Biol 10: e1001281.

Fineran, P. C., T. R. Blower, I. J. Foulds, D. P. Humphreys, K. S. Lilley, and G. P. Salmond (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A 106: 894-899.

Fontaine, F., R. T. Fuchs, and G. Storz (2011). Membrane localization of small proteins in *Escherichia coli*. J Biol Chem 286: 32464-32474.

Fozo, E. M., M. R. Hemm, and G. Storz (2008). Small toxic proteins and the antisense RNAs that repress them. Microbiol Mol Biol Rev 72: 579-589.

Fozo, E. M., M. Kawano, F. Fontaine, Y. Kaya, K. S. Mendieta, K. L. Jones, A. Ocampo, K. E. Rudd, and G. Storz (2008). Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. Mol Microbiol 70: 1076-1093.

Fozo, E. M., K. S. Makarova, S. A. Shabalina, N. Yutin, E. V. Koonin, and G. Storz (2010). Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Res 38: 3743-3759.

Gaballa, A., H. Antelmann, C. Aguilar, S. K. Khakh, K. B. Song, G. T. Smaldone, and J. D. Helmann (2008). The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. Proc Natl Acad Sci U S A 105: 11927-11932.

Garwin, J. L., and J. E. Cronan, Jr. (1980). Thermal modulation of fatty acid synthesis in *Escherichia coli* does not involve de novo enzyme synthesis. J Bacteriol 141: 1457-1459.

Gassel, M., T. Mollenkamp, W. Puppe, and K. Altendorf (1999). The KdpF subunit is part of the K(+)-translocating Kdp complex of *Escherichia coli* and is responsible for stabilization of the complex in vitro. J Biol Chem 274: 37901-37907.

Georg, J., and W. R. Hess (2011). cis-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev 75: 286-300.

Gerdes, K., P. B. Rasmussen, and S. Molin (1986). Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. Proc Natl Acad Sci U S A 83: 3116-3120.

Goldstein, J., N. S. Pollitt, and M. Inouye (1990). Major cold shock protein of *Escherichia coli*. Proc Natl Acad Sci U S A 87: 283-287.

Gonzalez Barrios, A. F., R. Zuo, Y. Hashimoto, L. Yang, W. E. Bentley, and T. K. Wood (2006). Autoinducer 2 controls biofilm formation in *Escherichia coli* through a novel motility quorum-sensing regulator (MqsR, B3022). J Bacteriol 188: 305-316.

Gonzalez-Pastor, J. E., E. C. Hobbs, and R. Losick (2003). Cannibalism by sporulating bacteria. Science 301: 510-513.

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

Greenfield, T. J., E. Ehli, T. Kirshenmann, T. Franch, K. Gerdes, and K. E. Weaver (2000). The antisense RNA of the par locus of pAD1 regulates the expression of a 33-amino-acid toxic peptide by an unusual mechanism. Mol Microbiol 37: 652-660.

Groisman, E. A. (2001). The pleiotropic two-component regulatory system PhoP-PhoQ. J Bacteriol 183: 1835-1842.

Gurnev, P. A., R. Ortenberg, T. Dorr, K. Lewis, and S. M. Bezrukov (2012). Persister-promoting bacterial toxin TisB produces anion-selective pores in planar lipid bilayers. FEBS Lett

Han, K., K. S. Kim, G. Bak, H. Park, and Y. Lee (2010). Recognition and discrimination of target mRNAs by Sib RNAs, a cis-encoded sRNA family. Nucleic Acids Res 38: 5851-5866.

Handler, A. A., J. E. Lim, and R. Losick (2008). Peptide inhibitor of cytokinesis during sporulation in *Bacillus subtilis*. Mol Microbiol 68: 588-599.

Havarstein, L. S., G. Coomaraswamy, and D. A. Morrison (1995). An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci U S A 92: 11140-11144.

Hazan, R., and H. Engelberg-Kulka (2004). *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol Genet Genomics 272: 227-234.

Hazan, R., B. Sat, and H. Engelberg-Kulka (2004). *Escherichia coli* mazEF-mediated cell death is triggered by various stressful conditions. J Bacteriol 186: 3663-3669.

Hemm, M. R., B. J. Paul, J. Miranda-Rios, A. Zhang, N. Soltanzad, and G. Storz (2010). Small stress response proteins in *Escherichia coli*: proteins missed by classical proteomic studies. J Bacteriol 192: 46-58.

Hemm, M. R., B. J. Paul, T. D. Schneider, G. Storz, and K. E. Rudd (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. Mol Microbiol 70: 1487-1501.

Hershberg, R., S. Altuvia, and H. Margalit (2003). A survey of small RNA-encoding genes in *Escherichia coli*. Nucleic Acids Res 31: 1813-1820.

Hobbs, E. C., J. L. Astarita, and G. Storz (2010). Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a bar-coded mutant collection. J Bacteriol 192: 59-67.

Hobbs, E. C., F. Fontaine, X. Yin, and G. Storz (2011). An expanding universe of small proteins. Curr Opin Microbiol 14: 167-173.

Holden, M. T., S. Ram Chhabra, R. de Nys, P. Stead, N. J. Bainton, P. J. Hill, M. Manefield, N. Kumar, M. Labatte, D. England, S. Rice, M. Givskov, G. P. Salmond, G. S. Stewart, B. W. Bycroft, S. Kjelleberg, and P. Williams (1999). Quorum-sensing cross talk: isolation and chemical characterization of cyclic dipeptides from *Pseudomonas aeruginosa* and other gram-negative bacteria. Mol Microbiol 33: 1254-1266.

Hong, S. H., X. Wang, H. F. O'Connor, M. J. Benedik, and T. K. Wood (2012). Bacterial persistence increases as environmental fitness decreases. Microb Biotechnol

Johansen, J., A. A. Rasmussen, M. Overgaard, and P. Valentin-Hansen (2006). Conserved small non-coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer membrane proteins. J Mol Biol 364: 1-8.

Kastenmayer, J. P., L. Ni, A. Chu, L. E. Kitchen, W. C. Au, H. Yang, C. D. Carter, D. Wheeler, R. W. Davis, J. D. Boeke, M. A. Snyder, and M. A. Basrai (2006). Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae. Genome Res 16: 365-373.

Kawamoto, H., Y. Koide, T. Morita, and H. Aiba (2006). Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. Mol Microbiol 61: 1013-1022.

Kim, Y., X. Wang, Q. Ma, X. S. Zhang, and T. K. Wood (2009). Toxin-antitoxin systems in *Escherichia coli* influence biofilm formation through YjgK (TabA) and fimbriae. J Bacteriol 191: 1258-1267.

Kim, Y., X. Wang, X. S. Zhang, S. Grigoriu, R. Page, W. Peti, and T. K. Wood (2010). *Escherichia coli* toxin/antitoxin pair MqsR/MqsA regulate toxin CspD. Environ Microbiol 12: 1105-1121.

Kodani, S., M. E. Hudson, M. C. Durrant, M. J. Buttner, J. R. Nodwell, and J. M. Willey (2004). The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in *Streptomyces coelicolor*. Proc Natl Acad Sci U S A 101: 11448-11453.

Koga, M., Y. Otsuka, S. Lemire, and T. Yonesaki (2011). *Escherichia coli* rnlA and rnlB compose a novel toxin-antitoxin system. Genetics 187: 123-130.

Kolodkin-Gal, I., B. Sat, A. Keshet, and H. Engelberg-Kulka (2008). The communication factor EDF and the toxin-antitoxin module mazEF determine the mode of action of antibiotics. PLoS Biol 6: e319.

Kolodkin-Gal, I., R. Verdiger, A. Shlosberg-Fedida, and H. Engelberg-Kulka (2009). A differential effect of *E. coli* toxin-antitoxin systems on cell death in liquid media and biofilm formation. PLoS One 4: e6785.

Koprivnjak, T., and A. Peschel (2011). Bacterial resistance mechanisms against host defense peptides. Cell Mol Life Sci 68: 2243-2254.

Kuipers, O. P., M. M. Beerthuyzen, P. G. de Ruyter, E. J. Luesink, and W. M. de Vos (1995). Autoregulation of nisin biosynthesis in *Lactococcus lactis* by signal transduction. J Biol Chem 270: 27299-27304.

Lazazzera, B. A., J. M. Solomon, and A. D. Grossman (1997). An exported peptide functions intracellularly to contribute to cell density signaling in *B. subtilis*. Cell 89: 917-925.

Leplae, R., D. Geeraerts, R. Hallez, J. Guglielmini, P. Dreze, and L. Van Melderen (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. Nucleic Acids Res 39: 5513-5525.

Lewis, K. (2010). Persister cells. Annu Rev Microbiol 64: 357-372.

Li, J., W. Wang, S. X. Xu, N. A. Magarvey, and J. K. McCormick (2011). *Lactobacillus reuteri*-produced cyclic dipeptides quench agr-mediated expression of toxic shock syndrome toxin-1 in staphylococci. Proc Natl Acad Sci U S A 108: 3360-3365.

Li, M., B. A. Diep, A. E. Villaruz, K. R. Braughton, X. Jiang, F. R. DeLeo, H. F. Chambers, Y. Lu, and M. Otto (2009). Evolution of virulence in epidemic community-associated methicillin-resistant *Staphylococcus aureus*. Proc Natl Acad Sci U S A 106: 5883-5888.

Link, A. J., K. Robison, and G. M. Church (1997). Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. Electrophoresis 18: 1259-1313.

Liu, M. Y., G. Gui, B. Wei, J. F. Preston, 3rd, L. Oakford, U. Yuksel, D. P. Giedroc, and T. Romeo (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. J Biol Chem 272: 17502-17510.

Liu, W. T., Y. L. Yang, Y. Xu, A. Lamsa, N. M. Haste, J. Y. Yang, J. Ng, D. Gonzalez, C. D. Ellermeier, P. D. Straight, P. A. Pevzner, J. Pogliano, V. Nizet, K. Pogliano, and P. C. Dorrestein (2010). Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. Proc Natl Acad Sci U S A 107: 16286-16290.

Loh, S. M., D. S. Cram, and R. A. Skurray (1988). Nucleotide sequence and transcriptional analysis of a third function (Flm) involved in F-plasmid maintenance. Gene 66: 259-268.

Lopez, D., M. A. Fischbach, F. Chu, R. Losick, and R. Kolter (2009). Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*. Proc Natl Acad Sci U S A 106: 280-285.

Lopez, D., and R. Kolter (2010). Functional microdomains in bacterial membranes. Genes Dev 24: 1893-1902.

Lopez, D., H. Vlamakis, R. Losick, and R. Kolter (2009). Cannibalism enhances biofilm development in *Bacillus subtilis*. Mol Microbiol 74: 609-618.

Lutz, R., and H. Bujard (1997). 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.

Maisonneuve, E., L. J. Shakespeare, M. G. Jorgensen, and K. Gerdes (2011). Bacterial persistence by RNA endonucleases. Proc Natl Acad Sci U S A 108: 13206-13211.

Mason, J. M., and P. Setlow (1986). Essential role of small, acid-soluble spore proteins in resistance of *Bacillus subtilis* spores to UV light. J Bacteriol 167: 174-178.

Masse, E., H. Salvail, G. Desnoyers, and M. Arguin (2007). Small RNAs controlling iron metabolism. Curr Opin Microbiol 10: 140-145.

Mehta, P., S. Goyal, and N. S. Wingreen (2008). A quantitative comparison of sRNA-based and protein-based gene regulation. Mol Syst Biol 4: 221.

Nariya, H., and M. Inouye (2008). MazF, an mRNA interferase, mediates programmed cell death during multicellular Myxococcus development. Cell 132: 55-66.

Navarre, C., P. Catty, S. Leterme, F. Dietrich, and A. Goffeau (1994). Two distinct genes encode small isoproteolipids affecting plasma membrane H(+)-ATPase activity of *Saccharomyces cerevisiae*. J Biol Chem 269: 21262-21268.

Ng, W. L., and B. L. Bassler (2009). Bacterial quorum-sensing network architectures. Annu Rev Genet 43: 197-222.

Novick, R. P., and E. Geisinger (2008). Quorum sensing in staphylococci. Annu Rev Genet 42: 541-564.

O'Connor, E. M., and R. F. Shand (2002). Halocins and sulfolobicins: the emerging story of archaeal protein and peptide antibiotics. J Ind Microbiol Biotechnol 28: 23-31.

Ogura, T., and S. Hiraga (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. Proc Natl Acad Sci U S A 80: 4784-4788.

Okada, M., I. Sato, S. J. Cho, H. Iwata, T. Nishio, D. Dubnau, and Y. Sakagami (2005). Structure of the *Bacillus subtilis* quorum-sensing peptide pheromone ComX. Nat Chem Biol 1: 23-24.

Otsuka, Y., and T. Yonesaki (2012). Dmd of bacteriophage T4 functions as an antitoxin against *Escherichia coli* LsoA and RnlA toxins. Mol Microbiol 83: 669-681.

Pandey, D. P., and K. Gerdes (2005). Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Res 33: 966-976.

Panina, E. M., A. A. Mironov, and M. S. Gelfand (2003). Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. Proc Natl Acad Sci U S A 100: 9912-9917.

Park, D. K., K. E. Lee, C. H. Baek, I. H. Kim, J. H. Kwon, W. K. Lee, K. H. Lee, B. S. Kim, S. H. Choi, and K. S. Kim (2006). Cyclo(Phe-Pro) modulates the expression of ompU in Vibrio spp. J Bacteriol 188: 2214-2221.

Pecota, D. C., and T. K. Wood (1996). Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. J Bacteriol 178: 2044-2050.

Pedersen, K., S. K. Christensen, and K. Gerdes (2002). Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. Mol Microbiol 45: 501-510.

Peng, J., J. Yang, and Q. Jin (2011). An integrated approach for finding overlooked genes in Shigella. PLoS One 6: e18509.

Peschel, A., and H. G. Sahl (2006). The co-evolution of host cationic antimicrobial peptides and microbial resistance. Nat Rev Microbiol 4: 529-536.

Ramage, H. R., L. E. Connolly, and J. S. Cox (2009). Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. PLoS Genet 5: e1000767.

Ramamurthi, K. S., S. Lecuyer, H. A. Stone, and R. Losick (2009). Geometric cue for protein localization in a bacterium. Science 323: 1354-1357.

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

Ren, D., L. A. Bedzyk, S. M. Thomas, R. W. Ye, and T. K. Wood (2004). Gene expression in *Escherichia coli* biofilms. Appl Microbiol Biotechnol 64: 515-524.

Romero, D., C. Aguilar, R. Losick, and R. Kolter (2010). Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. Proc Natl Acad Sci U S A 107: 2230-2234.

Romero, D., M. F. Traxler, D. Lopez, and R. Kolter (2011). Antibiotics as signal molecules. Chem Rev 111: 5492-5505.

Ruvolo, M. V., K. E. Mach, and W. F. Burkholder (2006). Proteolysis of the replication checkpoint protein Sda is necessary for the efficient initiation of sporulation after transient replication stress in *Bacillus subtilis*. Mol Microbiol 60: 1490-1508.

Saavedra De Bast, M., N. Mine, and L. Van Melderen (2008). Chromosomal toxin-antitoxin systems may act as antiaddiction modules. J Bacteriol 190: 4603-4609.

Santos Sierra, S., R. Giraldo, and R. Diaz Orejas (1998). Functional interactions between chpB and parD, two homologous conditional killer systems found in the *Escherichia coli* chromosome and in plasmid R1. FEMS Microbiol Lett 168: 51-58.

Schumacher, M. A., K. M. Piro, W. Xu, S. Hansen, K. Lewis, and R. G. Brennan (2009). Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. Science 323: 396-401.

Seyedsayamdost, M. R., M. F. Traxler, S. L. Zheng, R. Kolter, and J. Clardy (2011). Structure and biosynthesis of amychelin, an unusual mixed-ligand siderophore from *Amycolatopsis* sp. AA4. J Am Chem Soc 133: 11434-11437.

Shank, E. A., V. Klepac-Ceraj, L. Collado-Torres, G. E. Powers, R. Losick, and R. Kolter (2011). Interspecies interactions that result in *Bacillus subtilis* forming biofilms are mediated mainly by members of its own genus. Proc Natl Acad Sci U S A 108: E1236-1243.

Shevchenko, A., O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, and M. Mann (1996). Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc Natl Acad Sci U S A 93: 14440-14445.

Shimoni, Y., G. Friedlander, G. Hetzroni, G. Niv, S. Altuvia, O. Biham, and H. Margalit (2007). Regulation of gene expression by small non-coding RNAs: a quantitative view. Mol Syst Biol 3: 138.

Siezen, R. J., O. P. Kuipers, and W. M. de Vos (1996). Comparison of lantibiotic gene clusters and encoded proteins. Antonie Van Leeuwenhoek 69: 171-184.

Simmerman, H. K., J. H. Collins, J. L. Theibert, A. D. Wegener, and L. R. Jones (1986). Sequence analysis of phospholamban. Identification of phosphorylation sites and two major structural domains. J Biol Chem 261: 13333-13341.

Singh, R., C. E. Barry, 3rd, and H. I. Boshoff (2010). The three RelE homologs of *Mycobacterium tuberculosis* have individual, drug-specific effects on bacterial antibiotic tolerance. J Bacteriol 192: 1279-1291.

Staubitz, P., H. Neumann, T. Schneider, I. Wiedemann, and A. Peschel (2004). MprF-mediated biosynthesis of lysylphosphatidylglycerol, an important determinant in staphylococcal defensin resistance. FEMS Microbiol Lett 231: 67-71.

Suzuki, M., J. Zhang, M. Liu, N. A. Woychik, and M. Inouye (2005). Single protein production in living cells facilitated by an mRNA interferase. Mol Cell 18: 253-261.

Tamm, J., and B. Polisky (1985). Characterization of the ColE1 primer-RNA1 complex: analysis of a domain of ColE1 RNA1 necessary for its interaction with primer RNA. Proc Natl Acad Sci U S A 82: 2257-2261.

Tan, Q., N. Awano, and M. Inouye (2011). YeeV is an *Escherichia coli* toxin that inhibits cell division by targeting the cytoskeleton proteins, FtsZ and MreB. Mol Microbiol 79: 109-118.

Tang, T. H., N. Polacek, M. Zywicki, H. Huber, K. Brugger, R. Garrett, J. P. Bachellerie, and A. Huttenhofer (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Mol Microbiol 55: 469-481.

Thomassen, G. O., R. Weel-Sneve, A. D. Rowe, J. A. Booth, J. M. Lindvall, K. Lagesen, K. I. Kristiansen, M. Bjoras, and T. Rognes (2010). Tiling array analysis of UV treated *Escherichia coli* predicts novel differentially expressed small peptides. PLoS One 5: e15356.

Tillotson, R. D., H. A. Wosten, M. Richter, and J. M. Willey (1998). A surface active protein involved in aerial hyphae formation in the filamentous fungus Schizophillum commune restores the capacity of a bald mutant of the filamentous bacterium *Streptomyces coelicolor* to erect aerial structures. Mol Microbiol 30: 595-602.

Toledo-Arana, A., F. Repoila, and P. Cossart (2007). Small noncoding RNAs controlling pathogenesis. Curr Opin Microbiol 10: 182-188.

Tsilibaris, V., G. Maenhaut-Michel, N. Mine, and L. Van Melderen (2007). What is the benefit to *Escherichia coli* of having multiple toxin-antitoxin systems in its genome? J Bacteriol 189: 6101-6108.

Unoson, C., and E. G. Wagner (2008). A small SOS-induced toxin is targeted against the inner membrane in *Escherichia coli*. Mol Microbiol 70: 258-270.

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

Van Melderen, L., and M. Saavedra De Bast (2009). Bacterial toxin-antitoxin systems: more than selfish entities? PLoS Genet 5: e1000437.

Vazquez-Laslop, N., H. Lee, and A. A. Neyfakh (2006). Increased persistence in *Escherichia coli* caused by controlled expression of toxins or other unrelated proteins. J Bacteriol 188: 3494-3497.

Vlamakis, H., Kolter, R. "Biofilms." *Bacterial Stress Responses*. Ed. Storz, G.. Hengge, R. 2 ed. Washington: ASM Press, 2011. 365-373. Print.

Vogel, J., V. Bartels, T. H. Tang, G. Churakov, J. G. Slagter-Jager, A. Huttenhofer, and E. G. Wagner (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res 31: 6435-6443.

Vogel, J., and C. M. Sharma (2005). How to find small non-coding RNAs in bacteria. Biol Chem 386: 1219-1238.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J Mol Biol 225: 487-494.

Wadler, C. S., and C. K. Vanderpool (2007). A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc Natl Acad Sci U S A 104: 20454-20459.

Wang, X., Y. Kim, S. H. Hong, Q. Ma, B. L. Brown, M. Pu, A. M. Tarone, M. J. Benedik, W. Peti, R. Page, and T. K. Wood (2011). Antitoxin MqsA helps mediate the bacterial general stress response. Nat Chem Biol 7: 359-366.

Wang, X., Y. Kim, Q. Ma, S. H. Hong, K. Pokusaeva, J. M. Sturino, and T. K. Wood (2010). Cryptic prophages help bacteria cope with adverse environments. Nat Commun 1: 147.

Wang, X. and T. K. Wood (2011). Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. Appl Environ Microbiol 77: 5577-5583.

Wassarman, K. M. (2007). 6S RNA: a small RNA regulator of transcription. Curr Opin Microbiol 10: 164-168.

Wassarman, K. M., F. Repoila, C. Rosenow, G. Storz, and S. Gottesman (2001). Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev 15: 1637-1651.

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

Waters, L. S., M. Sandoval, and G. Storz (2011). The *Escherichia coli* MntR miniregulon includes genes encoding a small protein and an efflux pump required for manganese homeostasis. J Bacteriol 193: 5887-5897.

Waters, L. S., and G. Storz (2009). Regulatory RNAs in bacteria. Cell 136: 615-628.

Wozniak, R. A., and M. K. Waldor (2009). A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. PLoS Genet 5: e1000439.

Yamaguchi, Y., and M. Inouye (2011). Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. Nat Rev Microbiol 9: 779-790.

Yamaguchi, Y., J. H. Park, and M. Inouye (2009). MqsR, a crucial regulator for quorum sensing and biofilm formation, is a GCU-specific mRNA interferase in *Escherichia coli*. J Biol Chem 284: 28746-28753.

Yamanaka, K. (1999). Cold shock response in *Escherichia coli*. J Mol Microbiol Biotechnol 1: 193-202.

Yang, Y. L., Y. Xu, P. Straight, and P. C. Dorrestein (2009). Translating metabolic exchange with imaging mass spectrometry. Nat Chem Biol 5: 885-887.

Zhang, A., K. M. Wassarman, C. Rosenow, B. C. Tjaden, G. Storz, and S. Gottesman (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol Microbiol 50: 1111-1124.

Chapter 2:
Identification of a toxic peptide through bidirectional expression of small RNAs

2.1. Author's preface

This chapter describes our efforts in developing an antisense knockdown strategy to identify sRNAs found in *Escherichia coli* that may have essential functions in the cell. When our group first started this study in 2006, many laboratories around the globe had set out to uncover novel sRNAs across diverse microbial genomes. Upon finding these novel RNA molecules, much work is then required to elucidate their functions. In this project, we hoped to express the reverse complements, or antisense sequences, of a subset of sRNAs of unknown functions that were annotated in *E. coli* so that the antisense sequences could sequester the sRNAs and prevent them from interacting with their targets. Should the sRNAs have functions that were vital to cell survival, the antagonizing activity of their antisense counterparts was expected to result in growth suppression or cell death. Previously, this technique of screening for lethal phenotypes following the induction of antisense RNA sequences was used for the identification of essential genes in *Staphylococcus aureus*. To our knowledge, this study represents the first example in which antisense RNA technology is applied in the study of sRNAs.

Through our screen, we found that one antisense sequence, corresponding to the reverse complement of sRNA RygC, inhibited cell growth when its expression was induced. However, this sequence did not act by preventing the interactions between RygC and its cellular target. Rather, we found that it contained an ORF that encoded a small toxic peptide. RygC and its antisense counterpart were confirmed to form a type I TA pair by the Storz laboratory. The antitoxin sRNA was renamed SibC, while the toxic peptide was coined IbsC. Thus, the antisense sRNA expression technique implemented in this study may be utilized to discover additional type I TA pairs in bacterial genomes. Our

findings from this project steered us toward the study of TA systems, which have been implicated to have important regulatory functions, much like sRNAs.

This chapter is modified from an article that was previously published in Chembiochem. The project was initially conceptualized by Drs.Li, Brown, Navani, and Pathania. The plasmid used throughout the project was modified by Pascale Libront-Swanson. Initial work involving the cloning of sRNAs, their reverse complements, and the truncation variants of antisense *rygC* into the expression vector was carried out by Navani, Barker, Gu, and Zhu before I came on board to the project. When I took over the project, I conducted the lethality screens and growth assays to determine the effects of expressing the 12 sRNAs and their antisense sequences on the growth of *E. coli* with assistance from Sawchyn. We also carried out site-directed mutagenesis of antisense *rygC* to confirm that it coded for a toxic peptide. Sawchyn and myself performed all of the experiments used for all figures in the manuscript. I produced all of the figures, and I took on the leading role in writing this manuscript with helpful advice from Dr. Li. The citation for the original manuscript follows:

Mok, W. W., Navani, N. K., Barker, C., Sawchyn, B. L., Gu, J., Pathania, R., Zhu, R. D., Brown, E. D., and Li, Y. (2009) Identification of a toxic peptide through bidirectional expression of small RNAs. *Chembiochem 10*, 238-41.

2.2. Abstract

Noncoding RNA sequences residing in the intergenic regions of bacterial genomes were once neglected but are now gaining recognition as important regulatory elements of gene expression and stress adaptation. To date, over 80 sRNAs have been confirmed to exist in *E. coli* alone. Despite this progress, many of these sRNA sequences remain

uncharacterized. Here, we utilized an antisense RNA knockdown technology to examine 12 sRNAs, nine of which with unknown functions, in an effort to identify sequences whose functions are essential for cell viability. From this screen, we found that the overexpression of the reverse complements of the RygC sRNA resulted in growth suppression. This sequence was further characterized in order to determine its mechanism of action.

2.3. Introduction

Research in the field of noncoding functional RNA sequences has flourished over the past two decades, showcasing the utility of these nucleic acids that extends beyond their traditional roles as the workhorses behind protein translation. A search through the intergenic regions of prokaryotic genomes has unveiled a class of important regulatory elements, known as small RNAs (sRNAs). As their names imply, these RNA sequences are relatively short, typically ranging from 50 to 400 nucleotides (nt) in length. In the *E. coli* genome alone, over 80 sRNA sequences have been discovered using a combination of bioinformatics and experimental techniques, though the function of many of these sequences remain unknown [1]. Of the sRNAs that have been characterized to date, the majority of them interact with mRNA targets, thereby modulating mRNA translation initiation or stability upon binding [2]. Another subset of sRNAs can regulate the activity of protein targets or function as a part of a protein-sRNA complex [2]. Together, these tiny elements govern a number of bacterial stress response pathways [3] and they have been linked to pathogenicity in virulent species [3-7]. Due to the substantial role of sRNAs in maintaining cellular homeostasis and viability, disruption of their expression is expected to have deleterious effects on the cell.

Previously, screening for lethal or growth defective phenotypes upon the overexpression of antisense RNA sequences was used to identify essential genes in *Staphylococcus aureus* [8, 9]. Here, we implemented a similar approach in an effort to develop a method to isolate sRNA sequences with critical regulatory functions in *E. coli*.

2.4. Materials and methods

2.4.1. Oligonucleotides and reagents

All PCR primers and oligonucleotides used in this study were chemically synthesized by Integrated DNA Technologies (Coralville, IA, USA). The antibiotics used in this study were purchased from Sigma Aldrich (Oakville, ON, Canada). Kanamycin was used at a concentration of 50 μg mL^{-1}, while anhydrotetracycline (Atc) was supplemented at a concentration of 400 ng mL^{-1} on solid media or 200 ng mL^{-1} in liquid media. For molecular cloning, the High Fidelity PCR enzyme mix and T4 DNA ligase were purchased from Fermentas (Burlington, ON, Canada). Restriction enzymes were obtained through either Fermentas or New England Biolabs (Pickering, ON, Canada). For site-directed mutagenesis of the antisense *rygC*-encoding vector, PfuUltra High Fidelity DNA Polymerase and Dpn1 restriction enzyme from Stratagene (La Jolla, CA, USA) were used to amplify the mutated plasmids and cleave the template vector, respectively. Plasmid Mini-prep kits were purchased from either Qiagen (Mississauga, ON, Canada) or Promega (Madison, WI, USA).

2.4.2. Plasmids and bacterial strains

Molecular cloning steps, including PCR, restriction digestion, and ligation were conducted following established protocols provided by suppliers. Sense and antisense sRNA-encoding sequences used for the lethality screen were amplified from *E. coli*

DH5αZ1 genomic DNA by PCR. Likewise, truncated sense and antisense *rygC* sequences 1-4 were amplified by PCR using full-length antisense *rygC* as template. Truncated sequences 5 and 6 were obtained by hybridizing complementary synthetic oligonucleotides. These PCR amplified and synthetically prepared sRNA sequences were cloned into EcoRI and BamHI sites in pNYL-MCS11, a derivative of pZE21-MCS1 (courtesy of H. Bujard). Prior to cloning, pZE21-MCS1 was digested with restriction enzyme EcoRI in order to remove its optimal ribosome binding site. A new multiple cloning site was subsequently introduced at the EcoRI site, allowing for the replacement of the restriction sites that were lost upon RBS removal. Plasmids carrying sense and antisense sRNA sequences or truncated antisense *rygC* variants were transformed into *E. coli* strain DH5αZ1, which endogenously expresses the tetracycline repressor, by electroporation.

2.4.3. Lethality Screens

Transformants with plasmids containing sense and antisense sRNA-encoding inserts or truncated sense and antisense *rygC* variants were grown at 37 ^{o}C on solid Luria-Burtani (LB) medium supplemented with kanamycin in the absence (for uninduced conditions) or presence (for induced conditions) of Atc. The effects of overexpressing the various sRNA or *rygC*-encoding constructs were evaluated following overnight growth.

2.4.4. E. coli growth curve

Cells carrying pNYL-MCS11 or pNYL-MCS11 with an *rrygC* insert were grown overnight 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 5 mL of fresh LB-kanamycin media (for uninduced conditions) or in LB-kanamycin media supplemented

with Atc (for induced conditions). These cultures were incubated at 37 ^{o}C with shaking for 8 hours. Cell growth was monitored hourly by measuring the optical density at 600 nm (OD_{600}) for 0.2 mL of each culture using a SpectraMax Plus 384 plate reader (Molecular Devices, Sunnyvale, CA, USA). The assay was conducted in triplicate.

2.4.5. Mutagenesis of antisense rygC

The ribosome binding site, start codon(s), and stop codon of *rrygC* were mutated by site-directed mutagenesis following supplier-provided protocols (Stratagene). Briefly, forward and reverse mutagenic primers of 34-41 nt in length were designed, such that they would hybridize to the region that was to be changed and introduced point mutations at these sites. The *rrygC*-encoding vector was then amplified using these primers along with the PfuUltra High Fidelity DNA Polymerase (Stratagene). Following amplification, the methylated template plasmid was cleaved following two rounds of digestion using Dpn1 restriction enzymes (Stratagene). The reaction mixtures were subsequently transformed into *E. coli* strain XL-1 blue competent cells by electroporation. Plasmids isolated from the transformants were sequenced (Mobix Lab, McMaster Univeristy) to identify those carrying the desired mutations. The mutated plasmids were then transformed into *E. coli* DH5αZ1 cells. The lethality attributed to the overexpression of these sequences was screened on solid LB-kanamycin media in the absence or presence of Atc following the aforementioned protocol.

2.5. Results

2.5.1. Expression of sense and antisense sRNA sequences

We adopted a tetracycline inducible system to regulate the expression of the sense and antisense sequences of 12 sRNAs. These sequences were cloned into pNYL-MCS11,

a plasmid derived from pZE21-MCS1, which is described in an earlier study [10]. The optimized ribosome binding site from the plasmid was removed prior to cloning, such that the sense and antisense sRNAs would be transcribed but not translated when their expression was induced in the presence of anhydrotetracycline (Atc), an analog of tetracycline that appears to be a more potent yet less toxic inducer [11]. Analyses were subsequently performed in *E. coli* strain DH5αZ1, which has been engineered to endogenously express the TetR repressor (Figure 2.1) [10].

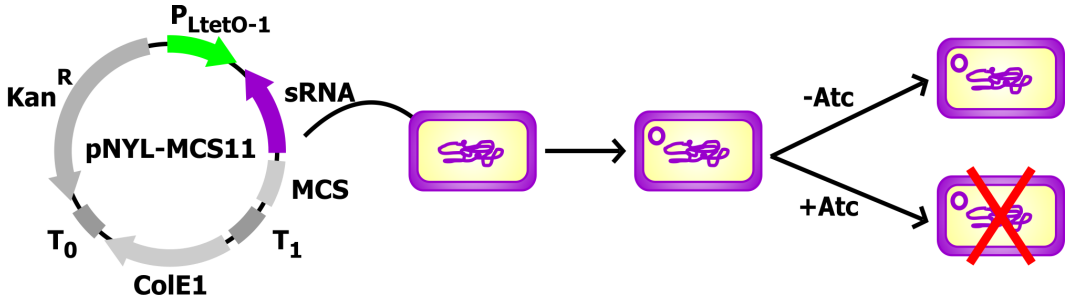


Figure 2.1. Overview of the screening approach adapted for the identification of sRNA sequences with critical functions. Sense and antisense sequences of 12 sRNAs were cloned into pNYL-MCS11 downstream of a tetracycline-inducible promoter (P_{LtetO1}). These constructs were then transformed into *E. coli* by electroporation. Lethal phenotypes linked to the overexpression of the sequences were screened by inducing their synthesis with Atc.

Of the sRNAs that were screened, three have been extensively studied, while the functions of the remaining nine have not yet been characterized (Table 2.1). Overexpression of all 12 sense sRNA sequences did not affect cell growth (Figure 2.2, top panel). On the other hand, inducing the expression of their antisense counterparts led to the discovery of one sequence that could suppress cell growth when it was overexpressed (Figure 2.2, bottom panel). This sequence corresponded to the antisense of the RygC sRNA (coined rRygC).

Table 2.1. List of sRNA sequences screened

	sRNA	Size (nt)	Flanking genes	Function	Refs.
1	SraD/ MicA	75	*ygaG/gshA*	Regulates expression of the OmpA outer membrane protein	[12, 15]
2	SraI/ RyhB	90	*yhhX/yhhY*	Regulates expression of iron-storage and iron-containing proteins	[12, 16]
3	SsrA/ tmRNA/ PsrD	363	*smpB/intA*	Rescue of stalled ribosomes in complex with SmpB	[12, 17]
4	RygC/ T27/ QUAD1c	140	*ygfA/serA*	Unknown	[12, 14]
5	SraB/Pke20	169	*yceF/yceD*	Unknown	[12]
6	SraC/RyeA/Tpke79 /ISO091	249	*pphA/yebY*	Unknown	[12, 14]
7	Tpke11	89	*dnaK/dnaJ*	Unknown	[12, 18]
8	C0067	125	*yafT/yafU*	Unknown	[12, 19]
9	C0293	73	*Icd/ymfD*	Unknown	[12, 19]
10	C0299	79	*hlyE/umuD*	Unknown	[12, 19]
11	C0465	78	*tar/chew*	Unknown	[12, 19]
12	C0719	222	*yghK/glcB*	Unknown	[12, 19]

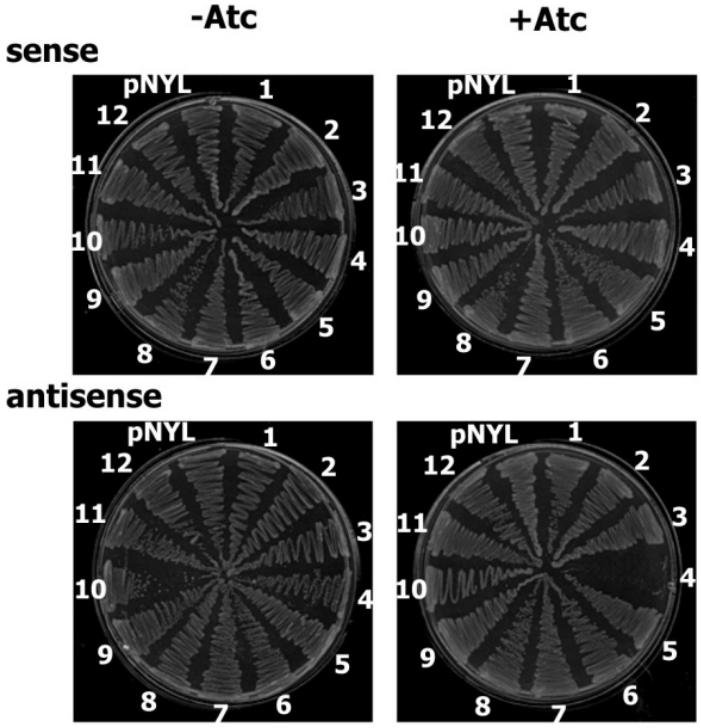


Figure 2.2. Effect of overexpressing sense and antisense variants of 12 sRNA sequences on the growth of *E. coli* strain DH5αZ1. Cells transformed with vectors encoding sense (top panel) and antisense (bottom panel) variants of 12 sRNAs were grown on solid media in the absence or presence of Atc. Through this screen, it was observed that the overexpression of the antisense variant of sequence 4, which corresponds to the RygC sRNA, led to growth inhibition.

2.5.2. Effects of rrygC overexpression on bacterial growth

In our initial screen, as shown in Figure 2.2, the consequences of antisense sRNA overexpression were qualitatively evaluated on solid media approximately 16 hours following induction by Atc. Monitoring the growth of *E. coli* in liquid media demonstrated that the growth inhibitory effect elicited by *rrygC* overexpression was not detected until 3 hours after induction (Figure 2.3). The growth of cells carrying pNYL-MCS11 without the *rrygC* insert was unaffected in the presence of 200 ng/mL of Atc. The delay in the onset of antisense- induced lethality may be a reflection of the time needed to

express a sufficient quantity of *rrygC*, allowing it to disrupt the regulatory function of sense RygC or other physiological activities.

