Toxin-antitoxin systems in bacteria: targets, mechanisms, and specificity

by

Christopher David Aakre

B.S. Biological Sciences
Stanford University, Stanford CA (2008)

SUBMITTED TO THE GRADUATE PROGRAM IN MICROBIOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2015 [June 2015]

Signature of Author: Signature redacted

Christopher David Aakre
Graduate Program in Microbiology
May 21, 2015

Certified by: Signature redacted

Michael T. Laub
Associate Professor of Biology
Thesis supervisor

Accepted by: Signature redacted

Michael T. Laub
Associate Professor of Biology
Co-Director, Microbiology Graduate Program

Toxin-antitoxin systems in bacteria: targets, mechanisms, and specificity

by

Christopher David Aakre

Submitted to the Graduate Program in Microbiology
on May 21, 2015 in partial fulfillment of the requirement for the degree of
Doctor of Philosophy in Microbiology at the Massachusetts Institute of Technology

ABSTRACT

Toxin-antitoxin (TA) systems are genetic modules widely present on bacterial chromosomes. These systems comprise a toxin and cognate antitoxin that are encoded together in an operon; normally, the toxin and antitoxin are synthesized and form a non-toxic complex. Under times of stress, however, the more labile antitoxin can be degraded, which frees the toxin to inhibit growth. TA systems have been implicated in a number of important processes, including plasmid stability, phage resistance, persistence, and virulence. Yet, there are a number of unanswered questions about these genetic modules. What are the cellular targets of toxins? How do antitoxins antagonize their cognate toxins? Do toxins and antitoxins interact in a one-to-one manner – one antitoxin for one toxin – or do they form large networks of cross-reacting systems?

To answer these questions, I have studied the targets, mechanisms, and specificity of TA systems in bacteria. For my first project, I identified SocAB, a novel TA system in the bacterium *Caulobacter crescentus*. Unlike canonical TA systems, in which the antitoxin is less stable than the toxin, I found that the toxin SocB is unstable and constitutively degraded by the protease ClpXP. This degradation requires its antitoxin, SocA, which acts a proteolytic adaptor. Furthermore, I found that SocB blocks replication progress through an interaction with the sliding clamp, thus expanding the number of known cellular targets for TA systems. For my second project, I studied interaction specificity in the ParDE TA family. I found that toxins and antitoxins in this family exhibit a strong preference for interacting with their cognate pair, and that specificity is determined by a small subset of coevolving residues at the interface of these two proteins. To understand how the identity of these coevolving residues controls interaction specificity, I generated a library of ~10^4 variants at these coevolving positions in the ParD antitoxin. By reacting this library against both cognate and non-cognate ParE toxins, I identified promiscuous ParD variants that are densely connected to specific variants in sequence space. These promiscuous states may facilitate changes in TA specificity and promote the expansion of these paralogous systems by duplication and divergence.

Thesis Supervisor: Michael T. Laub
Title: Associate Professor of Biology

"Let's go exploring!"

- Calvin & Hobbes

ACKNOWLEDGEMENTS

To my scientific mentor, Michael Laub, for your constant support and encouragement. Your enthusiasm for science is infectious, and you are one of the most brilliant science communicators I know. You also put up with my wacky workout crazes.

To my committee members, Alan Grossman and Wendy Gilbert, for traveling on this road with me. I appreciate your scientific, career, and personal guidance. Thank you Andrew Camilli for serving on my defense committee.

To members of the Laub lab. Thank you Diane for always having an open ear. You're one of the most brilliant and warm people I know. Thank you Anna for joining the lab with me and then abandoning me for greener and better pastures. I love getting coffee with you and pretending that we're adults. Thank you to the original room three crew, Josh and Andy, for the late nights. I'm sorry that I left you for my manifest destiny. Thank you Emma and Christos for rekindling my love of cats. Thank you Katie, Salazar, Peter, and Conor for our joint entrepreneurial brainstorming sessions. I'm sure one of our ventures will take off. Finally, thank you Anjana, Leonor, and Tung for keeping my science sharp.

To my mentees. Thank you Tanya for accompanying me on my first graduate journey. You were a joy to mentor, and I'm excited to see you progress through your own graduate career. Thank you David for the coffee and pastry breaks. You made me feel better by showing me that things could always be worse – I could be a medical student.

To my friends outside of the lab. Thank you Nicole for your sass and endless support. Thank you Peter for our case studying and thesis writing parties. You're the most serious friend that I don't take seriously. Thank you Tim and Adam for our impromptu brunches. You help keep my neuroses in check. Thank you to the Graduate Student Council folks – Kendall, Francesco, Shabnam, and others – for broadening my view of graduate life at MIT. Finally, thanks to friends from across the river at HBS – Caroline, Annie, and Alex. You always nodded understandingly when I talked about my research, which I took as a positive sign.

To my family. Thank you to my parents for understanding that calls are sometimes hard to pick up when I'm pipetting. Thank you to my sister, Lindsey, for starting a waffle tradition that will never die. I'll give you the secret waffle recipe before I leave Boston. Thank you to Dean and Tracy for treating me to wonderful dinners. Thank you to my soon-to-be new family, Nina, Phil, David, and Allie. You make holidays wonderful and are endlessly supportive.

To Jonathan. You were there for my last thesis, and now this one. You've kept me balanced and healthy when all my instincts point in the opposite direction. Without you, none of this would have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction: Bacteria and Toxin-Antitoxin Systems

I. Introduction

Bacteria are present everywhere in the environment – in the earth's soil and oceans, on our skin and in our guts. Often too small to be seen by the naked eye, over two million of them can fit on the head of a pin.[1] But collectively, bacteria pack a punch. Their mass – if we could collect all bacteria on the planet one-by-one and pile them onto a single, enormous scale – would surpass that of both plants and animals (Whitman et al., 1998). Their environmental impact is broad, as they contribute to global nutrient cycling through nitrogen fixation and the decomposition of organic matter. They also routinely impact human health. Bacteria in our gut have been shown to train the immune system to fight infections, aid in the breakdown of nutrients, and help protect us from metabolic disease (Fujimura et al., 2010). Bacteria can also make us sick. Tuberculosis alone kills 1.5 million people per year and has infected more than two billion people worldwide (Russell et al., 2010).

Ultimately, the ability of bacteria to perform these diverse tasks is dictated by their DNA, and a major challenge in microbiology is how to precisely map the relationship between bacterial DNA (genotype) and particular traits (phenotype). For example, which genes contribute to the ability of a bacterium to digest the sugar lactose? Or are required for the formation of a stress-resistant spore? How do bacteria measure their cell density? These types of fundamental questions have guided microbiology research over the last century, and we've made tremendous progress on many of them. For example, foundational work in *Escherichia coli* revealed the role of the *lac* operon in lactose fermentation; the *spo* genes required for sporulation have been mapped and temporally ordered in *Bacillus subtilis*; and the genes that contribute to the light-producing

[1] Typical pinhead is 2 mm^2, and the area of a typical *E. coli* cell is 1 μm^2.

quorum sensing system in *Vibrio fischeri* have been discovered (Errington, 1993; Miller and Bassler, 2001; Muller-Hill, 1996).

But what about genes that haven't been experimentally characterized as those above? Over the last decade, the advent of next-generation sequencing has resulted in an explosion of fully sequenced bacterial genomes, yet the number of genes that have been experimentally characterized represent only a fraction of the total. Even a decade after the genome sequencing of *Escherichia coli*, the most well-studied workhorse of molecular biology, 46% of its genes were "orphans" with no experimental evidence that pointed to a particular biological role (Riley et al., 2006). Thus, a remaining challenge is to understand the function of these uncharacterized genes in bacterial physiology, with a particular emphasis on large gene families of unknown function.

The focus of my thesis work has been on one such poorly characterized gene family – toxin-antitoxin (TA) genes in bacteria. Toxins are normally globular proteins that inhibit an essential process in the cell, and antitoxins can be proteins or RNA that neutralize the lethality of their cognate toxin. These systems were first identified on plasmids and found to contribute to plasmid stability (Ogura and Hiraga, 1983). However, they were subsequently found to be widely present on bacterial genomes, comprising up to 2.6% of the genome of certain free-living bacteria (Leplae et al., 2011). Despite the abundance of TA systems on bacterial chromosomes, their biological function is mostly unknown. Why are these systems present in so many copies in bacteria? When are they activated and how do they promote bacterial fitness? Additionally, how do toxins inhibit bacterial growth? How do antitoxins neutralize their cognate toxins? Are these systems insulated from each other, or do they form large, cross-reacting networks?

In this chapter, I review what we currently know about plasmid- and chromosomally-based TA systems. In subsequent chapters, I discuss how my thesis work has built on this foundation to begin answering the questions outlined above. Chapter 2 is devoted to my work on the TA system SocA-SocB, which broadened our knowledge of toxin targets and antitoxin mechanisms. Chapter 3 features my work on interaction specificity in the ParDE TA family, which revealed that antitoxins typically interact with only their cognate toxin. This work also revealed surprising insight into how promiscuous binding states may enable the duplication and divergence of paralogous protein families. And finally, in Chapter 4, I outline future research directions for these projects and for the toxin-antitoxin field in general.

II. Discovery and evolution of toxin-antitoxin systems

A. Discovery on plasmids

Plasmids, which are extrachromosomal, normally require mechanisms for accurate replication and partitioning to daughter cells. In the absence of these mechanisms, plasmids become unstable and can be rapidly lost following cell division. The first toxin-antitoxin system discovered, *ccdA-ccdB*, was identified by a group looking for genetic elements that promote the stability of the mini-F plasmid (Ogura and Hiraga, 1983). They found that the *ccdA-ccdB* genes from mini-F could promote the stability of an *oriC* plasmid when combined with an additional partitioning element. The *ccd* acronym stood for *coupled cell division*, as cells that lost plasmids containing *ccdA-ccdB* became filamentous and appeared to be inhibited for division. The *ccdB* gene (later known as the toxin) was required for the inhibition of division, whereas the *ccdA* gene (later known as the antitoxin) suppressed the inhibitory function of *ccdB*.