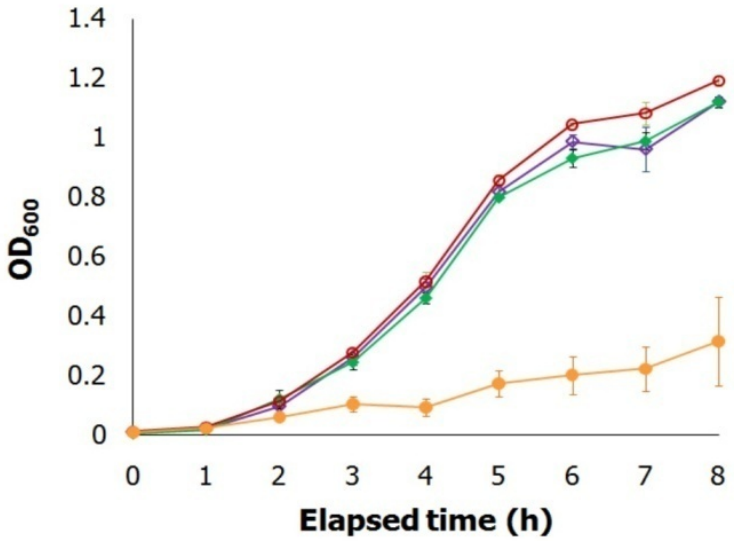


Figure 2.3. Growth of cells overexpressing antisense RygC. Cells carrying pNYL-MCS11 (denoted pNYL) and pNYL-MCS11 with the *rrygC* insert (*rrygC*) were grown overnight in Luria-Bertani (LB) media with 50 µg mL^{-1} of Kanamycin before being subcultured into fresh LB-Kanamycin media in the absence of Atc (-Atc) or in LB-Kanamycin supplemented with 200 ng mL^{-1} of Atc (+Atc). Growth of cells in each condition was monitored hourly for 8 hours by measuring their absorbance at 600 nm. This experiment was carried out in triplicates. (pNYL -Atc, pNYL +Atc, *rrygC* -Atc, *rrygC* +Atc)

2.5.3. Truncation of rrygC

As a first step towards understanding the mechanism behind rRygC induced growth suppression, we sought to determine the regions within its primary sequence that may be necessary for it to exert its inhibitory effect. This was achieved by truncating the DNA sequence encoding *rrygC* from the 5′ and 3′ ends, as well as introducing internal deletions (Figure 2.4A). These truncated sequences were subsequently cloned into pNYL-MCS11. Inducing the expression of the sense counterparts of these truncated sequences did not affect growth (Figure 2.4B, top panel). Removal of the first 22 nucleotides from the 5′

end of *rrygC*, which corresponded to the sequence complementary to *rygC*'s putative transcription terminator, did not diminish its inhibitory effect, as indicated by truncated sequences 1 and 2 (Figure 2.4B, bottom panel). Sequence 2 further demonstrated that the last 30 bases from the 3′ end in *rrygC* were dispensable. Overexpression of sequences harbouring deletions of the first 34 bases from the 5′ end, however, resulted in the loss of lethal phenotype, as illustrated by truncated sequences 3 to 5. Since we were interested in probing for the nucleotides that may be required for the rRygC transcript to elicit its growth-inhibitory function, we deleted 44 bases within its sequence along with the first 22 bases that were complementary to the proposed RygC terminator (sequence 6). Overexpression of this sequence did not interfere with the growth of *E. coli*, indicating that some of the deleted residues were vital for rRygC-associated lethality.

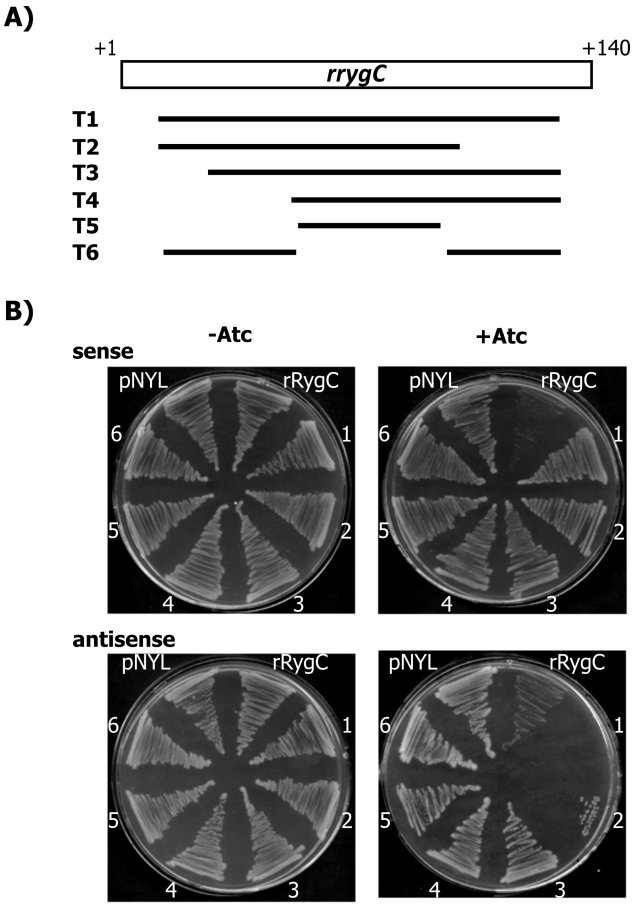


Figure 2.4. Truncation analysis of *rrygC*. A. Outline of truncated sequences 1-6. B. Effect of expressing the truncated sequences on the growth of *E. coli* DH5αZ1. It was observed that deletions from the 5' end of antisense *rygC* abolished the deleterious effects associated with the overexpression of this sequence.

2.5.4. rrygC encodes a toxic peptide

Further examination of the *rrygC* sequence revealed the presence of a ribosome binding site (RBS) 27 bases downstream of its transcription start site. This RBS was followed by two possible start codons. A stop codon was present 54 nucleotides following the first start codon. Together, these elements suggest that the antisense *rygC* transcript encodes a peptide of 18 or 19 amino acids. In agreement with this model, the overexpression of truncated sequences 3, 4, and 5 was no longer detrimental to cell growth as deletions introduced into these sequences would have contributed to the

removal of the RBS and/or start codon. As a result, the peptide could no longer be produced from these truncated transcripts. The internal deletions introduced into sequence 6 might have led to the removal of 33 nucleotides encoding 11 of the amino acids in the peptide in addition to the stop codon. Likewise, the deleterious effect to the cell when this sequence was induced was lost. To validate that *rrygC* encodes a peptide that compromises survival when it is abundant in the cell, we introduced mutations in the putative ribosome binding site and start codons via site-directed mutagenesis (Figure 2.5A). *rrygC* affiliated lethality was retained when the second ATG site was mutated (M2; Figure 2.5B). This indicates that this codon does not correspond to the start codon. However, when the first ATG codon (M1) and ribosome binding sites were mutated, the deleterious effect was abolished.

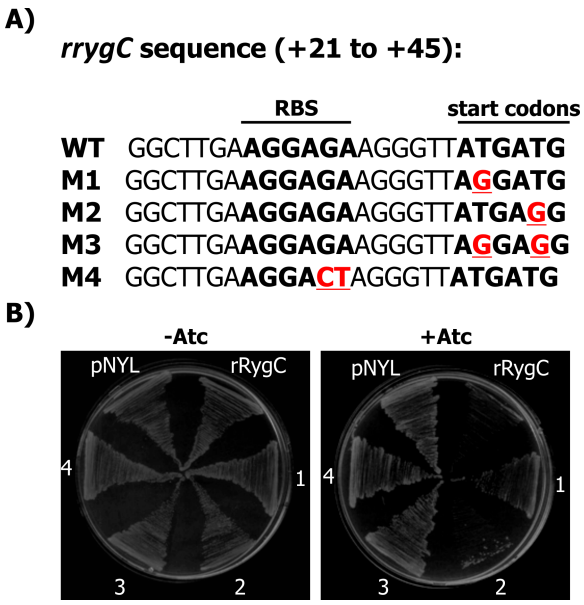


Figure 2.5. Mutagenesis of the putative RBS and start codon of antisense *rygC*. A. The predicted RBS and putative start codons encoded by antisense *rygC* are illustrated above. Mutant sequences M1 to M4 contain point mutations in these sites. B. Inducing the expression of sequences M1, M3 and M4 using Atc was no longer toxic to *E. coli.* Sequence M2 retained its deleterious effect, suggesting that the second ATG codon does not correspond to the start codon of *rrygC*.

2.6. Discussion

While we were preparing this manuscript, a study was published by Fozo and colleagues [20] revealing that *rrygC* and the reverse complements of other members of the QUAD 1 family of sRNAs harbours an ORF encoding a small hydrophobic peptide of 18 to 19 amino acids. Through this study in which the protein encoded by *rrygC* was extensively characterized, it was confirmed that this protein, referred to as IbsC, was a toxin whereas its corresponding sRNA was a transient antitoxin. Overexpression of IbsC resulted in depolarization of membrane potential, suggesting that this toxin may act by disrupting membrane integrity. The results of this study corroborate our observations that the lethality associated with *rrygC* overexpression is attributed to the increased production of a toxic peptide.

Although we now understand the mechanism behind *rrygC* induced growth suppression, the reason behind its synthesis *in vivo* and the mechanism behind its regulation remain unresolved. Since the RBS and start codon of this ORF overlaps with the transcribed region of sense *rygC*, we propose that the sRNA acts to suppress the translation of this peptide by occluding these translation initiation sites, thereby promoting cell survival (Figure 2.6A). Alternatively, similar to another subset of sRNAs that act on mRNA targets, RygC may destabilize the toxin-encoding transcript upon binding (Figure 2.6B).

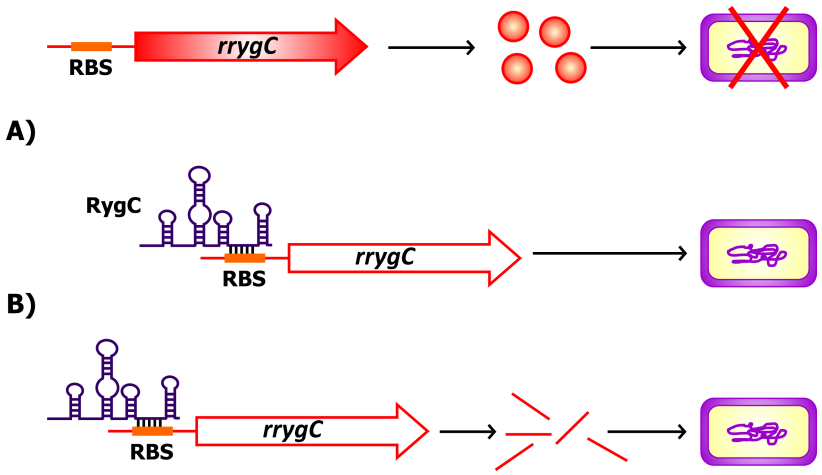


Figure 2.6. Hypotheses of *rrygC* regulation. A. *rrygC* encodes a toxic peptide, and the translation of this peptide is regulated by RygC. B. Interaction between the RygC sRNA and its antisense transcript leads to destabilization of the transcript and prevention of its translation.

Our initial objective for this study was to devise a screening approach based on antisense RNA knock-down technology that would enable for the identification of sRNA sequences that are critical for maintaining cellular integrity. We introduced sequences encoding the sense and antisense variants of 12 sRNAs in an RNA expression system and screened for sequences that elicited death or growth defects when their expression was

induced by Atc. Through this screen, we found that overexpressing the antisense counterpart of the RygC sRNA was inhibitory to cell growth. Closer examination of this sequence revealed that the inhibitory effect was not due to antisense knock-down of RygC. Rather, we serendipitously found that this antisense transcript contains an ORF, which encodes a toxic peptide of 19 amino acids. We believe that our method can be widely used for the identification of potential sRNA or other genes with intriguing functions, which may be overlooked using conventional techniques.

2.7. Acknowledgements

We thank Professor Herman Buzard for the gift of plasmids and *E. coli* DH5αZ1 strain. We thank Pascale Swanson for help in constructing pNYL-MCS11 and the Li Lab members for helpful discussions. This project is partially funded by the Ontario Premier Research Excellence Award and Canada Research Chairs Program. EDB and YL are Canada Research Chairs. WM holds an NSERC graduate student scholarship.

2.8. References

[1] J. Vogel, E. G. Wagner, Curr Opin Microbiol 2007, 10, 262.

[2] G. Storz, S. Altuvia, K. M. Wassarman, Annu Rev Biochem 2005, 74, 199.

[3] S. Gottesman, Trends Genet 2005, 21, 399.

[4] G. Padalon-Brauch, R. Hershberg, M. Elgrably-Weiss, K. Baruch, I. Rosenshine, H. Margalit, S. Altuvia, Nucleic Acids Res 2008, 36, 1913.

[5] D. H. Lenz, K. C. Mok, B. N. Lilley, R. V. Kulkarni, N. S. Wingreen, B. L. Bassler, Cell 2004, 118, 69.

[6] V. Pfeiffer, A. Sittka, R. Tomer, K. Tedin, V. Brinkmann, J. Vogel, Mol Microbiol 2007, 66, 1174.

[7] A. Toledo-Arana, F. Repoila, P. Cossart, Curr Opin Microbiol 2007, 10, 182.

[8] R. A. Forsyth, R. J. Haselbeck, K. L. Ohlsen, R. T. Yamamoto, H. Xu, J. D. Trawick, D. Wall, L. Wang, V. Brown-Driver, J. M. Froelich, K. G. C, P. King, M. McCarthy, C. Malone, B. Misiner, D. Robbins, Z. Tan, Z. Y. Zhu Zy, G. Carr, D. A. Mosca, C. Zamudio, J. G. Foulkes, J. W. Zyskind, Mol Microbiol 2002, 43, 1387.

[9] Y. Ji, B. Zhang, S. F. Van, Horn, P. Warren, G. Woodnutt, M. K. Burnham, M. Rosenberg, Science 2001, 293, 2266.

[10] R. Lutz, H. Bujard, Nucleic Acids Res 1997, 25, 1203.

[11] J. Degenkolb, M. Takahashi, G. A. Ellestad, W. Hillen, Antimicrob Agents Chemother 1991, 35, 1591.

[12] R. Hershberg, S. Altuvia, H. Margalit, Nucleic Acids Res 2003, 31, 1813.

[13] K. E. Rudd, Res Microbiol 1999, 150, 653.

[14] K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, Genes Dev 2001, 15, 1637.

[15] K. I. Udekwu, F. Darfeuille, J. Vogel, J. Reimegard, E. Holmqvist, E. G. Wagner, Genes Dev 2005, 19, 2355.

[16] E. Masse, S. Gottesman, Proc Natl Acad Sci U S A 2002, 99, 4620.

[17] P. W. Haebel, S. Gutmann, N. Ban, Curr Opin Struct Biol 2004, 14, 58.

[18] E. Rivas, R. J. Klein, T. A. Jones, S. R. Eddy, Curr Biol 2001, 11, 1369.

[19] B. Tjaden, R. M. Saxena, S. Stolyar, D. R. Haynor, E. Kolker, C. Rosenow, Nucleic Acids Res 2002, 30, 3732.

[20] E. M. Fozo, M. Kawano, F. Fontaine, Y. Kaya, K. S. Mendieta, K. L. Jones, A. Ocampo, K. E. Rudd, G. Storz, Mol Microbiol, 2008, 72, 579.

Chapter 3:
Decoding toxicity: Deducing the sequence requirements of IbsC, a type I toxin in *Escherichia coli*

3.1. Author's preface

Upon confirming that *sibC* (formerly known as *rygC*) and its antisense counterpart *ibsC* constitute a type I TA pair, we embarked on our journey toward elucidating its functions. This chapter details our work in characterizing IbsC, starting at the sequence level. The five Ibs homologs found in *E. coli* K-12 are very well conserved. At the start of this study, we pondered whether this sequence conservation is required for the function of these peptides or if mutations can be introduced into IbsC without compromising its toxicity. We developed a sequence randomization strategy, which enabled us to introduce 200 unique point mutations in IbsC. We also generated a series of truncation mutants in order to reduce the size of this 19 amino acid peptide. We found that IbsC can tolerate extensive mutations despite being well conserved and the first four residues at its N-terminus can be removed without compromising its toxicity. Thus, it can be minimized to 15 amino acids. We also gained structural insights on IbsC and its mutants through this study. We further found a correlation between the toxicity of IbsC mutants and their ability to act on the inner membrane of *E. coli*.

The sequence information gathered from this study can be used to establish and modify search parameters for novel type I TA pairs. Additionally, the mutants obtained through this screen can be used as tools in follow-up studies on *ibsC/sibC*. They can also be used in the engineering of molecular tools based on IbsC. These applications will be discussed in subsequent chapters of this dissertation.

For the experiments discussed in this chapter, I took a leading role in their conceptualization, design, and analysis with guidance from Dr. Li. I also wrote the manuscript. I conducted all of the experiments presented here, with exception of the

screening of IbsC double, triple, and quadruple mutants. These screens were performed by Nirav Patel, a senior undergraduate student who was working under my guidance. Circular dichroism spectroscopy was performed with kind assistance from Dr. Raquel Epand. This chapter is modified from a manuscript that was previously published in the Journal of Biological Chemistry. The citation of the article follows:

Mok, W.W., N.H. Patel, and Y. Li (2010) Decoding Toxicity: Deducing the sequence requirements of IbsC, a type I toxin in *Escherichia coli*. J Biol Chem 285: 41627-41636.

3.2. Abstract

Bacterial genomes encode a collection of small peptides that are deleterious to their hosts when overexpressed. The physiological relevance of the majority of these peptides is unknown at present, although many of them have been implicated in regulatory processes important for cell survival and adaptability. One peptide that is of particular interest to us is a 19 amino acid proteic toxin, coined IbsC, whose production is repressed by SibC, an RNA antitoxin. Together, IbsC and SibC constitute a type I toxin-antitoxin (TA) pair. To better understand the function of IbsC and to decipher the sequence determinants for its toxic phenotype, we carried out extensive sequence analyses of the peptide. We generated a series of truncation and single amino acid deletion mutants to determine the minimal sequence required for toxicity. We further probed into functionally relevant amino acids with a comprehensive set of IbsC mutants produced using a systematic sequence randomization strategy. We found that IbsC remained toxic in the presence of multiple deletions and single amino acid substitutions, despite being well-conserved in *E. coli* and across other Gram-negative bacteria. The

toxicity of this peptide was determined to be dependent on a stretch of highly hydrophobic residues near its centre. Our results defined sequence-function relationship of IbsC and offered additional insights into properties common to membrane-targeting type I toxins in *E. coli* and related species.

3.3. Introduction

Small peptides emerging from microbial genomes have been implicated in diverse catalytic and regulatory roles in cells. These functional peptides of only 16 to 60 amino acids are abundant in the genome of *Escherichia coli* K-12 where over 50 sequences have been detected in a recent study using a combination of biochemical and computational approaches (1). Additional sequences have also been confirmed to exist in other Gram-positive and Gram-negative species.

With their discovery, the array of functions carried out by these peptides also grows. Of the peptides that have been characterized thus far, a few have been demonstrated to function outside the cell as antimicrobial peptides or as secreted signalling peptides (2). A considerable proportion of peptides that remain in the cell have been found to be toxic to cells when they are overexpressed. Many of these toxic peptides are classified as a part of toxin-antitoxin (TA) systems.

Smaller, more hydrophobic peptides are typically found to be a part of type I TA systems, in which the translation of the toxin is suppressed by a noncoding small RNA (sRNA) antitoxin (3). This differs from type II TA systems where the larger and less hydrophobic toxins are antagonized by smaller and more labile proteic antitoxins (3). The most well-characterized type I TA pair identified to date is perhaps the *hok-sok* system initially discovered on the R1 plasmid of *E. coli*. Plasmid-encoded toxins contribute to

plasmid maintenance in bacteria, as the loss of such plasmids would lead to rapid degradation of the antitoxin and accumulation of the toxin, resulting in post-segregational cell death (4). In addition to plasmids, a number of type I TA pairs are found on the chromosome. However, the exact function of many chromosomally-encoded toxins is obscure. Hypotheses of their physiological roles may be drawn from those proposed for their more well-studied type II counterparts (as reviewed in (5-7)).

Similar to many plasmid-encoded TA pairs, TA pairs present on the chromosome have been implicated in the stabilization of conjugative transposons and phage DNA in the genome (8). Other TA systems have been linked to bacterial stress response. Toxin production is elevated in response to specific stress conditions, thereby reversibly suppressing cell growth and preserving energy to counteract these stressors until conditions improve (9,10). Two type I toxins in *E. coli*, TisB (11,12) and SymR (13), are induced by the SOS response, allowing for cells to cope with damaged RNA and DNA respectively. Consistent with this model, some TA pairs have been suggested to promote bacterial persistence, where the production of toxins in a subpopulation of cells causes them to become dormant in response to antibiotic treatments (14,15). Alternatively, TA systems have been postulated to mediate programmed cell death in a large fraction of cells within a population upon exposure to specific stress signals (16,17). The irreversible killing of the selected cells releases nutrients and other resources to sustain the surviving population. While the stress response model and the “altruistic killing” models seem contradictory, it is likely that the physiological role of chromosomally-encoded TA pairs may not be defined by a single unifying model, but individual TA systems may serve different purposes in the cell. It is also possible that the observed toxicity of these

peptides may be an artefact of their overexpression. They may have enzymatic or regulatory functions that are entirely independent of cell killing when expressed at endogenous levels (3). A recent global search for type I TA systems across bacterial species indicated that some TA families are well-conserved (18), implying that such TA systems may be important for bacterial fitness.

Among the six main families of type I TA pairs identified in *E. coli* K-12 is the Ibs/Sib family. Five Ibs/Sib homologs have been found in the *E. coli* K-12 genome, and they have been coined Ibs/SibA-E respectively (19). The *ibs* and *sib* genes are arranged on opposite strands, with the antitoxin encoding sequences completely overlapping the toxin open reading frames. These five TA repeats are distributed at three loci in the genome: *ibs*/*sibA* and *B* are encoded in tandem at the same intergenic region (IGR), while *ibs/sibD* and *E* are arranged in the same manner at another IGR. Here, we are primarily interested in *ibs/sibC*, the sole pair detected in the IGR between *fau* (formerly *ygfA*) and *serA*. We initially came across *ibs/sibC* in a screen for antisense sequences that can knock down the expression of sRNAs and elicit growth defects in *E. coli* (20). Through our screen, we observed growth suppression upon the induction of the reverse complement of the SibC sRNA. Our subsequent analyses demonstrated that the toxic phenotype was not attributed to sRNA interference. Rather, it was caused by the overproduction of a 19 amino acid peptide encoded in the antisense sequence. Our observations corroborates with the results reported in an earlier study published by the laboratory of Gisela Storz, who coined this peptide IbsC (for induction brings stasis) (19).

It has been postulated that IbsC, and other members of the Ibs/Sib family, localizes to the inner membrane following overexpression, and their accumulation contributes to

membrane depolarization (19). Like many type I TA pairs, the biological relevance and the exact mechanism of action of IbsC is currently unknown. It is also uncertain why this potentially deleterious element is maintained in the genome of various *E. coli* strains and in other proteobacterial species, including those from the Enterobacteriaceae, Pasteurellaceaes, and Helicobacteraceae families (18, 21). Herein, we have established a comprehensive set of IbsC mutants to be used to address questions pertaining to the sequence conservation and functionality of this enigmatic peptide.

We carried out extensive sequence truncation studies on IbsC to deduce the minimal sequence for toxicity. A sequence randomization strategy was implemented to systematically introduce mutations in IbsC, allowing us to deduce the amino acid requirements for toxicity. IbsC was found to tolerate high frequencies of amino acid substitutions, considering that a large proportion of mutants retained their toxicity and membrane depolarization capabilities. Our mutagenesis data suggests that the sequence space for toxic peptides is quite large. In general, amino acids with hydrophobic side chains were strongly preferred at multiple positions, and mutations disrupting consecutive hydrophobic residues near the core of the peptide gave rise to inactive mutants. Using IbsC as a model, we defined additional sequence requirements for toxic peptides and refined current parameters used to guide searches for these elements across genomes *in vivo* and *in silico*.

3.4. Materials and methods

3.4.1. Oligonucleotides

Sequences encoding IbsC and its derivatives were generated by PCR or by extending oligonucleotides with 15-20 nt complementary regions. The sequences of these

oligonucleotides are presented in Supplementary Table 1. Oligonucleotides were chemically synthesized by Integrated DNA Technologies (Coraville, IA, USA). Oligonucleotides used for the sequence randomization studies were synthesized such that each nucleotide within the codon(s) of interest was prepared using a mixture of 25 % deoxyadenine, deoxycytosine, deoxythymine, and deoxyguanine phosphoramidites. Oligonucleotides that were longer than 30 nt were purified by 10% (8 M urea) denaturing PAGE (polyacrylamide gel electrophoresis) prior to usage.

3.4.2. Peptide synthesis

Peptides, including wild-type IbsC and mutants IbsC (6-19), IbsC (L8G, L11L, S15A) and IbsC (L11R), were chemically synthesized by Genscript (Piscataway, NJ, USA).

3.4.3. Strains, plasmids, and growth conditions

Information pertaining to the *E. coli* strain and plasmid used in this study are presented in Supplementary Tables 2 and 3. The *E. coli* strain used in this study was DH5αZ1, which was previously described in (22). IbsC and its derivatives were cloned into pNYL-MCS11, a derivative of pZE21-MCS1 (courtesy of H. Bujard), downstream of the tetracycline-inducible promoter. Modifications were made to the parent vector, as previously described in (20). Briefly, we removed the ribosome binding site (RBS) present on pZE21-MCS1 and restored the multiple cloning sites that were lost during the RBS excision.

Molecular cloning procedures, including primer extension, restriction digestion, and ligation, were carried out following supplier-provided protocols. High Fidelity PCR Enzyme Mix, Klenow fragment (exo^-), and T4 DNA ligase were purchased from

Fermentas (Burlington, ON, Canada). EcoRI and BamHI restriction enzymes were purchased from New England Biolabs (Pickering, ON, Canada). Ligation products were transformed into DH5αZ1 cells by electroporation. With the exception of the mutants from the random sequence libraries, cloned constructs were confirmed by DNA sequencing at Mobix Lab (McMaster University).

Bacteria were cultured in Luria-Burtani (LB) media at 37 ^{o}C in a shaking incubator. Growth media were routinely supplemented with 50 μg mL^{-1} of kanamycin and spectinomycin (both purchased from Sigma-Aldrich, Oakville, ON, Canada). To induce the expression of *ibsC* and its derivatives, 200 ng mL^{-1} of anhydrotetracycline (Atc; Sigma-Aldrich) was added to the media.

3.4.4. Growth curves and lethality screens

Growth curves were used to assess the toxicity of IbsC truncation and deletion mutants. Cells carrying pNYL-MCS11 (the negative plasmid control) or plasmid with *ibsC* (the positive control) and its derivatives were cultured overnight. The cells were then subcultured in fresh media (1:400 dilution) in the absence (for uninduced conditions) or presence (for induced conditions) of Atc. These cultures were incubated at 37 ^{o}C with shaking at 260 rpm for 8 hours (h). Growth was monitored hourly by measuring the optical density at 600 nm (OD_{600}) of each sample using a VersaMax microplate reader (Molecular Devices, Sunnyvale, CA, USA). The assays were carried out in triplicate. To compare the toxicity associated with various IbsC mutants, the OD_{600} of each sample measured at t = 8 h was normalized against the OD_{600} of the plasmid control at this time point in order to derive values for "relative OD_{600}".

The toxicity of these constructs was also confirmed by lethality screens in which we plated cells carrying pNYL-MCS11 with *ibsC* mutants on LB agar with or without Atc. The effects of the overexpression of these constructs on growth were observed following overnight incubation at 37 ^{o}C.

3.4.5. Screening for toxic IbsC mutants in random sequence libraries

Upon transforming vectors with each pool of *ibsC* mutants with randomized codons into DH5αZ1 cells, we selected a collection of colonies to build the various IbsC mutant libraries. We picked around 100, 200, 350, and 450 colonies to establish each single, double, triple, and quadruple substitution library respectively. To screen for toxic IbsC derivatives in each of these libraries, each clone was cultured overnight and diluted 1:400 in LB media supplemented with antibiotics and Atc the following day. These samples were then incubated at 37 ^{o}C for 6 h with shaking. The growth assay was carried out in duplicate on two separate 96-well culture plates. OD_{600} measurements were taken at this time point because we consistently observed the greatest difference in growth between the plasmid and the IbsC controls at 6 h post-induction. OD_{600} measurements were normalized against the average OD_{600} corresponding to positive (μ_p) and negative (μ_n) controls. These values were denoted "Normalized Growth" and they were calculated using the equation:

$$\%N.G. = (OD_{600(mutant)} - \mu_p / |\mu_n - \mu_p|) \times 100$$

We sent all of the constructs from the libraries of mutants with single amino acid substitutions for sequencing (Bio Basic Inc., Markham, ON, Canada). We selected 19 mutants from each library covering all of the possible amino acid substitutions at each position. These mutants were subjected to follow-up growth assays in order to confirm

our observations from the initial screens. The assays were repeated at least three times. We considered mutants with %N.G. <40% to be active, as this value represented 3 standard deviation below the average toxicity of 25 negative control samples (*data not shown*). From the libraries of mutants with multiple amino acid substitutions, all active mutants were sent for sequencing. We also sequenced 10 inactive mutants from each library to confirm sequence coverage. Growth assays were repeated in triplicate for sequenced mutants.

3.4.6. Dye uptake assay

Cells carrying the plasmid control and the plasmid with *ibsC* or selected *ibsC* mutants were grown overnight and diluted 1:200 in fresh media. These cultures were grown at 37 ^{o}C with shaking until the OD_{600} of the samples were ~0.3. Cells were subsequently induced with Atc. At 3 h post-induction, cells were harvested and treated with 10 μg mL^{-1} $DiBAC_4(3)$ (Sigma-Aldrich). Cells were incubated with the dye for 20 min before being pelleted and washed with phosphate buffered saline (PBS). The final pellets were resuspended in different volumes of PBS to achieve uniform cell densities across samples. Fluorescence of each sample was measured using a Tecan Safire microplate reader (excitation at 490 ± 5 nm; emission at 516 ± 5 nm). Relative fluorescence of the mutants was calculated by dividing their fluorescence by the one corresponding to the plasmid control. The assay was repeated a minimum of 3 times.

3.4.7. Circular Dichroism Spectroscopy

CD spectra were collected with an AVIV 410 CD instrument (AVIV, Lakewood, NJ, USA). Scans were performed from 260 to 190 nm at a step resolution of 1 nm. A 1-mm pathlength quartz cuvette was used for the measurements. Sample temperature was

maintained at 25 ^{o}C using a thermostatically controlled cell holder. Peptides were solubilized in 2,2,2-trifluoroethanol (TFE; Sigma-Aldrich) at a concentration of 50 μM. Background signal contributed by the solvent was subtracted for each sample. The CD spectra were converted to mean residual ellipticity. The secondary structure content of each sample was estimated from each spectrum using the CONTIN method available through the CDPro software (23).

3.5. Results

3.5.1. Minimization of the IbsC toxin

With 19 amino acids, IbsC is one of the smallest bacterial toxins identified to date (1,19). We were interested in exploring whether this constitutes the lower size limit of type I toxins or whether shorter peptides can exhibit growth inhibitory potential in *E. coli*. We generated a series of terminal truncation mutants of IbsC by sequentially removing codons from the 5' and 3' ends of its open reading frame (ORF). We eliminated codons 2 through 7 from the 5' truncation variants (Figure 3.1A), while the 3' truncation variants lacked codons 19 through 13 (Figure 3.1B). The start codon was retained in the 5' truncation mutants to enable translation initiation. Each sequence contained the ribosome binding site native to IbsC. The 13 IbsC truncation sequences were introduced into pNYL-MCS11, such that their expression was driven by a tetracycline-inducible promoter ($P_{LtetO-1}$) (20,22). *E. coli* strain DH5αZ1, which constitutively expressed a tetracycline repressor to allow for tight regulation of genes governed by $P_{LtetO-1}$(22), were transformed with these constructs. The production of these IbsC derivatives was induced with anhydrotetracycline (Atc).

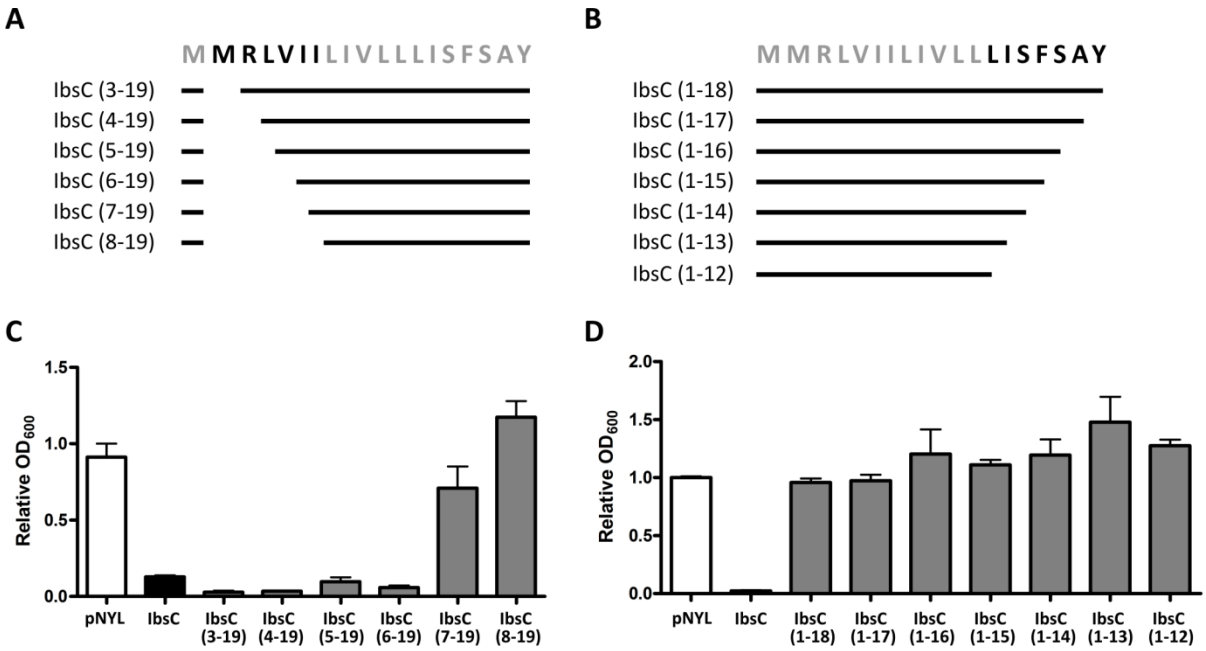


Figure 3.1. Toxicity of IbsC 5' and 3' truncation mutants. A. Amino acid sequence of wild-type IbsC and schematics of the 5' truncation mutants. Codons 2 to 7 were sequentially removed from the open reading frame (ORF) of *ibsC* in these constructs. B. Schematics of the 3' truncation mutants in which codons 19 through 13 were removed. C. Toxicity of 5' truncation mutants were assessed by comparing the relative OD_{600} of each mutant to that associated with the positive (IbsC) and negative (pNYL) controls at 8 h post-induction. Mutant IbsC (6-19), which lacked amino acids 2 to 5, was found to be the minimal toxic derivative of IbsC. D. Toxicity of 3' truncation mutants. Relative OD_{600} of each mutant was compared with that of the controls at 8 h post-induction. Removal of any of the amino acids near the C-terminus of IbsC was found to be deleterious to its toxicity. N = 3.

Growth of cells carrying plasmids expressing the IbsC truncation mutants was monitored over 8 h following the induction of each IbsC truncation mutant (Supplementary Figures S3-1 and S3-2). Cell density was determined by measuring the OD_{600}. To facilitate comparison of the toxicity of each mutant, the OD_{600} of each sample measured at t = 8 h was normalized with the OD_{600} of cells carrying pNYL-MCS11, which served as our negative control. As observed in Figure 3.1C, IbsC 5' truncation mutants lacking amino acids 2 to 5 remained growth suppressive, indicating that these

residues close to the N-terminus of the peptide are not essential for toxicity. Removal of amino acids beyond the fifth residue led to inactive constructs, as observed with mutants IbsC (7-19) and IbsC (8-19), which lacked amino acids 2 to 6 and amino acids 2 to 7 respectively. On the other hand, elimination of amino acids near the C-terminus was not tolerated, and cells expressing the 3' truncation mutants grew to similar extents as the plasmid control (Figure 3.1D). The data presented here suggests that the stretch of hydrophobic residues at the middle of the peptide and the few polar residues near the C-terminus may have functional significance for IbsC. We also demonstrated that this toxin may be minimized to 15 amino acids without resulting in significant loss of toxicity.

3.5.2. Contribution of individual amino acids to toxicity

Data from our sequence truncation studies suggested that amino acids from positions 6 to 19 of IbsC are required for its toxicity. However, we were uncertain if all of the amino acids in this region are required for toxicity. To probe into amino acids that are potentially important for the function of IbsC, we designed a set of 15 IbsC deletion mutants, each lacking an individual amino acid (Figure 3.2A). For amino acids that occur in tandem, such as the two isoleucine residues at positions 6 and 7, only the codon for the first amino acid was removed. Expression of the IbsC deletion mutants was induced in *E. coli* grown on LB-agar supplemented with Atc (Figure 3.2B). Consistent with results from our sequence truncation analysis, IbsC derivatives lacking any of the last 10 amino acids failed to suppress growth following overexpression (mutants Δ10 - Δ19), but amino acids M2, R3, and L4 can be deleted without interfering with the toxicity of IbsC (mutants Δ2 - Δ4). Interestingly, mutants lacking residues I6 (mutant Δ6) remained growth inhibitory when overexpressed although 5' truncation mutants lacking I6 were

nontoxic. While 5' truncation mutant IbsC (6-19), which lacked amino acids M2 to V5, was found to be toxic, removal of amino acid V5 on its own led to a loss of toxicity (mutant Δ5). This observation was confirmed upon monitoring the effect of overexpressing this deletion mutant in cells grown in liquid media over 8 h (Supplementary Figure S3-3). It is possible that the removal of residue V5 perturbed the hydrophobicity in its adjacent region or the active structure of the peptide.

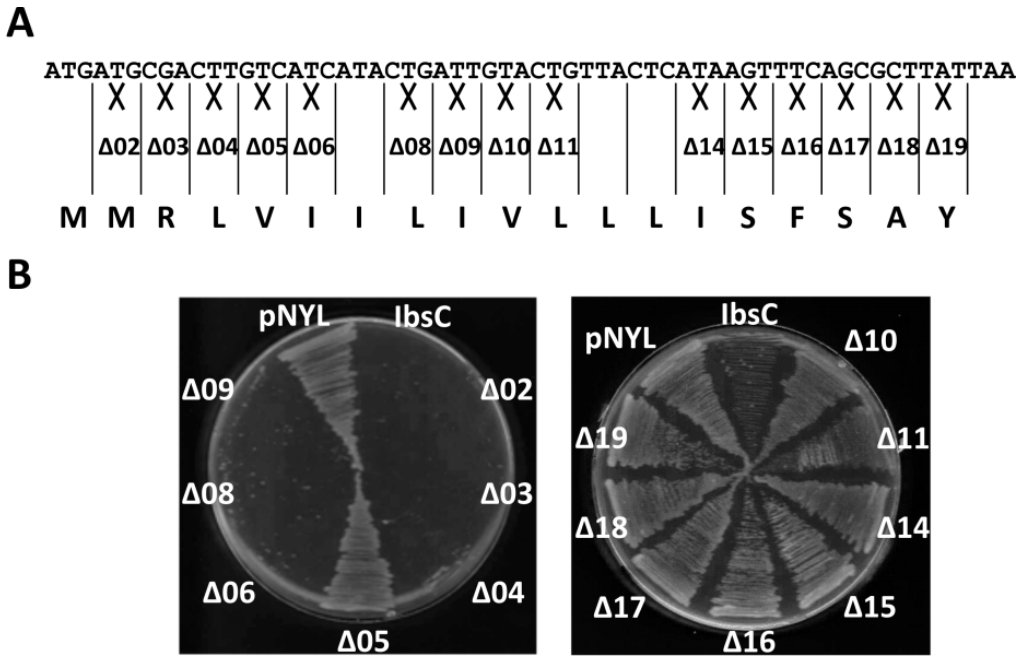


Figure 3.2. Effect of single amino acid deletions on the toxicity of IbsC. A. Design of IbsC deletion mutants. Starting from codon 2, codons were individually removed from *ibsC*. For amino acids that occurred in tandem, such as I6 and I7, only the codon of the first amino acid was removed. B. Phenotypic changes associated with the overexpression of each deletion mutant in *E. coli* DH5αZ1. Bacteria carrying pNYL-MCS11 with the deletion mutants were plated on LB-agar with Atc and incubated at 37 ^{o}C overnight. With the exception of amino acid V5, elimination of any amino acids from positions 2 to 9 did not seem to affect the toxicity of IbsC (left panel). Removal of any of the last 10 amino acids led to a loss of toxicity (right panel).

3.5.3. Variability of amino acid V5

The significance of residue V5 in full-length IbsC was examined using a sequence randomization strategy. The sequence encoding the open reading frame of *ibsC* was designed with mutations at codon 5, such that each nucleotide in the codon was synthesized using a mixture of 25% deoxyadenine, deoxycytosine, deoxythymine and deoxyguanine phosphoramidites (Figure 3.3A). This pool (or library) of *ibsC* mutants was subsequently cloned into pNYL-MCS11, downstream of P_{LtetO1}. *E. coli* DH5αZ1 cells carrying these recombinant vectors were grown in liquid media supplemented with Atc in order to screen for growth suppression associated with the overexpression of these IbsC derivatives. Based on our design, we would theoretically need to sample 64 (4^3) sequences in order to cover all the possible mutations in this codon. If the randomization was unbiased, examining 64 mutants would encompass all 20 amino acids and the three stop codons. In order to account for unligated vectors and the isolation of identical mutants, we selected ~100 clones to be screened using a growth assay.

In this assay, the OD_{600} of each sample ($OD_{600(mutant)}$) was measured at 6 h post-induction. These values were normalized against the average OD_{600} corresponding to cells carrying pNYL-MCS11 (the negative control, μ_n) and to cells carrying pNYL-MCS11 with wild-type IbsC (the positive control, μ_p). The normalized OD_{600} was denoted "Normalized Growth" (%N.G.) and was calculated using the following equation:

$$\%N.G. = (OD_{600(mutant)} - \mu_p / |\mu_n - \mu_p|) \times 100$$

We observed that ~70% of the sequences that were sampled in this screen appeared active, as their overexpression suppressed growth to levels that were comparable to the overexpression of wild-type IbsC. Following the screen, the constructs were sequenced

and amino acid substitutions at position 5 were determined. We isolated IbsC derivatives with substitutions for each of the 20 amino acids at this position. We selected single mutants for each substitution and subjected them to a follow-up screen in order to confirm their toxicities. As detailed in Figures 3.3B and 3.3C, position 5 of IbsC can withstand substitution by most amino acids regardless of their hydrophobicity, size, and charge. However, glycine substitution was not tolerated at this position and overexpression of the V5G mutant did not suppress growth. The loss of toxicity in this mutant and in the Δ5 deletion mutant suggests that the presence of a functional group in the amino acid at the fifth position may be important to the activity of IbsC.

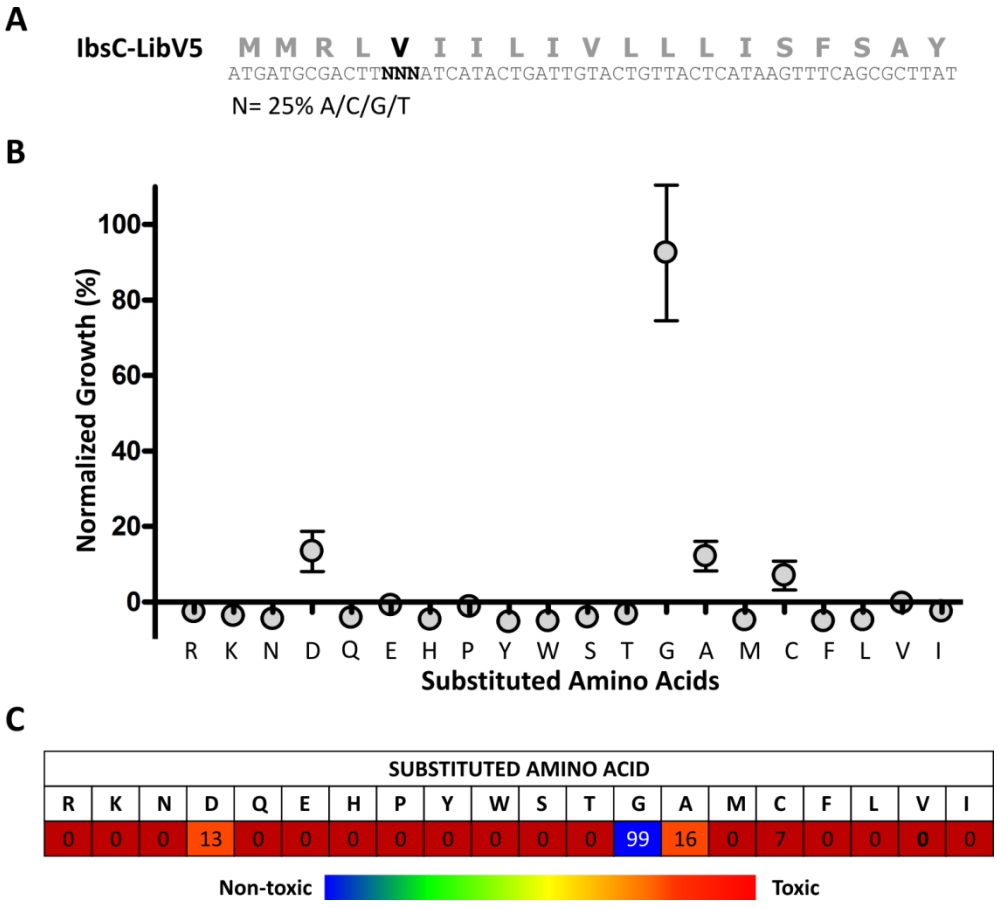


Figure 3.3. Randomization of the codon encoding amino acid V5 in IbsC. A. Design of the IbsC-V5 library. Each nucleotide in codon 5 was synthesized with 25% deoxyadenine, deoxyguanine, deoxycytosine, and deoxythymine phosphoramidites (denoted by "N"). B. Growth assay of IbsC-V5 mutants with the 20 amino acid substitutions. One mutant with each amino acid substitution was selected from the library and was subjected to a follow-up growth assay in which *E. coli* expressing each mutant was grown for 6 h. OD_{600} of each sample was measured thereafter, and these values were normalized against the growth of bacteria expressing IbsC and the plasmid control (referred to as "normalized growth" or %N.G.). The data points and error bars represent average %N.G. and standard deviation calculated from triplicate. C. Average toxicities of IbsC mutants with each of the 20 amino acid substitutions at position V5. The amino acids substituted at this position are noted in row 2 of the table. The number in each box in row 3 represents the average %N.G. associated with the overexpression of each mutant calculated from 3 experiments. Amino acid substitutions giving rise to active sequences (%N.G. <10%) are depicted by red boxes, those that led to peptides with intermediate toxicities (%N.G. between 10-20%) are indicated by orange boxes, and those that led to a loss of toxicity (%N.G. > 40%) are shown in blue.

3.5.4. Variability of C-terminal amino acids

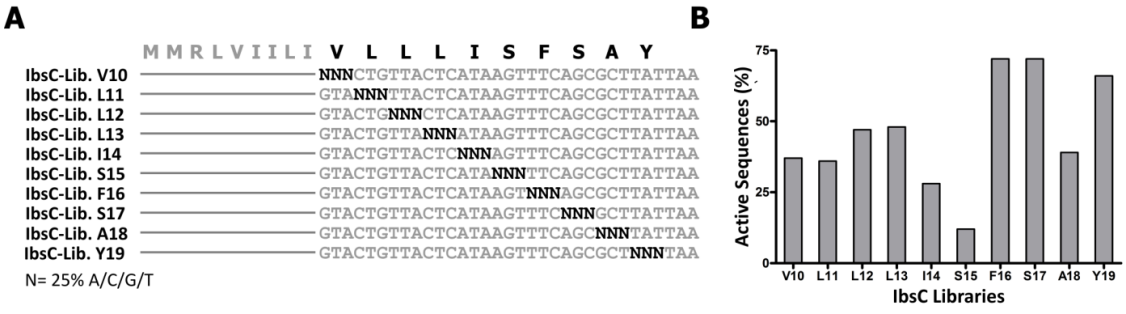
In addition to amino acid V5, our single amino acid deletion experiments revealed that the loss of any of the last 10 amino acids in IbsC was deleterious to its toxicity. As such, we carried out a complete substitution analysis of each amino acid in this region using the aforementioned sequence randomization strategy in order to deduce the amino acid requirements in this region of IbsC (Figure 3.4A). Following their synthesis, the 10 pools of mutant *ibsC* sequences were separately cloned into pNYL-MCS11. Each group of recombinant plasmids was then transformed into DH5αZ1 cells. Approximately 100 constructs were selected from each pool to establish our 10 libraries of IbsC derivatives, and the toxicity of sequences from each library was independently assessed using growth assays. Active sequences were defined as those that can suppress growth by >60% when overexpressed (%N.G. <40%).

From our initial screen, we observed that amino acid substitutions were tolerated at most positions in this segment of IbsC. An average of 40% of the sequences across the 10

libraries were found to be active (Figure 3.4B). Amino acids near the C-terminus of the peptide, namely positions F16, S17, and Y19, were more mutatable than the rest of the sequence. Libraries corresponding to these amino acids contained nearly 70% active sequences. On the other hand, the library corresponding to position S15 displayed a much lower frequency of active sequences (12%). We sequenced all of the constructs that were examined. Upon analyzing the amino acid substitutions in each library, we noticed that our randomization approach generated near comprehensive sequence coverage. Amino acids that were absent from each library were often coded by non-degenerate codons, such as methionine and tryptophan (*data not shown*). We rationally designed and synthesized *ibsC* mutants that were not isolated from the initial screen in order to complete the set of substitutions. To verify the activity of each mutant, we selected 20 mutants from each library, each with a unique amino acid substitution, and repeated growth assays using these constructs (Supplementary Figure S3-4).

As presented in Figure 3.4C, amino acids located near the centre of IbsC, principally the ones at positions 10 to 14, exhibited a strong preference for hydrophobic residues (with L, F, M>V>I). Hydrophilic and charged residues are generally not favoured at these positions. Substitutions by charged amino acids near the core of IbsC often led to a complete loss of toxicity. In contrast, positions 16, 17, and 19 demonstrated a propensity to accept hydrophilic and ionic residues. However, strongly hydrophobic residues, such as leucine, valine, and isoleucine, are not favourable substitutions at position 16 and 17. Consistent with the high percentage of active sequences observed in our initial screen, positions 16, 17, and 19 were able to withstand more than 15 of the 19 amino acid substitutions. Position 15, which gave rise to few active sequences in the screen, had only

three neutral amino acid substitutions. Like serine, the parent amino acid at this position, the accepted substitutions were small and mostly polar (T, C, and A).



C

Original Amino Acid		SUBSTITUTED AMINO ACID																		
		R	K	N	D	Q	E	H	P	Y	W	S	T	G	A	M	C	F	L	V	I
10	V	80	100	31	91	90	100	95	98	69	25	44	3	37	2	66	46	33	7	0	77
11	L	100	100	85	100	70	100	100	100	87	100	57	51	17	26	1	1	4	0	20	21
12	L	100	100	69	100	36	100	48	100	26	7	7	13	12	18	8	2	6	0	16	45
13	L	87	100	69	76	66	100	72	77	58	35	8	0	26	1	0	1	0	0	0	0
14	I	100	100	100	100	100	100	94	97	29	1	87	12	66	28	0	10	0	0	5	0
15	S	74	62	22	60	61	49	50	63	100	79	0	0	71	0	75	0	76	76	100	100
16	F	28	4	7	4	3	14	0	0	0	0	7	1	0	0	0	31	0	21	47	14
17	S	1	0	1	3	3	1	7	1	0	69	0	25	7	8	15	0	44	76	53	100
18	A	74	93	61	66	70	64	77	100	63	70	24	2	47	0	2	6	82	5	4	29
19	Y	0	27	1	39	12	31	2	90	0	13	12	6	33	2	12	0	31	6	0	1

Non-toxic — Toxic

Figure 3.4. Complete amino acid substitution of the last 10 amino acids of IbsC. A. Randomization of codons encoding amino acids V10 to Y19. Each nucleotide in the codon to be randomized (denoted by "N") was synthesized with 25% deoxyadenine, deoxyguanine, deoxycytosine, deoxyand thymine phosphoramidites. Using this approach, 10 IbsC random sequence libraries were generated. B. Percentage of active sequences observed in each library. Approximately 100 mutants were present in each library. The effect of overexpressing each mutant in the libraries was determined by measuring the OD_{600} of cells 6 h following their induction. Growth of each sample was normalized against the OD_{600} of cells expressing IbsC and the plasmid control. Mutants whose overexpression led to normalized growth (%N.G.) of 40% or less in the two growth assays were considered as active and were tabulated. C. Average toxicities of IbsC mutants with each of the 20 amino acid substitutions at positions 10-19. One mutant with each of the 19 amino acid substitutions was selected from each library to assemble a comprehensive set of single amino acid substitution mutants. The effect of overexpressing these mutants on the growth of *E. coli* was evaluated 6 h post-induction. In the table, the first two columns depict the position and original amino acid at the residue of interest.

Row 2 depicts the amino acid substitution at each position. The number in each box represent the average %N.G. calculated from three growth assays. Toxicity of each mutant is color coded: active sequences (%N.G. <10%) are depicted by red boxes, mutants with intermediate activities (%N.G. between 10-20%) are indicated by orange boxes, weakly active mutants (%N.G. between 20-40%) are highlighted in amber/yellow and inactive mutants (%N.G. > 40%) are shown in green/blue.

3.5.5. Multiple amino acid substitutions in IbsC

We have established that single amino acid substitutions are generally well-tolerated in IbsC. Here, we seek to determine whether this peptide can accept mutations at multiple locations. We are interested in examining if the identity of an amino acid at one location depends on that at another. Due to its hydrophobicity and its propensity to interact with the inner membrane (19), IbsC is hypothesized to adopt an α-helical structure. As such, we have applied the sequence randomization approach to mutate multiple amino acids on the same face of the predicted helix.

We first mutagenized the isoleucine residues located at positions 7 and 14, which are predicted to reside on one face of the helix (Figure 3.5A). We carried out growth assays with 232 constructs that make up the double mutation library. Compared to the screens conducted with the single mutation libraries, we observed remarkably fewer active sequences. We sequenced 35 constructs from this library. This encompasses all of the constructs that exhibited full and partial toxicities in addition to 10 constructs that were found to be nontoxic. We removed sequences with base insertions or deletions, as well as those that appeared to be contaminated with more than one population of mutants. The remaining sequences were subjected to additional growth assays to confirm our observations. From this library, only five sequences with toxicities comparable to IbsC were isolated (Supplementary Figure S3-5A). The %N.G. associated with these mutants

was <10%. We also isolated three sequences that displayed intermediate activity (with %N.G. between 10-40%). From these sequences, we note that hydrophobic residues (C, V, L, and I) are favoured at both position 7 and 14. When the amino acids at these two loci were replaced by more hydrophilic residues, the toxicity of the resulting peptide diminished, although these mutants were still able to suppress growth by 60% or more. We also obtained the sequences of 10 inactive mutants. Most of these constructs were found to contain charged residues (K, R, D, or E) at one or both of these positions.

In addition to generating a library of double mutants, we screened triple mutants by randomizing the codons encoding amino acids L8, L11, and S15 (Figure 3.5B). These amino acids are hypothesized to be on the side of the putative α-helix opposite of I7 and I14. We expected to detect very few toxic triple mutants from our screen. Thus, we subjected nearly 350 constructs to growth assays. From this screen, we identified one active sequence with the expected length (Supplementary Figure S3-5B). However, further sequence analysis indicated that this mutant only contained two amino acid substitutions (L8G and S15A). The remaining mutants that we sequenced were inactive. Many of these mutants contained proline at positions 8 and 15. Consistent with the inactive mutants we examined in the double mutant library, we observed that positively and negatively charged residues were generally disfavoured in the triple mutants.

From our single amino acid substitution study, we observed that positions 11 to 14 of IbsC demonstrated a strong preference for hydrophobic residues (see Figure 3.4C). We speculated that the presence of these consecutive hydrophobic amino acids is required for the toxicity of IbsC. They may be involved in mediating the interaction between IbsC and the inner membrane. Thus, we randomized the 12 nucleotides corresponding to codons 11

to 14 to examine the amino acid requirements in this region (Figure 3.5C). Due to the scarcity of active sequences isolated from our double and triple mutant libraries, we examined the effect of the overexpression of around 450 sequences on the growth of *E. coli*. We did not isolate any active clones through this screen. Twenty constructs from this library were selected and sequenced. We eliminated sequences that appeared to have base insertions, deletions, or contaminations, and the remaining sequences were subjected to additional growth assays. We noticed that hydrophilic and charged residues were prevalent in these inactive mutants (Supplementary Figure S3-5C). One mutant, IbsC (L11V, L12L, L13I, I14V), was found to be inactive even though the amino acids at positions 11 to 14 of this mutant remained highly hydrophobic, akin to the residues in the original sequence. This suggests that drastic changes to the sequence of IbsC are deleterious to its toxicity.

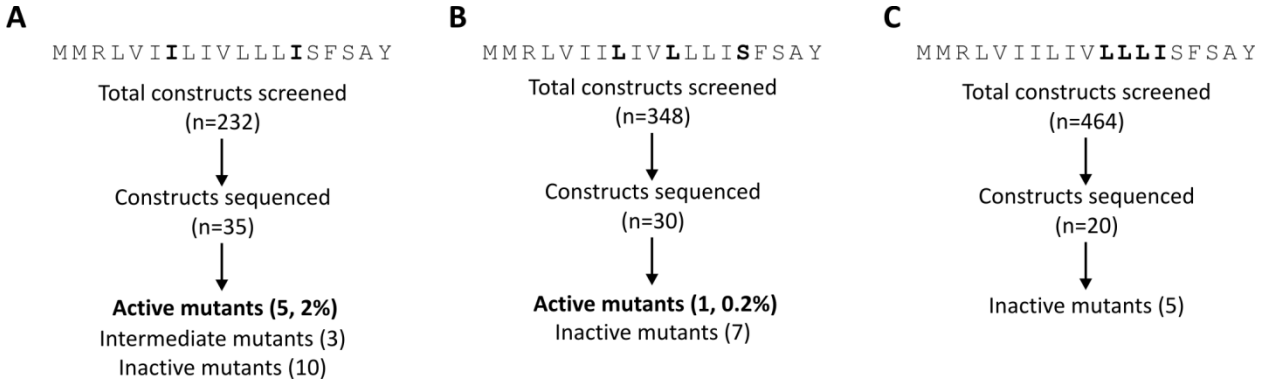


Figure 3.5. Toxicities of IbsC mutants with multiple amino acid substitutions. A. Library of IbsC double mutants in which codons encoding amino acids I7 and I14 were randomized using the sequence randomization approach. B. Library of IbsC triple mutants with randomized codons L8, L11, and S15. C. Library of IbsC quadruple mutants with randomized codons L11, L12, L13, and I14. Amino acids encoded by the randomized codons in each library are indicated in black. The number of constructs that were screened and sequenced is presented in the flowcharts. The number of sequenced mutants displaying high (%N.G. <10%), intermediate (%N.G. between 10-40%), and low (%N.G. >40%) toxicities in each library are indicated.

3.5.6. Structural analysis of IbsC and selected derivatives

To further examine whether the activity of IbsC mutants is dependent on their structures, circular dichroism spectroscopy was carried out. Wild-type peptide and mutants IbsC (6-19) and IbsC (L8G, L11L, S15A) were chemically synthesized and subjected to CD analysis. The non-toxic mutant, IbsC (L11R), was also analysed for comparison. With hydrophobicities of >75 %, the peptides were insoluble in aqueous solvents. As such, 2,2,2-trifluoroethanol (TFE) was used as the solvent. The CD spectra collected for each peptide was analysed using a secondary structure prediction program. As hypothesized, IbsC predominantly existed as an α-helix (Supplementary Figures S6A-D). Interestingly, truncation mutant IbsC (6-19) and triple mutant IbsC (L8G, L11L, S15A), which displayed comparable toxicities as IbsC, were comprised of β-aggregates along with α-helices. On the other hand, IbsC (L11R), the inactive point mutant, also seemed to favour the α-helical conformation under these conditions. Our structural analysis suggests that IbsC does preferentially adopt an α-helical conformation. However, the toxicities associated with these peptides may not be solely dependent on their secondary structures.

3.5.7. Overexpression of IbsC and its toxic derivatives causes membrane depolarization

IbsC has been proposed to interact with the inner membrane of *E. coli*. Its overexpression has been proposed to induce pore formation and elicit membrane depolarization (19). It is uncertain whether toxic IbsC derivatives disrupt the integrity of the inner membrane like their wild-type counterpart and if non-toxic variants have lost their ability cause membrane defects. To examine the mechanism of toxicity of IbsC mutants, we subjected *E. coli* expressing a subset of these mutants to a dye uptake assay.

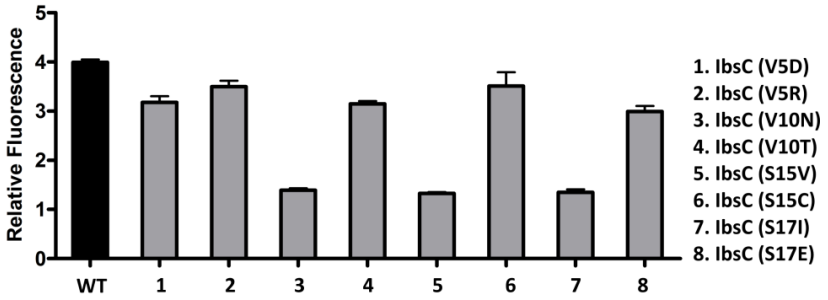
We monitored the ability of cells to take up a fluorescent dye, $DiBAC_4(3)$ [Bis-(1,3-dibarbituric acid)-trimethine oxonal], following the overexpression of 14 active and inactive IbsC variants. Following membrane depolarization, $DiBAC_4(3)$ enters cells, localizes to highly hydrophobic environments, and interacts with cytoplasmic proteins (24). This subsequently results in an increase in fluorescence, which is expected to be proportional to the change in membrane potential (24). In this assay, we induced the expression of IbsC and its derivatives in DH5αZ1 cells for 3 h before they were incubated with $DiBAC_4(3)$. We chose to assess dye uptake at 3 h post-induction, because maximal increase in fluorescence was detected at this time point following the induction of wild-type IbsC (data not shown).

We subjected eight IbsC point mutants to the dye uptake assay. This set of mutants contained constructs that exhibited full activity [IbsC (V5R, V10T, S15C, and S17E)], intermediate activity [IbsC (V5D and V10N)], and no activity [IbsC (S15V and S17I)] in our growth assays. IbsC overexpression led to a 4-fold increase in fluorescence relative to the plasmid control (pNYL). Comparable to the wild-type toxin, active mutants led to a 3- to 4-fold fluorescence increase (Figure 3.6A). The two inactive mutants demonstrated a slight increase in fluorescence (~1.4-fold relative to the negative control). It is possible that the accumulation of these small peptides resulted in stress on the inner membrane, yet the defect may not be severe enough to cause growth suppression. Overexpression of the IbsC (V10N) mutant did not result in apparent damage of the inner membrane. In previous growth assays, we saw that there is often variability in the toxicity of intermediately active constructs (Supplementary Figure S3-3). Here, our results suggest

that this variability in toxicity may be attributed to impaired membrane interaction or membrane penetration by these mutants.

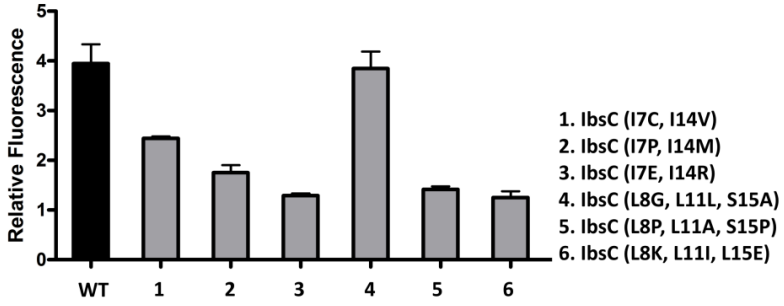
We further selected six double and triple mutants with varying toxicities and examined their effect on the integrity of the inner membrane (Figure 3.6B). Change in fluorescence observed following the induction of the three inactive mutants approximated the change associated with the plasmid control. The expression of mutant IbsC (I7P, I14M), which displayed moderate toxicity, led to a two-fold increase in fluorescence. Overexpression of mutant IbsC (I7C, I14V) resulted in fluorescence increase of ~2.5-fold. Consistent with this observation, this mutant was 10% less toxic than IbsC. We also examined one mutant, IbsC (L8G, L11L, S15A), which appeared as toxic as IbsC in our initial screen. Similar to IbsC, this construct elicited a 4-fold increase in fluorescence after 3 h of Atc induction. Overall, the extent of membrane depolarization associated with the overexpression of these mutants coincided with their toxicities.

A



Constructs	WT	1	2	3	4	5	6	7	8
% N.G.	0	13	0	31	3	100	0	100	0.7
R.F.	4.0	3.2	3.5	1.4	3.5	1.3	3.5	1.4	3.0

B



Constructs	WT	1	2	3	4	5	6
% N.G.	0	8	48	95	0	92	100
R.F.	4.0	2.4	1.8	1.3	3.8	1.4	1.2

Figure 3.6. Overexpression of toxic IbsC derivatives is disruptive to the integrity of the inner membrane. A. Effect of IbsC point mutants on membrane potential. Eight IbsC point mutants with different levels of observed toxicity were overexpressed in *E. coli* DH5αZ1. At 3 h after inducing the expression of each mutant, *E. coli* was incubated with $DiBAC_4(3)$, a potential-sensitive fluorescent dye. $DiBAC_4(3)$ penetrates cells with compromised inner membranes, resulting in an increase in green fluorescence (λ_{em} = 516 nm). Similar to the overexpression of IbsC, cells expressing toxic mutants with V5D, V5R, V10T, S15C, and S17E substitutions were approximately 3- to 4-fold more fluorescent relative to the plasmid control. B. Change in membrane polarization following the overexpression of IbsC combination mutants. We selected 3 IbsC double mutants and 3 IbsC triple mutants with varying toxicities and tested their effect on the inner membrane using the $DiBAC_4(3)$ uptake assay. Inducing the expression of active mutants IbsC (I7C, I14V) and IbsC (L8G, L11L, and S15A) dissipated the proton motive force, resulting in dye uptake and fluorescence enhancements of ~2.5 to 4-fold. The tables below each graph compare the toxicities (denoted by %N.G.) with the relative fluorescence (R.F.) observed following the overexpression of each construct. N = 3.

3.6. Discussion

The Ibs family of toxins is well-conserved in *E. coli* and other families of Gram-negative bacteria. Nevertheless, our study revealed that IbsC can withstand extensive mutations. Despite having only 19 amino acids, we found, through sequence truncation experiments, that the length of this peptide can be further minimized to 15 residues. With the exception of amino acid V5 and those at the last 10 positions of the peptide, amino acids can be individually removed without compromising the IbsC toxicity. We systematically introduced mutations in the codons encoding these 11 amino acids using a sequence randomization strategy, and generated over 200 IbsC point mutants. Through growth and dye uptake assays, we found that a number of IbsC mutants exhibited growth inhibition and membrane disruption activities that are comparable to that of their wild-type counterpart when they are overexpressed. We are curious as to why amino acid substitutions are not more frequently observed in the native Ibs homologs. The sequences of the five Ibs homologs in *E. coli* K-12 (19) and those derived from *Salmonella typhimurium* LT2 (25) are very well conserved. Amino acid substitutions that are found at the more variable positions often share similar properties (*e.g.* R being substituted by K; L being substituted by V or I).

Considering that the Sib antitoxins are encoded on the antisense strand of Ibs, the nucleotide sequences of both genes may have coevolved in a manner that can favour the structure of the sRNA antitoxins and prevent non-cognate sRNA-mRNA interactions. In a recent study, it was reported that codons encoding amino acid S17 of IbsC, which are not conserved among the five Ibs homologs in *E. coli* K-12*,* serve as a part of a recognition site between cognate *sib-ibs* pairs (26). In our present study, where the SibC antitoxin is

not overexpressed, we also observed a high frequency of amino acid substitutions at this position in our toxic mutants.

The sequence conservation may also imply a possible sequence-function relationship in the peptide. The toxic phenotype we observed may be a consequence of IbsC overexpression. Under physiological conditions, the expression of the Ibs peptides may rarely surpass the threshold at which it starts to become deleterious to the cell. As such, the native Ibs sequences may be favoured because they are better suited for their potential biological functions and for protein-protein or protein-membrane interactions.

At present, the exact cellular target(s) of IbsC and its mechanism of action are unclear. Our mutagenesis data provided some insight into the residues that may be important for the function of this peptide. IbsC was proposed to adopt an α-helical conformation and insert into membranes (19). Structural data presented in our present study supports that hypothesis. Growth analyses conducted with the IbsC point mutants suggest that the hydrophobic residues spanning the middle of the peptide may be required for the proposed transmembrane localization of IbsC. From the combination mutants, we found that valine, leucine, and isoleucine are highly favoured at multiple positions. From our mutagenesis studies, we noticed that polar and ionisable amino acids are well-tolerated at positions flanking the hydrophobic residue-rich region in the middle of the peptide. If IbsC is indeed a transmembrane peptide, it is possible that these hydrophilic and charged residues can facilitate initial contact between the peptide and polar head groups of the bilayer. These residues may further play a role in anchoring the peptide and stabilizing its interaction with the inner membrane. While we observed that an inactive point mutant with an arginine substitution at the core of the peptide was capable of adopting an α-

helical structure in solvent, the presence of this charged residue may impede its interaction with the lipid bilayer. In the same experiments, we noticed that mutants IbsC (6-19) and IbsC (L8G, L11L, S15A) were not as helical as the wild-type toxin or as the IbsC (L11R) mutant in TFE. However, they retain their hydrophobic cores and may stably insert into the phospholipid bilayer when placed in the context of the membrane. In addition to causing structural changes and disrupting membrane insertion, mutations giving rise to non-toxic mutants may also perturb the peptides' ability to self-aggregate or interact with other cellular targets.

Type I toxins have been implicated in regulatory processes that are important for the individual and communal survival and adaptability of bacteria. These elements are suggested to be abundant in microbial genomes. Currently, a number of biochemical approaches and computer algorithms are implemented in searches for peptides and sRNAs that comprise putative type I toxin-antitoxin pairs (18,27). As more TA pairs are identified and characterized, parameters guiding their searches are refined. Using IbsC as a model, we developed a better understanding of the sequence requirements for small peptide toxins found in prokaryotes. From our sequence truncation analysis, we learned that IbsC can be minimized to 15 amino acids and remain active, suggesting that the lower size limit of toxins can be set at 15 residues. Based on the minimal active truncation mutant, IbsC (6-19), we postulate that this toxin requires a minimum of 10 amino acids with highly hydrophobic side chains (*e.g.* M, F, L, V, or I) to retain its toxicity. Growth assays carried out with our IbsC point mutants indicated that only substitutions by other hydrophobic residues are tolerated at these positions near the core of the peptide. This is consistent with previous studies on transmembrane proteins, which

suggest that 9 to 11 residues are sufficient to promote helix translocation across the membrane (28-30). At positions flanking the hydrophobic core, amino acids with different hydrophilic side chains were found to be able to substitute for native hydrophilic residues. This again demonstrates that many of IbsC's residues can be substituted with similar functional groups. This knowledge of the general architecture of IbsC may aid in the design and search for novel toxin/ antitoxin systems.

Searches for novel TA systems are further fuelled by the therapeutic potential of these elements. For example, expression of the HokC toxin from *E. coli* was induced in melanoma (31), breast (32), and lung (33) cancer cell lines. This toxin showed promise in hindering the growth and proliferation of these malignant cells. Type I toxins may also be considered as possible antimicrobial targets. Recent studies have shown that treating *E. coli* with certain antibiotics, such as ciprofloxacin, can induce the expression of toxins that have been linked to the SOS response, thereby promoting the formation of persister cells and enhancing tolerance to these drugs (34). In another study, it was found that the overexpression of membrane-targeting peptides, including IbsC, could increase the sensitivity of bacteria toward tobramycin and other aminoglycosides (35). Given the possible applications of toxin-antitoxin systems, it would be of interest to further probe into the regulation and function of the Ibs family of toxins in the cell and to examine whether IbsC and the derivatives we generated through these studies would be suited for therapeutic purposes.

3.7. Acknowledgements

We are thankful to Li Lab members for their helpful discussions with regards to the project design and to the manuscript. We are also tremendously grateful to Drs. Richard and Raquel Epand, Dr. Giuseppi Melacini, and Julijana Milojevic for their time, use of equipment, and advice on structural studies. We thank Professor Herman Bujard for the gift of plasmids and the *E. coli* DH5αZ1 strain. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Y.L. is supported by the Canada Research Chairs program. W.K.M. is an NSERC CGS Doctoral Scholarship recipient.