Following the identification of *ccdA-ccdB*, another plasmid-stabilizing gene pair was identified on plasmid R1 named *hok-sok* (Gerdes et al., 1986a). Similar to *ccdA-ccdB*, the *hok-sok* genes promoted plasmid stability, and the authors found that the toxicity of *hok* (*host killing*) could be neutralized by the presence of the *sok* gene (*suppressor of killing)* (Gerdes et al., 1988). However, there were some notable differences between *ccdA-ccdB* and *hok-sok*. First, the systems did not bear any homology to one another, indicating that they most likely belonged to different protein families. Second, cells killed by the *hok* gene did not filament, but rather became pale and "ghost-like" by phase microscopy, suggesting that *ccdB* and *hok* have different cellular targets.

The *ccdA-ccdB* and *hok-sok* pairs represent the first toxin-antitoxin systems to be discovered, although they were not yet named "TA systems" in the literature. Rather, they were known as genetic elements that promoted plasmid stability. A number of additional TA systems were identified soon thereafter, all of which had plasmid-stabilizing functions. These included *pemK-pemI* of plasmid R100 (Tsuchimoto et al., 1988), *parD-parE* of plasmid RK2 (Roberts and Helinski, 1992; Roberts et al., 1990), and *phd-doc* of the plasmid-borne prophage P1 (Lehnherr et al., 1993). These systems did not appear related to one another and the toxins often had different effects on cell physiology upon activation.

How did these diverse, unrelated gene pairs promote plasmid stability? The authors who identified *ccdA-ccdB* initially proposed that *ccdB* becomes activated only in cells with a single copy of the plasmid; *ccdB* would then inhibit cell division, giving the plasmid time to replicate before the cell eventually divides (Ogura and Hiraga, 1983). However, this hypothesis was later revised when *ccdB* activation was found to result in cell death and not the transient inhibition of cell division (Jaffé et al., 1985). In this revised model, the plasmid-stabilizing function of *ccdA-*

ccdB is due to the selective killing of plasmid-free cells in the population. These findings were later replicated for the *hok-sok* system, as artificially promoting loss of the R1 plasmid resulted in cell death and a "ghost"-like appearance for the majority of cells (Gerdes et al., 1986a). To highlight the fact that these systems are promoting plasmid maintenance through the selective killing of plasmid-free cells, rather than by promoting plasmid stability per se, the term "post-segregational killing" was proposed to describe their common mechanism of action.

How, then, do toxins only become activated in cells that lose the TA-encoding plasmid? For the *ccdA-ccdB* system, this selective activation is based on a difference in stability between the toxin and antitoxin (Van Melderen et al., 1994). The antitoxin CcdA is more unstable than its toxin CcdB, and as a result *ccdA* must be continually transcribed in order for CcdA levels to be high enough to neutralize CcdB. Cells that lose the mini-F plasmid no longer produce CcdA, and the more stable CcdB is then freed to inhibit growth of the cell. For the *hok-sok* system, this selectivity is not based on a difference in protein stability, but rather RNA stability. The *sok* antitoxin encodes an unstable antisense RNA that inhibits the translation of the *hok* mRNA. Upon loss of plasmid R1, the *sok* RNA is rapidly degraded, which frees cells to begin producing the Hok toxin (Gerdes et al., 1988). The differential stability of antitoxins and toxins is thought to be a common feature shared by many TA systems (Lehnherr and Yarmolinsky, 1995; Tsuchimoto et al., 1992).

B. Discovery on bacterial chromosomes

The discovery of TA systems on plasmids was followed shortly thereafter by their identification on bacterial chromosomes. The first chromosomal homolog identified was the *E. coli relF* gene, later renamed *hokD*, which bore similarity to the *hok* toxin from plasmid R1 (Gerdes et al.,

1986b). Similar to *hok,* transcriptional induction of *relF* resulted in cell death and a "ghost-like" cell appearance. A subsequent homology search revealed the presence of five *hok*-like loci in *E. coli*, all of which appear to be inactivated by either an insertion element, point mutation, or major genetic rearrangement (Pedersen and Gerdes, 1999). The authors noted that these systems may be widely distributed in Gram-negative bacteria, as homologous *hok* sequences were also found in *Agrobacterium* and *Rhizobium* species (Poulsen et al., 1989). The discovery of *hok* homologs was followed by the identification of the *chpA* and *chpB* loci in *E. coli*, which are homologous to the *pemI-pemK* TA system from plasmid R100 (Masuda et al., 1993).

The physiological role of chromosomal TA systems was unclear, as deletions had no apparent phenotype and, unlike their plasmid counterparts, they were not always capable of stabilizing plasmids (Gerdes et al., 1986b; Poulsen et al., 1989). By analogy to plasmid-based TA systems, some authors proposed that these chromosomal systems can prevent the loss of large regions of chromosomal DNA in the absence of selection (Szekeres et al., 2007). Nonetheless, this function could not be ascribed to all chromosomal TA systems, as deletions of chromosomal TA systems are often easily obtained (Fiebig et al., 2010). Following their discovery, chromosomal TA systems have been implicated in a number of important cellular processes, including stress responses, phage resistance, persister cell formation, and the promotion of virulence (Helaine et al., 2014; Maisonneuve et al., 2011; Sberro et al., 2013; Vesper et al., 2011). I delay the further discussion of physiological functions of TA systems until Section V of this chapter.

C. Modern discovery approaches

The success of using homology to identify chromosomal TA systems led to the first comprehensive search for TA loci in 126 sequenced prokaryotic genomes (Pandey and Gerdes,

2005). Using a simple E-value cutoff for BLAST analyses, the authors identified a total of 671 predicted TA loci belonging to seven different families. These systems were broadly distributed among both Gram-positive and Gram-negative bacteria, and were often found clustered next to mobile genetic elements or inside mega-integrons. The authors proposed that this clustering may result in TA systems frequently moving between bacterial chromosomes (Pandey and Gerdes, 2005).

Later work refined this computational approach by using additional information to identify TA systems, including gene size, operon structure, overlap between toxin and antitoxin genes, and association with other known TA systems (Guglielmini et al., 2008; Leplae et al., 2011; Makarova et al., 2009; Sevin and Barloy-Hubler, 2007). These studies expanded known TA system diversity to 12 toxin and 20 antitoxin families, and revealed that TA systems are widely distributed in all bacteria. Interestingly, obligate intracellular bacteria tend to lack TA systems, leading some authors to speculate that the selective pressure to maintain TA systems is lost for bacteria living in constant environments (Pandey and Gerdes, 2005).

Toxin-antitoxin systems have also been identified more recently using a homology-blind experimental approach (Sberro et al., 2013). Microbial genomes used to be sequenced by shotgun cloning, in which random fragments of DNA were cloned into plasmids, propagated in *E. coli,* and then Sanger sequenced. The authors noted that functional toxin-antitoxin systems should exhibit a unique cloning pattern. Clones spanning only the toxin should be absent, as they would result in the production of a toxin without its cognate antitoxin. However, clones that span only the antitoxin or both the toxin and antitoxin should be unaffected. This approach led to the identification of 400 TA systems, of which 123 belonged to eight different families that were

previously uncharacterized. Six of these eight families were experimentally verified to encode functional TA systems, and one was shown to function as an abortive infection system against phage T7. These results highlight the power of experimental approaches to identify novel TA families that are not obvious homologs of those identified on plasmids.

D. Evolution of TA systems

Following the identification of TA systems on chromosomes, one question was how these systems are transmitted. Are they inherited in a lineage-specific fashion? Or are they subject to horizontal gene transfer? There are two lines of evidence that suggest that these systems are primarily spread through horizontal gene transfer. First, TA systems are often found closely associated with recombination sites and mobile genetic elements (Pandey and Gerdes, 2005). For example, all 13 TA systems in *Vibrio cholerae* are located in the super-integron of chromosome II and are flanked by *attC* sites that permit cassette integration. Second, a comprehensive analysis of TA systems revealed that their distribution rarely follows species phylogeny. The order and number of *vapB-vapC* systems is not conserved in the closely related *S. solfataricus* and *S. tokodaii* species (Pandey and Gerdes, 2005), nor are the distribution of TA systems conserved among closely related *E. coli* species and *Salmonella enterica* serovars (Lobato-Márquez et al., 2015; Van Melderen and Saavedra De Bast, 2009).

Another outstanding question is how different toxin and antitoxin families have paired with each other over the course of evolution. Recent evidence has suggested that TA systems have been formed by a "mix and match" of toxins and antitoxins of different folds (Arbing et al., 2010). For example, RelE/YoeB-family toxins are found paired with cognate antitoxins from three structurally unrelated families. The pairing of toxins and antitoxins of different folds makes TA

nomenclature difficult, as by convention, the operon is named after the toxin family. As such, antitoxins from different structural families may have the same name. There are four ParD-ParE pairs in the *Caulobacter* genome, for example, but only two of the ParD antitoxins appear to belong to the ParD family from plasmid RK2 (Fiebig et al., 2010).[2]

III. Classification of toxin-antitoxin systems

There are certain features common to all TA systems. These include that the toxin and antitoxin exist in an operon, that the toxin is a protein capable of blocking growth, and that the antitoxin neutralizes the effect of its cognate toxin. However, this simplification masks much of the underlying diversity present in TA systems. For example, the antitoxin can be an RNA or a protein; additionally, the antitoxin can inhibit its toxin by masking its active site or by degrading the toxin transcript, among other mechanisms. These differences have led to the grouping of TA systems into a total of five mechanistic types (Figure 1.1). As more systems are characterized, the number of types is expected to grow. It is interesting to note, however, that there remains ambiguity what constitutes a TA system. For example, why are restriction-modification systems, which share many of these features, not considered TA systems? This ambiguity will be discussed in more detail at the end of this section.

Type I: An RNA antitoxin that blocks toxin translation

The defining feature of Type I TA systems is the presence of a small RNA antitoxin, encoded on the opposite strand of the toxin, that binds to the toxin transcript to block its translation (Fozo et al., 2008). The mechanism of translation inhibition varies, but the small RNA antitoxin will often

[2] Sequences appear unrelated, although this does not rule out similarity at the structural level.

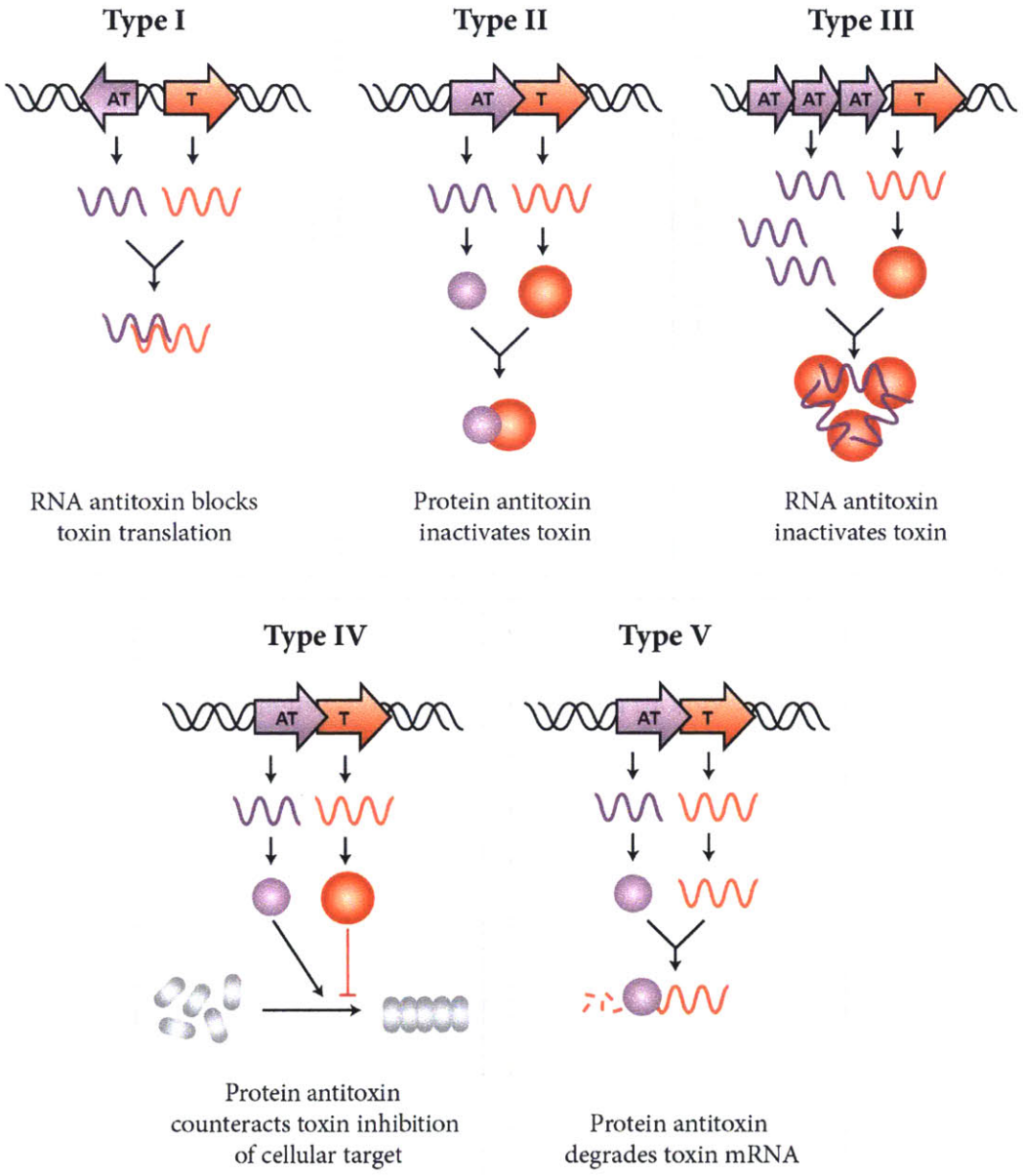


Figure 1.1 The five types of toxin-antitoxin systems.

The different TA system differ by the mechanism of antitoxin. In Type I and Type III TA systems, the antitoxin is an RNA that inactivates its toxin through translational inhibition or by binding the toxin protein. In Type II, IV, and V TA systems, the antitoxin is a protein that inhibits its toxin through sequestration, counteracting the effect of the toxin on its target, or by degrading toxin mRNA. Adapted from (Markovski and Wickner, 2013).

block ribosome binding or promote cleavage of the toxin transcript. For example, the *symR* antitoxin has been shown to base pair with the ribosome binding site of the *symE* transcript, which blocks translation of the toxin and leads to a decrease in transcript stability (Kawano et al., 2007). A more complicated example is the *hok-sok* system. The toxin transcript contains two open reading frames – *mok*, whose translation is required for the translation of *hok*, and *hok*, which encodes the toxin. The *sok* antitoxin competes for binding to the ribosome binding site of *mok*, which then indirectly blocks translation of the *hok* toxin (Thisted and Gerdes, 1992). Binding of *sok* to the *mok-hok* transcript also results in transcript cleavage by RNase III, indicating that antitoxin binding can result in translational inhibition at multiple, independent levels (Gerdes et al., 1992).

The toxins from Type I TA systems are typically small, hydrophobic proteins that contain α-helical transmembrane domains. Overexpression of these toxins can lead to membrane depolarization and cell death (Gerdes et al., 1986b; Weaver et al., 2003). Fractionating experiments revealed that these proteins localize to the inner membrane (Unoson and Wagner, 2008), suggesting that the toxins create pores in the inner membrane that lead to cell lysis. The exception to this model of Type I toxins is SymE, which is a large, non-hydrophobic protein that promotes RNA cleavage (Kawano et al., 2007).

Type II: A protein antitoxin that inactivates the toxin

Type II toxins are the most well-characterized family of TA systems and are defined by a protein antitoxin that forms a stable complex with its cognate toxin. In general, the toxin encodes a globular protein, and the antitoxin encodes an unstable protein that wraps around the toxin to inactivate it. Crystal structures of TA systems from the type II family have helped elucidate how

antitoxins inhibit toxin activity. Some of the antitoxin mechanisms include steric occlusion of the toxin's active site, mimicking of the toxin substrate, or allosteric inhibition of the toxin (Blower et al., 2011a).

The most straightforward mechanism involves direct binding of the antitoxin to the active site of the toxin. For example, the ParD antitoxin binds to conserved hydrophobic grooves on the outer surface of the ParE toxin, which masks hydrophobic residues in the ParE β-sheet that are required to bind to its target, gyrase (Barbosa et al., 2012; Dalton and Crosson, 2010). This binding does not induce conformational changes in the ParE toxin, suggesting that the primary mode of action is steric occlusion. Another interesting example is the YefM antitoxin, which forms a dimer that binds the catalytic site of the endoribonucleases YoeB (Kamada and Hanaoka, 2005). YefM binding results in a conformational change in the active site which presumably abolishes its activity. A final example is the MazE antitoxin, which wraps around the surface of the endoribonucleases MazF. The C-terminus of MazE is thought to act as single-stranded RNA decoy (Kamada et al., 2003). This binding forces out the S1-S2 loop in the active site of MazF that provides the stabilizing histidine of the catalytic triad.

Another mechanism of toxin inactivation involves the allosteric modulation of toxin activity, represented by the CcdA-CcdB family. The toxin CcdB alternates between two conformational states: one that can bind its cellular target, GyrA, and another that can bind its antitoxin, CcdA (De Jonge et al., 2009). Furthermore, CcdA binds to CcdB using two overlapping sites that differ strongly in their affinity. Binding of CcdA to CcdB using its first and high-affinity site results in a conformational change in CcdB that releases CcdB from GyrA in a process known as

"rejuvenation". Binding of CcdA to CcdB using its second and lower-affinity site results in its oligomerization, which is required for transcriptional regulation of the *ccd* operon.

Type III: An RNA antitoxin that inactivates the toxin

The first member of the Type III family was ToxI-ToxN, which was identified based on its homology to phage resistance systems (Fineran et al., 2009). The defining feature of this family is the presence of an RNA antitoxin that binds and neutralizes the toxin protein. This is contrast to Type I systems, in which the RNA antitoxin blocks translation of the toxin transcript. A crystal structure of ToxI-ToxN revealed that three ToxI antitoxin monomers bind to three ToxN toxin monomers to form a trimeric complex (Blower et al., 2011b). ToxI makes extensive RNA-protein interactions with ToxN, including residues in the active site of the endoribonucleases ToxN. More recently, a structure-based computational approach identified a total of three Type III families (Blower et al., 2012a). These families appear to be widely distributed among bacteria, with examples in the Proteobacteria and Firmicutes.