3.8. References

1. Hemm, M. R., Paul, B. J., Schneider, T. D., Storz, G., and Rudd, K. E. (2008) *Mol. Microbiol.* **70**, 1487-1501
2. Alix, E., and Blanc-Potard, A. B. (2009) *Mol. Microbiol.* **72**, 5-11
3. Fozo, E. M., Hemm, M. R., and Storz, G. (2008) *Microbiol. Mol. Biol. Rev.* **72**, 579-589
4. Gerdes, K., Rasmussen, P. B., and Molin, S. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3116-3120
5. Hayes, F. (2003) *Science* **301**, 1496-1499
6. Magnuson, R. D. (2007) *J. Bacteriol.* **189**, 6089-6092
7. Van Melderen, L., and Saavedra De Bast, M. (2009) *PLoS Genet.* **5**, e1000437
8. Wozniak, R. A., and Waldor, M. K. (2009) *PLoS Genet.* **5**, e1000439
9. Pedersen, K., Christensen, S. K., and Gerdes, K. (2002) *Mol. Microbiol.* **45**, 501-510
10. Ramage, H. R., Connolly, L. E., and Cox, J. S. (2009) *PLoS Genet.* **5**, e1000767
11. Darfeuille, F., Unoson, C., Vogel, J., and Wagner, E. G. (2007) *Mol. Cell* **26**, 381-392

12. Unoson, C., and Wagner, E. G. (2008) *Mol. Microbiol.* **70**, 258-270

13. Kawano, M., Aravind, L., and Storz, G. (2007) *Mol. Microbiol.* **64**, 738-754

14.Buts, L., Lah, J., Dao-Thi, M. H., Wyns, L., and Loris, R. (2005) *Trends Biochem. Sci.* **30**, 672-679

15. Rotem, E., Loinger, A., Ronin, I., Levin-Reisman, I., Gabay, C., Shoresh, N., Biham, O., and Balaban, N. Q. (2010) *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12541-12546

16. Amitai, S., Yassin, Y., and Engelberg-Kulka, H. (2004) *J. Bacteriol.* **186**, 8295-8300

17. Amitai, S., Kolodkin-Gal, I., Hananya-Meltabashi, M., Sacher, A., and Engelberg-Kulka, H. (2009) *PLoS Genet.* **5**, e1000390

18. Fozo, E. M., Makarova, K. S., Shabalina, S. A., Yutin, N., Koonin, E. V., and Storz, G. (2010) *Nucleic Acids Res.* **38**, 3743-3759

19. Fozo, E. M., Kawano, M., Fontaine, F., Kaya, Y., Mendieta, K. S., Jones, K. L., Ocampo, A., Rudd, K. E., and Storz, G. (2008) *Mol. Microbiol*. **70**, 1076-1093

20. Mok, W. W., Navani, N. K., Barker, C., Sawchyn, B. L., Gu, J., Pathania, R., Zhu, R. D., Brown, E. D., and Li, Y. (2009) *Chembiochem* **10**, 238-241

21. Hershberg, R., Altuvia, S., and Margalit, H. (2003) *Nucleic Acids Res.* **31**, 1813-1820

22. Lutz, R., and Bujard, H. (1997) *Nucleic Acids Res.* **25**, 1203-1210

23. Sreerama, N., and Woody, R. W. (2000) *Anal. Biochem.* **287**, 252-260

24. Epps, D. E., Wolfe, M. L., and Groppi, V. (1994) *Chem. Phys. Lipids* **69**, 137-150

25. Papenfort, K., Pfeiffer, V., Lucchini, S., Sonawane, A., Hinton, J. C., and Vogel, J. (2008) *Mol. Microbiol.* **68**, 890-906

26. Han, K., Kim, K. S., Bak, G., Park, H., and Lee, Y. (2010) *Nucleic Acids Res*. **38**, 5851-5866

27. Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001) *Genes Dev.* **15**, 1637-1651

28. Calamia, J., and Manoil, C. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4937-4941

29. Gurezka, R., and Langosch, D. (2001) *J. Biol. Chem.* **276**, 45580-45587

30. Krishnakumar, S. S., and London, E. (2007) *J. Mol. Biol.* **374**, 671-687

31. Boulaiz, H., Prados, J., Melguizo, C., Marchal, J. A., Carrillo, E., Peran, M., Rodriguez-Serrano, F., Martinez-Amat, A., Caba, O., Hita, F., Concha, A., and Aranega, A. (2008) *Br. J. Dermatol.* **159**, 370-378

32. Boulaiz, H., Prados, J., Melguizo, C., Garcia, A. M., Marchal, J. A., Ramos, J. L., Carrillo, E., Velez, C., and Aranega, A. (2003) *Br. J. Cancer* **89**, 192-198

33. Prados, J., Melguizo, C., Rama, A., Ortiz, R., Boulaiz, H., Rodriguez-Serrano, F., Caba, O., Rodriguez-Herva, J. J., Ramos, J. L., and Aranega, A. (2008) *Int. J. Oncol.* **33**, 121-127

34. Dorr, T., Vulic, M., and Lewis, K. (2010) *PLoS Biol.* **8**, e1000317

35. Lee, S., Hinz, A., Bauerle, E., Angermeyer, A., Juhaszova, K., Kaneko, Y., Singh, P. K., and Manoil, C. (2009) *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14570-14575

3.9. Supplementary Figures

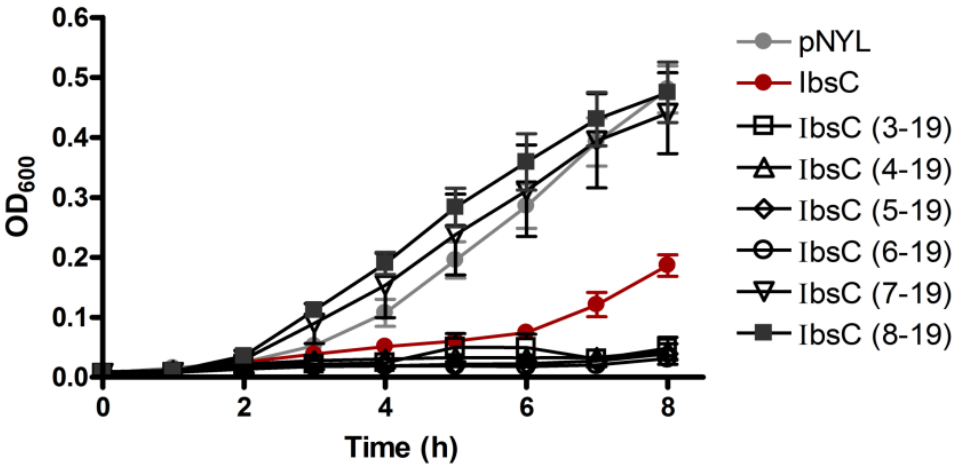


Figure S3-1. Effect of overexpressing IbsC 5' truncation mutants. Growth of *E. coli* DH5αZ1 expressing each 5' truncation mutant was evaluated over 8 hours (h) in triplicate. The optical density at 600 nm (OD_{600}) was measured hourly for each culture. Growth of cells producing these IbsC derivatives was compared to cells with the plasmid control (pNYL, indicated in gray) and those expressing IbsC (WT-IbsC, indicated in red). We observed that mutants lacking amino acids 2 to 5 remained growth suppressive.

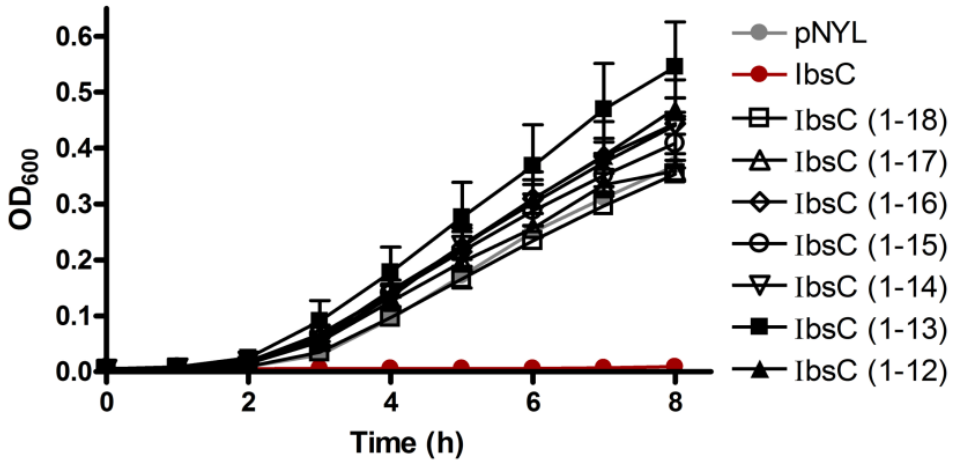


Figure S3-2. Effect of overexpressing IbsC 3' truncation mutants. The OD_{600} of *E. coli* expressing each 3' truncation mutant was measured hourly for 8 h in triplicate. Growth of bacteria overexpressing the truncation mutants were compared to the growth of the plasmid control (pNYL, indicated in gray) and to the growth of cells overexpressing IbsC (WT-IbsC, indicated in red). Removal of amino acids near the C-terminus of IbsC led to a loss of toxicity.

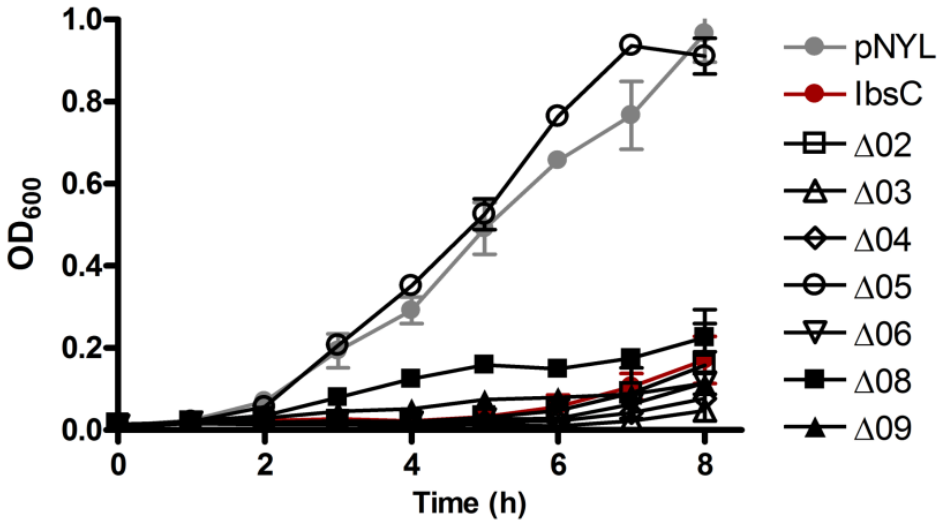


Figure S3-3. Toxicity of IbsC single amino acid deletion mutants. IbsC deletion mutants Δ02-Δ09, which lacked amino acids M2 to I9, were overexpressed in *E. coli*. Growth following overexpression was monitored over 8 h, and the OD_{600} of each sample was measured hourly. With the exception of construct Δ05, which lacked amino acid V5, overexpression of the deletion mutants was inhibitory to the growth of *E. coli*. As controls for this growth curve experiment, the growth of cells carrying pNYL-MCS11 (pNYL, indicated in gray) and pNYL-MCS11 with IbsC (WT-IbsC, indicated in red) was also monitored. These assays were repeated at least three times.

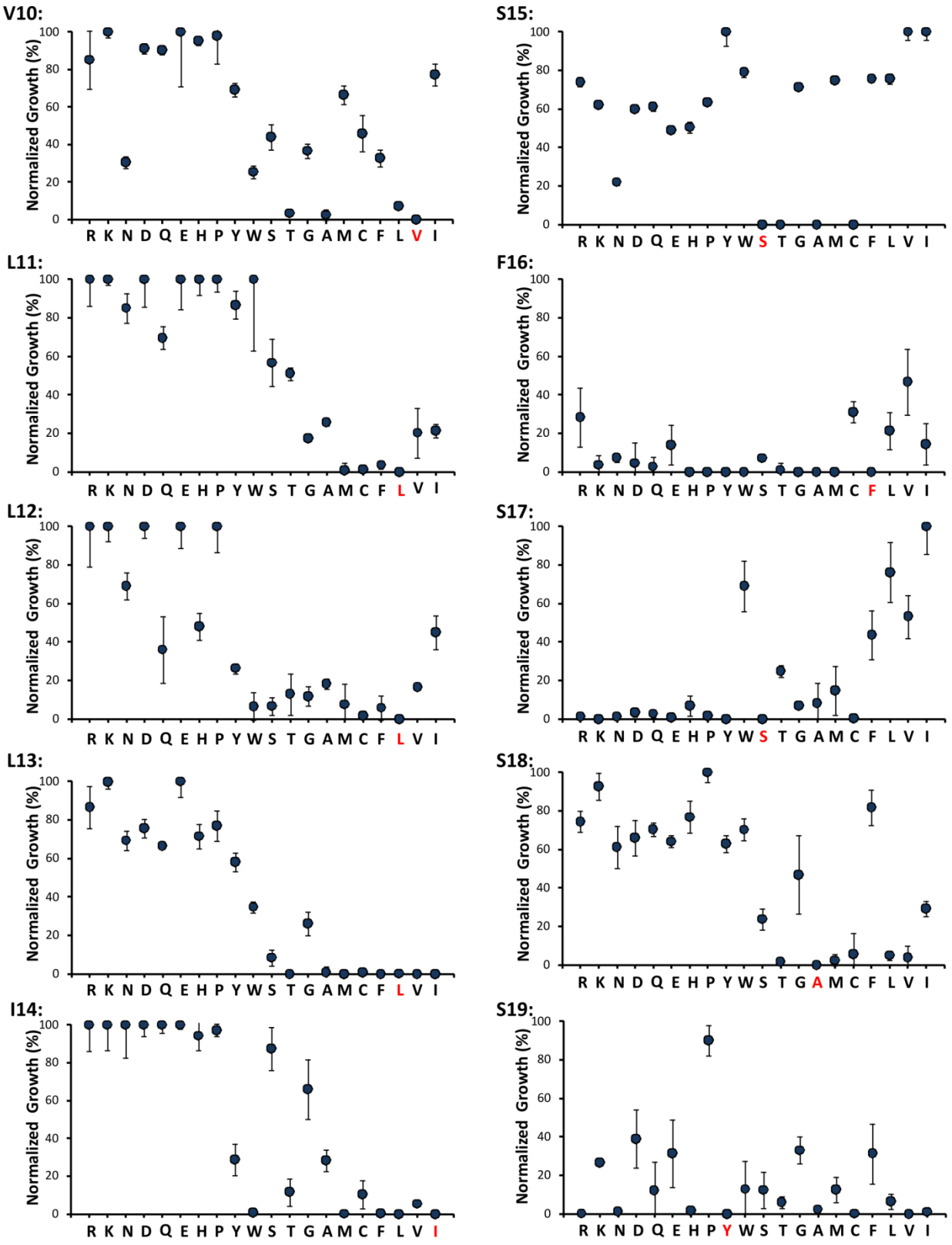

V10:
S15:
L11:
F16:
L12:
S17:
L13:
S18:
I14:
S19:
Normalized Growth (%)
0
20
40
60
80
100
R K N D Q E H P Y W S T G A M C F L V I

Figure S3-4. Toxicity of IbsC mutants with single amino acid substitutions. IbsC derivatives with the 20 possible amino acid substitutions were selected from each of the 10 IbsC random sequence libraries. The name of the library is indicated above the graph to which it corresponds. Growth defects associated with the overexpression of each mutant were evaluated 6 h after Atc induction. The OD_{600} of each sample was normalized against those of the plasmid control and the IbsC control. These values were referred to as "Normalized Growth" (%N.G.). The x-axis of the graphs shows the amino acid substitutions observed at the position of interest in each library. These amino acids are arranged in increasing hydrophobicity. %N.G. values are indicated on the y-axis. Average %N.G. calculated from 3 experiments are plotted on the graphs.

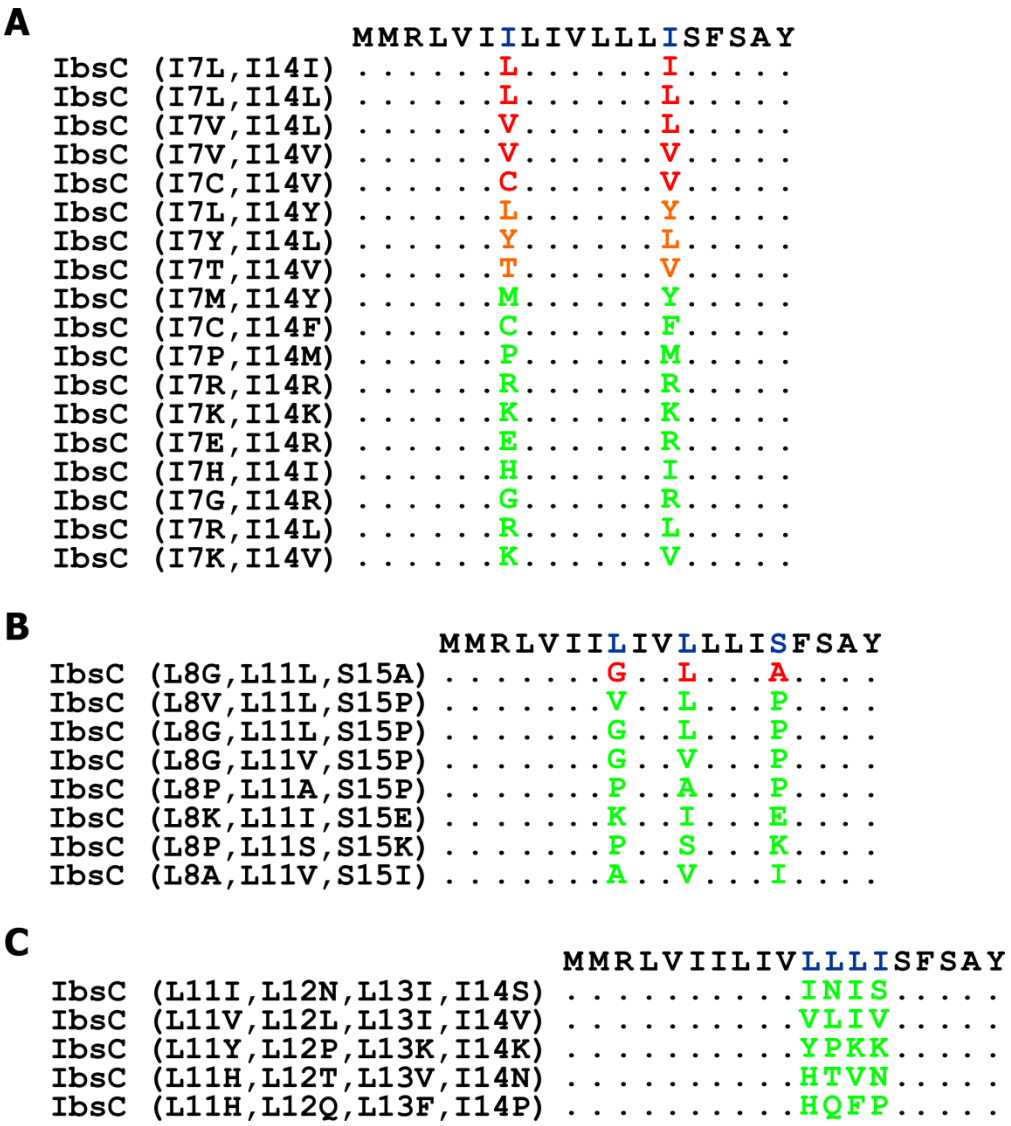
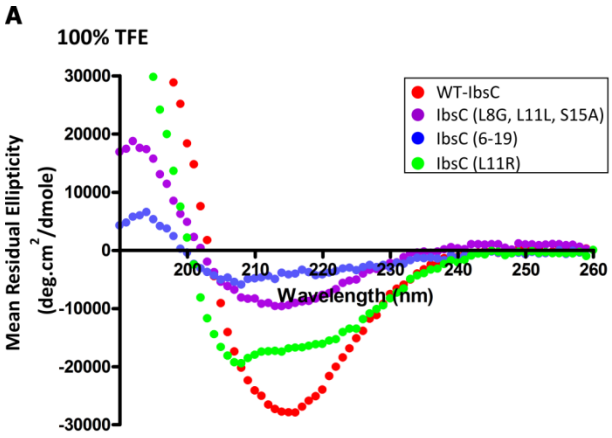
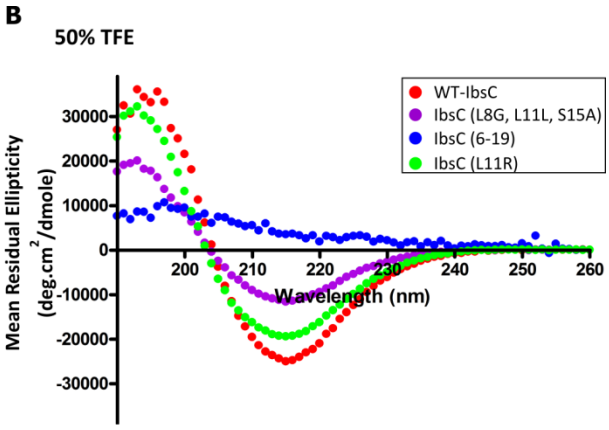


Figure S3-5. Toxicities of sequenced IbsC mutants with multiple amino acid substitutions. A. Library of IbsC double mutants in which codons encoding amino acids I7 and I14 (indicated in blue) were randomized using the sequence randomization approach. B. Library of IbsC triple mutants with randomized codons L8, L11, and S15. C. Library of IbsC quadruple mutants with randomized codons L11, L12, L13, and I14. The toxicities of constructs from each of the three combination mutant libraries were examined through growth assays. All active sequences whose overexpression corresponded to normalized growth (%N.G.) of ≤40% were sequenced. We also selected several inactive mutants (%N.G. of approximately 100%) for sequencing. Amino acid substitutions giving rise to toxic mutants (%N.G. <10%) are indicated in red. Substitutions giving rise to mutants with intermediate toxicities (%N.G. between 10-40%) are indicated in orange, while those that led to a loss of toxicity (%N.G. > 40%) are indicated in green.





C

SAMPLE	HELIX	STRAND	TURN	RANDOM COIL
WT	0.484	0.088	0.149	0.279
L8G,L11L,S15A	0.205	0.319	0.190	0.289
6-19	0.173	0.306	0.229	0.291
L11R	0.492	0.103	0.155	0.249

D

SAMPLE	HELIX	STRAND	TURN	RANDOM COIL
WT	0.397	0.048	0.200	0.354
L8G,L11L,S15A	0.159	0.370	0.210	0.262
6-19	0.005	0.494	0.237	0.264
L11R	0.377	0.216	0.167	0.240

Figure S3-6. Structural analysis IbsC and its derivatives by circular dichroism (CD) spectroscopy. CD spectra were obtained using 50 µM of peptides analysed in a 1-mm cell at 25 ^{o}C. The buffers are comprised of A. 100% 2,2,2-trifluoroethanol (TFE) or B. 50% TFE. Proportion of secondary structures of each peptide observed in C. 100% TFE and in D. 50% TFE as predicted using the CDPro software.

3.10. Supplementary tables

Table S3-1. Oligonucleotides used to generate *ibsC* and its derivatives

Name	Sequence (5'→ 3')	Purpose
IbsC-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGA	PCR of WT-IbsC; PCR of 3' truncation mutants
IbsC-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAG	PCR of WT-IbsC
IbsC (3-19)-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGCGACTTGTCATCATACTGATTGTACTGTT	5' truncation
IbsC (3-19)-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGAT	5' truncation
IbsC (4-19)-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGCTTGTCATCATACTGATTGTACTGTTAC	5' truncation
IbsC (4-19)-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATG	5' truncation
IbsC (5-19)-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGGTCATCATACTGATTGTACTGTTACT	5' truncation
IbsC (5-19)-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTAT	5' truncation
IbsC (6-19)-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATCATACTGATTGTACTGTTACTCA	5' truncation
IbsC (6-19)-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGT	5' truncation
IbsC (7-19)-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATACTGATTGTACTGTTACTCAT	5' truncation
IbsC (7-19)-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAG	5' truncation
IbsC (8-19)-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGCTGATTGTACTGTTACTCATAA	5' truncation
IbsC (8-19)-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGTACAATC	5' truncation
IbsC (1-18)-rev	TCAGCAG GGATCC TTAAGCGCTGAAACTTATGAGTAA	3' truncation
IbsC (1-17)-rev	TCAGCAG GGATCC TTAGCTGAAACTTATGAGTAACAG	3' truncation
IbsC (1-16)-rev	TCAGCAG GGATCC TTAGAAACTTATGAGTAACAGTAC	3' truncation
IbsC (1-15)-rev	TCAGCAG GGATCC TTAACTTATGAGTAACAGTACAAT	3' truncation
IbsC (1-14)-rev	TCAGCAG GGATCC TTATATGAGTAACAGTACAATCAG	3' truncation
IbsC (1-13)-rev	TCAGCAG GGATCC TTAGAGTAACAGTACAATCAGTAT	3' truncation
IbsC (1-12)-rev	TCAGCAG GGATCC TTATAACAGTACAATCAGTATGAT	3' truncation
IbsC Δ3-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCTTGTCATCATACTGATTGTACTGTTACTCATAAG	Deletion mutants
IbsC Δ3-7-	TCAGCAG GGATCC	Deletion

rev	TTAATAAGCGCTGAAACTTATGAGTAACAGTACAAT	mutants
IbsC Δ4-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGAGTCATCATACTGATTGTACTGTTAC TCATAAG	Deletion mutants
IbsC Δ5-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGACTTATCATACTGATTGTACTGTTAC TCATAAG	Deletion mutants
IbsC Δ6-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGACTTGTCATACTGATTGTACTGTTAC TCATAAG	Deletion mutants
IbsC Δ8-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGACTTGTCATCATAATTGTACTGTTAC TCATAAG	Deletion mutants
IbsC Δ9-19 fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGACTTGTCATCATACTG	Deletion mutants
IbsC Δ9-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGTACCAGTATGATGACA AGTCGCA	Deletion mutants
IbsC Δ10-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGAATCAGTATGATGACA AGTCGCA	Deletion mutants
IbsC Δ11-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTTATGAGTAATACAATCAGTATGATGACA AGTCGCA	Deletion mutants
IbsC Δ14-rev	TCAGCAG GGATCC TTAATAAGCGCTGAAACTGAGTAACAGTACAATCAGTATGATGACA AGTCGCA	Deletion mutants
IbsC Δ15-rev	TCAGCAG GGATCC TTAATAAGCGCTGAATATGAGTAACAGTACAATCAGTATGATGACA AGTCGCA	Deletion mutants
IbsC Δ16-rev	TCAGCAG GGATCC TTAATAAGCGCTACTTATGAGTAACAGTACAATCAGTATGATGACA AGTCGCA	Deletion mutants
IbsC Δ17-rev	TCAGCAG GGATCC TTAATAAGCGAAACTTATGAGTAACAGTACAATCAGTATGATGACA AGTCGCA	Deletion mutants
IbsC Δ18-rev	TCAGCAG GGATCC TTAATAGCTGAAACTTATGAGTAACAGTACAATCAGTATGATGACA AGTCGCA	Deletion mutants
IbsC-V5-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGACTTNNNATCATACTGATTGTACTG	Library
IbsC-V5-rev	CTGCTGA GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGAT	
IbsC-V10-Y19-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGACTTGTCATCATACTGATT	
IbsC-V10-rev	CTGCTGA GGATCC TTAATAAGCGCTGAAACTTATGAGTAACAGNNNAATCAGTATGATG ACAAG	
IbsC-L11-rev	CTGCTGA GGATCC TTAATAAGCGCTGAAACTTATGAGTAANNNTACAATCAGTATGATG ACAAG	

IbsC-L12-rev	CTGCTGA GGATCC TTAATAAGCGCTGAAACTTATGAGNNNCAGTACAATCAGTATGATG ACAAG
IbsC-L13-rev	CTGCTGA GGATCC TTAATAAGCGCTGAAACTTATNNNTAACAGTACAATCAGTATGATG ACAAG
IbsC-I14-rev	CTGCTGA GGATCC TTAATAAGCGCTGAAACTNNNGAGTAACAGTACAATCAGTATGATG ACAAG
IbsC-S15-rev	CTGCTGA GGATCC TTAATAAGCGCTGAANNNTATGAGTAACAGTACAATCAGTATGATG ACAAG
IbsC-F16-rev	CTGCTGA GGATCC TTAATAAGCGCTNNNACTTATGAGTAACAGTACAATCAGTATGATG ACAAG
IbsC-S17-rev	CTGCTGA GGATCC TTAATAAGCNNNGAAACTTATGAGTAACAGTACAATCAGTATGATG ACAAG
IbsC-A18-rev	CTGCTGA GGATCC TTAATANNNGCTGAAACTTATGAGTAACAGTACAATCAGTATGATG ACAAG
IbsC-Y19-rev	CTGCTGA GGATCC TTANNNAGCGCTGAAACTTATGAGTAACAGTACAATCAGTATGATG ACAAG
IbsC-I7I14-fwd	CTAGTAG GAATTC AGGAGAAGGGTTATGATGCGACTTGTCATCNNNCTGATTGTACTGT TACTC
IbsC-I7I14-rev	CTGCTGA GGATCC TTAATAAGCGCTGAAACTNNNGAGTAACAGTACAATCAG
IbsC-L8L11S15-fwd	**CTAGTAG GAATTC** AGGAGAAGGGTTATGATGCGACTTGTCATCATA
IbsC-L8L11S15-rev	**CTGCTGA GGATCC** TTAATAAGCGCTGAANNNTATGAGTAANNNTACAATNNN TATGATGACAAGTCGCAT
IbsC-11-14-fwd	**CTAGTAG GAATTC** AGGAGAAGGGTTATGATGCGACTTGTCATCATACTGATTGTA
IbsC-11-14-rev	**CTGCTGA GGATCC** TTAATAAGCGCTGAAACTNNNNNNNNNNNNTACAATCAGTATGAT GAC

- Underlined bases denote regions of complementarity between forward and reverse oligonucleotides that were used for Klenow reaction.
- "N" denotes nucleotides synthesized with 25% adenine, guanine, cytosine, and thymine phosphoramidites.

Table S3-2. *E. coli* strain used in this study.

Name	Description	Reference
DH5αZ1	Derivative of DH5αZ, *lacR-tetR::spec*^{R} inserted at the lambda attachment site *attB* between *bio* and *gal*	Lutz and Bujard (1997)

Table S3-3. Plasmid used to express *ibsC* and its derivatives in this study.

Name	Description	Reference
pNYL-MCS11	Derivative of pZE21-MCS1 (Lutz and Bujard, 1997), which contains a tetracycline-inducible promoter ($P_{LtetO-1}$), ColE1 origin of replication, and *kan*^{R}; ribosome binding site (RBS) removed	Lutz and Bujard (1997); Mok *et. al.* (2009)

Chapter 4:
Multiple factors contributing to the regulation of the IbsC toxin in *Escherichia coli*

4.1. Author's preface

In Chapter 3, we explored the sequence requirements for the toxicity of IbsC. In this chapter, we examine sequence elements involved in the regulation of *ibsC*. As the overexpression of *ibsC* is growth suppressive, its production is expected to be tightly repressed when the toxin is not required in the cell. We hypothesized that this strong repression of *ibsC* expression is achieved through multiple levels of regulation. To examine the transcription of the toxin gene and to probe for transcription factor binding sites, we fused sequences upstream and downstream of the *ibsC* core promoter (-35 to -10 region) to a reporter gene encoding the green fluorescent protein (GFP). Through fluorescence assays and immunoblotting experiments carried out with cells with these promoter-reporter constructs, we identified a putative inhibitory regulatory element around the -60 position of *ibsC*. Our data also suggest that the transcription of toxin is low in cells growing in nutrient-rich environments.

Once produced, *ibsC* transcripts are further repressed at the post-transcriptional level through their interactions with SibC. In a previous study, recognition and initial binding between the toxin and antitoxin transcripts were demonstrated to be mediated by two target recognition domains (TRDs). On *ibsC*, one of these TRDs is located near its translation initiation region, while the other is in its open reading frame (ORF). Upon mutating the two TRDs in SibC, we found that six nucleotides complementary to the TRD at the ORF are particularly important for *ibsC* repression, because mutations at these positions resulted in a loss of activity. Three of the nucleotides at this TRD are especially variable between *ibs/sib* homologs found in *E. coli* K-12. Thus our data suggest that these

nucleotides may be involved in ensuring that interactions only occur between cognate pairs of *ibs* and Sib RNAs.

The regulation of *ibsC* is multi-layered, and many details pertaining to this complex picture remain to be unveiled. The promoter-reporter constructs and SibC mutants engineered in this study can be applied to investigate these unresolved aspects of *ibsC* regulation. For example, the promoter-reporter constructs can be used to deduce the conditions that can trigger an upregulation of *ibsC* expression. Using the SibC mutagenesis data, mutations that are complementary to those observed in selected SibC mutants can be introduced into *ibsC*. Through growth recovery assays carried out with these SibC and *ibsC* mutants, we can probe into the evolution of this TA system and examine whether certain base-pairs are favoured in naturally occurring *ibs/sib* pairs. For the constructs described herein, I played the principle role in their design. I also conducted the experiments, analysed the data, and composed the manuscript. Dr. Li has provided helpful input and suggestions throughout this process. This chapter is modified from a manuscript that is being prepared for publication.

4.2. Abstract

The genomes of free-living bacteria harbour a collection of toxin-antitoxin (TA) pairs, where toxins have the potential to disrupt cellular processes and inhibit growth when their expression is induced. Under most growth conditions, the deleterious activities of these toxic proteins are repressed by specific protein or RNA antitoxins. The expression of some of these toxins has been shown to be beneficial for adaptation and survival in cells grown under certain conditions. However, toxin regulation and growth conditions that can trigger their expression remain a mystery. In this study, we examine

the factors that regulate the expression of a type I TA pair, IbsC/SibC, found in *Escherichia coli* K-12. We found that the promoter of *ibsC*, which encodes the toxin component of this TA pair, may contain regulatory sequences that contribute to the downregulation of this gene under nutrient rich conditions. Through mutagenesis studies carried out with the SibC antitoxin, we observed that six nucleotides that are complementary to two codons in the open reading frame of *ibsC* are especially important for the interactions between antitoxin and toxin transcripts. Overall, our findings suggest that the expression of *ibsC* is tightly regulated at the transcriptional and post-transcriptional levels.

4.3. Introduction

Recent surveys of microbial genomes have unveiled an abundance of two-component systems referred to as toxin-antitoxin (TA) systems (Fozo et al., 2010; Leplae et al., 2011). TA pairs contain a stable protein toxin, which can perturb target enzyme activity or cellular architecture when it is expressed, thereby eliciting cell death (Fozo, Hemm, et al., 2008; Yamaguchi & Inouye, 2009; Syed & Levesque, 2012). The expression or action of the toxin can be antagonized by its cognate antitoxin via several different mechanisms. As observed in type I TA pairs, non-coding antitoxin small RNAs (sRNAs) can bind toxin mRNAs in order to prevent toxin translation initiation (Fozo, Hemm, et al., 2008). Alternatively, protein-protein interactions between protein toxins and antitoxins (type II TA pairs) (Syed & Levesque, 2012) or protein-RNA interactions between protein toxins and RNA antitoxins (type III TA paris) (Fineran et al., 2009) have been observed. In all cases, antitoxin binding of the toxin sequesters it away from its cellular targets.

TA systems were initially identified on plasmids, where they contribute to plasmid maintenance or addiction as the loss of TA-encoding plasmids often resulted in post-segregational cell death (Ogura & Hiraga, 1983). By comparison, the function of chromosomally encoded TA genes and the rationale for maintaining these potentially deleterious elements in the genome is less clear. From TA systems that have been more extensively characterized thus far, many have been linked to functions that are beneficial for the survival and adaptation of bacteria, including the maintenance of genome integrity, stress response, and multidrug tolerance (Van Melderen & Saavedra De Bast, 2009; Wang & Wood, 2011). This suggests that TA elements may play physiologically important roles in the bacteria.

Compared with type II and III TA systems, the functions of type I TA pairs are less well studied. Among the type I TA pairs with yet to be determined functions are those belonging to the Ibs/Sib family, one of six type I TA families that have been annotated in *E. coli* K-12 thus far. Five copies of *ibs/sib* genes, coined *ibs/sibA-E* respectively, have been identified in *E. coli* K-12 (Fozo, Kawano, et al., 2008). Additional homologs have further been found in related Proteobacterial species (Fozo et al., 2010). In this TA system, cognate pairs of *sib* antitoxin genes and *ibs* toxin genes are encoded on opposite strands at the same locus. Thus the antitoxin sRNA and the toxin mRNA are near perfect reverse complements of each other, and the sRNA spans the entire length of the translation initiation region (TIR) and open reading frame (ORF) on the toxin transcript. While *ibs/sib* TA pairs identified in *E. coli* K-12 are highly homologous, each *ibs* mRNA, with the exception of *ibsA*, can only be repressed by their cognate *sib* sRNA (Fozo, Kawano, et al., 2008). In a study in which *ibs* and *sib* genes were expressed in

trans from two separate plasmids, cognate *ibs* and *sib* RNAs first recognize and interact with each other via two target recognition domains (TRDs), which are located in two highly variable loop regions in the five Sib/*ibs* sequences (Han et al., 2010). In *ibs* mRNAs, the principle TRD, coined TRD1, is found in their ORFs, while the secondary TRD, TRD2, is positioned near their TIRs. The initial interactions between Sib and *ibs* RNAs at these two sites allow them to form a transient kissing complex, which then promotes further base-pairing between the two RNA species. Ultimately, the complex is proposed to be targeted for degradation.

Although details pertaining to the post-transcriptional regulation of *ibs* transcripts by Sib antitoxins are beginning to emerge, the regulation of *ibs* genes at the transcriptional level remains obscure. It was previously reported that SibC expression is detected in cells growing in exponential and stationary phases regardless of whether they were cultured in rich or minimal media (Fozo, Kawano, et al., 2008). Under similar conditions, a low level of *ibsC* expression is observed only when *sibC* expression is repressed via promoter deletion. As elevated levels of IbsC can disrupt the integrity of the inner membrane and impair growth, it is expected that *ibsC* is strongly repressed by SibC when its production is not needed. Other factors may also contribute to its repression at transcriptional, post-transcriptional, and translational levels. Moreover, the growth conditions that can trigger toxin production and antitoxin repression are also unknown. The elucidation of such conditions along with potential regulatory factors governing *ibs* expression would provide insight into the biological relevance and function of this family of TA systems.

As a step toward solving the puzzle pertaining to the function and regulation of *ibsC*, we set out to identify potential sequence elements that contribute to toxin regulation at the transcriptional and post-transcriptional levels. Using a series of promoter-reporter gene fusion constructs, we pinpointed a possible negative regulatory region in the promoter of *ibsC* (P_{ibsC}). This suggests that the transcription of *ibsC* under nutrient rich conditions may be repressed by transcription factors binding to this sequence. To explore the sequence requirements for the regulatory elements involved in the interaction between *ibsC* and SibC, we introduced mutations to the TRDs of SibC. We found that TRD1 is the dominant regulatory sequence, as mutations introduced to this site resulted in a loss of repression. On the other hand, mutations in TRD2 did not have significant effects on antitoxin activity. Our data further demonstrated that mutations in the sequence around TRD1 disrupted antitoxin function, suggesting that the structure of the stem loop housing TRD1 is important for its activity.

4.4. Materials and methods

4.4.1. Oligonucleotides

All oligonucleotides used in this study are shown in Supplementary Table 1 and were chemically synthesized by Integrated DNA Technologies (Coralville, IA, USA). Oligonucleotides that were longer than 40 nucleotides (nt) were purified by 10% (8 M urea) denaturing polyacrylamide gel electrophoresis (PAGE) prior to use. For the synthesis of *sibC* mutants with randomized target recognition domains, each nucleotide in the randomized region was prepared using a mixture of 25% deoxyadenosine, deoxyguanosine, deoxycytidine, and deoxythymidine phosphoramidites.

4.4.2. Growth media, enzymes, and reagents

Bacteria were routinely cultured in Luria Burtani (LB) broth, and when indicated, 50 μg mL^{-1} of kanamycin or 25 μg mL^{-1} 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.

Kanamycin, chloramphenicol, Atc, and IPTG were purchased from Sigma-Aldrich (Oakville, ON, Canada). High Fidelity PCR Enzyme Mix, Klenow fragment, and T4 DNA ligase were purchased from MBI-Fermentas (Burlington, ON, Canada). Restriction enzymes were purchased from New England Biolabs (Pickering, ON, Canada).

4.4.3. Bacterial strains and plasmids

All molecular cloning procedures were carried out in *E. coli* DH5αZ1. *ibsC* placed under the control of a tetracycline inducible promoter (P_{LtetO1}) was introduced into the chromosome of *E. coli* DH5αZ1, producing *E. coli* DH5αZ1*ibsC*. *E. coli* DH5αZ1*ibsC* was subsequently used in SibC-*ibsC* interaction screens. *E. coli* MG1655 was used for fluorescence assays and immunoblotting experiments.

P_{ibsC} and derivatives along with *gfp* were cloned into pNYL-MCSII (Mok et al., 2009). pNYL-MCSII also served as the backbone for the *sibC* mutant constructs. However, the *sibC* mutants were cloned into a derivative of pNYL-MCSII where the kanamycin resistance marker was replaced with one that confers resistance toward chloramphenicol. This vector was coined pNYLcat. For the integration of P_{LtetO1}-*ibsC* into the chromosome of *E. coli* DH5αZ1, pBS-araBADflank*kan* and pKD46 were used (Campbell & Brown, 2002; Datsenko & Wanner, 2000).