Type IV: A protein antitoxin that counteracts the inhibition of cellular target

In Type IV systems, the antitoxin and toxin proteins do not directly interact. Rather, the antitoxin counteracts the activity of the toxin on its cellular target. The only known example of Type IV systems is YeeU-YeeV, which was discovered based on its similar operon structure to that of other TA systems in *E. coli* (Brown and Shaw, 2003). Like most TA systems, the YeeV toxin inhibited cell growth and the YeeU antitoxin rescued this growth inhibition. YeeV was found to bind the essential cytoskeleton proteins, FtsZ and MreB, and inhibit their polymerization (Tan et al., 2011). However, quite unexpectedly, the toxin and antitoxin proteins were not found to directly interact. Rather, the YeeU antitoxin was found to bind FtsZ and MreB

and counteract the effects of the YeeV toxin (Masuda et al., 2012). Interestingly, YeeU expression lowers YeeV levels to almost undetectable levels (Brown and Shaw, 2003), suggesting that by counteracting its activity, YeeU may make YeeV may be more susceptible to proteolysis.

Type V: A protein antitoxin that degrades the toxin transcript

A search for genes that are required for MqsR-mediated persistence led to the identification of the GhoT-GhoS TA system (Wang et al., 2012). Similar to Type I systems, the toxin GhoT encodes a small hydrophobic protein that disrupts cell membrane polarity. However, the antitoxin GhoS is not an RNA, but rather a protein that specifically cleaves the GhoT transcript. Structural analysis of GhoS revealed that it is an endoribonuclease with a ferrodoxin-like fold. Unlike endoribonuclease toxins, however, GhoS must exhibit high sequence specificity for the GhoT transcript, as its production does not block cell growth. GhoS-GhoT is currently the only described TA system from the Type V family.

Ambiguity in the classification of TA systems

There are other genetic modules that bear a striking similarity to toxin-antitoxin systems, but for historical reasons are not traditionally considered bona fide TA pairs. For example, restriction-modification systems encode an endonuclease, which cleaves DNA at a particular recognition sequence, and a methyltransferase, which catalyzes the addition of methyl groups to the same recognition sequence (Wilson and Murray, 1991). The methylation of the recognition sequence prevents cleavage by the endonuclease, thus making the methyltransferase an "antitoxin" to the endonuclease "toxin." Furthering their similarity to TA systems, restriction-modification pairs are encoded in an operon, promote plasmid stability, are found on both plasmids and chromosomes, and protect against phage infection (Naito et al., 1995; Wilson and Murray, 1991).

However, these systems were discovered in the 1960s before the identification of TA systems (Meselson and Yuan, 1968), and thus have not been re-characterized as TA systems following the development of the current mechanistic TA types.

Other groups have proposed that metabolic modules can also be considered as a type of TA system (Amato et al., 2013). The alarmone "toxin" (p)ppGpp slows down bacterial growth, and the enzyme "antitoxin" SpoT can break down (p)ppGpp to relieve the blocking of growth. This broader definition of TA systems has not gained widespread use in the literature.

IV. Cellular targets of toxins

Toxins target a range of processes within the cell in order to block bacterial growth. These include essential functions such as DNA replication, translation, and cell wall growth (Figure 1.2). Some of the essential processes that are targeted by TA systems are reviewed below.

A. DNA gyrase inhibitors

The first confirmed target of a toxin was DNA gyrase, and this discovery occurred almost a decade after TA systems were initially characterized (Bernard and Couturier, 1992). A selection for mutants resistant to the CcdB toxin identified mutations in *gyrA*, which encodes one of the subunits of DNA gyrase that catalyzes the ATP-dependent negative supercoiling of DNA. The authors found that CcdB binding traps gyrase in a "cleavable complex", which is an intermediate state that occurs after gyrase-mediated DNA cleavage but prior to re-ligation. Trapping of gyrase in the cleavable complex leads to double strand DNA breaks, a mechanism analogous to that of the quinolone class of antibiotics (Bernard et al., 1993). However, mutants resistant to CcdB are

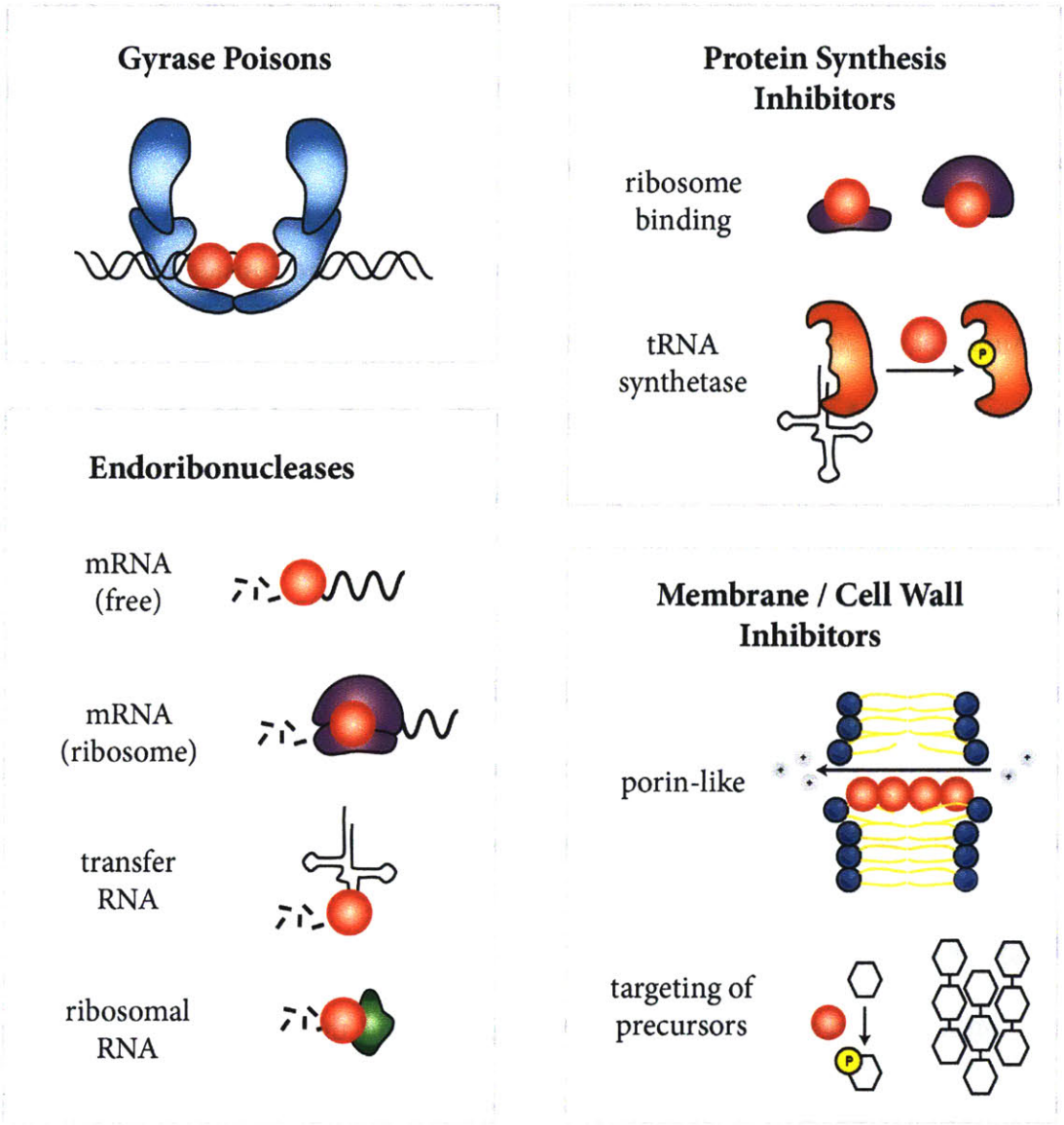


Figure 1.2 Cellular targets of toxins

Toxins target essential cellular processes such as DNA replication, translation, or cell wall biogenesis. Gyrase poisons include toxins such as CcdB and ParE. Toxins that cleave RNA, or endoribonucleases, fall into different classes depending on whether they target free mRNA (MazF), ribosome-associated mRNA (RelE), transfer RNA (VapC), or ribosomal RNA (MazF). Protein synthesis inhibitors can bind directly to ribosomes (Doc) or inhibit the charging of tRNAs by phosphorylating tRNA synthetase (HipA). Toxins that target the cell wall can disrupt membrane polarity by acting like a porin (all type V toxins, most type I toxins) or by phosphorylating peptidoglycan precursors (PezT).

not resistant to quinolones and vice versa, suggesting that the two gyrase poisons may target different regions of the protein.

Another toxin that targets DNA gyrase is ParE from the plasmid RK2 (Jiang et al., 2002). ParE was shown to inhibit replication of *oriC* DNA, and further analysis revealed that ParE blocked the supercoiling activity of DNA gyrase *in vitro*. Similar to quinolones and CcdB, ParE appears to trap DNA gyrase in a cleavable complex (Yuan et al., 2010). However, in contrast to other gyrase poisons, ParE requires ATP to inhibit gyrase, and mutants of gyrase that are resistant to quinolones or CcdB are not resistant to ParE. These results highlight that there are a number of distinct surfaces by which toxins inhibit DNA gyrase, but that each toxin described to date traps DNA gyrase in the cleavable complex , resulting in the formation of DNA double strand breaks.

B. Endoribonucleases

The most commonly studied class of toxins are endoribonucleases, which are toxins that cleave cellular RNAs to block translation. To date, there are three classes of RNA molecules targeted by endoribonucleases: messenger RNAs (ribosome-associated or unbound), transfer RNAs, and ribosomal RNA. Depending on the cleavage specificity of the toxin, targeting of these RNAs can result in either a partial or complete block in translation.