4.4.4. Molecular cloning

The P_{ibsC}-*gfp* promoter-reporter constructs were produced as follows: P_{ibsC} were synthesized either by amplifying the promoter regions of *ibsC* by PCR using genomic DNA extracted from *E. coli* MG1655 as template or by extension of two overlapping oligonucleotides encoding the promoter using the Klenow reaction. The promoters were fused to *gfp* by crossover PCR. Purified inserts and pNYL-MCSII were digested using XhoI and BamHI, purified by agarose gel electrophoresis, and ligated following manufacturers' protocols. The introduction of P_{ibsC}-*gfp* constructs into pNYL-MCSII effectively replaces the tetracycline-inducible promoter (P_{LtetO1}) originally present on the plasmids. Ligation products were transformed into competent *E. coli* DH5αZ1 by electroporation and successful clones were confirmed by sequencing performed at Mobix Lab (McMaster University). Plasmids containing the correct inserts were subsequently transformed into *E. coli* MG1655.

As positive controls for promoter-reporter assays, pNYL-MCSII containing *gfp* under the control of a synthetic strong constitutive promoter, which we have coined P_{opt} for optimal promoter, was generated. The P_{opt} sequence was derived from the Anderson promoter collection (iGEM, Berkley, accession number BBa_J23100) and was synthesized by annealing two complementary oligonucleotides. Briefly, the oligonucleotides were combined at equimolar concentrations, heated at 90 ^{o}C for 1 min, and cooled 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, thereby replacing P_{LtetO1}. *gfp* was cloned downstream of this promoter at EcoRI and BamHI sites. To produce fusion promoters containing wild-type and mutant derivatives of

the -60 to -35 region of P_{ibsC} placed upstream of P_{opt}, two complementary oligonucleotides encoding the fusion promoters were annealed. In the "scrambled" mutants, the nucleotides corresponding to the -60 to -56 region of P_{ibsC} were exchanged with those located at positions -55 to -51, while nucleotides at positions -50 to -46 were interchanged with those at positions -45 to -41. In the $P_{ibsC(MUT)}$ and the $P_{opt(MUT)}$ mutants, the following point mutations were introduced: A-59G, T-57C, A-56T, A-51G, and G-48C. $P_{opt(SCRM)}$ and $P_{opt(MUT)}$ were then cloned into pNYL-MCSII at XhoI and EcoRI restriction sites akin to P_{opt}. The ligation products in all cases were transformed into electrocompetent *E. coli* DH5αZ1. Successful clones were confirmed by sequencing and were subsequently transformed into *E. coli* K-12.

For the SibC-*ibsC* interaction screens, *sibC* and its mutants were 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* and its derivatives were synthesized using Klenow extension of oligonucleotides with overlapping regions. P_{LlacO1} was subsequently fused to *sibC* and its mutants 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 *E. coli* DH5αZ1*ibsC*. P_{LlacO1}-*sibC*mutTRD1, P_{LlacO1}-*sibC*mutTRD2, and P_{LlacO1}-*sibC*mutTRD12 ligation products were directly transformed into *E. coli* DH5αZ1*ibsC by* electroporation.

4.4.5. Integration of ibsC into E. coli DH5αZ1

ibsC was integrated into the genome of *E. coli* DH5αZ1 following protocols derived from those described by Datsenko and Wanner (2000) and Campbell and Brown (2002) (Campbell & Brown, 2002; Datsenko & Wanner, 2000). P_{LtetO1} was synthesized by annealing of complementary oligonucleotides, while *ibsC* was synthesized by Klenow extension of overlapping oligonucleotides. The promoter was then fused to the toxin-encoding gene by crossover PCR. The products were then digested with PmeI and BamHI restriction enzymes and ligated into pBS-*araBAD*flankKan such that the arabinose inducible promoter originally present on the plasmid is replaced by P_{LtetO1} and *ibsC*. Ligation products were transformed into *E. coli* DH5αZ1 and successful clones were confirmed by sequencing. Using the plasmids as templates, the sequence encoding 500 bp from *polB*, the kanamycin resistance marker, P_{LtetO1}-*ibsC*, and 500 bp from *araC* was amplified by PCR. 50 ng of this 2.5 kb piece of linear DNA was transformed into *E. coli* DH5αZ1 harbouring pKD46, which encode phage λ Red recombinases. Transformants were then plated on LB agar supplemented with kanamycin and cultured at 37 ^{o}C in order to select for integrants with P_{LtetO1}-*ibsC* and the kanamycin resistance marker integrated at the *araBAD* locus as well as for the loss of the temperature sensitive plasmid. Successful integrants were confirmed by colony PCR in which the *polB*-kan^R-P_{LtetO1}-*ibsC*-*araC* fragments were amplified. The amplicons were further validated by sequencing.

4.4.6. Fluorescence assays

E. coli MG1655 harbouring plasmid-borne promoter-reporter constructs were propagated overnight in LB broth supplemented with kanamycin at 37 ^{o}C with shaking at 260 rpm. As negative and positive controls for the assay, cells carrying pNYL-MCSII and

P_{opt}-*gfp* in pNYL-MCSII were grown in the same manner. The bacteria were then diluted in fresh media (1:200 dilution) and grown for an additional 6 h. The optical density at 600 nm (OD_{600}) of each sample was measured using a VersaMax microplate reader (Molecular Devices, Sunnyvale, CA, USA). One millilitre of each culture was subsequently harvested, pelleted, and washed with 1× phosphate buffered saline (PBS). The final pellets were resuspended in 1 mL of 1× PBS, and 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 and were further normalized against the average fluorescence signals obtained from the negative pNYL (μ_n) and positive P_{opt} (μ_p) controls. These values were denoted as "relative fluorescence" (RF) and were calculated using the following equation:

$$\%R.F. = (\text{Growth-normalized fluorescence} - \mu_n / |\mu_p - \mu_n|) \times 100$$

These assays were repeated at least three times.

4.4.7. Western analyses

E. coli MG1655 carrying pNYL-MCSII, P_{opt}-*gfp* in pNYL-MCSII, or promoter-reporter constructs in pNYL-MCSII were cultured overnight in LB broth with kanamycin before being subcultured in fresh media (1:200 dilution). Cells were grown for 6 additional hours before 1 mL of each culture was harvested and pelleted. The pellets were resuspended in 100 µL of Tris-EDTA buffer and 100 µL of 2× loading buffer. Cells were subsequently lysed by heating at 90 ^{o}C for 20 min. Total protein from each sample was resolved on a 15% SDS-PAGE gel before being transferred to a 0.22 µm PVDF

membrane (Bio-Rad, Mississauga, ON, Canada) at 75 V for 1 to 1.5 h. Membranes were incubated in BLOTTO (5% skim milk powder in PBS with 0.1% Tween 20) at 4 ^{o}C overnight. Monoclonal anti-GFP antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA) diluted 1:1000 in BLOTTO was then added. As a loading control, the expression of GroEL was probed on a second membrane using monoclonal anti-GroEL antibody (Sigma-Aldrich), which was diluted 1: 100 000 in BLOTTO. After 1 h of incubation, the membranes were washed with PBS with 0.1% Tween 20 and with PBS. The membrane previously treated with anti-GFP antibody is then incubated with goat anti-mouse IgG-HRP antibody (Santa Cruz Biotechnology), which was diluted 1:5000 in PBS. The membrane treated with anti-GroEL antibody is incubated with goat anti-rabbit IgG-HRP antibody (Santa Cruz Biotechnology), which was also diluted 1:5000 in PBS. After 1 h, the membranes were washed as described and the blots were developed using the ECL Prime Western Blotting Detection Reagent (GE Healthcare, Baie d'Urfé, PQ, Canada) following manufacturer's protocols.

4.4.8. SibC-ibsC interaction assays

To examine the interactions between mutant SibC and the *ibsC* transcript, 300 colonies of *E. coli* DH5αZ1*ibsC* carrying pNYLcat with *sibC*mutTRD1, 100 colonies with *sibC*mutTRD2, and 100 colonies with *sibC*mutTRD12 were selected. The bacteria were cultured overnight in a 96-well plate in LB supplemented with chloramphenicol at 37 ^{o}C with shaking at 260 rpm. Cells were then subcultured (1:400 dilution) in LB-chloramphenicol in the absence of inducers, in the presence of IPTG, Atc, or both IPTG and Atc. As negative and positive controls for the assays, cells carrying pNYLcat and wild-type *sibC* in pNYLcat were also cultured. Cell density of each sample was measured

at 600 nm at 6 h post-induction using a VersaMax microplate reader. The assays were carried out in duplicate. OD_{600} measurements were normalized against the average OD_{600} corresponding to positive (μ_p) and negative (μ_n) controls. These values were denoted "Normalized Growth" (N.G.) and they were calculated using the equation:

$$\%N.G. = (OD_{600(mutant)} - \mu_n / |\mu_p - \mu_n|) \times 100$$

Following the assays, all of the constructs that were examined were sent for sequencing at BioBasic Inc. (Markham, ON, Canada).

4.5. Results

4.5.1. Transcriptional regulation of ibsC

It was previously reported that *ibsC* transcripts are not detected by northern analysis unless the promoter of *sibC* is deleted or mutated (Fozo, Kawano, et al., 2008). However, the regulation of *ibsC* at the transcriptional level remains enigmatic. To assess the level of *ibsC* expression under nutrient rich growth conditions and to identify possible regulatory elements governing the expression of this toxin, sequences upstream and downstream of the -35 to -10 region, or the core promoter, of *ibsC* were appended to reporter gene *gfp*. These promoter-reporter fusion products were coined P_{ibsC}-*gfp*. The promoter region incorporated in each construct is denoted in parentheses in their names. Using *E. coli* harbouring these promoter-reporter constructs, fluorescence assays were carried out to evaluate *gfp* expression. We first fused the -35 to +10 region of *ibsC* to *gfp* (Figure 4.1A). In this construct, a consensus ribosome binding site (RBS) was introduced between the promoter and *gfp* to enable translation. With $P_{ibsC(-35+10)}$-*gfp*, low levels of reporter gene expression was detected when compared to *gfp* expression regulated by a

strong constitutive promoter P_{opt} (Figure 4.1B). This was confirmed by western analysis in which GFP production can be observed upon analysing protein extracted from $P_{ibsC(-35+10)}$-*gfp*-containing cells. To examine the effects of sequences upstream of the core promoter, $P_{ibsC(-70+10)}$-*gfp* was constructed. The inclusion of these upstream nucleotides resulted in complete repression of *gfp* expression with fluorescence reduced to background levels, as seen with the pNYL control, which lacks *gfp*. The band corresponding to GFP was no longer detected by immunoblotting (Figure 4.1B). To determine whether sequences downstream of the transcription start site were needed for expression, the aforementioned clones were extended to include the entire 5'-untranslated region (5'-UTR) of *ibsC*, creating $P_{ibsC(-70+60)}$-*gfp* and $P_{ibsC(-35+60)}$-*gfp*. In these constructs, *gfp* is translated using the RBS of *ibsC* as the translation initiation sequence of *ibsC* is positioned directly upstream of the start codon of *gfp*. The fluorescence of $P_{ibsC(-70+60)}$-*gfp* remained close to background levels. On the other hand, a three-fold enhancement in fluorescence was observed with $P_{ibsC(-35+60)}$-*gfp* relative to $P_{ibsC(-35+10)}$-*gfp*. This suggests that the sequence encoding the 5'-UTR of *ibsC* is required to promote its expression and future *ibsC* constructs were designed to contain this region.

The reduced levels of *gfp* expression associated with the presence of the -70 to -35 sequence of *ibsC* points to possible negative regulatory elements upstream of the core promoter. To further characterize and pinpoint the exact location of these elements, we synthesized a set of P_{ibsC} derivatives by incorporating sequences of varying lengths upstream of the core promoter. We also expanded the P_{ibsC} sequence to encompass more upstream sequence in an effort to deduce distal regulatory sequences. One of these constructs, coined $P_{ibsC(IGR)}$-*gfp*, consists of the 168 nt intergenic region between *ibsC* and

its upstream gene, *serA* (Figure 4.1C). We further generated a series of truncations variants by removing ten nucleotides at a time starting from the -100 position of *ibsC*. These mutants are referred to as $P_{ibsC(-100)}$-*gfp* to $P_{ibsC(-40)}$-*gfp*. From fluorescence assays carried out with *E. coli* containing these constructs, we observed a stepwise decrease in *gfp* expression as upstream regions are added to $P_{ibsC(-35)}$-*gfp* up to the -60 position (Figure 4.1D; Supplementary Figure S4-1). As more nucleotides are added, we observed a gradual increase in *gfp* expression, with construct $P_{ibsC(-100)}$-*gfp* exhibiting about one third of the *gfp* expression level seen in the $P_{ibsC(-35+60)}$-*gfp* construct. This was confirmed by immunoblotting experiments, which showed a gradual decrease in GFP production as nucleotides were appended from -35 to -60, followed by a gradual increase as nucleotides were added from -70 to -100 (Figure 4.1D). Taken together, these observations suggest that there may be multiple regulatory elements upstream of the *ibsC* core promoter, with a negative element centred around -60 and possible positive regulatory elements further upstream.

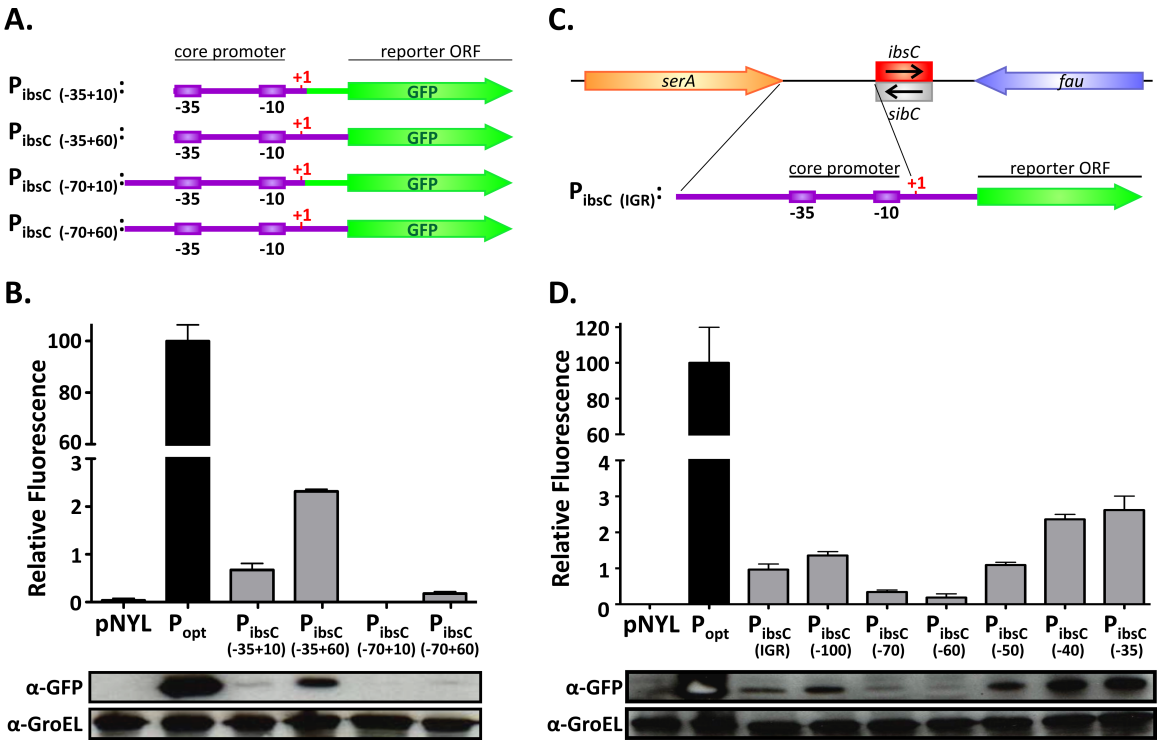


Figure 4.1. Delineating the promoter and regulatory regions of *ibsC*. A. Isolating the promoter of *ibsC*. The -35 to +10, -35 to +60, -70 to +10, and -70 to +60 regions of *ibsC* were fused to reporter gene *gfp*. These sequences were then cloned into pNYL-MCSII. The +1 to +60 sequence contains the entire 5'-UTR, including the RBS, of *ibsC*. In promoter-reporter constructs containing only the +1 to +10 region of *ibsC*, as in $P_{ibsC(-35+10)}$ and $P_{ibsC(-70+10)}$, a consensus RBS sequence (5'-AGGAGGTTCTTTA) was placed upstream of the ORF of *gfp*. B. Activities of putative *ibsC* promoters. Fluorescence assays were conducted with *E. coli* K-12 carrying the promoter-reporter constructs described in panel A of this figure in order to assess *gfp* expression. Cells carrying plasmids lacking *gfp* (pNYL) and plasmids with *gfp* placed under the control of P_{opt} (P_{opt}) served as negative and positive controls, respectively. The assay was conducted in triplicate. Total proteins isolated from these cells were further subjected to immunoblotting experiments in order to probe for GFP production using anti-GFP antibodies. The expression of GroEL is shown as a loading control. C. Design of *ibsC* promoter truncation variants. The P_{ibsC} truncation variants consist of different sequences upstream of the transcription start site of *ibsC* and its 5'-UTR, located from nucleotides +1 to +60. For example, construct $P_{ibsC(IGR)}$ contains the full IGR between *serA* and *ibsC*. The number of nucleotides upstream of the transcription start site of *ibsC* present in the other P_{ibsC} constructs are denoted by the numbers in brackets. These promoter sequences are fused to the ORF encoding GFP. D. Activities of P_{ibsC} truncation variants. Fluorescence emanating from *E. coli* K-12 carrying the promoter-reporter constructs described above were measured as described in panel A. Total protein isolated from these cells was subjected to immunoblot analysis to probe for GFP expression using anti-GFP antibodies. GroEL was detected as a loading control.

As the $P_{ibsC(-60)}$ construct dramatically reduced the expression from the *ibsC* core promoter, we next investigated the nature of this negative regulatory element to confirm that its down-regulatory effect was related to the sequence of this region and was not an artefact of cloning. Point mutations were first introduced at five randomly selected bases in this region, producing the $P_{ibsC(MUT)}$ mutant (Figure 4.2A). These mutations restored *gfp* expression to about one third of $P_{ibsC(-35)}$ levels (Figure 4.2B). More extensive alterations were then applied to this sequence, where the nucleotides in the -60 to -56 region were exchanged with those in the -55 to -51 region and nucleotides in the -50 to -46 region were exchanged with the -45 to -41 region. This "scrambled" mutant, coined $P_{ibsC(SCRM)}$, restored *gfp* expression to levels that were comparable to those observed when the reporter gene was placed under the control of $P_{ibsC(-35)}$. The presence of $P_{ibsC(SCRM)}$ and $P_{ibsC(-35)}$ upstream of *gfp* also produced similar amounts of GFP when probed by immunoblotting (Figure 4.2B).

To investigate whether the -60 to -36 region of P_{ibsC} was an independent regulatory element that could function in other sequence contexts, this region and its mutants were appended upstream of P_{opt} (Figure 4.2C). The presence of the wild-type -60 to -36 sequence resulted in a five-fold decrease in *gfp* expression relative to P_{opt} (Figure 4.2D). Similar to the results obtained with $P_{ibsC(MUT)}$ and $P_{ibsC(SCRM)}$, replacing wild-type -60 to -36 sequence with its point mutation and scrambled derivatives restored *gfp* expression to 40% and 80% relative to P_{opt}, respectively. These results show that the -60 to -36 sequence of P_{ibsC} can contribute to the downregulation of associated promoters, possibly through the binding of negative regulatory factors. Since the introduction of

point mutations failed to completely alleviate this inhibitory effect, the factor(s) interacting with this sequence may have loose sequence specificity for binding.

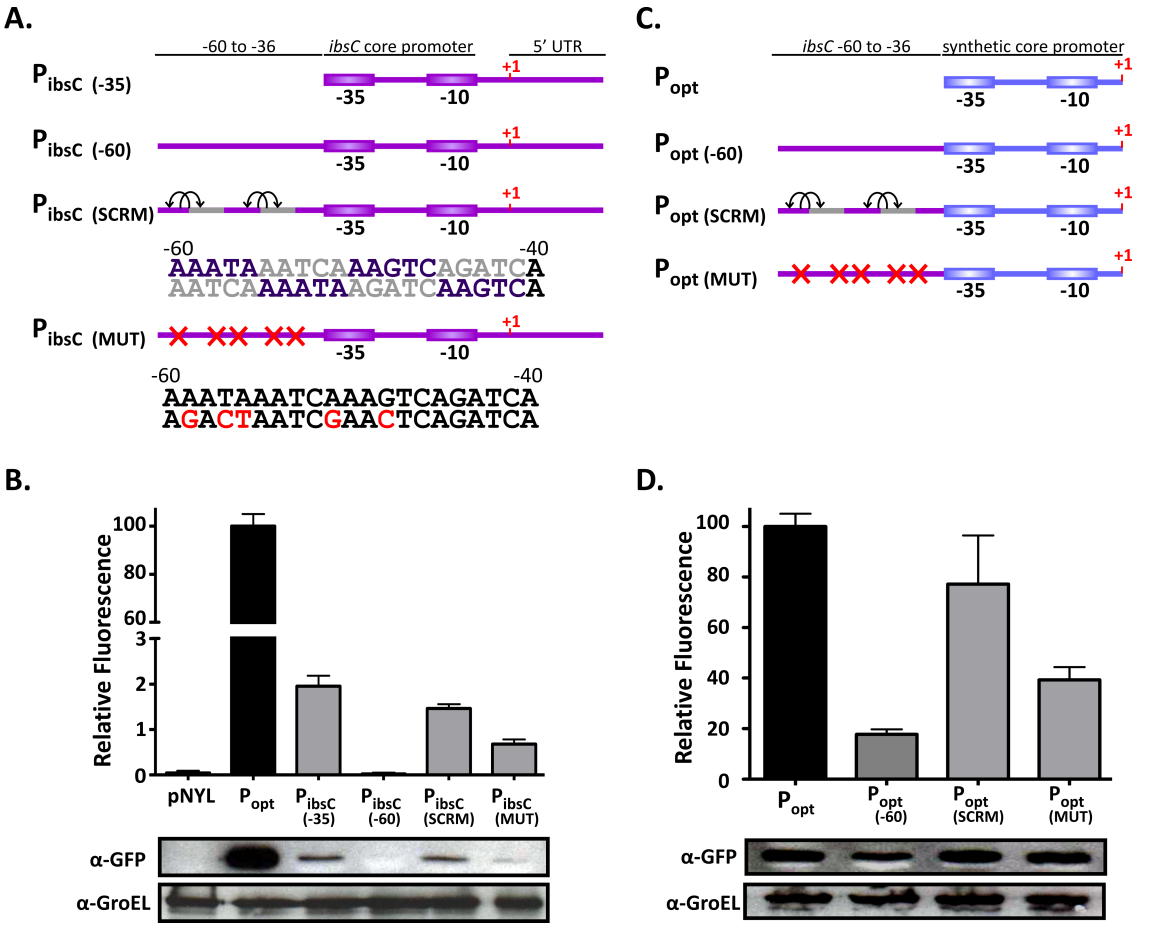


Figure 4.2. Isolating potential negative regulatory element in P_{ibsC}. A. Design of P_{ibsC} mutants. In mutant $P_{ibsC(SCRM)}$, nucleotides -60 to -56 were exchanged with nucleotide -55 to -51, while nucleotides -50 to -46 were exchanged with nucleotides -45 to -41. The wild-type -60 to -40 sequence of *ibsC* is shown below the diagram, and the mutant sequence is presented underneath it. In mutant $P_{ibsC(MUT)}$, five point mutations were introduced at positions -59, -57, -56, -51, and -48 as shown in red. These constructs were fused to *gfp*. B. Activities of P_{ibsC} mutants. The activities of $P_{ibsC(SCRM)}$ and $P_{ibsC(MUT)}$ were assessed with fluorescence assays carried out in *E. coli* K-12. Fluorescence signals from $P_{ibsC(-35)}$ and $P_{ibsC(-60)}$ were measured as comparisons. Cells carrying pNYL-MCSII and *gfp* placed under the control of P_{opt} served as negative and positive controls, respectively. The assay was completed in triplicate. The production of GFP in these cells was further probed by immunoblotting using anti-GFP antibodies. The expression of GroEL was probed as a loading control. C. Inhibitory effects of the -60 to -36 sequence of *ibsC* on P_{opt}. The wild-type -60 to -36 sequence of *ibsC* along with its scrambled and mutated derivatives were placed upstream of P_{opt}. These constructs were cloned upstream of *gfp*

whose translation is mediated by a consensus RBS sequence (5'-AGGAGGTTCTTTA). D. Activity of P_{ibsC}-P_{opt} chimeric promoters. The activities of P_{opt}, P_{opt} with the -60 to -36 sequence of *ibsC* ($P_{opt(-60)}$), $P_{opt(SCRM)}$, and $P_{opt(MUT)}$ were assessed with fluorescence assays and Western analyses as described in panel B.

4.5.2. Post-transcriptional regulation of ibsC

As shown in our promoter characterization and in studies by other laboratories, in the absence of external stresses, the transcription of *ibsC* appears to be low. A second level of regulation involves the sequestration of *ibsC* mRNAs that are still transcribed by SibC, which inhibits translation of *ibsC* mRNAs into IbsC toxins. It was previously reported that the initial contacts between cognate pairs of *ibs* and Sib RNAs is mediated by two TRDs, which act independently of each other. While the five *ibs/sib* homologs in *E. coli* K-12 share over 75% homology, these domains are situated in regions that are highly variable. To investigate whether certain sequence motifs are favoured in these domains and if SibC-*ibsC* binding requires perfect sequence complementarity in these regions, we examined the effects of mutations in the TRDs of SibC on its ability to repress *ibsC*.

We initially attempted to express *ibsC* and *sibC* from the same locus on pNYL-MCSII in order to maintain the *cis*-encoded orientation as observed in the *E. coli* genome. In this design (Supplementary Figure S4-2A), *ibsC* is placed under the control P_{LtetO1} on the sense strand, while *sibC* is placed under the control of P_{LlacO1} on the antisense strand. This construct was introduced into *E. coli* DH5αZ1, a derivative of *E. coli* K-12 that was previously engineered to express a Tet repressor and a Lac repressor constitutively (Lutz & Bujard, 1997), allowing for tight regulation of genes downstream of P_{LtetO1} and P_{LlacO1}. However, the expression of *sibC* from this system did not protect cells from the toxicity

of *ibsC* expression (Supplementary Figure S4-2B). These data suggest that an excess of SibC production relative to *ibsC* is required in order for the antitoxin to adequately repress the toxin. To increase the ratio of SibC to *ibsC* RNA, we designed a system in which P_{LtetO1}-*ibsC* is incorporated into the chromosome (Supplementary Figure S4-2C) and P_{LlacO1}-*sibC* is incorporated into a pNYL-MSCII vector. As multiple copies of pNYL-MCSII will be present in the cell compared to the single chromosomally encoded copy of *ibsC*, the system enables the production of an excess of SibC relative to *ibsC*. Inducing *ibsC* expression with Atc in this strain was found to cause cell death (Supplementary Figure S4-2D). This growth suppressive phenotype was no longer observed when *sibC* was induced by IPTG, showing that sufficient SibC RNA was being produced to inhibit *ibsC* toxicity. This system was used for subsequent assays.

The sequence specificity at TRD1 was examined by introducing random mutations at nucleotides 49 to 54 of *sibC*, which correspond to codons 16 and 17 in the ORF of *ibsC*. To disrupt the sequence of TRD2, nucleotides 107 to 113 of *sibC,* complementary to the region near the RBS in *ibsC*, were mutagenized. In the aforementioned study by Han and colleagues, the exchange of these elements from SibC with those from SibD resulted in a loss of interaction between the chimeric SibC and *ibsC*, while repression of *ibsD* was achieved (Han et al., 2010). We also assessed the impact associated with the concurrent mutation of both TRDs.

4.5.3. Sequence requirements at TRD1

We subjected 300 colonies of *E. coli* DH5αZ1*ibsC* carrying *sibC* with random mutations at TRD1 to growth assays (Figure 4.3A). Following the induction of *ibsC* and *sibC,* we observed that 12% of the sequences were active and able to restore growth to

50% above background (Figure 4.3B and Supplementary Figure S4-3A). The remaining 88% of the constructs failed to rescue cells from *ibsC* overexpression. The constructs were then sequenced (Supplementary Figure S4-3B). Upon excluding constructs that were not successfully sequenced, those that lacked a *sibC* mutant insert, and those that harboured extensive deletions or mutations outside TRD1, we observed that all of the active sequences corresponded to wild-type *sibC*. The dominance of the wild-type sequence at TRD1 in the active constructs isolated from this screen and the lack of active mutants, despite our extensive efforts to isolate them, indicate that this motif is optimal for mediating the interaction between SibC and *ibsC* and is not tolerant to mutations.

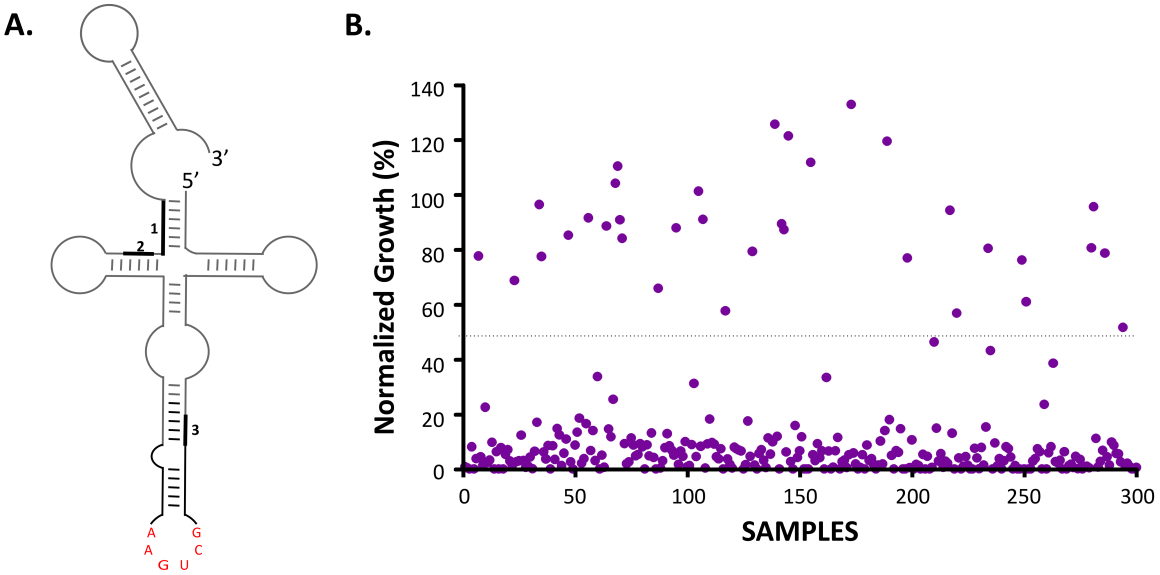


Figure 4.3. Mutagenesis of SibC TRD1. A. Location of TRD1 in SibC. The six nucleotides at TRD1 that were subjected to mutagenesis are highlighted in red. The sequences that are complementary to the RBS, the start codon, and stop codon on *ibsC* are indicated as 1, 2, and 3, respectively. B. Activity of SibC TRD1 mutants. 300 colonies of *E. coli* DH5αZ1*ibsC* carrying different *sibC* TRD1 mutants were cultured overnight and subjected to growth assays following the induction of *ibsC* with Atc and mutant *sibC* with IPTG. The growth of each sample was normalized by that of cells not expressing the antitoxin and those that were expressing wildtype *sibC*. The normalized growth values are depicted here. This assay was performed in duplicate. The cut-off for active mutants was set at 50% or above and N.G. of 50% is indicated by the dashed line on the graph.

4.5.4. TRD2 displayed low sequence specificity

We then mutated seven nucleotides in TRD2 (Figure 4.4A). In contrast to TRD1, the mutations introduced in TRD2 did not appear to perturb SibC activity. Of the 100 sequences that were examined, 95% of SibC mutants restored growth of *E. coli* to over 50% above background (Figure 4.4B; Supplementary Figure S4-4A). Sequence motifs observed at TRD2 of active SibC mutants are shown in Supplementary Figure S4-4B. Only two clones failed to rescue cells from *ibsC*, and another three were found to exhibit low activity, restoring growth to around 40% above background. Sequencing results obtained for the inactive sequences indicate that they did not contain a *sibC* insert. One of the constructs with low activity was observed to contain multiple mutations, while another was found to contain an A insertion between stems P4 and P5 of SibC. This single insertion might have disrupted the structure of stem P5, which in turn can perturb the loop with TRD1 (Figure 4.4C). Insertions and deletions in SibC were routinely observed in active sequences with lower activity (%N.G. <70%). The last mutant with low activity (%N.G.=40%) contained mutations $G_{107}CGCCCC_{113}$ at TRD2, which differed from the original $C_{107}UCCUUC_{113}$ by five bases. It is possible that this GC-rich sequence has perturbed the structure of SibC at TRD2.

Examining the active mutants of SibC obtained through this screen, dominant or recurring sequence motifs were not observed. Looking at the occurrence of individual nucleotides at each position of TRD2, guanine residues appeared to be strongly disfavoured at position 111 and its occurrence is generally low at positions 108, 109, and 112 (Figure 4.4D). Cytosine, which is dominant at TRD2 of SibC, is preferred at positions 109 to 112. On the other hand, we did not detect preference for uracil, the other

nucleotide present in TRD2. While TRD2 has been reported to be an independent element important for *ibsC* recognition, the lack of a definitive sequence motif at TRD2 observed in active SibC mutants isolated from this screen indicate that TRD2 is not essential for *ibsC* interactions and the function of TRD2 may be secondary to that of TRD1. As long as wild-type TRD1 is present, mutations at TRD2, regardless of extent, do not inhibit the ability of SibC to recognize and repress *ibsC*. Conversely, mutations at TRD1 result in a loss of SibC activity despite the presence of TRD2.

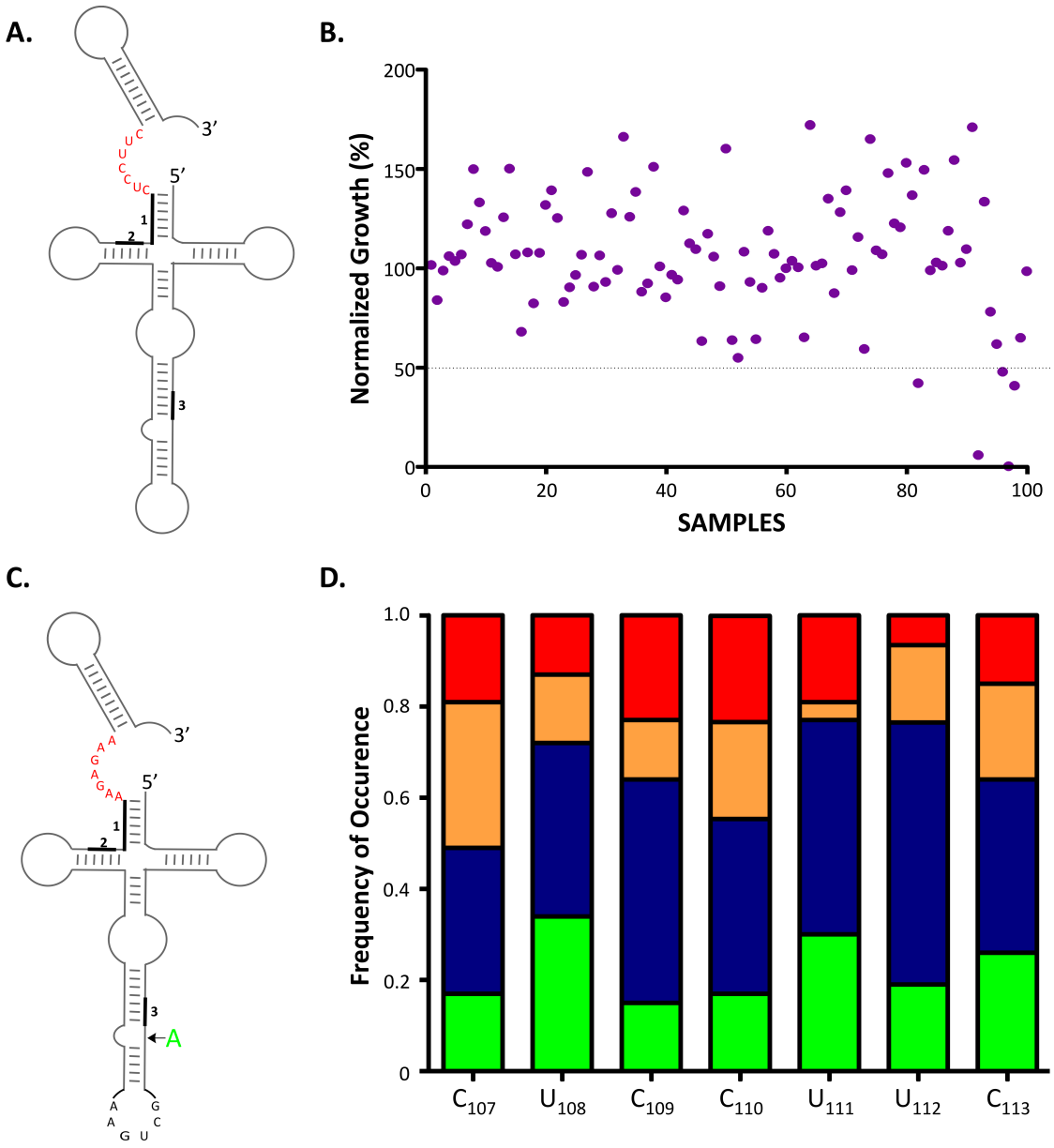


Figure 4.4. Mutagenesis of SibC TRD2. A. Location of TRD2 in SibC. The seven nucleotides at TRD2 that were subjected to mutagenesis are highlighted in red. The sequences that are complementary to the RBS, the start codon, and stop codon on *ibsC* are indicated as 1, 2, and 3, respectively. B. Activity of SibC TRD2 mutants. 100 colonies of *E. coli* DH5αZ1*ibsC* carrying different *sibC* TRD2 mutants were cultured in media containing Atc and IPTG in order to induce the expression of *ibsC* and *sibC*, respectively. The growth of each sample was assessed after 6 h by measuring their OD_{600}. These values were normalized by the growth of cells carrying pNYLcat (negative control) and the growth of cells expressing wild-type *sibC* (positive control). The normalized values are shown on this graph. The dashed line represents %N.G. of 50%, which was established as the cut-off for active mutants. This assay was conducted in duplicate. C. A partially active

SibC TRD2 mutant. One of the SibC mutants isolated from the screen described in Part B of this figure exhibited low activity despite carrying a wild-type TRD1 due to a single adenosine insertion in the stem adjacent to the loop housing TRD1 (indicated in green). D. Frequency of occurrence of each nucleotide in active TRD2 mutants. The frequency of occurrence of adenosine (in green), cytosine (in blue), guanine (in orange), and thymine (in red) observed at each of the seven nucleotides in TRD2 (indicated along the *x*-axis) are shown.

4.5.5. Mutating TRDs 1 and 2

When mutations were introduced at both TRD1 and TRD2 (Figure 4.5A), we did not isolate any active mutants after screening 100 constructs (Figure 4.5B; Supplementary Figure S4-5A). The majority of sequences evaluated did not exhibit activity and were associated with %N.G. of less than 10%. Upon examining complete sequences of these mutants that did not contain any nucleotide insertions or deletions, it was revealed that many were drastically different from wild-type *sibC* at TRD1 and TRD2 (Supplementary Figure S4-6B). We isolated five sequences that bore similarities to SibC from this screen. One of these mutants contained wild-type TRD1 and TRD2 sequences (Figure 4.5C). However, this inactive SibC mutant lacked the U_{44} residue in stem P5, suggesting that mutations in the stem-loop containing TRD1 may disrupt the structure of this motif and ultimately impairing its function. Three mutants with wild-type TRD2 but mutated TRD1 were isolated (Supplementary Figure S4-5B, first three sequences). Consistent with the data acquired from our TRD1 mutagenesis study, the presence of TRD2 alone is insufficient for SibC to antagonize *ibsC*. Furthermore, we identified one mutant that restored growth to 37.5% relative to SibC (Supplementary Figure S4-5B, fourth sequence). While all seven nucleotides at TRD2 of this mutant, with a sequence of $A_{107}GGACCU_{113}$, are different from that of SibC, TRD1 ($G_{49}CUGGG_{54}$) only differs by two nucleotides resulting from two transitions at A53G and A54G. Thus the partial

complementation between the TRD1 of this SibC derivative and that of *ibsC* may allow for weak associations between the toxin and mutant antitoxin and for the degradation of targeted toxin transcripts. Considering that this mutant is not likely able to bind with the *ibsC* mRNA as efficiently and as tightly as SibC, it was not capable of fully restoring the growth of *E. coli*. Results from this screen involving mutagenesis of both TRD1 and TRD2 further confirm our earlier observations regarding the importance of TRD1 in the ability of SibC to target the *ibsC* mRNA.

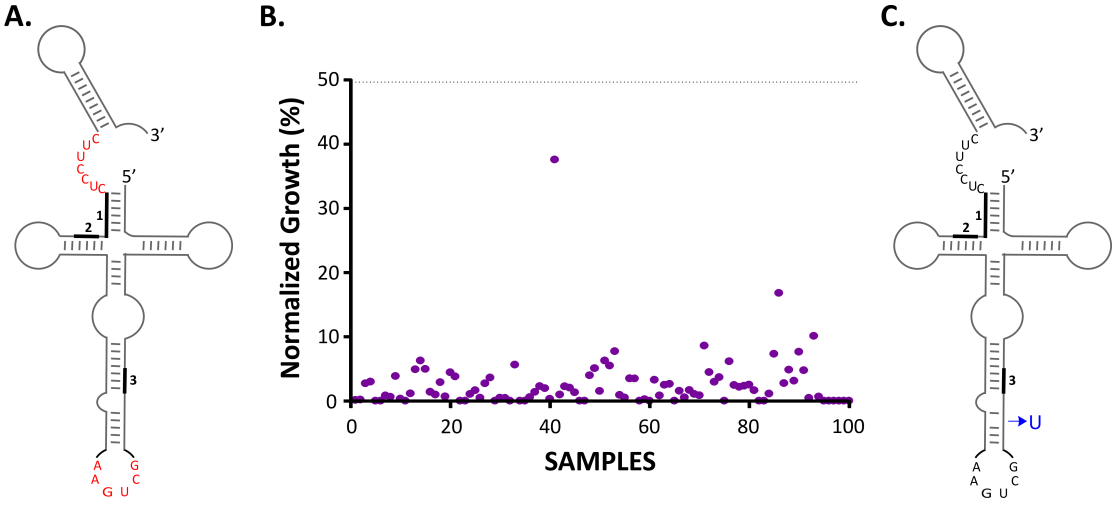


Figure 4.5. Mutagenesis of SibC TRD1 and TRD2. A. Location of TRDs 1 and 2 in SibC. The six nucleotides at TRD1 and the seven nucleotides at TRD2 that were subjected to mutagenesis are highlighted in red. The sequences that are complementary to the RBS, the start codon, and stop codon on *ibsC* are indicated as 1, 2, and 3, respectively. B. Activity of SibC TRDs 1 and 2 mutants. 100 colonies of *E. coli* DH5αZ1*ibsC* carrying different *sibC* TRDs 1 and 2 mutants were grown in the presence of Atc and IPTG to induce the expression of *ibsC* and mutant *sibC*, respectively. The growth of each sample was normalized by that of cells carrying pNYLcat (negative control) and those with P_{LlacO1}-*sibC* in pNYLcat (positive control). The normalized growth values are depicted here. This assay was performed in duplicate. The cut-off for active mutants was set at %N.G. of 50%, which is indicated by the dashed line on the graph. C. An inactive SibC mutant. A mutant with wild-type TRD1 and TRD2 sequences was isolated from the screen described in Part B of this figure. However, this mutant has a uridine deletion in the stem connected to the TRD1 loop, which may have perturbed the structure of the loop.

4.6. Discussion

In this study, we explored potential factors that may be involved in transcriptional and post-transcriptional regulation of the *ibsC* gene in *E. coli* K-12. As the production of this toxic peptide can greatly impact the physiology and survival of the bacteria, it is not surprising for its expression to be stringently governed at multiple levels. In examining the promoter of *ibsC,* the sequence around the -60 position of P_{ibsC} was inhibitory to the expression of downstream genes in cells grown under nutrient rich conditions. The repressive nature of this region was validated when it was appended upstream of a synthetic strong promoter, P_{opt}. Due to the proximity of this inhibitory sequence to the -35 element of P_{ibsC}, it is possible for this region to alter the binding of RNA polymerase holoenzymes to the promoter. Alternatively, it may serve as a binding site for transcription factors that downregulate the expression of *ibsC* when transcription of the toxin mRNA is not required. Regulatory factors binding to this region may be isolated via pull-down assays and confirmed by electromobility shift assays using the -70 to -50 portion of P_{ibsC} as bait. Our promoter analyses also revealed that the region encoding the 5'-UTR of *ibsC*, consisting of nucleotides +1 to +60, may be important for the expression of downstream transcripts, because little *gfp* expression was detected when the reporter gene was fused to the promoter construct consisting of only the +1 to +10 sequence downstream of the core promoter. Whether this region contains binding sites for protein factors that promote the transcription of *ibsC* or if the presence of the 5'-UTR improves the stability or translation initiation of associated transcripts remains to be investigated.

In establishing an assay platform for studying the interactions between SibC and *ibsC*, we observed that expression of inducible copies of toxin and antitoxin from the

same plasmid failed to rescue cells from the growth defects associated with toxin production. Toxin repression was achieved when *ibsC*, under the control of P_{LtetO1}, was expressed from the chromosome of *E. coli* K-12 and P_{LlacO1}-regulated *sibC* was expressed from pNYLcat, a high copy plasmid. Therefore, an excess of SibC over *ibsC* is necessary in order for the antitoxin to antagonize toxin expression. This is consistent with previous reports that for another type I TA pair, *symER,* the antitoxin and toxin transcripts are present in a ratio of approximately 10-to-1 (Kawano et al., 2007). Regulatory antisense RNAs have been demonstrated to interact stoichiometrically and non-catalytically (Georg & Hess, 2011). In incidences where the interactions lead to the degradation of mRNA targets, the sRNAs are expected to be degraded as well. Thus, maintaining an elevated level of antitoxin sRNA relative to toxin mRNA is beneficial to the cell, because it ensures that the toxin transcripts are efficiently and reliably repressed once toxin production is no longer needed.

The recognition and initial interactions between SibC and the *ibsC* mRNA is mediated through two main domains, TRD1 and TRD2, which act independently of each other (Han et al., 2010). The sequence of TRD1, which targets a loop region in the ORF of *ibsC*, is sufficient for *ibsC* binding. On the other hand, TRD2, which targets the translational initiation region of *ibsC*, was proposed to require a specific structure in addition to its recognition sequence for it to function. From our mutagenesis studies, however, we found that TRD1 alone was sufficient to mediate the interactions between *ibsC* and SibC, as extensive mutations can be introduced into TRD2 without compromising the activity of such SibC mutants. In contrast, mutations at the stem-loop of TRD1 produced inactive SibC mutants, despite the presence of wild-type TRD2 in

these derivatives. Thus, our data suggests that TRD1, which is located in a more accessible loop compared with TRD2, is the dominant sequence needed for SibC and *ibsC* targeting. Upon mutating TRD1, we repeatedly isolated active mutants with the wild-type sequence. As the libraries of mutant SibC sequences were constructed from synthetic oligonucleotides rather than site-directed mutagenesis of wild-type *sibC*, it is not possible for this observation to be a result of sequence contamination by sequences that failed to be mutated. This further suggests that wild-type TRD1 is optimal for *ibsC* targeting and binding.

Based on our findings, we hypothesize that the initial interactions between the toxin mRNA and the antitoxin sRNA at TRD1 may not block the translation initiation of *ibsC* if a high level of toxin transcript exist in the cell. Rather, they prevent the complete translation of the *ibsC* as the predominant mechanism of inhibition. In our previous study, we found that deletion of any of the amino acids from the C-terminus of IbsC rendered the peptide non-toxic (Mok et al., 2010). Thus the truncated IbsC variant that may arise from the aborted translation of its mRNA would no longer be able to permeate the inner membrane of *E. coli* and elicit cell death. The duplex formed near TRD1 following SibC-*ibsC* binding may further serve as a substrate for RNase III or other RNases, resulting in irreversible inhibition of *ibsC*. In mutating TRDs 1 and 2, we found that an inactive mutant with wild-type TRD1 and TRD2 sequences. This mutant, however, lacked nucleotide U_{44} at stem P5, which is adjacent to the loop containing TRD1. This single deletion might have distorted the structure of loop P5, thereby impairing the interactions between this SibC mutant and *ibsC*. This suggests that the structure of the stem loop

supporting TRD1 in SibC may be important for its function. Further work is needed to deduce sequence and structural motifs needed to mediate toxin-antitoxin interactions.

It has been proposed that the five *ibs/sib* homologs in *E. coli* K-12 may have evolved through duplication events, and mutations may have accumulated in each homolog over time. Based on the outcomes of our screens, mutations that preserved the structures of stem-loops containing the TRDs and promoted independent regulation of each *ibs/sib* pair while encoding a toxic Ibs peptide would be favoured through evolution. The five homologs may be maintained in the genome due to their divergent functions. They may each be responsive to a different environmental cue. Using the promoter-reporter constructs engineered in this study, we can begin to deduce the conditions that trigger IbsC production. Such findings can provide a better understanding of the biological function of this enigmatic TA pair.

It has been reported that weakening the inner membrane of bacteria through IbsC induction can enhance the sensitivity of bacteria toward certain classes of antibiotics, such as aminoglycosides (Lee et al., 2009). Thus, knowledge of when *ibsC* is expressed in the cell and factors that participate in its regulation can allow for the therapeutic potential of this TA pair to be explored and for agents that can act synergistically with existing antibiotics via the induction of *ibsC* to be discovered. Taken together, further studies into the regulation of this TA pair can yield insight into its biological relevance and into potential therapeutics that can be derived from IbsC/SibC.

4.7. Acknowledgements

We are also grateful to Li Lab members past and present, particularly Simon McManus, for helpful discussions. This work is supported by a research grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

4.8. References

Campbell, T. L., and E. D. Brown (2002). 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.

Datsenko, K. A., and B. L. Wanner (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640-6645.

Fineran, P. C., T. R. Blower, I. J. Foulds, D. P. Humphreys, K. S. Lilley, and G. P. Salmond (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A 106: 894-899.

Fozo, E. M., M. R. Hemm, and G. Storz (2008). Small toxic proteins and the antisense RNAs that repress them. Microbiol Mol Biol Rev 72: 579-589, Table of Contents.

Fozo, E. M., M. Kawano, F. Fontaine, Y. Kaya, K. S. Mendieta, K. L. Jones, A. Ocampo, K. E. Rudd, and G. Storz (2008). Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. Mol Microbiol 70: 1076-1093.

Fozo, E. M., K. S. Makarova, S. A. Shabalina, N. Yutin, E. V. Koonin, and G. Storz (2010). Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Res 38: 3743-3759.

Georg, J., and W. R. Hess (2011). cis-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev 75: 286-300.

Han, K., K. S. Kim, G. Bak, H. Park, and Y. Lee (2010). Recognition and discrimination of target mRNAs by Sib RNAs, a cis-encoded sRNA family. Nucleic Acids Res 38: 5851-5866.

Kawano, M., L. Aravind, and G. Storz (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Mol Microbiol 64: 738-754.

Lee, S., A. Hinz, E. Bauerle, A. Angermeyer, K. Juhaszova, Y. Kaneko, P. K. Singh, and C. Manoil (2009). Targeting a bacterial stress response to enhance antibiotic action. Proc Natl Acad Sci U S A 106: 14570-14575.

Leplae, R., D. Geeraerts, R. Hallez, J. Guglielmini, P. Dreze, and L. Van Melderen (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. Nucleic Acids Res 39: 5513-5525.

Lutz, R., and H. Bujard (1997). 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.

Mok, W. W., N. K. Navani, C. Barker, B. L. Sawchyn, J. Gu, R. Pathania, R. D. Zhu, E. D. Brown, and Y. Li (2009). Identification of a toxic peptide through bidirectional expression of small RNAs. Chembiochem 10: 238-241.

Mok, W. W., N. H. Patel, and Y. Li (2010). Decoding toxicity: deducing the sequence requirements of IbsC, a type I toxin in *Escherichia coli*. J Biol Chem 285: 41627-41636.

Ogura, T., and S. Hiraga (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. Proc Natl Acad Sci U S A 80: 4784-4788.

Syed, M. A., and C. M. Levesque (2012). Chromosomal bacterial type II toxin-antitoxin systems. Can J Microbiol 58: 553-562.

Van Melderen, L., and M. Saavedra De Bast (2009). Bacterial toxin-antitoxin systems: more than selfish entities? PLoS Genet 5: e1000437.

Wang, X., and T. K. Wood (2011). Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. Appl Environ Microbiol 77: 5577-5583.

Yamaguchi, Y., and M. Inouye (2009). mRNA interferases, sequence-specific endoribonucleases from the toxin-antitoxin systems. Prog Mol Biol Transl Sci 85: 467-500.

4.9. Supplementary figures

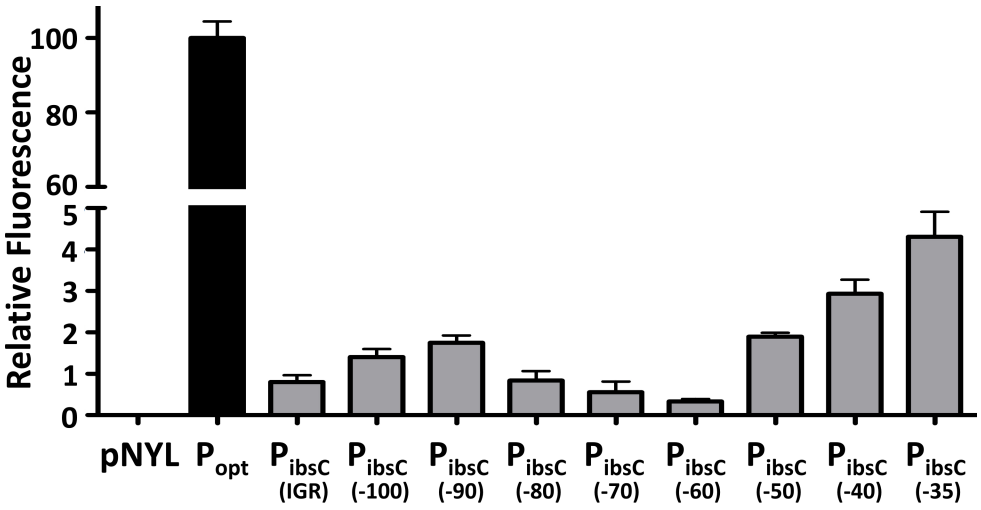


Figure S4-1. Activities of P_{ibsC} truncation variants. Construct $P_{ibsC(IGR)}$ consists of the 168 nt between *serA* and *ibsC*, as well as the 5'-UTR (from nucleotides +1 to +60) of *ibsC*. This sequence was fused to the ORF of *gfp*. In the subsequent constructs, each number in the brackets denotes the number of nucleotides upstream of the transcription start site of *ibsC* included in each sequence. All of these constructs also contain the 5'-UTR of *ibsC* and are cloned upstream of the ORF of *gfp*. Fluorescence assays were carried out with *E. coli* K-12 harbouring these constructs in triplicate. Bacteria carrying pNYL-MCSII lacking *gfp* (pNYL) and *gfp* placed under the regulation of a strong constitutive promoter (P_{opt}) served as negative and positive controls, respectively.

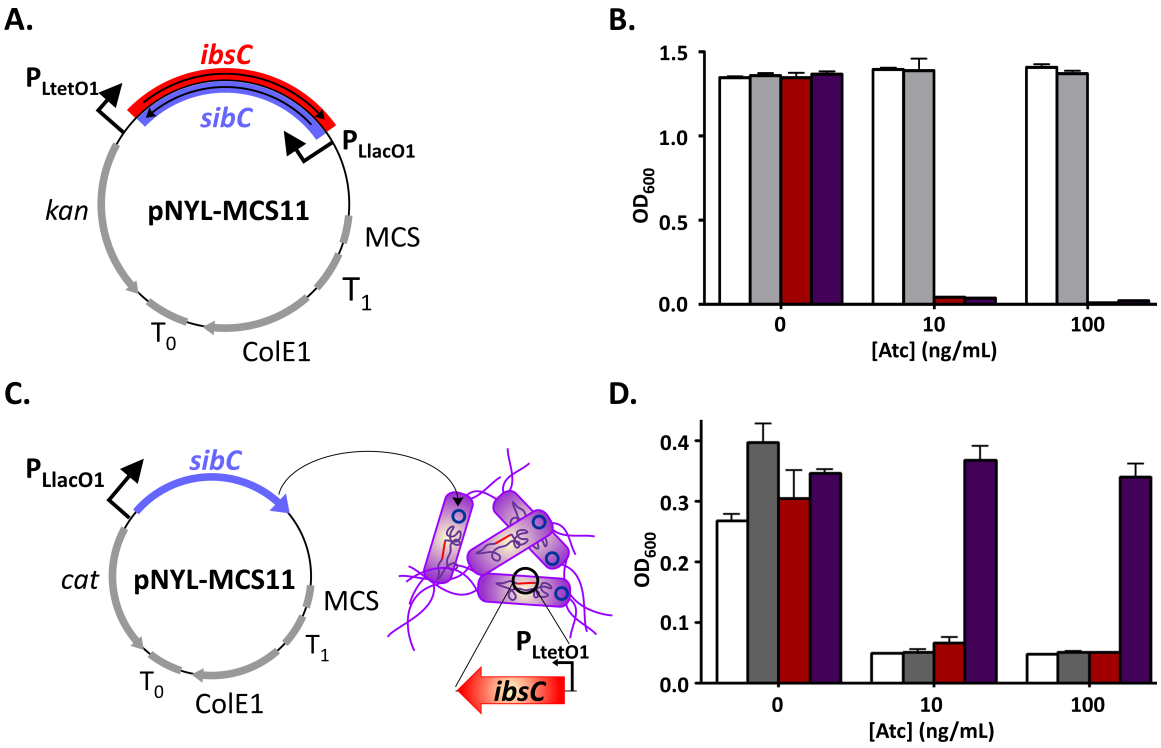
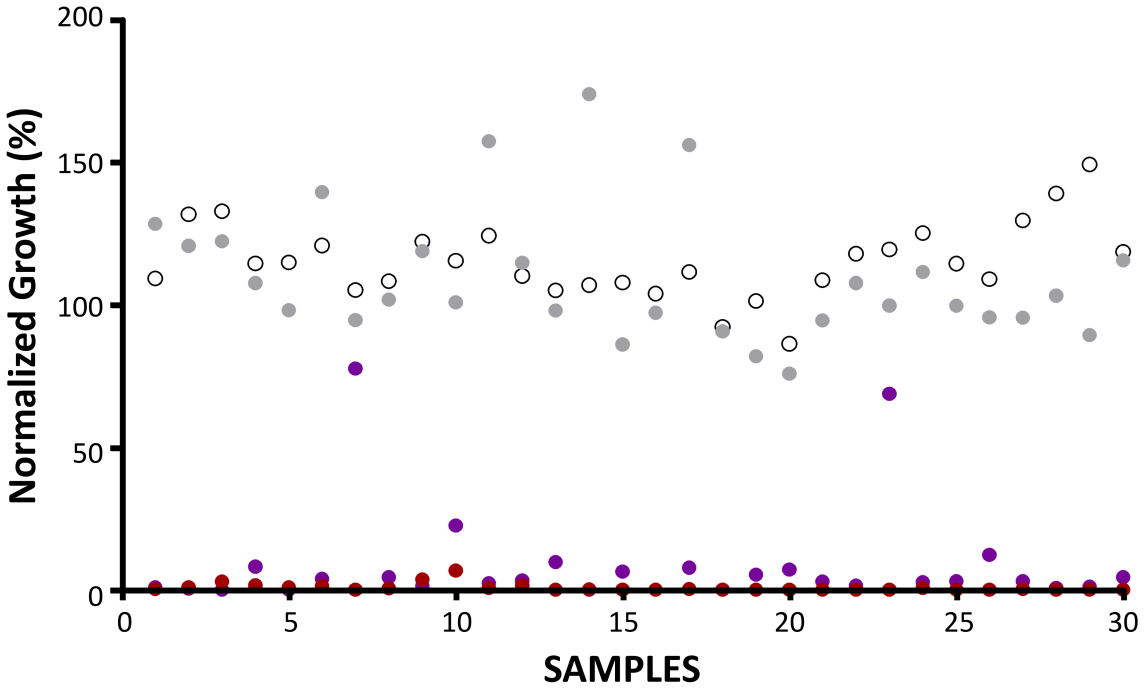


Figure S4-2. Design of system for *ibsC*-SibC interaction assays. A. Plasmid-based expression system. In this design, *ibsC* and *sibC* are found at the same locus in pNYL-MCSII. The expression of *ibsC*, which is encoded on the sense strand, is regulated by P_{LtetO1}, while *sibC*, encoded on the antisense strand, is regulated by P_{LlacO1}. B. Regulation of *ibsC* by SibC in the plasmid-based system. *E. coli* DH5αZ1 carrying the plasmid with *sibC* and *ibsC* were cultured in the absence of inducers (white bars) or in the presence of IPTG (grey bars), Atc (red bars), or IPTG and Atc (purple bars). Atc concentrations of 0 ng mL^{-1}, 10 ng mL^{-1}, and 100 ng mL^{-1} were used for these assays. It was observed that the induction of both *ibsC* and *sibC* (purple bars) failed to rescue the cells from the deleterious effects of the toxin. This assay was done in triplicate. C. Use of an *ibsC* knock-in strain. P_{LtetO1}-regulated *ibsC* was stably introduced into the chromosome of *E. coli* DH5αZ1, producing *E. coli* DH5αZ1*ibsC*. P_{LlacO1}-regulated *sibC* is supplied through a derivative of pNYL-MCSII. As the chromosome of *E. coli* DH5αZ1*ibsC* encodes a kanamycin resistance gene as a result of the integration, the kanamycin resistance gene on pNYL-MCSII has been replaced with one that confers resistance to chloramphenicol (*cat*). D. Repression of *ibsC* by SibC. *E. coli* DH5αZ1*ibsC* carrying the *sibC*-encoding plasmid was cultured in media as described in panel B. It was observed that the induction of *sibC* was able to rescue bacteria from *ibsC* expressed from the chromosome (purple bars).

A.

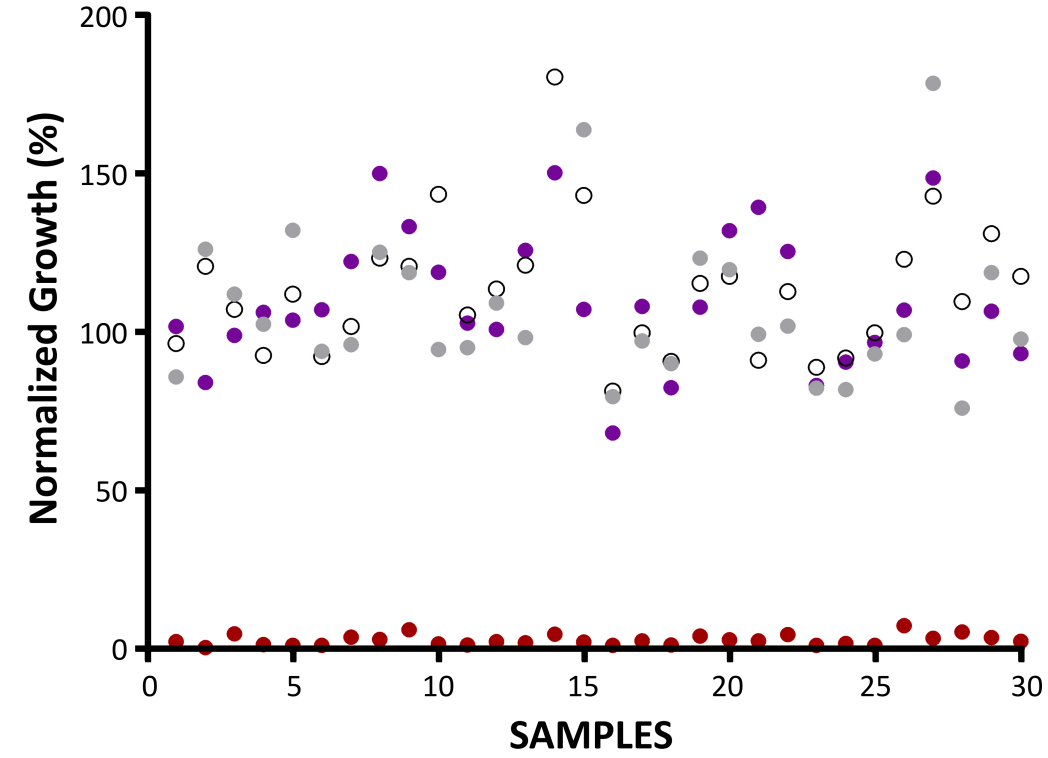


B.

G C U G A A	G C G A U G	G A A A U G	A G G C G G
A G U G A A	G C G C G G	G A C A C C	A G G C G U
	G A A U A C	G A G C U C	A G C A U U
G C G G G G	G A G G U U	G G G U G C	A U G A C U
G C G G C G	G A C U U A	A C C U C G	C A A C C U
G C G U A G	G G U A G U	A G A G G U	C G G A U G
G G U G G G	G G U C G G	A G G G U G	C U G U U U
G G U G G U	G G G G G U	A U G G C U	U G G C G C
G G G G A G	G G C C A G	C C C A G G	U U A A G G
G G G G G A	G U G G G G	C C G U U C	U U G G C U
G G A G A G	A A A G A G	C U G G G G	
G A C G A G	A C G U G A	U G A C G A	
G U G G G A	A G G G G A	U G C G G U	
A C U U U A	A G C G C A	U G C A A C	
A C A G A G	A G U G G G	U G C A G A	
A G G G A A	C C U A G G	U G G G G G	
C C U U A G	C G G A A A	U U C G C U	
C C G G G A	U G U C G A		

Figure S4-3. Activities and sequences of SibC TRD1 mutants. A. Normalized growth of cells expressing SibC variants with mutations at TRD1. 30 representatives of the 300 samples of *E. coli* DH5αZ1*ibsC* carrying plasmids encoding different SibC TRD1 mutants that were screened are shown in this graph. The cells were grown in the absence of inducers (white circles) or in the presence of IPTG (grey cirlces), Atc (red circles), or both IPTG and Atc (purple circles). The growth of each sample was normalized against that of cells not expressing an antitoxin and those that were expressing wild-type SibC, and the normalized growth values are plotted on this graph. This assay was done in duplicate. B. Sequences of SibC TRD1 mutants. The mutations observed at TRD1 in SibC mutants that were successfully sequenced are shown here. The wild-type sequence is shown in red. In each sequence, nucleotides that were not mutated are highlighted in green. All of the SibC mutants that are shown here did not rescue *E. coli* DH5αZ1*ibsC* from the toxicity of IbsC.

A.



B.

C U C C U U C

C U C U U A U
C U C A G C C
C A C C A G C
C A C C U C G
C G C C A C C

C U U C A G A
C U U A U C U
C C C C C G A
C C C U C C C
A C C C C C C
G A C C U C A
G C C C C C C
G C G C U C C
G G C C C A C
G G C C C C C
U G C C C C C
U G U C A U C

C A C G A A A
C A G C A C G
C A U G U C G
C C C G A C U
C C C G G C G
C C U U U A A
A A G U C U U
A C U C C C C
G A C A A A C
G A C A A U U
U A U C A C C
U C C U C G C
U U A A C G C

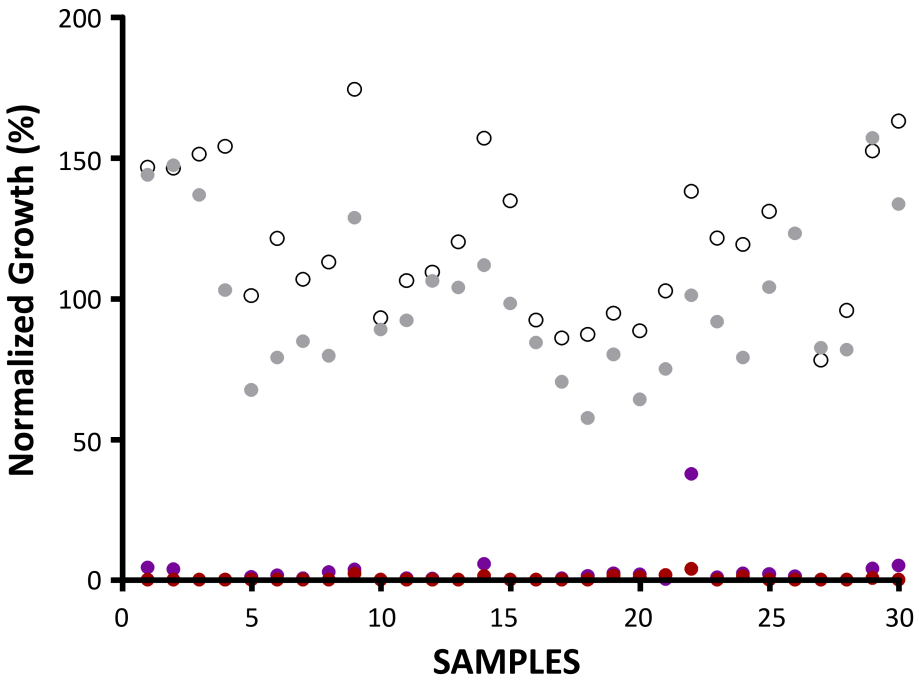
C G U A A C A
A C C G C G A
A U U G C C U
G A C G A C G
G C A C C A G

G C A U U G G
G G A G A A C
U C A U U C A

A A A A C A A
A A U G C C A
A C U U C C G
G A A U A A A
U A G U C G U
U C G A C C G
U C G U C C A

Figure S4-4. Activities and sequences of SibC TRD2 mutants. A. Normalized growth of cells expressing SibC variants with mutations at TRD2. 100 samples of *E. coli* DH5αZ1*ibsC* carrying plasmids encoding different SibC TRD2 mutants were grown in the absence of inducers (white circles) or in the presence of IPTG (grey cirlces), Atc (red circles), or both IPTG and Atc (purple circles). The growth of each sample was normalized against that of cells not expressing an antitoxin and those that were expressing wild-type SibC. Normalized growth values obtained from 30 samples examined through the screens are plotted on this graph. This assay was done in duplicate. B. Sequences of SibC TRD2 mutants. The mutations observed at TRD2 in SibC mutants that were successfully sequenced are shown here. The wild-type sequence is shown in red. All of the SibC mutants that are shown here were active and were able to rescue *E. coli* DH5αZ1*ibsC* from the toxicity of IbsC. In each sequence, nucleotides that were not mutated (the same as wildtype SibC) are highlighted in purple.

A.



B.

TRD1 (49-54)	TRD2 (107-113)
G C U G A A	C U C C U U C
C C U A G G	C U C C U U C
G C G A U G	C U C C U U C
C U G G G G	C U C C U U C
G C U G G G	A G G A C C U
U G U G G G	C U C U U U C
G G G A U C	C C C C U C C
G G G C G G	C U C G C U C
U G U C G A	C C C C C C C
G A G C A C	C C C G C G C
G A G G G G	C C C C C C G
G G A G C A	A C G C U C U
G G C A U G	A U C C C A C
G U C A U A	A C C G U U G
G U G G G G	C G C C A C A
U G C G A G	G G C C G C C
G C G G C U	G C G C A A C
G G A C G A	A C C G A G C
G G G A G C	G C C C C C C
G G G G C C	C C C U A C G
G G G G G G	G U C A G G A
G U U U U A	C A A A G G U

TRD1 (49-54)	TRD2 (107-113)
C G G G C G	C C C U C A C
C U U U G G	C C G C U C G
U U A U A U	G U C G G C C
G A A C G A	U A A C A G A
G A A G C G	U C A U C C C
G G G U G U	A U A G A C C
A A U C G G	G U C U A A G
A U A U A G	A U G G G G C
A U U G U U	G G C A G A U
C G U G C A	A A G A G C A
U A A C G G	A A C A U C C
G C C G G G	U C U A G A U
A G A G U A	A C G U A G U
A G G G G A	A A A G G G U
C G A G G G	A G C G C C G
C G G G G G	C G U U C A G
A A A U U C	U A A G C C C
A G G A G C	G C G A C C C
C A G U U G	G C C A U G U
U U G U A U	U C U A G A U

Figure S4-5. Activities and sequences of SibC mutants with mutations at TRDs 1 and 2. A. Normalized growth of cells expressing SibC variants with mutations at TRD1 and TRD2. Data from 30 of the 100 samples of *E. coli* DH5αZ1*ibsC* carrying plasmids encoding different SibC TRD1 and 2 mutants examined in our screen are shown. The bacteria were grown in the absence of inducers (white circles) or in the presence of IPTG (grey cirlces), Atc (red circles), or both IPTG and Atc (purple circles) in order to induce *ibsC* and/or *sibC* expression. The growth of each sample was normalized against that of cells not expressing an antitoxin and those that were expressing wild-type SibC, and the normalized growth values are plotted on this graph. This assay was done in duplicate. B. Sequences of SibC TRD1 and 2 mutants. The mutations observed at TRD1 and at TRD2 in SibC mutants that were successfully sequenced are shown here. The wild-type sequence is shown in red. All of the SibC mutants that are shown here were inactive. In each sequence, with the exception of sequence 4, nucleotides that were not mutated (the same as wildtype SibC) are highlighted in green. In sequence 4, which displayed partial activity (N.G. of 37.5%), the conserved nucleotides are highlighted in orange.

Table S4-1. Oligonucleotides used to generate P_{ibsC} derivatives and *sibC* mutants. Restriction sites present in each oligonucleotide are underlined. Nucleotides that were randomly mutated for the synthesis of SibC mutant TRD1, TRD2, and TRD1 and 2 libraries are denoted by "**N**".