Toxins that cleave mRNAs can be divided into the ribosome-dependent (requires binding to ribosome to cleave mRNAs) and the ribosome-independent (can cleave free mRNAs). The best characterized ribosome-dependent endoribonuclease is RelE, which was hypothesized to be a translational inhibitor based on the phenotype of mutants in its antitoxin *relB* (Gotfredsen and Gerdes, 1998). RelE was later shown to associate with ribosomes and block translation (Galvani

et al., 2001; Pedersen et al., 2002). The mechanism of translation inhibition was refined when RelE was shown to occupy the ribosomal A site and cleave mRNAs between the second and third site in the codon (Neubauer et al., 2009; Pedersen et al., 2003). Other endoribonucleases do not appear to require the ribosome for their function. For example, induction of MazF blocks global translation in a reversible manner and results in mRNA cleavage (Pedersen et al., 2002; Zhang et al., 2003). However, MazF was shown to cleave target transcripts *in vitro* in the absence of ribosomes, indicating that MazF can cleave unbound mRNAs (Zhang et al., 2003).

Beyond targeting mRNAs, endoribonucleases are also known to block translation by targeting tRNA molecules. For example, the VapB-VapC family is one of the most abundant on bacterial chromosomes, and the VapC toxin was shown to inhibit translation (Winther and Gerdes, 2009). In contrast to other endoribonucleases, however, VapC was not found associated with ribosomes, nor did its production result in the cleavage of mRNA, rRNA, or tmRNA (Winther and Gerdes, 2011). Instead, VapC was shown to cleave the initiator transfer RNA ($tRNA^{fMet}$) in its anticodon stem-loop. This cleavage was specific for the initiator tRNA, as the ectopic overproduction of $tRNA^{fMet}$ was sufficient to counteract the toxicity of VapC *in vivo*. This mechanism may be shared by many VapC toxins, as VapC from *Leptospira interrogans* was also found to cleave the initiator tRNA (Lopes et al., 2014).

The last known target of endoribonucleases is ribosomal RNA. Recently, MazF was shown to cleave 16S rRNA at the 3' terminus, possibly resulting in ribosomes that are missing the anti-Shine-Dalgarno sequence (Vesper et al., 2011). The authors proposed that mRNAs are also cleaved downstream of their Shine-Dalgarno sequence, resulting in a sub-class of leaderless transcripts that are translated specifically by ribosomes with truncated 16S rRNAs. MazF from

Mycobacterium tuberculosis also cleaves ribosomal RNA to block translation, but targets the 23S rRNA instead of 16S rRNA (Schifano et al., 2013). Cleavage occurs at a single position in the ribosomal A site, destabilizing the association between the 50S and 30S ribosomal subunits.

C. Protein Synthesis Inhibitors

Endoribonuclease toxins block translation through the cleavage of RNA, but there are also toxins that inhibit translation through other mechanisms. For example, the toxin Doc blocks translation through an interaction with the 30S subunit but does not cleave mRNA (Liu et al., 2008). Rather, Doc binding results in the stabilization of polysomes and mRNA. The authors proposed that Doc arrests translation during elongation, and consistent with this model, Doc appears to share a binding site on the ribosome with another replication elongation inhibitor, hygromycin B. Another translation inhibitor is HipA, which is a kinase that was originally thought to phosphorylate EF-Tu (Schumacher et al., 2009). Two groups later contradicted these findings, demonstrating that HipA instead phosphorylates GltX, the enzyme responsible for charging $tRNA^{Glu}$ (Germain et al., 2013; Kaspy et al., 2013). This phosphorylation leads to the accumulation of uncharged tRNAs and activation of the stringent response.

D. Membrane and Cell Wall Inhibitors

Another essential cell process targeted by toxins is membrane integrity and cell wall biosynthesis. Most toxins of the type I class are small, hydrophobic proteins that localize to the inner cell membrane (Fozo et al., 2008; Unoson and Wagner, 2008). Induction of these toxins results in loss of membrane integrity and a cell "ghosting" phenotype, which suggests that they insert themselves into cell membranes using a porin-like mechanism (Gerdes et al., 1986b; Weaver et al., 2003). This mechanism appears to be shared by GhoT, which is a toxin of the type V class

(Wang et al., 2012). In addition to inhibiting membrane integrity, toxins can also target cell wall biosynthesis. The main example of this class is the zeta toxin. The crystal structure of this toxin revealed that zeta adopts a phosphotransferase-like fold, and that the toxicity is dependent on residues in the active site of the enzyme (Meinhart et al., 2003). Lysis of cells following zeta induction suggested that the toxin targets the cell wall, but the precise target was unclear. A crucial hint came with the observation that the zeta toxin is also toxic to eukaryotes such as *S. cerevisiae (Zielenkiewicz et al., 2009),* suggesting that zeta targets a cell wall component shared by bacteria and eukaryotes. One such component is the peptidoglycan precursor uridine diphosphate-*N*-acetylglucosamine (UNAG), which was shown to be directly phosphorylated by the PezT toxin (a member of the zeta family) (Mutschler et al., 2011). Interesting, phosphorylated UNAG is not only inactive as a peptidoglycan precursor, but is also capable of inhibiting MurA, an essential enzyme in the peptidoglycan biosynthesis pathway. Thus, it appears that targeting UNAG not only depletes metabolic precursors for peptidoglycan biosynthesis but also forms a competitive inhibitor for the pathway itself.

V. Functions of toxin-antitoxin systems

Toxin-antitoxin systems from TA systems were originally identified based on their ability to stabilize plasmids, but an unresolved question is whether they are performing other, currently undescribed functions. Do plasmid-based TA systems perform other functions? What about chromosomal systems, which are presumably not required for the stable maintenance of bacterial chromosomes? Why are chromosomal TA systems present in so many seemingly redundant copies? These questions are currently under debate, but a number of models have emerged that attempt to explain the ubiquitous presence of these systems on bacterial chromosomes.

A. Abortive infection systems

Bacteriophage are viruses that infect bacteria, and bacteria have a number of mechanisms to protect themselves from phage infection. One such mechanism is abortive infection, in which an infected cell altruistically kills itself in order to prevent the virus from spreading in the population (Samson et al., 2013). The first experimental evidence linking TA systems to abortive infection was for the *hok-sok* system of plasmid R1. High-copy production of *hok-sok* was found to provide partial protection against T4 phage infection (Pecota and Wood, 1996). A later search for genes with homology to abortive infection systems resulted in the identification of ToxI-ToxN, a TA system from an *Erwinia carotovora* plasmid (Fineran et al., 2009). ToxI-ToxN was found to provide broad resistance to a number of different *E. carotovora* phages tested (13 out of 25). The authors hypothesized that phage-mediated inhibition of transcription would block production of the ToxI antitoxin, thus leading to ToxN-mediated cell death of infected cells. Interestingly, phages appear to have evolved mechanisms to evade ToxI-ToxN-mediated abortive infection. Certain phages produce a molecular mimic of the ToxI RNA, which ensures that the ToxN toxin remains inactive (Blower et al., 2012b). Another example is phage T4, which encodes an antitoxin mimic capable of binding and inactivating multiple toxins (Otsuka and Yonesaki, 2012).

B. Promotion of physiological downshifts

A recurring model in the literature is that toxins are activated in times of physiological downshifts such as starvation or translational stress. This model is based on the observation that most antitoxins are quite unstable; as such, blocking new RNA or protein production should result in the rapid clearance of antitoxins from the cell and the activation of toxins. In support of

this hypothesis, starving cells or blocking translation results in the transcriptional activation of the *relBE* system in *E. coli* (Christensen et al., 2001). Transcriptional activation can be a readout for TA system activation, as the toxin-antitoxin complex is often a repressor of its own transcription (Li et al., 2008). Interestingly, the starvation-induced activation of *relBE* is dependent on the protease Lon, suggesting that Lon degrades the antitoxin RelB in times of stress. Bulk translation levels during starvation are also higher in Δ*relBE* cells, suggesting that RelE activation during starvation can dampen translation at a global level (Christensen et al., 2001). Transcriptional activation during starvation is also observed for a number of other TA systems in *E. coli* (Christensen et al., 2003; Christensen-Dalsgaard et al., 2010). However, the deletion of these TA systems has not yet been shown to increase fitness during starvation, so the purpose or advantage of this activation is not yet clear.

C. Generation of persister cells

Persister cells are those that exhibit a transient, non-heritable resistance to antibiotic exposure. They were first identified by Joseph Bigger in his study of *Staphylococcus* sensitivity to penicillin (Bigger, 1944).[3] A screen for mutants that exhibit a higher level of persistence identified the *hipA7* mutant, which generated ampicillin-resistant persisters at a frequency of 10^{-2} compared to 10^{-6} for wild-type cells (Moyed and Bertrand, 1983). Interestingly, *hipA* encodes the toxin of the *hipB-hipA* TA system, which provided the first link between TA systems and bacterial persistence (Black et al., 1994). The high-persistence phenotype of *hipA7* was dependent on the stringent

[3] Bigger observed that a very small fraction of *Staphylococcus* cells survive exposure to penicillin, and these survivors are just as sensitive to antibiotics upon outgrowth and re-exposure. The survivors must have therefore been transiently resistant to the effects of the antibiotic.

response alarmone, (p)ppGpp, suggesting that the stringent response and bacterial persistence may be somehow connected (Korch et al., 2003).