Name	Sequence (-5'→ 3')	Purpose
Popt-AN- Fwd	TCATCATCTCGAGTTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCGAATTCCTAGTAG	Synthesis of P_{opt} by annealing
Popt- AN-Rev	CTACTAGGAATTCGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAACTCGAGATGATGA	Synthesis of P_{opt} by annealing
GFP1-PCR-Fwd	CGATGGCAGGGCAGCATGGGTAAGGAGGTTTCTTTAATG	PCR/ Cross-over PCR of *gfp* (-with $P_{ibsC(-35+10)}$ or $P_{ibsC(-70+10)}$)
GFP2-PCR-Fwd	GGGCTTGAAGGAGAAGGGTTATGGCTAGCAAAGGAGAAGA	PCR/ Cross-over PCR of *gfp* (-with all other PibsC constructs except for those mentioned above)
GFP-PCR-Rev	CTAGTAGGGATCCTTAGCAGCCCGATCC	PCR/ Cross-over PCR of *gfp*
$P_{ibsC(-35+10)}$-AN-Fwd	TTGTGTTTATTTAATGGGGATTTACGATGGCAGGGCAGCATGGGTAAGGAGGTTTCTTTAATG	Synthesis of $P_{ibsC(-35+10)}$ by annealing
$P_{ibsC(-35+10)}$-AN-Rev	CATTAAAGAAACCTCCTTACCCATGCTGCCCTGCCATCGTAAATCCCCATTAAATAAACACAA CTCGAGATGATGA	Synthesis of $P_{ibsC(-35+10)}$ by annealing
$P_{ibsC(-35+10)}$-XOPCR-Fwd	TCATCATCTCGAGTTGTGTTTATTTAATGGGGA	Cross-over PCR of $P_{ibsC(-35+10)}$ with *gfp*
$P_{ibsC(-70+10)}$-KL-Fwd	CTGCTCCAGAAAATAAATCAAAGTCAGATCAATGCGTTGTGTTTATTTAATGGGGATTT	Synthesis of $P_{ibsC(-70+10)}$ by Klenow
$P_{ibsC(-70+10)}$ -KL-Rev	CATTAAAGAAACCTCCTTACCCATGCTGCCCTGCCATCGTAAATCCCCATTAAATAAACACAACG	Synthesis of $P_{ibsC(-70+10)}$ by Klenow
$P_{ibsC(-70+10)}$-XOPCR-Fwd	TCATCATCTCGAGCTGCTCCAGAAAATAAATCA	Cross-over PCR of $P_{ibsC(-70+10)}$ with *gfp*
$P_{ibsC(IGR)}$-PCR-Fwd	CTAGTAGCTCGAGTTCCCCTTCTCTGAAAATCA	PCR of $P_{ibsC(IGR)}$
P_{ibsC}-PCR-Rev	TCTTCTCCTTTGCTAGCCATAACCCTTCTCCTTCAAGCCC	PCR of P_{ibsC} constructs
$P_{ibsC(-100)}$-	CTAGTAGCTCGAGTTGCCGCGTTGCGCATTCTA	PCR of $P_{ibsC(-}$

PCR-Fwd		$_{100)}$
$P_{ibsC(-90)}$ -PCR-Fwd	CTAGTAGCTCGAGGCGCATTCTACTTTGCGAGT	PCR of $P_{ibsC(-90)}$
$P_{ibsC(-80)}$ -PCR-Fwd	CTAGTAGCTCGAGCTTTGCGAGTCTGCTCCAGA	PCR of $P_{ibsC(-80)}$
$P_{ibsC(-70)}$-KL-Fwd	CTGCTCCAGAAAATAAATCAAAGTCAGATCAATGCGTTGTGTTTAT TTAATGGGGATTTACGA TGGCAGGGCAGCATGGGGC	Synthesis of $P_{ibsC(-70)}$ by Klenow
$P_{ibsC(-70)}$-KL-Rev	TCTTCTCCTTTGCTAGCCATAACCCTTCTCCTTCAAGCCCTCGCTTC GGTGAGGGCTTTACCG TTACAGCCCCATGCTGCCCTGCCAT	Synthesis of $P_{ibsC(-70)}$ by Klenow
$P_{ibsC(-70)}$-XOPCR- Fwd	TCATCATCTCGAGCTGCTCCAGAAAATAAATCA	Cross-over PCR of $P_{ibsC(-70)}$ with *gfp*
$P_{ibsC(-60)}$-KL-Fwd	AAATAAATCAAAGTCAGATCAATGCGTTGTGTTTATTTAATGGGG ATTTACGATGGCAGGGCA GCATGGGGCTGTAACGG	Synthesis of $P_{ibsC(-60)}$ by Klenow
$P_{ibsC(-60)}$-KL-Rev	TCTTCTCCTTTGCTAGCCATAACCCTTCTCCTTCAAGCCCTCGCTTC GGTGAGGGCTTTACCG TTACAGCCCCATGCTGC	Synthesis of $P_{ibsC(-60)}$ by Klenow
$P_{ibsC(-60)}$-XOPCR- Fwd	CTAGTAGCTCGAGAAATAAATCAAAGTCAGATC	Cross-over PCR of $P_{ibsC(-60)}$ with *gfp*
$P_{ibsC(-50)}$-KL-Fwd	AAGTCAGATCAATGCGTTGTGTTTATTTAATGGGGATTTACGATGG CAGGGCAGCATGGGGCT GTAACGGTAAAG	Synthesis of $P_{ibsC(-50)}$ by Klenow
$P_{ibsC(-50)}$-KL-Rev	TCTTCTCCTTTGCTAGCCATAACCCTTCTCCTTCAAGCCCTCGCTTC GGTGAGGGCTTTACCG TTACAGCCCCATGCTGC	Synthesis of $P_{ibsC(-50)}$ by Klenow
$P_{ibsC(-50)}$-XOPCR- Fwd	CTAGTAGCTCGAGAAGTCAGATCAATGCGTTGT	Cross-over PCR of $P_{ibsC(-50)}$ with *gfp*
$P_{ibsC(-40)}$-KL-Fwd	AATGCGTTGTGTTTATTTAATGGGGATTTACGATGGCAGGGCAGCA TGGGGCTGTAACGGTAAA GCCCTC	Synthesis of $P_{ibsC(-40)}$ by Klenow
$P_{ibsC(-40)}$-KL-Rev	TCTTCTCCTTTGCTAGCCATAACCCTTCTCCTTCAAGCCCTCGCTTC GGTGAGGGCTTTACCGT TACAGCCCCATGCTGC	Synthesis of $P_{ibsC(-40)}$ by Klenow
$P_{ibsC(-40)}$-XOPCR- Fwd	CTAGTAGCTCGAGAATGCGTTGTGTTTATTTAA	Cross-over PCR of $P_{ibsC(-40)}$ with *gfp*
$P_{ibsC(-35)}$-KL-Fwd	TTGTGTTTATTTAATGGGGATTTACGATGGCAGGGCAGCATGGGGC TGTAACGGTAAAGCCCTC	Synthesis of $P_{ibsC(-35)}$ by Klenow
$P_{ibsC(-35)}$-KL-Rev	TCTTCTCCTTTGCTAGCCATAACCCTTCTCCTTCAAGCCCTCGCTTC GGTGAGGGCTTTACCGT TACAGC	Synthesis of $P_{ibsC(-35)}$ by Klenow
$P_{ibsC(-35)}$-XOPCR- Fwd	TCATCACTCGAG TTGTGTTTATTTAATGGGGA	Cross-over PCR of $P_{ibsC(-35)}$ with *gfp*
$P_{ibsC(SCRM)}$-KL-Fwd	CTAGTAG**CTCGAG**AATCAAAATAAGATCAAGTCAATGCGTTGTGT TTATTTAATGGGGATTTACGAT	Synthesis of $P_{ibsC(SCRM)}$ by Klenow

$P_{ibsC(MUT)}$-KL-Fwd	CTAGTAG **CTCGAG**A**G**A**CT**AATC**G**AA**C**TCAGATCAATGCGTTGTGTTTATTTA ATGGGGATTTACGAT	Synthesis of $P_{ibsC(MUT)}$ by Klenow
$P_{opt(-60)}$-AN-Fwd	**CTAGTAG**CTCGAGAAATAAATCAAAGTCAGATCAATGCGTTGACG GCTAGCTCAGTCCTAGGTACAGTGCTAGCGAATTC**GCATCAT**	Synthesis of $P_{opt(-60)}$ by annealing
$P_{opt(-60)}$-AN-Rev	ATGATGCGAATTCGCTAGCACTGTACCTAGGACTGAGCTAGCCGT CAACGCATTGATCTGACTTTGATTTATTTCTCGAGCTACTAG	Synthesis of $P_{opt(-60)}$ by annealing
$P_{opt(SCRM)}$-AN-Fwd	CTCGAGAATCAAAATAAGATCAAGTCAATGCGTTGACGGCTAGCT CAGTCCTAGGTACAGTGCTAGCGAATTC	Synthesis of $P_{opt(SCRM)}$ by annealing
$P_{opt(SCRM)}$-AN-Rev	GAATTCGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAACGCA TTGACTTGATCTTATTTTGATTCTCGAG	Synthesis of $P_{opt(SCRM)}$ by annealing
$P_{opt(MUT)}$-AN-Fwd	CTCGAGA**G**A**CT**AATC**G**AA**C**TCAGATCAATGCGTTGACGGCTAGCT CAGTCCTAGGTACAGTGCTAGCGAATTC	Synthesis of $P_{opt(MUT)}$ by annealing
$P_{opt(MUT)}$-AN-Rev	GAATTCGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAACGCA TTGATCTGAGTTCGATTAGTCTCTCGAG	Synthesis of $P_{opt(MUT)}$ by annealing
P_{LtetO1}-*ibsC*-Fwd	CTAGTAGGGATCCTCCCTATCAGTGATAGAGATTGACATCCCTATC AGTGATAGAGATACTGAGCACA	PCR of P_{LtetO1}-*ibsC*
P_{LtetO1}-*ibsC*-Rev	ATGATGAGTTTAAACAAGGGTAAGGGAGGATTGCT	PCR of P_{LtetO1}-*ibsC*
P_{LtetO1}-*ibsC*-KI-Fwd	CGACTTCAGACGGGCATTAACGATAGTG	PCR of linear DNA for knock-in of *ibsC*
P_{LtetO1}-*ibsC*-KI-Rev	ACCGCGAATGGTGAGATTGAGAATATAACC	PCR of linear DNA for knock-in of *ibsC*
P_{LlacO1}-AN-Fwd	ATAAATGTGAGCGGATAACATTGACATTGTGAGCGGATAACAAGA TACTGAGCACAAGGGTAAGGGAGGATTGCTCCTCCCCTG	Synthesis of P_{LlacO1} by annealing
P_{LlacO1}-AN-Rev	CAGGGGAGGAGCAATCCTCCCTTACCCTTGTGCTCAGTATCTTGTT ATCCGCTCACAATGTCAATGTTATCCGCTCACATTTAT	Synthesis of P_{LlacO1} by annealing
sibC-KL-Fwd	AGGGTAAGGGAGGATTGCTCCTCCCCTGAGACTGACTGTTAATAA GCGCTGAAACTTATGAGTAACAGTACAATCAGTAT	Synthesis of *sibC* by Klenow
sibC-KL-Rev	GGTAAAGCCCTCACCGAAGCGAGGGCTTGAAGGAGAAGGGTTATG ATGCGACTTGTCATCATACTGATTGTACTGTTACT	Synthesis of *sibC* by Klenow
sibCμTRD1-KL-Fwd	AGGGTAAGGGAGGATTGCTCCTCCCCTGAGACTGACTGTTAATAA GC**NNNNNN**ACTTATGAGTAACAGTACAATCAGTAT	Synthesis of *sibCμTRD1* mutants by Klenow
sibCμTRD1-KL-Rev	GGTAAAGCCCTCACCGAAGCGAGGGCTTGAAGGAGAAGGGTTATG ATGCGACTTGTCATCATACTGATTGTACTGTTACT	Synthesis of *sibCμTRD1 mutants* by Klenow

sibCμTRD2-KL-Fwd	AGGGTAAGGGAGGATTGCTCCTCCCCTGAGACTGACTGTTAATAA GCGCTGAAACTTATGAGTAACAGTACAATCAGTAT	Synthesis of *sibCμTRD2* mutants by Klenow
sibCμTRD2-KL-Rev	GGTAAAGCCCTCACCGAAGCGAGGGCTT**NNNNNNN**AAGGGTTATG ATGCGACTTGTCATCATACTGATTGTACTGTTACT	Synthesis of *sibCμTRD2 mutants* by Klenow
sibCμTRD1&2 -KL-Fwd	AGGGTAAGGGAGGATTGCTCCTCCCCTGAGACTGACTGTTAATAA GC**NNNNNN**ACTTATGAGTAACAGTACAATCAGTAT	Synthesis of *sibCμTRD1&2* mutants by Klenow
sibCμTRD1&2 -KL-Rev	GGTAAAGCCCTCACCGAAGCGAGGGCTT**NNNNNNN**AAGGGTTATG ATGCGACTTGTCATCATACTGATTGTACTGTTACT	Synthesis of *sibCμTRD1&2 mutants* by Klenow
P_{LlacO1}-*sibC*-XOPCR-Fwd	CTAGTAGCTCGAGATAAATGTGAGCGGATAACA	Cross-over PCR of P_{LlacO1} and *sibC* or *sibC* mutants
P_{LlacO1}-*sibC*-XOPCR-Rev	ATGATGAGGATCCGGTAAAGCCCTCACCGAAGC	Cross-over PCR of P_{LlacO1} and *sibC* or *sibC* mutants

Chapter 5:
Using a type I toxin
to improve the efficiency and selectivity of molecular cloning strategies

5.1. Author's preface

In Chapter 3 of this dissertation, our work toward elucidating the sequence specificities and mechanism of toxicity of IbsC was showcased. Here, we used the sequence information amassed from that study to engineer molecular cloning vectors containing *ibsC* or an active *ibsC* mutant, which were incorporated to reduce the background frequently observed with current cloning methods. Since its conception, molecular cloning has revolutionized the field of biology, opening the floodgate to a wide array of topics that range from the study of individual genes to systems biology. However, its success is often hampered by high background, attributed especially to the uptake of undigested plasmids by transformants. Thus, cloning can be a costly and time-consuming endeavour.

Using IbsC for positive selection, transformants that have taken up plasmids that were not properly digested and consequently have retained *ibsC* will be killed as a result of toxin production. Thus, transformants that are able to form colonies are expected to contain the insert of interest. We have cloned reporter genes using this system and achieved over 95% positive clones. This is comparable to the efficiency achieved with commercially available cloning kits. The IbsC-based technique shown in this study does not require the use of proprietary enzymes. The cloning vector can be easily propagated in a specific *E. coli* strain, while cloning may be carried out in any strain of interest. Therefore, the use of this cloning system is considerably less expensive compared with other commercial systems. As purification steps are not required between digestion and ligation, cloning can be completed in less than three hours. This timescale is either comparable or shorter than that of various cloning kits. Considering that *ibsC* and its

variants are relatively short, they can be easily transferred into other cloning vectors and used for positive selection with other cloning systems, whether they are ligation-dependent or ligation-free. Thus we have developed an efficient, economic, and versatile cloning system using IbsC and its mutants.

This chapter is modified from a manuscript that is being prepared for publication. I took a leading role in designing the vectors and the experiments conducted to examine their selectivity. I also performed the experiments described herein, analysed the data, and composed the manuscript. Dr. Li has provided valuable advice and input throughout the process.

5.2. Abstract

We engineered a series of cloning vectors that can greatly minimize the background observed in conventional ligase-based molecular cloning in order to improve the success of this method. Variants of IbsC, a type I toxin derived from *Escherichia coli* K-12, were incorporated into these vectors. The lethal phenotype associated with toxin production is lost once transformants take up recombinant plasmids containing the desired insert, replacing *ibsC* or its variant. Using this positive-selection system, purification steps between digestion and ligation are no longer necessary. Once a PCR product is obtained, total time between digestion and plating of transformants can be reduced to as little as three hours. As demonstrated with the cloning of reporter genes, this system consistently produced over 95% positive clones. Thus, these IbsC-based cloning vectors are as reliable as commercially available cloning systems, yet they have the advantage of being more time-efficient and cost-effective.

5.3. Introduction

Molecular cloning is a long established and frequently used technique in biology and biotechnology. The use of this versatile method can account for much progress in biology, enabling studies that range from the characterization of countless genes and gene products (Schall et al., 1990; Wank et al., 1992; Dewhirst et al., 2010) to the assembly of synthetic organisms (Gibson et al., 2010). As interests in novel organisms expand and as fields such as synthetic biology emerge, there is an increasing demand for more efficient technologies to clone genes, assemble recombinant DNAs, and build genomic libraries. In addition to reducing the time and labour required to clone desired genes, new strategies should be aimed at minimizing the background that plague current cloning experiments. It is estimated that less than 10% of vectors circularize with the target DNA insert using conventional molecular cloning approaches, which inevitably increases the amount of work and resources required to screen and sift through a pool of "empty" vectors in order to isolate a single successful clone (Stieber et al., 2008).

Over the past two decades, a number of cloning techniques designed to improve the efficacy of the procedure have been developed (Bernard et al., 1994; Bernard, 1995; Szpirer & Milinkovitch, 2005; Mandi et al., 2009). These include PCR-based methods (Shuldiner et al., 1990; Bryksin & Matsumura, 2010), T-vector cloning (Marchuk et al., 1991), and recombinase-dependent cloning (Walhout et al., 2000). Commercially available systems that have incorporated bacteria toxin-antitoxin (TA) pairs to facilitate the selection of positive clones are also at our disposal. A prominent example of this is the GATEWAY cloning system, which contains the type II bacterial toxin-antitoxin (TA) pair CcdAB (Walhout et al., 2000). While these cloning kits are effective in reducing the

background observed with conventional cloning and some systems even circumvent the use of antibiotics for selection and plasmid stabilization altogether (Szpirer & Milinkovitch, 2005), they carry some constraints. The main disadvantage of these systems is their cost, especially due to the requirement of specific bacterial strains, proprietary enzymes, and the need to stimulate toxin production with inducers. Many of these systems are still time-consuming, taking at least a day to produce a clone containing an insert of interest. Herein, we have developed positive selection vectors using the IbsC toxin found in *E. coli* that are inexpensive and flexible to use.

The IbsC toxin was first identified in *E. coli* MG1655 and has since been annotated in other *E. coli* strains and in other proteobacterial species (Fozo et al., 2008; Fozo et al., 2010). The translation of the toxin mRNA was repressed by a noncoding small RNA (sRNA) antitoxin, referred to as SibC, encoded at the same locus but on the opposite strand (Fozo et al., 2008; Han et al., 2010). Due to this sRNA-mRNA mode of toxin-antitoxin interaction, IbsC-SibC is classified as a type I TA pair (Van Melderen, 2010). This differs from type II TA systems, such as CcdAB, where a protein toxin is repressed by a protein antitoxin (Van Melderen, 2010). The ectopic overexpression of *ibsC* was deleterious to *E. coli* (Fozo et al., 2008; Mok et al., 2009). This highly hydrophobic 19-amino acid peptide acts on the inner membrane of the bacterium, compromising its integrity and resulting in membrane depolarization when produced at high levels (Fozo et al., 2008). While IbsC and its homologs are well conserved, this toxin can tolerate extensive mutations (Mok et al., 2010). In addition to point mutations, it can be truncated and short peptide tags as well as entire proteins can be appended to its

N-terminus without compromising its toxicity (*Mok and Li, unpublished results*). These features render IbsC amenable to changes to suit different cloning needs.

We incorporated sequences encoding the open reading frame (ORF) of *ibsC* and one of its active mutants into a ligation-dependent cloning vector. Considering that ligation remains a popular method of cloning and collections of fast and efficient restriction enzymes have been developed by a number of manufacturers, we have decided to use this cloning strategy to demonstrate that IbsC can be used in a positive selection system. We demonstrated that the IbsC-based cloning system offers high selectivity, comparable to the success reported for commercially available cloning systems, and inserts of interest can be cloned, transformed, and plated in three hours. As IbsC can easily be incorporated into any plasmid vector, IbsC and its derivatives can be used in conjunction with many of the aforementioned cloning approaches.

5.4 Materials and methods

5.4.1. Oligonucleotides and reagents

All PCR primers and oligonucleotides used in this study were chemically synthesized by Integrated DNA Technologies (Coralville, IA, USA). Oligonucleotides longer than 40 nucleotides (nt) were purified by 10% (8 M urea) denaturing PAGE (polyacrylamide gel electrophoresis) prior to use. Kanamycin, Anhydrotetracycline (Atc), and 5-bromo-4-chloro-3-indolyl-galactopyranoside (X-gal) were purchased from Sigma-Aldrich (Oakville, ON, Canada). Kanamycin was used at a concentration of 50 μg mL^{-1}, Atc was used at 200 ng mL^{-1}, while X-gal was used at 40 μg mL^{-1}. *ibsC* and *ibsCHK* fused to flanking MCS regions were synthesized using the Klenow fragment from MBI-Fermentas (Burlington, ON, Canada). Likewise, the enzymes for molecular cloning,

including the High Fidelity PCR enzyme mix and T4 DNA ligase were purchased from MBI-Fermentas. Restriction enzymes were either purchased from MBI-Fermentas or from New England Biolabs (Pickering, ON, Canada). Plasmids were isolated from bacteria using the PureYield Miniprep Kit manufactured by Promega (Madison, WI, USA).

5.4.2. Strains and plasmids

ibsC and *ibsCHK* flanked by two MCS regions were cloned into pNYL-MCSII (Mok et al., 2009). The inserts were introduced downstream of the tetracycline-inducible promoter P_{LtetO1}. The vectors were initially transformed into *E. coli* DH5αZ1 (courtesy of H. Bujard) by electroporation. Competition assays and molecular cloning experiments were conducted in *E. coli* MG1655 and *E. coli* BW25113, respectively.

5.4.3. Synthesis of pNYLibsC and pNYLibsCHK

Sequences encoding *ibsC* or its mutant *ibsCHK* flanked by multiple cloning sites were generated by extending oligonucleotides with 20 nt overlapping regions via the Klenow reaction following the manufacturer's protocol. These sequences were cloned into pNYL-MCSII between the EcoRI and BamHI restriction sites. Ligation products were transformed into *E. coli* DH5αZ1 by electroporation. The identities of the clones were validated by DNA sequencing carried out by Mobix Lab (McMaster University).

5.4.4. Growth assays

Growth assays were carried out to assess the toxicity of *ibsC* and its derivative *ibsCHK* when expressed from the cloning vectors. *E. coli* DH5αZ1 transformed with pNYL-MCSII, which served as a negative control, pNYLibsC, and pNYLibsCHK were propagated overnight at 37 ^{o}C with shaking at 260 rpm. The cells were then subcultured

(1:400 dilution) in fresh LB broth supplemented with kanamycin in the absence (for uninduced conditions) or presence (for induced conditions) of Atc. The cells were cultured for 8 h at 37 ^{o}C, shaking at 260 rpm, and cell density was monitored hourly by measuring the optical density at 600 nm (OD_{600}) of each sample using a VersaMax microplate reader (Molecular Devices, Sunnyvale, CA, USA). The assays were carried out in triplicate.

5.4.5. Co-transformation assay

The selectivity of pNYLibsC was assessed with co-transformation assays in which different ratios of pNYL-MCSII containing *gfp* cloned downstream of P_{LtetO1} (pNYLgfp) and pNYLibsC were transformed into *E. coli* MG1655 by electroporation. pNYLgfp and pNYLibsC were mixed together in ratios of 0:1, 1:99, 1:9, 1:1, and 1:0, and 5 ng of plasmid DNA was transformed into 40 µL of electrocompetent bacteria. As negative controls, pNYLgfp and pNYL-MCSII were combined in the aforementioned ratios and transformed into *E. coli* under identical conditions. Transformants were grown on LB agar supplemented with kanamycin overnight. Fluorescent colonies were observed and counted under the Safe Imager Blue Light Transilluminator (Invitrogen Corp., Grand Island, NY, USA). The percentage of fluorescent colonies on each plate was then calculated. This assay was completed in triplicate.

5.4.6. Molecular cloning experiments

To demonstrate the selectivity of pNYLibsC in molecular cloning, reporter genes *gfp* and *lacZ* were cloned into pNYLibsC and pNYL-MCSII (negative control). *gfp* and *lacZ* were amplified by PCR using the High Fidelity PCR Enzyme Mix (MBI-Fermentas) and desired restriction sites were introduced into their 5' and 3' ends. The PCR products

were resolved by agarose gel electrophoresis and were purified using the GenElute Gel Extraction Kit (Sigma-Aldrich). The inserts were then digested. For the data shown in Figure 5.3B, inserts and vectors were digested using high fidelity EcoRI and XmaI from New England Biolabs following manufacturer's protocols. For the data shown in Figure 5.3C, *lacZ* and vectors were digested using KpnI and HindIII. In additional cloning experiments, the inserts and vectors were digested with enzymes from either MBI-Fermentas or New England Biolabs. When using enzymes from MBI-Fermentas, the reactions were incubated at 37 ^{o}C for 20 min before being deactivated at 80 ^{o}C for 10 min. The digested DNA sequences were immediately used for ligation, where the reactions were incubated at room temperature for 1 h. To improve transformation efficiency, the ligation reaction mixtures were purified using the Sigma-Aldrich GenElute Gel Extraction Kit, though this step is optional. Briefly, the 10 µL ligation reaction mixtures were mixed with 90 µL of ddH_2O before the DNA was bound, washed, and eluted from the columns following manufacturer's directions. The ligation mixtures were transformed into *E. coli* BW25113 by electroporation. For the cloning of *gfp*, transformants were plated onto LB-agar with kanamycin and incubated at 37 ^{o}C overnight. Forty-five colonies were selected from each plate of transformants and cultured overnight in LB broth with kanamycin at 37 ^{o}C with shaking at 260 rpm in a 96-well plate. Following overnight growth, 100 µL of each sample was transferred to a 96-well clear bottom black plate (Corning, Corning, NY, USA) and fluorescence from each culture was assessed by scanning the plate using the Typhoon 9400 variable mode imager (λ_{ex}= 488 nm) (GE Healthcare, Baie d'Urfé, PQ, Canada). Percentage of fluorescent colonies from the cloning with pNYLibsC and with pNYL-MCSII were tabulated. The

experiments were carried out in triplicate. For the cloning of *lacZ*, transformants were plated on LB-agar supplemented with kanamycin and X-gal for blue-white screening. The experiments were repeated at least three times and the percentages of blue colonies were tabulated.

5.5 Results

5.5.1. Design of IbsC-based cloning vectors

To engineer an IbsC-based positive selection cloning vector, the coding sequence of IbsC derived from *E. coli* MG1655 along with its RBS and translation initiation sequence were cloned into pNYL-MCSII, a plasmid derived from pZE21-MCSI (Figure 5.1A) (Lutz & Bujard, 1997). The resulting vector is coined pNYLibsC. *ibsC* is flanked by multiple cloning sites, allowing the toxin ORF to be replaced with an insert of interest once it is successfully ligated. In pNYLibsC, *ibsC* is cloned downstream of a tetracycline inducible promoter (P_{LtetO1}) (Lutz & Bujard, 1997). Therefore, the expression of the toxin can be induced using tetracycline or its analog Atc when the vector is introduced into *E. coli* DH5αZ1, which was engineered to encode a constitutively expressed tetracycline repressor (TetR) from its chromosome (Lutz & Bujard, 1997). As observed in Supplementary Figure S5-1, when the expression of *ibsC* is induced using Atc in *E. coli* DH5αZ1 carrying pNYLibsC, growth of the bacteria is suppressed. In addition to wild-type *ibsC*, mutations were introduced in the *ibsC* ORF to produce a derivative that contains NsiI and HindIII restriction sites in the ORF (Figure 5.1B). These mutations resulted in M2H and R3K substitutions in the peptide. Thus the mutant was coined IbsCHK and the vector with this mutant is referred to as pNYLibsCHK. Cloning of a sequence at the NsiI site of this mutant allows the insert to be fused downstream of a RBS

and a start codon, while the presence of an additional HindIII site enables non-directional cloning to be carried out. Growth assays carried out with *E. coli* DH5αZ1 expressing this sequence also demonstrated the expression of *ibsCHK* is deleterious to the bacteria (Supplementary Figure S5-1).

Using different strains for vector propagation and positive clone selection, it is possible to use this system without the need for inducers. In *E. coli* DH5αZ1 and strains that express TetR, it is necessary to induce the expression of *ibsC* and its mutants using Atc. In other *E. coli* strains, however, the expression of the P_{LtetO1}-regulated toxin is derepressed due to the absence of TetR. Therefore, the cloning vectors can be inexpensively propagated in *E. coli* DH5αZ1 in the absence of Atc. For positive clone selection, the clones can be transformed into the TetR-lacking strains. In these strains the negative clones will express *ibsC* without the need to add an inducer.

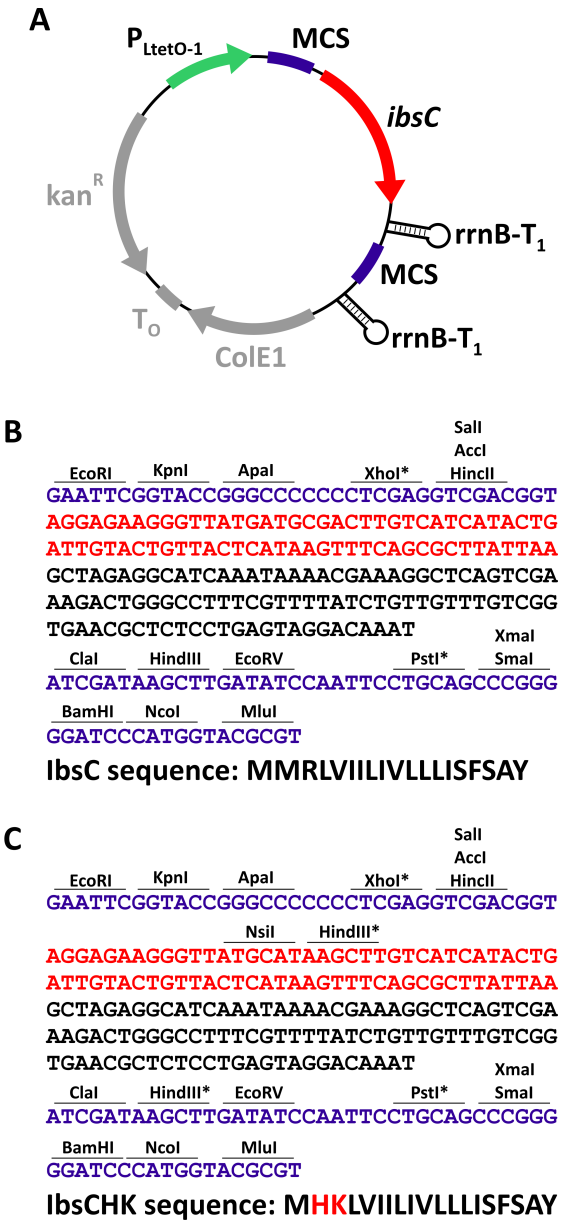


Figure 5.1. Design of *ibsC*-based cloning vectors. A. Vector map of pNYLibsC. *ibsC* and its derivative are cloned downstream of a tetracycline-inducible promoter (P_{LtetO1}). The *ibsC* ORF is immediately followed by a transcription terminator derived from *rrnB* and is flanked by two multiple cloning site (MCS) regions. B. Partial sequence of pNYLibsC. The multiple cloning site regions are shown in blue, *ibsC* is shown in red, and the *rrnB*-derived transcription terminator is shown in black. Restriction sites that occur more than once in the cloning vector are indicated by asterisks (*). C. Partial sequence of pNYLibsCHK. M2H and R3K mutations have been introduced into IbsC in order to generate NsiI and HindIII restriction sites in the plasmid. The colour schemes are the same as indicated for part B.

5.5.2. Selectivity of pNYLibsC

To examine the selectivity of the *ibsC*-based cloning vectors, we carried out a co-transformation assay in which pNYLibsC and pNYLgfp, where *gfp* was cloned downstream of P_{LtetO1} in place of *ibsC*, were transformed into *E. coli* MG1655 in different ratios (Figure 5.2). As negative controls, pNYLgfp and the parent vector, pNYL-MCSII, were transformed into the bacteria. Results from this assay demonstrated that at a low pNYLgfp: pNYLibsC ratio of 1:99, nearly 70% of the colonies were fluorescent. This indicates that there is selectivity against pNYLibsC, as this plasmid represented 99% of the DNA transformed into cells, yet only 30% of the resulting colonies contained this plasmid. As the ratio was increased to 1:9, pNYLibsC is further selected against, as 97% of the resulting colonies were fluorescent. When the two vectors were transformed in equal amounts, nearly all of the colonies were fluorescent. In contrast, when pNYL-MCSII was used in place of pNYLibsC, less than 2%, 9% and 52% of fluorescent colonies were observed when the vectors were transformed at pNYLgfp: pNYL-MCSII of 1:99, 1:9, and 1:1, respectively. Our data demonstrates that pNYLgfp does not have any inherent transformation advantage, and that the selectivity observed is contributed by the toxicity of IbsC, which stunted the growth of cells transformed with pNYLibsC. Together, these observations suggest that the use of the *ibsC*-based vector offers significant selectivity for the clones of interest, which were cells carrying pNYLgfp in this case, even if they are present in low quantities relative to pNYLibsC.

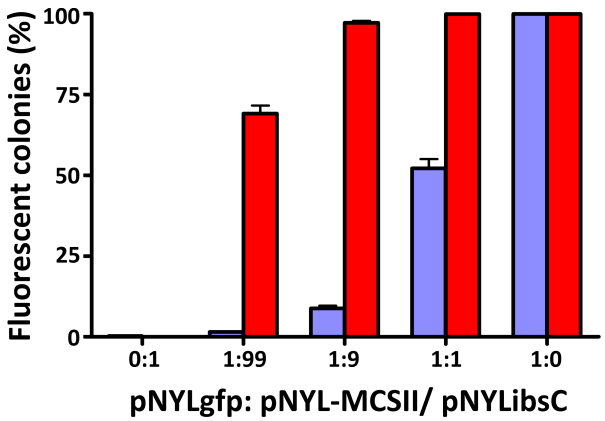


Figure 5.2. Co-transformation of pNYLgfp and pNYLibsC. pNYL-MCSII containing *gfp* cloned downstream of P_{LtetO1} (pNYLgfp) was transformed in combination with pNYLibsC into *E. coli* MG1655 in ratios of 0:1, 1:99, 1:9, 1:1, and 1:0. The transformants were then plated on LB-agar plates supplemented with kanamycin. The assay was completed in triplicate. The percentage of fluorescent colonies on each agar plate following overnight incubation was tabulated (shown in red bars). As negative controls, pNYLgfp was also co-transformed with pNYL-MCSII in the same ratios using the same methods (shown in blue bars).

5.5.3. Selection of positive clones with pNYLibsC

The utility of the *ibsC*-based vectors in molecular cloning was demonstrated via the cloning of two reporter genes, *gfp* and *lacZ*. Following digestion of pNYLibsC, the insert and vectors were ligated in a ratio of 1:1 and the products were transformed into *E. coli* BW25113, which allowed for blue-white screening following *lacZ* cloning. Purification of inserts and vectors following digestion was not necessary, neither were kinase and alkaline phosphatase treatments of DNA. Thus, this entire procedure can be carried out in two to three hours (Figure 5.3A). As negative controls, the reporter genes were also cloned into pNYL-MCSII. For the cloning of *gfp*, the transformants were plated on Luria Burtani (LB) agar supplemented with kanamycin. Following overnight incubation, 45 colonies from each plate were selected and cultured. Fluorescence of each sample was then detected (Figure 5.3B). From this assay, it was observed that the use of

pNYLibsC produced 96% of fluorescent colonies, while the use of pNYL-MCSII only resulted in 42% of fluorescent colonies. This shows that the isolation of clones lacking inserts is a frequent occurrence using the non-selective system, and that the number of clones lacking inserts is almost eliminated when the pNYLibsC vector is used. To demonstrate the selectivity of pNYLibsC when larger inserts are introduced in the 2.3 kb plasmid, *lacZ*, which is over 3 kb long, was cloned following the aforementioned procedures. The transformants are plated on LB agar supplemented with kanamycin and X-gal, allowing for blue-white screening to be conducted. Similar to the results observed with the cloning of *gfp*, cloning of *lacZ* into pNYLibsC yielded 85% blue colonies, which are indicative of successful cloning (Figure 5.3C). The use of pNYL-MCSII yielded only 12% of successfully ligated constructs. This shows that the IbsC cloning system can be used to convert an inefficient process into a procedure that produces a majority of successful clones.

It has been our observation that cloning using pairs of restriction enzymes containing cohesive ends that are partially complementary are predisposed to self-ligation, resulting in the ligation of vectors without taking up an insert. With the use of pNYLibsC, we were able to produce 30 to 60% more positive clones compared with pNYL-MCSII when cloning with these problematic restriction pairs (*data not shown*).

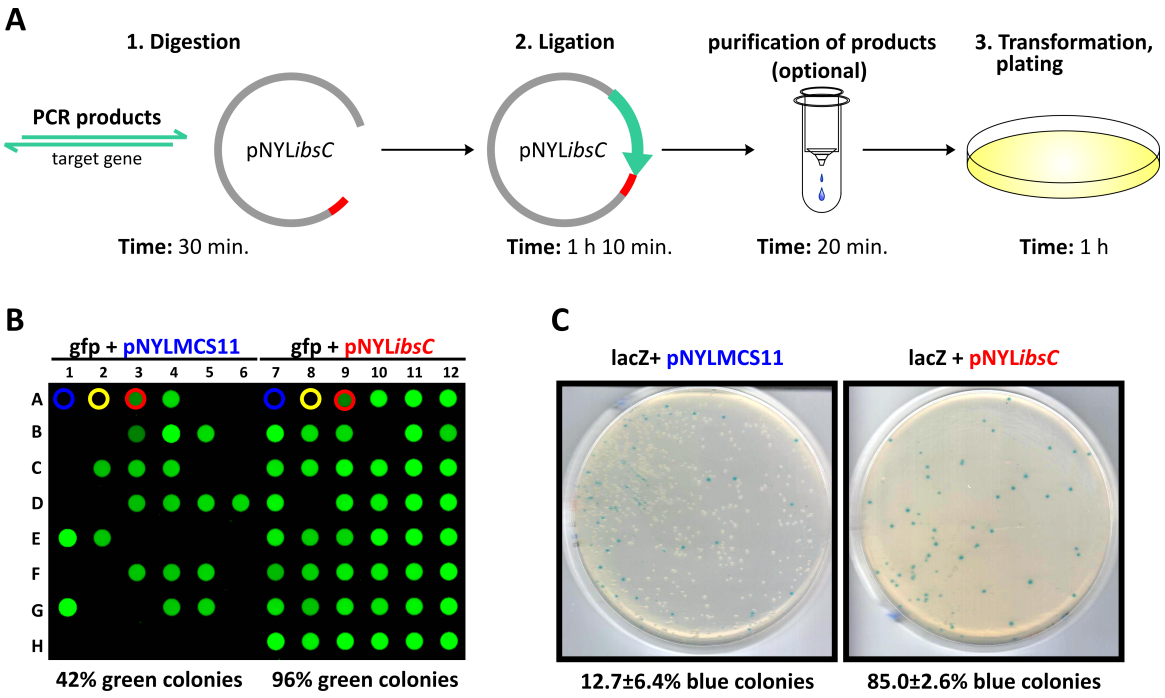


Figure 5.3. Molecular cloning with pNYLibsC. A. A scheme depicting the steps involved in generating successful clones using pNYLibsC or pNYLibsCHK. The time required for each step indicated coincide with the use of FastDigest enzymes manufactured by MBI-Fermentas (refer to materials and methods) and ligation at room temperature. To improve transformation efficiency, the ligation products may be purified prior to electroporation. Following transformation, the bacteria are incubated at 37 ^{o}C for 1 h before plating on agar supplemented with kanamycin. B. Cloning of *gfp* into pNYLibsC. Following the cloning of *gfp* into pNYLibsC and transformation into *E. coli* BW25113 using the procedures shown in part A of this figure, 45 colonies from a plate of transformants were cultured overnight and the fluorescence of each sample was detected the following day. As a negative control for the experiment, *gfp* was also cloned into pNYL-MCSII following identical procedures. The percentage of fluorescent colonies observed with each plasmid was calculated. As controls, fluorescence emanating from the growth media (blue circle), from cells transformed with pNYLibsC (yellow circles), and from cells transformed with pNYLgfp (red circles) are also shown. C. Cloning of *lacZ* into pNYLibsC. *lacZ* was cloned into pNYLibsC and transformed into *E. coli* BW25113. The transformants were then plated on LB-agar supplemented with kanamycin and X-gal, allowing for blue-white screening. As a negative control, *lacZ* was also cloned into pNYL-MCSII. The percentage of blue colonies appearing on each plate was calculated. The assays were carried out in triplicate.

5.6. Discussion

Currently on the market, several cloning systems based on the CcdAB TA pair are available. Among them is the GATEWAY cloning system, which is widely used in molecular biology (Walhout et al., 2000). The GATEWAY system requires the insert of interest to be first introduced into an "entry clone" before it is transferred into the "destination vector". These steps depend on two proprietary enzymes, coined "clonases", to catalyse recombination events. These criteria increase both the time and cost of cloning using this system.