A more concrete model linking TA systems to persister cell generation was provided by the Gerdes group in a series of recent publications (Germain et al., 2015; Maisonneuve et al., 2011, 2013). Their key observation is that cells lacking ten of the *E. coli* endoribonuclease toxins (Δ10 TA) have lower levels of persistence than wild-type cells, suggesting that stochastic activation of toxins contributes to persister cell formation (Maisonneuve et al., 2011). They later reported that toxin activation is dependent on a signaling hierarchy that starts with the alarmone (p)ppGpp, and then cascades through polyphosphate, the protease Lon, and antitoxin degradation (Maisonneuve et al., 2013). This model, if true, also nicely explains why the HipA toxin has been linked to persistence. HipA phosphorylates glutamyl-tRNA synthetase, which leads to the accumulation of uncharged $tRNA^{Glu}$ and the production of (p)ppGpp (Germain et al., 2013; Kaspy et al., 2013). The production of (p)ppGpp would then lead to the activation of TA systems, as per the Gerdes model (Germain et al., 2015).

The attractiveness of this model notwithstanding, there are a few caveats in interpreting these results. First, these persister assays are performed at high density in LB medium when cells are undergoing a growth rate transition. *E. coli* growth slows down around an OD of 0.3 in LB due to the lack of a utilizable carbon source (Sezonov et al., 2007). This growth reduction is problematic for persister assays, as *E. coli* cells are known to increase persister cell formation by orders of magnitude during carbon source transitions (Amato et al., 2013). Second, genome sequencing of the Δ10 TA strain revealed the presence of an additional 10 kb deletion not reported in the

literature (unpublished data).[4] If the deletion affects the rate at which these cells enter stationary phase, this would complicate the measurement of persister frequencies. Third, other groups have attempted to replicate the Gerdes persister measurements in more a controlled, steady-state environment with no success (L. Van Melderen, personal communication), suggesting that their results may be specific to carbon source transitions. As such, the relationship between TA systems and persistence is still under investigation.

D. Promotion of virulence

A search for toxin-antitoxin systems in *Mycobacterium tuberculosis* identified over 88 predicted systems, which is among the highest for sequenced bacteria genomes (Ramage et al., 2009). There are three observations that suggest that these TA systems may be involved in *M. tuberculosis* pathogenicity. First, these systems are conserved in members of the virulent *M. tuberculosis* complex, but largely absent from non-pathogenic *Mycobacterium* species (Ramage et al., 2009). Second, many of these systems are transcriptionally induced during *in vivo* infection (Singh et al., 2010). Finally, deletion of multiple *mazEF* systems reduces survival of *M. tuberculosis* during nutritional and oxidative stress, and also impairs *M. tuberculosis* virulence in a guinea pig model of infection (Tiwari et al., 2015). The mechanism that leads to an increase in survival is not known, but the authors speculate that TA systems may help active the stress regulon and promote dormancy in the host.

TA systems may also promote virulence in a number of other species. In uropathogenic *E. coli*, the *yefM-yoeB* and *ybaJ-hha* TA systems promote colonization of the bladder (Norton and Mulvey, 2012). In addition, TA systems appear to promote the survival of *Haemophilus*

[4] The missing 10 kb region includes the genes *djlB, ybeT, ybeU, djlC, hscC, rihA, gltL, gltK, gltJ,* and *gltI*

influenzae and *Salmonella enterica* during infection (Lobato-Márquez et al., 2015; Ren et al., 2012). It was recently proposed that the internalization of Salmonella by macrophages may induce persister formation, and this induction was partially dependent on each of the 14 TA systems present (Helaine et al., 2014). The identification of a phenotype for so many single TA deletion mutants was surprising, as most phenotypes have only been observed for strains deleted of multiple TA systems (Maisonneuve et al., 2011).

VI. Specificity in toxin-antitoxin systems

The abundance of TA systems in bacteria, often in multiple copies per chromosomes, raises an interesting question: to what extent are these systems specific for their cognate partners? This specificity could be encoded at multiple levels. For example, TA systems could exhibit specificity at the level of toxin-antitoxin interaction (one toxin for one antitoxin, and vice versa) or at the level of cross activation (activation of one TA system does not lead to the activation of other TA systems). I review some of the evidence for specificity at each of these levels below.

A. Toxin-antitoxin interaction specificity

For type II systems, which are composed of a protein toxin and antitoxin, binding between cognate pairs has been well documented (Aizenman et al., 1996; Galvani et al., 2001; Maki et al., 1996). Crystal structures of type II complexes demonstrate that toxins and antitoxins make extensive contact across their interface, sometimes totaling over 6000 $Å^2$ of shared surface area (Dalton and Crosson, 2010). Given the extent of these contacts, are toxins and antitoxins specific for their cognate pairs? Or are they capable of binding other toxins and antitoxins present in the cell?

Genetic data suggests that these interactions may be specific, as deleting an antitoxin is generally lethal unless it is supplied in *trans* on a plasmid (Fiebig et al., 2010). Certain antitoxins have been observed to only neutralize their cognate toxins, although these analyses are often limited to only three or four cognate pairs (Hallez et al., 2010; Ramage et al., 2009). In contrast, other groups have found that cognate and non-cognate pairs are capable of interacting *in vivo* and *in vitro* (Yang et al., 2010; Zhu et al., 2010). These non-cognate interactions often occur between toxin families, leading the authors to hypothesize that toxins and antitoxins may form large, promiscuous networks that alter cellular physiology in response to environmental cues (Zhu et al., 2010). The conflicting evidence for interaction specificity will thus require more large-scale approaches to understanding whether toxin-antitoxin interactions are specific.

B. Toxin-antitoxin cross activation

In addition to interactions between non-cognate toxins and antitoxins, there are other potential opportunities for cross-reactivity between TA systems. For example, the activation of one toxin could lead to the activation of other TA systems within the cell. Evidence for this model came from a recent study that looked at the transcriptional activation of TA systems in response to amino acid starvation. Many TA complexes negatively regulate their own promoter, and as such, transcriptional activation is often an indicator of toxin activation (Cataudella et al., 2012; Overgaard et al., 2008). The authors found that while both *mazEF* and *relBE* are transcriptionally activated during starvation, the activation of *mazEF* did not occur in a Δ*relBE* mutant (Kasari et al., 2013). Furthermore, ectopic production of the endoribonuclease RelE resulted in the transcriptional activation of many different toxins, including *mazF, yoeB,* and *yhaV*. This cross-activation did not appear to be a general function of shutting down translation, as strains

deficient in proteases showed the same effect. Rather, these toxins were found to cleave the transcripts of other TA systems, resulting in the specific accumulation of toxin-encoding fragments. Endoribonuclease toxins thus have the potential to form complicated cross-reacting networks with unknown implications for bacterial physiology.

VII. Conclusion

Toxins and antitoxins exhibit diverse mechanisms and targets within the cell in order to modulate bacterial physiology. At the beginning of my graduate career, I started by working on the mechanisms and targets of TA systems. I describe this work in Chapter 2, which focuses on the SocA-SocB TA system in *Caulobacter crescentus*. I found that the antitoxin SocA neutralizes its toxin by a completely novel mechanism, and that the toxin SocB targets the replicative sliding clamp to block replication progression. Following my work on SocA-SocB, I worked on characterizing the specificity of protein-protein interactions in TA systems. I describe this work in Chapter 3, which focuses on the ParD-ParE family of TA systems. I found that interactions in this family are incredibly specific, and I identified a small set of coevolving residues that are sufficient, when mutated, to reprogram interaction specificity. Additionally, I made a library in these coevolving residues in the ParD antitoxin and mapped the sequence space of functional variants. This work revealed an abundance of promiscuous ParD variants that are densely connected in sequence space. In Chapter 4, I speculate on future research directions for these projects and the toxin-antitoxin field.

References

Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (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.

Amato, S.M., Orman, M.A., and Brynildsen, M.P. (2013). Metabolic Control of Persister Formation in Escherichia coli. Mol. Cell *50*, 475–487.

Arbing, M.A., Handelman, S.K., Kuzin, A.P., Verdon, G., Wang, C., Su, M., Rothenbacher, F.P., Abashidze, M., Liu, M., Hurley, J.M., et al. (2010). Crystal structures of Phd-Doc, HigA, and YeeU establish multiple evolutionary links between microbial growth-regulating toxin-antitoxin systems. Struct. Lond. Engl. 1993 *18*, 996–1010.

Barbosa, L.C.B., Garrido, S.S., Garcia, A., Delfino, D.B., Santos, L. do N., and Marchetto, R. (2012). Design and synthesis of peptides from bacterial ParE toxin as inhibitors of topoisomerases. Eur. J. Med. Chem. *54*, 591–596.

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

Bernard, P., Kézdy, K.E., Van Melderen, L., Steyaert, J., Wyns, L., Pato, M.L., Higgins, P.N., and Couturier, M. (1993). The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase. J. Mol. Biol. *234*, 534–541.

Bigger, J.W. (1944). Treatment of staphylococcal infections with penicillin by intermittent treatment. Lancet 497–500.

Black, D.S., Irwin, B., and Moyed, H.S. (1994). Autoregulation of hip, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis. J. Bacteriol. *176*, 4081–4091.

Blower, T.R., Salmond, G.P.C., and Luisi, B.F. (2011a). Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin-antitoxin partners. Curr. Opin. Struct. Biol. *21*, 109–118.

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

Blower, T.R., Short, F.L., Rao, F., Mizuguchi, K., Pei, X.Y., Fineran, P.C., Luisi, B.F., and Salmond, G.P.C. (2012a). Identification and classification of bacterial Type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes. Nucleic Acids Res. *40*, 6158–6173.

Blower, T.R., Evans, T.J., Przybilski, R., Fineran, P.C., and Salmond, G.P.C. (2012b). Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. PLoS Genet. *8*, e1003023.

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

Cataudella, I., Trusina, A., Sneppen, K., Gerdes, K., and Mitarai, N. (2012). Conditional cooperativity in toxin-antitoxin regulation prevents random toxin activation and promotes fast translational recovery. Nucleic Acids Res. *40*, 6424–6434.