In comparison to existing commercial cloning kits, the IbsC-based cloning system offers similar efficiency and selectivity (>95%), but it can be used in conjunction with generic restriction and ligase enzymes. In recent years, many of these enzymes, such as the FastDigest enzymes offered by Fermentas Inc., have been modified to exhibit improved speed and fidelity. Moreover, the plasmids, pNYLibsC and pNYLibsCHK, can be easily propagated in *E. coli* strains expressing *tetR*, such as *E. coli* DH5αZ1, when cultured under non-inducing conditions. Molecular cloning can then be carried out in other strains devoid of *tetR*, and the expression of *ibsC* and its derivatives offer selection against cells that have been transformed with an "empty" vector. Thus, the use of the IbsC-based cloning vectors presented here can be rapid and cost-effective.

As the *ibsC* and *ibsCHK* ORFs are relatively small, they can easily be transferred to other commonly used cloning vectors. They can be used with any cloning platform of choice, be it ligation-based or recombinase-based. We have observed that the expression of P_{LtetO1} regulated *ibsC* from the chromosome of *E. coli* DH5αZ1 is deleterious to the bacteria (*data not shown*). This suggests that *ibsC* and *ibsCHK* may be used with lower

copy number plasmids or weaker inducible promoters and the selectivity they offer would still be conferred. Furthermore, the N-termini of IbsC and its derivatives can tolerate mutations, truncations, and expansions (Mok et al., 2010). We have found that protein tags, such as the FLAG-tag, and proteins, such as GFP, can be appended to this end without compromising its toxicity (*data not shown*). Therefore, we can envision engineering sequences encoding protein tags in the 5' end of *ibsCHK* to produce cloning vectors that are amenable for protein expression and purification.

The intracellular target of IbsC renders it less prone to resistance, which can make this toxin a better positive selection marker compared with others that are presently being used. For instance, the CcdB toxin used in Gateway systems targets the DNA-gyrase complex, thereby inhibiting DNA replication and killing the bacteria. It has been demonstrated that a single point mutation in DNA gyrase can prevent CcdB binding and confer resistance toward the toxin (Williams & Hergenrother, 2012). As IbsC acts on the inner membrane of *E. coli*, resistance toward the toxin is not expected to arise readily, as the restructuring of the membrane is costly to the organism. As for the toxin, IbsC has been shown to be able to withstand extensive mutations without losing its toxicity.

We have developed cloning vectors using the IbsC toxin and its mutant, IbsCHK, which can effectively facilitate the positive selection of successful clones. Using this system with ligation-based cloning, clones can be generated from PCR products in 3 h, and only 5 to 10 min of hands-on time is needed. In our cloning experiments, we consistently observe cloning efficiencies of over 80-90%. Collectively, these features render the IbsC-based cloning system a rapid, efficient, and cost-effective method to improve current molecular cloning strategies.

5.7. Acknowledgement

We would like to thank Professor Herman Bujard for the gift of pZE21-MCSI and *E. coli* DH5αZ1. We are also grateful to Li Lab members past and present for helpful discussions. This work supported by a research grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

5.8. References

Bernard, P. (1995). New ccdB positive-selection cloning vectors with kanamycin or chloramphenicol selectable markers. Gene 162: 159-160.

Bernard, P., P. Gabant, E. M. Bahassi, and M. Couturier (1994). Positive-selection vectors using the F plasmid ccdB killer gene. Gene 148: 71-74.

Bryksin, A. V., and I. Matsumura (2010). Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48: 463-465.

Dewhirst, F. E., T. Chen, J. Izard, B. J. Paster, A. C. Tanner, W. H. Yu, A. Lakshmanan, and W. G. Wade (2010). The human oral microbiome. J Bacteriol 192: 5002-5017.

Fozo, E. M., M. Kawano, F. Fontaine, Y. Kaya, K. S. Mendieta, K. L. Jones, A. Ocampo, K. E. Rudd, and G. Storz (2008). Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. Mol Microbiol 70: 1076-1093.

Fozo, E. M., K. S. Makarova, S. A. Shabalina, N. Yutin, E. V. Koonin, and G. Storz (2010). Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Res 38: 3743-3759.

Gibson, D. G., J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z. Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, 3rd, H. O. Smith, and J. C. Venter (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329: 52-56.

Han, K., K. S. Kim, G. Bak, H. Park, and Y. Lee (2010). Recognition and discrimination of target mRNAs by Sib RNAs, a cis-encoded sRNA family. Nucleic Acids Res 38: 5851-5866.

Lutz, R., and H. Bujard (1997). 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.

Mandi, N., P. Kotwal, and S. Padmanabhan (2009). Construction of a novel zero background prokaryotic expression vector: potential advantages. Biotechnol Lett 31: 1905-1910.

Marchuk, D., M. Drumm, A. Saulino, and F. S. Collins (1991). Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res 19: 1154.

Mok, W. W., N. K. Navani, C. Barker, B. L. Sawchyn, J. Gu, R. Pathania, R. D. Zhu, E. D. Brown, and Y. Li (2009). Identification of a toxic peptide through bidirectional expression of small RNAs. Chembiochem 10: 238-241.

Mok, W. W., N. H. Patel, and Y. Li (2010). Decoding toxicity: deducing the sequence requirements of IbsC, a type I toxin in *Escherichia coli*. J Biol Chem 285: 41627-41636.

Schall, T. J., M. Lewis, K. J. Koller, A. Lee, G. C. Rice, G. H. Wong, T. Gatanaga, G. A. Granger, R. Lentz, H. Raab, and et al. (1990). Molecular cloning and expression of a receptor for human tumor necrosis factor. Cell 61: 361-370.

Shuldiner, A. R., L. A. Scott, and J. Roth (1990). PCR-induced (ligase-free) subcloning: a rapid reliable method to subclone polymerase chain reaction (PCR) products. Nucleic Acids Res 18: 1920.

Stieber, D., P. Gabant, and C. Szpirer (2008). The art of selective killing: plasmid toxin/antitoxin systems and their technological applications. Biotechniques 45: 344-346.

Szpirer, C. Y., and M. C. Milinkovitch (2005). Separate-component-stabilization system for protein and DNA production without the use of antibiotics. Biotechniques 38: 775-781.

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

Walhout, A. J., G. F. Temple, M. A. Brasch, J. L. Hartley, M. A. Lorson, S. van den Heuvel, and M. Vidal (2000). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol 328: 575-592.

Wank, S. A., R. Harkins, R. T. Jensen, H. Shapira, A. de Weerth, and T. Slattery (1992). Purification, molecular cloning, and functional expression of the cholecystokinin receptor from rat pancreas. Proc Natl Acad Sci U S A 89: 3125-3129.

Williams, J. J., and P. J. Hergenrother (2012). Artificial activation of toxin-antitoxin systems as an antibacterial strategy. Trends Microbiol 20: 291-298.

5.9. Supplementary figures

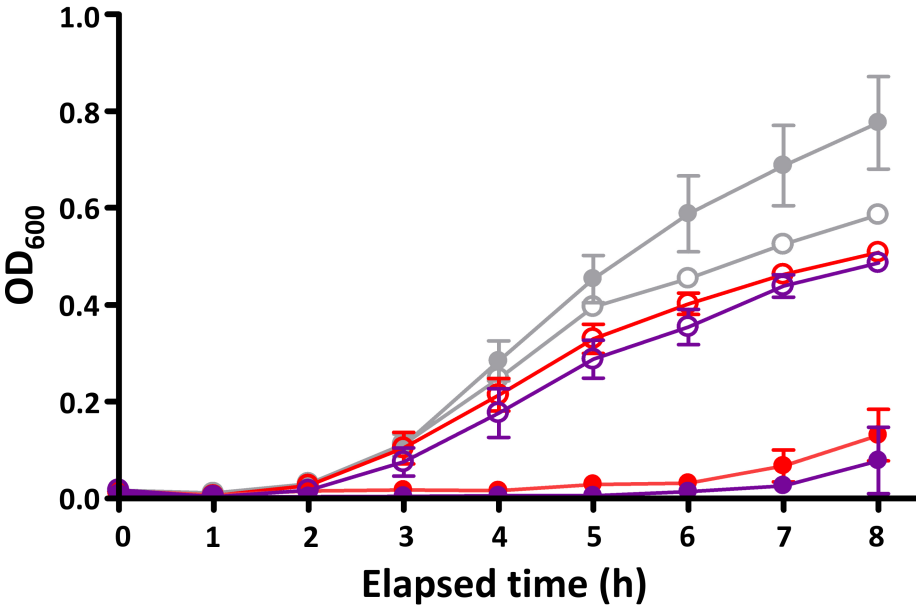


Figure S5-1. Growth of *E. coli* carrying pNYLibsC and pNYLibsCHK. *E. coli* DH5αZ1 carrying pNYLibsC, pNYLibsCHK, and pNYL-MCSII (negative control) were cultured for 8 h in LB media supplemented with kanamycin in the absence (uninduced conditions, as depicted by open circles) or in the presence (induced conditions, as depicted by closed circles) of Atc. Growth of each sample was monitored hourly by measuring their absorbance at 600 nm. *N=3*. (Grey circles denote cells carrying pNYL-MCSII; red circles denote cells carrying pNYLibsC; purple circles denote cells carrying pNYLibsCHK)

Chapter 6:
Discussions and future directions

6.1. Summary of key findings

The Ibs/Sib family of type I TA systems are abundant in *E. coli* K-12 and are highly conserved in related Proteobacterial species (Fozo et al., 2008; Fozo et al., 2010). Surprisingly, these genetic elements avoided detection until recently and few details concerning the evolution, regulation, and function of this group of enigmatic TA pairs are known. Our quest toward characterizing the IbsC/SibC type I TA pair, as detailed in this dissertation, started with its serendipitous identification in the genome of an *E. coli* K-12 variant and focuses on our efforts in uncovering the characteristics of the IbsC toxin and its regulatory factors. Our findings have led us to better understand the function of this novel genetic element.

Through our sequence analysis of IbsC, as detailed in Chapter 3, we discovered that while the Ibs toxins are well conserved in *E. coli* K-12 and across various proteobacterial species, the toxin is tolerant of many amino acid substitutions. A stretch of amino acids at its N-terminus were dispensable, thus the length of the peptide can be minimized to 15 amino acids without compromising its toxicity. However, the hydrophobic residues near the core of the peptide and the bulky polar residues at the C-terminus are required for its toxicity. Based on these observations, we hypothesized that the hydrophobic core of the peptide is inserted into the membrane, with the C-terminus facing the periplasmic space and the N-terminus in the cytoplasm. Therefore, the positively-charged arginine residue, which may have mediated the initial interactions between IbsC and negatively-charged head groups of the phospholipid bilayer, is expected to be found close to the inner leaflet of the membrane, in compliance with the

"positive-inside" rule (von Heijne, 1992). Our mechanistic studies further revealed a correlation between the toxicity of IbsC mutants and their ability to permeabilize the inner membrane of *E. coli*.

As IbsC is highly toxic to cells and can rapidly inhibit growth once it is produced, its expression is expected to be tightly regulated. It was previously reported that *ibsC* transcripts are not detected with northern blotting unless the expression of *sibC* is inhibited via promoter mutations or deletions (Fozo et al., 2008). Therefore, under the growth conditions examined, *ibsC* is strongly repressed by SibC. A study by Han and colleagues implicated two regions in SibC, coined TRD1 and TRD2, in being the key determinants for mediating the initial recognition and interactions between cognate pairs of Sib and *ibs* RNAs in *E. coli* K-12 (Han et al., 2010). Our study involving the mutagenesis of the nucleotides constituting TRD1 and TRD2 of SibC showcases TRD1 as the dominant element involved in toxin-antitoxin recognition. Active sequences isolated from our screen were repeatedly found to contain wild-type TRD1. Mutating six nucleotides in TRD1, which are complementary to codons 16 and 17 of *ibsC*, resulted in reduced or loss of activity. In contrast, mutation at TRD2, which is complementary to the translation initiation region of *ibsC*, produced many functional mutants. In addition to post-transcriptional regulation, the expression of *ibsC* is likely to be repressed at the transcriptional level as well. Data from our analyses of the *ibsC* promoter hint at possible negative regulatory elements. The exact factors governing the expression of this toxin remains to be elucidated, but multiple layers of regulation are expected to be found.

The sequence, mechanistic, and structural insights on IbsC and SibC acquired through this work has furthered our understanding of the function of this TA system.

Taking advantage of the toxicity of IbsC, we can utilize this peptide and its derivatives in the design and engineering of molecular tools. For example, using the information gathered from the sequence characterization of IbsC, cloning vectors that improved the efficiency of current molecular cloning strategies by facilitating positive selection were developed, as discussed in Chapter 5. The sequence information gathered from the aforementioned study can also be applied in the design of potential therapeutics based on IbsC. Some of the *ibsC* and *sibC* mutants along with promoter-reporter constructs developed in our studies can be used as tools to address outstanding questions pertaining to the function and regulation of this TA system. Such questions and possible experiments that can be conducted to resolve them are discussed in subsequent sections of this Chapter.

6.2. Antisense RNA regulation and Type I TA Pairs

As described in Chapter 2 of this thesis, we first came across *ibsC/sibC* through a screen for antisense RNAs that can sequester and inhibit the functions of regulatory sRNAs in *E. coli*. Instead of identifying an antisense sequence that inhibited the action of an sRNA leading to growth suppression, we found that the reverse complement of *sibC*, *ibsC*, caused cell death as it encoded a small, hydrophobic, and toxic peptide. In this case, *sibC* was the gene that encoded an antisense RNA that repressed the toxin mRNA. While we did not identify antisense regulatory RNA sequences through our screen, *cis*-encoded regulatory RNAs are suggested to represent a large proportion of the bacterial transcriptome (Georg & Hess, 2011). This mechanism of post-transcriptional regulation has been demonstrated to be advantageous to cells as it is faster compared with transcriptional regulation (Shimoni et al., 2007). In incidences where *de novo* synthesis of

regulator following a sudden onset of stress is required, RNA-based regulation is more efficient compared with post-translational, protein-mediated regulation (Shimoni et al., 2007). This rapid response is particularly beneficial to the cell when the expression of the target can drastically affect the physiology of the cell, as in the case of toxins. Furthermore, *cis*-encoded RNA regulators are space-efficient, because they are present at the same genetic locus as their target and do not require an increase in genome size to incorporate these regulators. At present, more type II TA pairs, in which the antitoxins and toxins act via protein-protein interactions, have been annotated across diverse bacterial genomes compared with type I pairs. This disparity may reflect the lack of attention toward type I systems until recent years. As searches for type I TA pairs in different microorganisms continue, we can expect a surge in their number. The sequence properties of type I toxins deduced from the study described in Chapter 3, such as a minimum length of 15 amino acids and a consecutive stretch of 10 hydrophobic amino acids, may be used to define criteria in searches for such toxic peptides.

6.3. Outstanding questions concerning ibsC regulation

While we know that the expression of *ibsC*, akin to many type I toxins, is antagonized by its cognate antitoxin sRNA, many questions concerning its regulation remain. As discussed in Chapter 4, following the recognition and contact between *ibsC* and SibC, which is primarily mediated by TRD1, translation of the toxin transcript is proposed to be inhibited. The duplex is also hypothesized to be degraded, rendering the repressive effects of SibC on *ibsC* irreversible. However, the RNase responsible for this degradation has yet to be identified. As the expression of *ibsC* can have detrimental effects on the cell, it is expected to be tightly regulated, not only at the post-

transcriptional level by SibC, but at the transcriptional level as well. In Chapter 4 of this thesis, we discussed our work in delineating putative regulatory sequences in the promoter of *ibsC*. Specifically, we demonstrated that the sequence around the -60 position of *ibsC* may harbour repressive regulatory elements. Whether this sequence interferes with binding of the RNA polymerase machinery or serves as a binding site for transcription factors remains to be elucidated.

To identify potential regulators that can act on this region of P_{ibsC}, pull-down assays using this sequence as a probe can be carried out in order to isolate regulatory factors binding to this region, and their identities can be determined using mass spectrometry. The interactions between P_{ibsC} and these regulators, in addition to the effects of such interactions on the expression of downstream genes can be evaluated using techniques such as electromobility shift assays and reporter assays, respectively. It should also be noted that autoregulation has been demonstrated in some type II TA systems. For example, the RelB antitoxin and the RelE toxin form a complex that binds its own promoter and represses the transcription of its own operon when antitoxin production is high (Cataudella et al., 2012). When toxin production is elevated, the $RelB_2RelE$ complex is destabilized, and the repression is alleviated. This results in enhanced antitoxin production. In another type II TA system, MqsRA, the MqsA antitoxin binds to the promoter of its own operon in order to regulate its expression (Brown et al., 2011). It may be possible for IbsC to be involved in its own regulation, binding its promoter as an IbsC complex or forming a DNA-binding complex with another protein.

6.4. Evolution of ibs/sib Homolog in E. coli

The reason for bacteria to maintain multiple copies of the *ibs/sib* TA pairs in their genomes is unknown. It has been suggested that the five homologs present in *E. coli* K-12 were acquired through gene duplication at two loci of the genome. Over time, they evolved by accumulating mutations, particularly in the TRDs, allowing for each Sib sRNA to discriminate its cognate target from the others (Han et al., 2010). In our sequence analyses of IbsC, we observed that the amino acids at positions 16 and 17, which are encoded by nucleotides constituting the TRD1 loop, were especially variable compared with other residues in the peptide. Our data suggests that certain codons may be selected for at these positions so that toxin and antitoxin functions at the protein and at the RNA levels are supported. At the protein level, the resulting peptide has to remain growth suppressive or toxic to cells. At the RNA level, the two RNA species must still be able to adopt structures that are permissive to their interactions. To examine this hypothesis, SibC variants with mutations at TRD1 that were incapable of inactivating wild-type *ibsC* in our screen could be tested for their ability to repress toxic *ibsC* derivatives with complementary sequences.

Mutations could also have occurred in each of their promoters, allowing for each Ibs toxin to be differentially expressed. In a preliminary experiment completed by our lab, the core promoter and sequences encoding the 5'-UTR of each *ibs* gene derived from *E. coli* K-12 were fused to reporter gene *lacZ* (*Mok, Fonseka, and Li, unpublished data*). β-galactosidase assays carried out with cells harbouring these constructs indicated that promoters derived from *ibs* genes from each of the three genetic loci displayed different activity. This suggests that the five homologs may be differentially expressed. Each may

be triggered by a unique environmental cue and may contribute to a different function in the cell.

6.5. Potential Biological Function of ibsC/sibC

Although the effects of overexpressing Ibs toxins on the architecture and survival of the cell are understood, its consequences at endogenous levels of expression remain a mystery. It may be possible that the outcome of *ibsC* expression depends on the concentration of toxin molecules produced. At low levels of IbsC, the peptide may compromise the inner membrane and dissipate the proton motive force. This may decrease the energy production in the cell and slow down metabolic processes. At elevated concentrations, the peptide may further deplete cells of ATP, causing them to enter a state of dormancy. If IbsC is produced to levels comparable to that achieved via its ectopic expression, it could lead to the irreversible destruction of the inner membrane and cell death.

It has been demonstrated that another type I TA pair, TisAB, has been linked to bacterial persistence. Persister cells are a small subpopulation of dormant cells that form stochastically in response to stress (Lewis, 2010). These phenotypic variants are genotypically identical to their siblings, however the differential expression of specific genes has caused these cells to enter a reversible dormant state. Persistence is a major contributor to multidrug tolerance and recalcitrance of chronic bacterial infections. Expression of the TisB toxin is stimulated by the induction of the SOS response following treatment with Ciprofloxacin (Dorr et al., 2009). Similar to IbsC, TisB is a small hydrophobic peptide of 29 amino acids, which acts on the inner membrane of *E. coli* (Gurnev et al., 2012). In doing so, it dissipates the proton motive force and decreases

ATP production (Unoson & Wagner, 2008). TisB production in bacteria is suggested to be heterogeneous, fluctuating around a mean level of expression (Lewis, 2010). Cells expressing a higher level of TisB are proposed to become dormant persisters and are less susceptible to antibiotics treatment compared with faster growing siblings. As multiple genes, including many TA encoding ones, have been linked to persister cell formation, it is possible for IbsC and its homologs to be involved in bacterial persistence.

To determine whether *ibsC* is differentially expressed amongst bacteria in the same population, cells carrying plasmids with the *ibsC* promoter fused to *gfp*, as described in Chapter 4, can be cultured under different growth conditions. They can then be subjected to fluorescence-assisted cell sorting (FACS). Fluctuations in GFP production in individual cells cannot be measured via fluorescence assays carried out with the bulk population, but it is achievable using FACS. FACS would enable cells expressing higher levels of the fluorescent protein relative to the rest of the population to be isolated. This experiment would also allow conditions that may induce IbsC expression to be identified.

The effects of IbsC on the metabolic state of cells can be assessed by measuring protein synthesis in *E. coli* using a strain engineered to express a less stable GFP variant regulated by a growth rate-dependent promoter derived from the *rrnB* P1 core promoter from *E. coli* (Bartlett & Gourse, 1994; Sternberg et al., 1999; Andersen et al., 1998). In previous studies, it has been observed that as the growth rates of these cells decreases when cells enter dormancy, the ribosomal promoter becomes inactive and transcription of the mutant *gfp* diminishes (Shah et al., 2006). As the remaining GFP variants degrade, the cells become dim. Should IbsC cause *E. coli* to become dormant, its expression in the reporter strain would cause the fluorescence detected from these cells to decrease. FACS

can be employed to sort the bright members of the population from dim ones. To ascertain that the production of IbsC and its subsequent disruptions on the electron transport chain are causing the cells to enter into a dormant state, thereby slowing transcription and translation, inactive mutants of IbsC can also be expressed in this strain. These mutants are not expected to produce dormant cells, resulting in predominantly bright cells in the population. The tolerance of the dimmer population produced following *ibsC* induction toward antibiotics, such as ciprofloxacin, can then be evaluated in order to examine whether this population function as persisters.

6.6. Therapeutic Potential of ibsC/sibC

While type I toxin production has been linked to bacterial persistence during treatment with certain classes of antibiotics, the induction of IbsC has also been documented to enhance the sensitivity of bacteria toward other classes of antibacterial agents. Expression of IbsC at sublethal concentrations in *Pseudomonas* mimics the effects of membrane-disrupting peptides produced by mistranslation following exposure to aminoglycosides, thereby enhancing the uptake of this class of antibiotics and increasing the sensitivity toward these drugs (Lee et al., 2009). Based on these accounts, compounds that can induce *ibsC* expression or suppress *sibC* expression have the potential to act synergistically when co-administered with aminoglycosides. The P_{ibsC}-*gfp* promoter-reporter constructs developed for the study detailed in Chapter 4 can be coupled with high-throughput screening to identify small molecules that can augment IbsC production from compound libraries. To ascertain that such a compound is capable of inducing sufficient toxin production to compromise the membrane, the identification of compounds that can concomitantly induce all five *ibs* genes would be ideal. Such compounds may

help preserve the efficacy of existing compounds and provide a new line of defence toward antibiotic resistant pathogens.

In addition to exploring the therapeutic potential of compounds that can induce toxin expression, IbsC itself may be investigated as a possible lead for new antimicrobial agents. IbsC shares similar physical properties with antimicrobial peptides (also known as host-defence peptides), which constitute an integral part of innate immunity of many eukaryotic organisms. Antimicrobial peptides are generally short (with 10 to 50 amino acids) and cationic (with an overall charge of +2 to +9) (Hancock & Sahl, 2006). They also carry over 50% of hydrophobic residues. The composition of amino acids with ionic and nonpolar side chains allows the peptides to fold and partition into separate domains, especially when in contact with amphipathic membrane bilayers. The cationic peptides first make contact with their target microorganisms by binding to negatively charged lipopolysaccharide moieties on the outer membrane of Gram-negative bacteria or to lipoteichoic acids in the cell wall of Gram-positive activity, thus enabling their "self-promoted uptake" (Hancock & Chapple, 1999; Sawyer et al., 1988). Once they have compromised the first barrier, the peptides further disrupt the cytoplasmic membrane by means of inducing pore formation, membrane leakage, or catastrophic dissolution of the membrane, thus resulting in cell death (Wimley & Hristova, 2011). In addition to the membrane, these peptides have been shown to act on intracellular targets and perturb key cellular processes, including the DNA, RNA, protein, and cell wall synthesis (Brogden, 2005). Moreover, antimicrobial peptides have immunomodulatory functions, allowing the host cell to clear an infection while dampening the adverse effects of pro-inflammatory

responses (Easton et al., 2009). With this mélange of activities, these peptides have the potential to be powerful antimicrobial agents.

To explore the therapeutic potential of IbsC and other type I toxins, our group and others examined the effects of the exogenous administration of these peptides on the physiology and survival of various bacteria. Pecota and colleagues synthesized the 52 amino acid Hok toxin, part of the *hok/sok* type I TA system found in *E. coli*, which also kills the cell via membrane depolarization (Pecota et al., 2003). They treated six microorganisms, including *E. coli*, *B. subtilis*, and *P. aeruginosa*, with purified Hok. However, the application of 200 μg ml^{-1} of Hok failed to produce significant cell killing. Similar results were obtained when truncated derivatives of Hok was used. Our group has exposed *E. coli* to wild-type IbsC, truncation mutant IbsC(6-19), and a number of point mutants (*Mok and Li, unpublished data*). Akin to the Pecota study, we did not observe significant cell killing even with the application of 1 mg mL^{-1} of each peptide. Nevertheless, as IbsC is amenable to extensive mutations, as demonstrated in Chapter 3 of this thesis, it may be possible to modify the sequence and balance charge and hydrophobicity in the peptide in order to drive is self-promoted uptake.

One major roadblock presently hindering the development of peptide therapeutics lies in the cost of their production. The synthesis of a 1 g of a 19-mer such as IbsC can cost over 100 CAD (Hancock & Sahl, 2006). Although high-throughput synthesis approaches exist and have been shown to considerably reduce the cost and labour of peptide library production, the process remains expensive (Hilpert et al., 2005; Hilpert et al., 2007). Moreover, the sequence space coverage offered by these methods remains low. One possible method to reduce the cost of peptide screening is to use bacterial expression

and secretion systems in combination with *in vitro* selection strategies to produce these peptides. For over two decades, *in vitro* selection has been applied in the discovery of functional nucleic acids, including ligand binding aptamers as well as catalytic DNA and RNA sequences (Ellington & Szostak, 1990; Tuerk & Gold, 1990). Through iterative rounds of selection, nucleic acid sequences with targeted functions are isolated and enriched from a pool or library of up to 10^{16} random sequences (Figure 6.1A). As DNA synthesis is considerably less expensive compared with peptide synthesis, random DNA libraries encoding a peptide of interest and its mutants can be produced at a fraction of the cost of an equivalent peptide library.

In the isolation of potent bacteria-penetrating IbsC derivatives, DNA libraries encoding IbsC and its mutants can be synthesized and cloned into expression vectors, allowing the peptides to be expressed in *E. coli*. To facilitate the secretion of these peptides into the extracellular milieu, sequences encoding secretion signal peptides, such as the Sec-dependent pathway signal peptides, may be appended to the variable N-terminus of the peptides (Oka et al., 1985). As the periplasmic release of proteins is limited by periplasmic volume, one can make use of leaky mutant *E. coli* strains with defective outer membrane lipoproteins to promote secretion of peptides into the culture media (reviewed in (Yoon et al., 2010)). Alternatively, the peptides can be co-expressed with YebF, an extracellular protein, in order for them to be more effectively exported into the extracellular media (Zhang et al., 2006). To select for active antimicrobial peptides, the *E. coli* producing strains can be co-cultured with an organism of interest. Upon identifying secreted peptides that can inhibit the growth of the target organism, vectors can be isolated from the producing cells. The peptide-encoding sequences could also be

amplified by hypermutagenic PCR, allowing more potent derivatives of these candidates to be generated. This pool of sequences is once again expressed in *E. coli*, so that secreted IbsC mutants with high antimicrobial activity and self-promoted uptake capability can be produced following repeated rounds of selection. A general scheme depicting this peptide production and screening approach is illustrated in (Figure 6.1B).

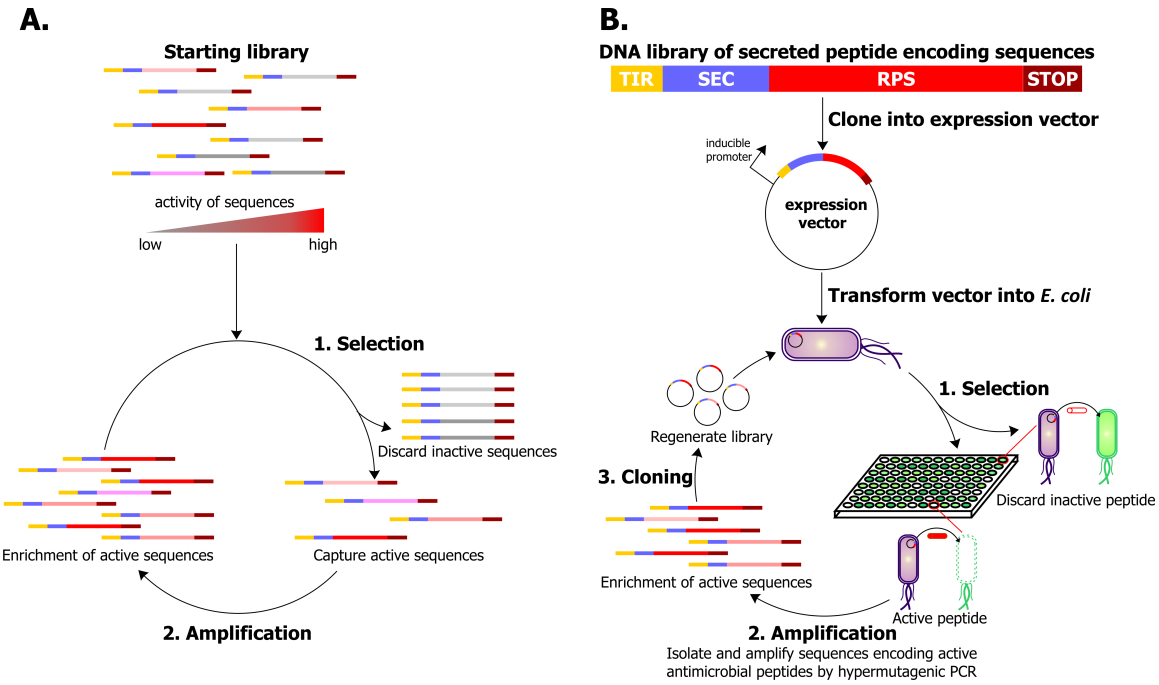


Figure 6.1. *In vitro* selection and peptide library screening. A. General scheme of *in vitro* selection. B. Scheme of IbsC mutant production and screening using an *E. coli*-based expression and secretion system in conjunction with *in vitro* selection. A library of DNA sequences encoding a translation initiation region (TIR) with consensus ribosome binding sites and start codons; a Sec-dependent pathway signal peptide (SEC); randomized peptide-encoding sequence (RPS) encoding IbsC and its derivatives; and a stop codon (STOP) is synthesized. The nucleotides in the RPS are randomized such that each base has a 25% chance of being either adenine, cytosine, guanine, or thymine. The library is cloned into an inducible expression vector compatible with a suitable laboratory *E. coli* strain, and the vectors are subsequently transformed into *E. coli*. The peptide-secreting *E. coli* is co-cultured with a target microorganism under inducing conditions, allowing the antimicrobial activities of peptides to be assessed in large-scale, high-throughput growth assays. Recombinant *E. coli* that produce active, growth-inhibiting IbsC derivatives are isolated and vectors from these cells are purified. The peptide-coding DNA sequences are then amplified via hypermutagenic PCR, creating clustered sequence libraries based on the initial active sequences. The libraries are then cloned into the expression vectors again and selection is repeated. After several rounds of selection, IbsC mutants with high potency and bacteria-penetrating activity may be isolated.

6.7. Concluding remarks

Our efforts in conjunction with work done by other laboratories in the community have unveiled details pertaining to the consequences of IbsC overexpression, the mechanism of toxin action, and the interplay between toxin and antitoxin in the IbsC/SibC TA system. However, we are only at the threshold of an exciting phase of TA research. Many interesting questions surrounding the function of this TA pair and related type I TA systems remain to be answered. Many TA pairs have been shown to play roles that are vital to the adaptation and survival of bacteria. IbsC/SibC, whose biological function remains to be uncovered, may be of equal importance. Details concerning the regulation of *ibsC* and *sibC* at the transcription level in addition to the sequence of events following SibC-*ibsC* binding are unknown. Apart from the endogenous function and actions of this TA pair, its therapeutic potential is another stimulating and largely unexplored line of research. As the production of IbsC is disruptive to the cytoplasmic membrane of *E. coli*, this toxin, its active mutants, and its non-toxic derivatives can be investigated as drugs, drug targets, and drug carriers. The findings concerning the mechanism, regulation, and sequence requirements of IbsC and SibC presented in this thesis will lay the foundation for such future studies.

6.8. References

Andersen, J. B., C. Sternberg, L. K. Poulsen, S. P. Bjorn, M. Givskov, and S. Molin (1998). New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. Appl Environ Microbiol 64: 2240-2246.

Bartlett, M. S., and R. L. Gourse (1994). Growth rate-dependent control of the rrnB P1 core promoter in *Escherichia coli*. J Bacteriol 176: 5560-5564.

Brogden, K. A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3: 238-250.

Brown, B. L., T. K. Wood, W. Peti, and R. Page (2011). Structure of the *Escherichia coli* antitoxin MqsA (YgiT/b3021) bound to its gene promoter reveals extensive domain rearrangements and the specificity of transcriptional regulation. J Biol Chem 286: 2285-2296.

Cataudella, I., A. Trusina, K. Sneppen, K. Gerdes, and N. Mitarai (2012). Conditional cooperativity in toxin-antitoxin regulation prevents random toxin activation and promotes fast translational recovery. Nucleic Acids Res. Epub ahead of print.

Dorr, T., K. Lewis, and M. Vulic (2009). SOS response induces persistence to fluoroquinolones in *Escherichia coli*. PLoS Genet 5: e1000760.

Easton, D. M., A. Nijnik, M. L. Mayer, and R. E. Hancock (2009). Potential of immunomodulatory host defense peptides as novel anti-infectives. Trends Biotechnol 27: 582-590.

Ellington, A. D., and J. W. Szostak (1990). In vitro selection of RNA molecules that bind specific ligands. Nature 346: 818-822.

Fozo, E. M., M. Kawano, F. Fontaine, Y. Kaya, K. S. Mendieta, K. L. Jones, A. Ocampo, K. E. Rudd, and G. Storz (2008). Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. Mol Microbiol 70: 1076-1093.

Fozo, E. M., K. S. Makarova, S. A. Shabalina, N. Yutin, E. V. Koonin, and G. Storz (2010). Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Res 38: 3743-3759.

Georg, J., and W. R. Hess (2011). cis-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev 75: 286-300.

Gurnev, P. A., R. Ortenberg, T. Dorr, K. Lewis, and S. M. Bezrukov (2012). Persister-promoting bacterial toxin TisB produces anion-selective pores in planar lipid bilayers. FEBS Lett

Han, K., K. S. Kim, G. Bak, H. Park, and Y. Lee (2010). Recognition and discrimination of target mRNAs by Sib RNAs, a cis-encoded sRNA family. Nucleic Acids Res 38: 5851-5866.

Hancock, R. E., and D. S. Chapple (1999). Peptide antibiotics. Antimicrob Agents Chemother 43: 1317-1323.

Hancock, R. E., and H. G. Sahl (2006). Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat Biotechnol 24: 1551-1557.

Hilpert, K., R. Volkmer-Engert, T. Walter, and R. E. Hancock (2005). High-throughput generation of small antibacterial peptides with improved activity. Nat Biotechnol 23: 1008-1012.

Hilpert, K., D. F. Winkler, and R. E. Hancock (2007). Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. Nat Protoc 2: 1333-1349.

Lee, S., A. Hinz, E. Bauerle, A. Angermeyer, K. Juhaszova, Y. Kaneko, P. K. Singh, and C. Manoil (2009). Targeting a bacterial stress response to enhance antibiotic action. Proc Natl Acad Sci U S A 106: 14570-14575.

Lewis, K. (2010). Persister cells. Annu Rev Microbiol 64: 357-372.

Oka, T., S. Sakamoto, K. Miyoshi, T. Fuwa, K. Yoda, M. Yamasaki, G. Tamura, and T. Miyake (1985). Synthesis and secretion of human epidermal growth factor by *Escherichia coli*. Proc Natl Acad Sci U S A 82: 7212-7216.

Pecota, D. C., G. Osapay, M. E. Selsted, and T. K. Wood (2003). Antimicrobial properties of the *Escherichia coli* R1 plasmid host killing peptide. J Biotechnol 100: 1-12.

Sawyer, J. G., N. L. Martin, and R. E. Hancock (1988). Interaction of macrophage cationic proteins with the outer membrane of *Pseudomonas aeruginosa*. Infect Immun 56: 693-698.

Shah, D., Z. Zhang, A. Khodursky, N. Kaldalu, K. Kurg, and K. Lewis (2006). Persisters: a distinct physiological state of E. coli. BMC Microbiol 6: 53.

Shimoni, Y., G. Friedlander, G. Hetzroni, G. Niv, S. Altuvia, O. Biham, and H. Margalit (2007). Regulation of gene expression by small non-coding RNAs: a quantitative view. Mol Syst Biol 3: 138.

Sternberg, C., B. B. Christensen, T. Johansen, A. Toftgaard Nielsen, J. B. Andersen, M. Givskov, and S. Molin (1999). Distribution of bacterial growth activity in flow-chamber biofilms. Appl Environ Microbiol 65: 4108-4117.

Tuerk, C., and L. Gold (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505-510.

Unoson, C., and E. G. Wagner (2008). A small SOS-induced toxin is targeted against the inner membrane in *Escherichia coli*. Mol Microbiol 70: 258-270.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J Mol Biol 225: 487-494.

Wimley, W. C., and K. Hristova (2011). Antimicrobial peptides: successes, challenges and unanswered questions. J Membr Biol 239: 27-34.

Yoon, S. H., S. K. Kim, and J. F. Kim (2010). Secretory production of recombinant proteins in *Escherichia coli*. Recent Pat Biotechnol 4: 23-29.

Zhang, G., S. Brokx, and J. H. Weiner (2006). Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. Nat Biotechnol 24: 100-104.