Christensen, S.K., Mikkelsen, M., Pedersen, K., and Gerdes, K. (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., Pedersen, K., Hansen, F.G., and Gerdes, K. (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.

Christensen-Dalsgaard, M., Jørgensen, M.G., and Gerdes, K. (2010). Three new RelE-homologous mRNA interferases of Escherichia coli differentially induced by environmental stresses. Mol. Microbiol. *75*, 333–348.

Dalton, K.M., and Crosson, S. (2010). A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex. Biochemistry (Mosc.) *49*, 2205–2215.

Errington, J. (1993). Bacillus subtilis sporulation: regulation of gene expression and control of morphogenesis. Microbiol. Rev. *57*, 1–33.

Fiebig, A., Castro Rojas, C.M., Siegal-Gaskins, D., and Crosson, S. (2010). Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin-antitoxin systems. Mol. Microbiol. *77*, 236–251.

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., and Salmond, G.P.C. (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., Hemm, M.R., and Storz, G. (2008). Small toxic proteins and the antisense RNAs that repress them. Microbiol. Mol. Biol. Rev. MMBR *72*, 579–589, Table of Contents.

Fujimura, K.E., Slusher, N.A., Cabana, M.D., and Lynch, S.V. (2010). Role of the gut microbiota in defining human health. Expert Rev. Anti Infect. Ther. *8*, 435–454.

Galvani, C., Terry, J., and Ishiguro, E.E. (2001). Purification of the RelB and RelE proteins of Escherichia coli: RelE binds to RelB and to ribosomes. J. Bacteriol. *183*, 2700–2703.

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

Gerdes, K., Bech, F.W., Jørgensen, S.T., Løbner-Olesen, A., Rasmussen, P.B., Atlung, T., Boe, L., Karlstrom, O., Molin, S., and von Meyenburg, K. (1986b). Mechanism of postsegregational

killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon. EMBO J. *5*, 2023–2029.

Gerdes, K., Helin, K., Christensen, O.W., and Løbner-Olesen, A. (1988). Translational control and differential RNA decay are key elements regulating postsegregational expression of the killer protein encoded by the parB locus of plasmid R1. J. Mol. Biol. *203*, 119–129.

Gerdes, K., Nielsen, A., Thorsted, P., and Wagner, E.G. (1992). Mechanism of killer gene activation. Antisense RNA-dependent RNase III cleavage ensures rapid turn-over of the stable hok, srnB and pndA effector messenger RNAs. J. Mol. Biol. *226*, 637–649.

Germain, E., Castro-Roa, D., Zenkin, N., and Gerdes, K. (2013). Molecular mechanism of bacterial persistence by HipA. Mol. Cell *52*, 248–254.

Germain, E., Roghanian, M., Gerdes, K., and Maisonneuve, E. (2015). Stochastic induction of persister cells by HipA through (p)ppGpp-mediated activation of mRNA endonucleases. Proc. Natl. Acad. Sci. U. S. A.

Gotfredsen, M., and Gerdes, K. (1998). The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family. Mol. Microbiol. *29*, 1065–1076.

Guglielmini, J., Szpirer, C., and Milinkovitch, M.C. (2008). Automated discovery and phylogenetic analysis of new toxin-antitoxin systems. BMC Microbiol. *8*, 104.

Hallez, R., Geeraerts, D., Sterckx, Y., Mine, N., Loris, R., and Van Melderen, L. (2010). New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in Escherichia coli O157:H7. Mol. Microbiol. *76*, 719–732.

Helaine, S., Cheverton, A.M., Watson, K.G., Faure, L.M., Matthews, S.A., and Holden, D.W. (2014). Internalization of Salmonella by Macrophages Induces Formation of Nonreplicating Persisters. Science *343*, 204–208.

Jaffé, A., Ogura, T., and Hiraga, S. (1985). Effects of the ccd function of the F plasmid on bacterial growth. J. Bacteriol. *163*, 841–849.

Jiang, Y., Pogliano, J., Helinski, D.R., and Konieczny, I. (2002). ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase. Mol. Microbiol. *44*, 971–979.

De Jonge, N., Garcia-Pino, A., Buts, L., Haesaerts, S., Charlier, D., Zangger, K., Wyns, L., De Greve, H., and Loris, R. (2009). Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. Mol. Cell *35*, 154–163.

Kamada, K., and Hanaoka, F. (2005). Conformational Change in the Catalytic Site of the Ribonuclease YoeB Toxin by YefM Antitoxin. Mol. Cell *19*, 497–509.

Kamada, K., Hanaoka, F., and Burley, S.K. (2003). Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. Mol. Cell *11*, 875–884.

Kasari, V., Mets, T., Tenson, T., and Kaldalu, N. (2013). Transcriptional cross-activation between toxin-antitoxin systems of Escherichia coli. BMC Microbiol. *13*, 45.

Kaspy, I., Rotem, E., Weiss, N., Ronin, I., Balaban, N.Q., and Glaser, G. (2013). HipA-mediated antibiotic persistence via phosphorylation of the glutamyl-tRNA-synthetase. Nat. Commun. *4*, 3001.

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

Korch, S.B., Henderson, T.A., and Hill, T.M. (2003). Characterization of the hipA7 allele of Escherichia coli and evidence that high persistence is governed by (p)ppGpp synthesis. Mol. Microbiol. *50*, 1199–1213.

Lehnherr, H., and Yarmolinsky, M.B. (1995). Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *92*, 3274–3277.

Lehnherr, H., Maguin, E., Jafri, S., and Yarmolinsky, M.B. (1993). Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. J. Mol. Biol. *233*, 414–428.

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

Li, G.-Y., Zhang, Y., Inouye, M., and Ikura, M. (2008). Structural mechanism of transcriptional autorepression of the Escherichia coli RelB/RelE antitoxin/toxin module. J. Mol. Biol. *380*, 107–119.

Liu, M., Zhang, Y., Inouye, M., and Woychik, N.A. (2008). Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. Proc. Natl. Acad. Sci. U. S. A. *105*, 5885–5890.

Lobato-Márquez, D., Moreno-Córdoba, I., Figueroa, V., Díaz-Orejas, R., and García-Del Portillo, F. (2015). Distinct type I and type II toxin-antitoxin modules control Salmonella lifestyle inside eukaryotic cells. Sci. Rep. *5*, 9374.

Lopes, A.P.Y., Lopes, L.M., Fraga, T.R., Chura-Chambi, R.M., Sanson, A.L., Cheng, E., Nakajima, E., Morganti, L., and Martins, E.A.L. (2014). VapC from the leptospiral VapBC toxin-antitoxin module displays ribonuclease activity on the initiator tRNA. PloS One *9*, e101678.

Maisonneuve, E., Shakespeare, L.J., Jørgensen, M.G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. Proc. Natl. Acad. Sci. *108*, 13206–13211.

Maisonneuve, E., Castro-Camargo, M., and Gerdes, K. (2013). (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. Cell *154*, 1140–1150.

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2009). Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. Biol. Direct *4*, 19.

Maki, S., Takiguchi, S., Horiuchi, T., Sekimizu, K., and Miki, T. (1996). Partner switching mechanisms in inactivation and rejuvenation of Escherichia coli DNA gyrase by F plasmid proteins LetD (CcdB) and LetA (CcdA). J. Mol. Biol. *256*, 473–482.

Markovski, M., and Wickner, S. (2013). Preventing bacterial suicide: a novel toxin-antitoxin strategy. Mol. Cell *52*, 611–612.

Masuda, H., Tan, Q., Awano, N., Wu, K.-P., and Inouye, M. (2012). YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in Escherichia coli. Mol. Microbiol. *84*, 979–989.

Masuda, Y., Miyakawa, K., Nishimura, Y., and Ohtsubo, E. (1993). chpA and chpB, Escherichia coli chromosomal homologs of the pem locus responsible for stable maintenance of plasmid R100. J. Bacteriol. *175*, 6850–6856.

Meinhart, A., Alonso, J.C., Sträter, N., and Saenger, W. (2003). Crystal structure of the plasmid maintenance system epsilon/zeta: functional mechanism of toxin zeta and inactivation by epsilon 2 zeta 2 complex formation. Proc. Natl. Acad. Sci. U. S. A. *100*, 1661–1666.

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

Van Melderen, L., Bernard, P., and Couturier, M. (1994). Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. Mol. Microbiol. *11*, 1151–1157.

Meselson, M., and Yuan, R. (1968). DNA restriction enzymes from E. coli. Nature 1110–1114.

Miller, M.B., and Bassler, B.L. (2001). Quorum Sensing in Bacteria. Annu. Rev. Microbiol. *55*, 165–199.

Moyed, H.S., and Bertrand, K.P. (1983). hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. J. Bacteriol. *155*, 768–775.

Muller-Hill, B. (1996). The Lac Operon: A Short History of a Genetic Paradigm.

Mutschler, H., Gebhardt, M., Shoeman, R.L., and Meinhart, A. (2011). A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. PLoS Biol. *9*, e1001033.

Naito, T., Kusano, K., and Kobayashi, I. (1995). Selfish behavior of restriction-modification systems. Science *267*, 897–899.

Neubauer, C., Gao, Y.-G., Andersen, K.R., Dunham, C.M., Kelley, A.C., Hentschel, J., Gerdes, K., Ramakrishnan, V., and Brodersen, D.E. (2009). The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE. Cell *139*, 1084–1095.

Norton, J.P., and Mulvey, M.A. (2012). Toxin-Antitoxin Systems Are Important for Niche-Specific Colonization and Stress Resistance of Uropathogenic Escherichia coli. PLoS Pathog *8*, e1002954.

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

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

Overgaard, M., Borch, J., Jørgensen, M.G., and Gerdes, K. (2008). Messenger RNA interferase RelE controls relBE transcription by conditional cooperativity. Mol. Microbiol. *69*, 841–857.

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

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

Pedersen, K., and Gerdes, K. (1999). Multiple hok genes on the chromosome of Escherichia coli. Mol. Microbiol. *32*, 1090–1102.

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

Pedersen, K., Zavialov, A.V., Pavlov, M.Y., Elf, J., Gerdes, K., and Ehrenberg, M. (2003). The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. Cell *112*, 131–140.

Poulsen, L.K., Larsen, N.W., Molin, S., and Andersson, P. (1989). A family of genes encoding a cell-killing function may be conserved in all gram-negative bacteria. Mol. Microbiol. *3*, 1463–1472.

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

Ren, D., Walker, A.N., and Daines, D.A. (2012). Toxin-antitoxin loci vapBC-1 and vapXD contribute to survival and virulence in nontypeable Haemophilus influenzae. BMC Microbiol. *12*, 263.

Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K.B., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., et al. (2006). Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. Nucleic Acids Res. *34*, 1–9.

Roberts, R.C., and Helinski, D.R. (1992). Definition of a minimal plasmid stabilization system from the broad-host-range plasmid RK2. J. Bacteriol. *174*, 8119–8132.

Roberts, R.C., Burioni, R., and Helinski, D.R. (1990). Genetic characterization of the stabilizing functions of a region of broad-host-range plasmid RK2. J. Bacteriol. *172*, 6204–6216.

Russell, D.G., VanderVen, B.C., Lee, W., Abramovitch, R.B., Kim, M.-J., Homolka, S., Niemann, S., and Rohde, K.H. (2010). Mycobacterium tuberculosis wears what it eats. Cell Host Microbe *8*, 68–76.

Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. *11*, 675–687.

Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U., and Sorek, R. (2013). Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. Mol. Cell *50*, 136–148.

Schifano, J.M., Edifor, R., Sharp, J.D., Ouyang, M., Konkimalla, A., Husson, R.N., and Woychik, N.A. (2013). Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. Proc. Natl. Acad. Sci. U. S. A. *110*, 8501–8506.

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

Sevin, E.W., and Barloy-Hubler, F. (2007). RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. Genome Biol. *8*, R155.

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

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

Szekeres, S., Dauti, M., Wilde, C., Mazel, D., and Rowe-Magnus, D.A. (2007). Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. Mol. Microbiol. *63*, 1588–1605.

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

Thisted, T., and Gerdes, K. (1992). Mechanism of post-segregational killing by the hok/sok system of plasmid R1. Sok antisense RNA regulates hok gene expression indirectly through the overlapping mok gene. J. Mol. Biol. *223*, 41–54.

Tiwari, P., Arora, G., Singh, M., Kidwai, S., Narayan, O.P., and Singh, R. (2015). MazF ribonucleases promote Mycobacterium tuberculosis drug tolerance and virulence in guinea pigs. Nat. Commun. *6*.

Tsuchimoto, S., Ohtsubo, H., and Ohtsubo, E. (1988). Two genes, pemK and pemI, responsible for stable maintenance of resistance plasmid R100. J. Bacteriol. *170*, 1461–1466.

Tsuchimoto, S., Nishimura, Y., and Ohtsubo, E. (1992). The stable maintenance system pem of plasmid R100: degradation of PemI protein may allow PemK protein to inhibit cell growth. J. Bacteriol. *174*, 4205–4211.

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

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H., and Moll, I. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli. Cell *147*, 147–157.

Wang, X., Lord, D.M., Cheng, H.-Y., Osbourne, D.O., Hong, S.H., Sanchez-Torres, V., Quiroga, C., Zheng, K., Herrmann, T., Peti, W., et al. (2012). A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. Nat. Chem. Biol. *8*, 855–861.

Weaver, K.E., Weaver, D.M., Wells, C.L., Waters, C.M., Gardner, M.E., and Ehli, E.A. (2003). Enterococcus faecalis plasmid pAD1-encoded Fst toxin affects membrane permeability and alters cellular responses to lantibiotics. J. Bacteriol. *185*, 2169–2177.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. U. S. A. *95*, 6578–6583.

Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. Annu. Rev. Genet. *25*, 585–627.

Winther, K.S., and Gerdes, K. (2009). Ectopic production of VapCs from Enterobacteria inhibits translation and trans-activates YoeB mRNA interferase. Mol. Microbiol. *72*, 918–930.

Winther, K.S., and Gerdes, K. (2011). Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. Proc. Natl. Acad. Sci. U. S. A. *108*, 7403–7407.

Yang, M., Gao, C., Wang, Y., Zhang, H., and He, Z.-G. (2010). Characterization of the interaction and cross-regulation of three Mycobacterium tuberculosis RelBE modules. PloS One *5*, e10672.

Yuan, J., Sterckx, Y., Mitchenall, L.A., Maxwell, A., Loris, R., and Waldor, M.K. (2010). Vibrio cholerae ParE2 poisons DNA gyrase via a mechanism distinct from other gyrase inhibitors. J. Biol. Chem. *285*, 40397–40408.

Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G., and Inouye, M. (2003). MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. Mol. Cell *12*, 913–923.

Zhu, L., Sharp, J.D., Kobayashi, H., Woychik, N.A., and Inouye, M. (2010). Noncognate Mycobacterium tuberculosis toxin-antitoxins can physically and functionally interact. J. Biol. Chem. *285*, 39732–39738.

Zielenkiewicz, U., Kowalewska, M., Kaczor, C., and Ceglowski, P. (2009). In vivo interactions between toxin-antitoxin proteins epsilon and zeta of streptococcal plasmid pSM19035 in Saccharomyces cerevisiae. J. Bacteriol. *191*, 3677–3684.

Chapter 2

A Bacterial Toxin Inhibits DNA Replication Elongation Through a Direct Interaction with the β Sliding Clamp

This work was published as Aakre, C.D., Phung, T.N., Huang, D., and Laub, M.T. Mol. Cell *52*, 617–628 on December 12, 2013.

Summary

Toxin-antitoxin (TA) systems are ubiquitous on bacterial chromosomes, yet the mechanisms regulating their activity, and the molecular targets of toxins, remain incompletely understood. Here, we identify SocAB, a new TA system in *Caulobacter crescentus*. Unlike canonical TA systems, the toxin SocB is unstable and constitutively degraded by the protease ClpXP; this degradation requires the antitoxin, SocA, which acts as a proteolytic adaptor. We find that the toxin, SocB, blocks replication elongation through an interaction with the sliding clamp, driving replication fork collapse. Mutations that suppress SocB toxicity map to either the hydrophobic cleft on the clamp that binds DNA polymerase III or a clamp-binding motif in SocB. Thus, our results suggest that SocB disrupts replication by outcompeting other clamp-binding proteins. Collectively, our results expand the diversity of mechanisms employed by TA systems to regulate toxin activity and to inhibit bacterial growth, and they further suggest that inhibiting clamp function may be a generalizable antibacterial strategy.

Introduction

Toxin-antitoxin (TA) systems are genetic modules that are widely present on plasmids and bacterial chromosomes, with some species encoding more than 50 TA pairs (Pandey and Gerdes, 2005). Each TA system is typically comprised of a toxin and its cognate antitoxin that are encoded together in an operon; normally, both the toxin and antitoxin are synthesized and form a stable, non-toxic complex. However, under stressful conditions, the more labile antitoxin can be degraded, freeing the stable toxin to inhibit bacterial growth (Wang et al., 2011). TA systems are implicated in a number of important cellular processes, including plasmid stability (Ogura and Hiraga, 1983), bacterial persistence (Maisonneuve et al., 2011), stress responses (Vesper et al., 2011) and resistance to phage (Sberro et al., 2013).

The toxins from characterized TA systems inhibit growth by targeting a relatively limited set of critical cellular functions such as translation, DNA replication, or cell wall growth. For instance, the toxin Doc inhibits translation by binding the 30S ribosomal subunit (Liu et al., 2008), whereas other toxins, such as MazF, block translation by cleaving mRNAs (Zhang et al., 2003). The toxins CcdB and ParE inhibit replication by binding DNA gyrase, which is required to relieve supercoiling that occurs ahead of the replication fork (Yuan et al., 2010). Additionally, the PezT toxin blocks cell wall growth by phosphorylating peptidoglycan precursors, thereby inhibiting the initial step in peptidoglycan synthesis (Mutschler et al., 2011). Strikingly, small-molecule antibiotics inhibit cellular proliferation by targeting a very similar set of processes, including translation (aminoglycosides), DNA replication (fluoroquinolones), and cell wall growth (β-lactams) (Walsh, 2003). The recent rise in antibiotic resistance (Bush et al., 2011), and the lack of diversity of current antibiotics (Coates et al., 2011), highlight the need to identify new

targets within bacteria to inhibit their proliferation. The study of bacterial toxins may help elucidate such targets for the development of novel small-molecule therapeutics.

The DNA replication machinery, or replisome, could be a prime target, but is surprisingly not targeted by any known TA systems or antibiotics in clinical use (Robinson et al., 2012). Bacterial DNA replication is catalyzed by a multi-component complex known as DNA polymerase III (Pol III). The Pol III core (subunit composition αεθ) is weakly processive on its own and can only incorporate 1-10 nucleotides per binding event (Johnson and O'Donnell, 2005). To increase its processivity, Pol III core associates with the β sliding clamp, DnaN, a ring-shaped protein that encircles DNA topologically. Binding to DnaN increases the processivity of Pol III over three orders of magnitude (Johnson and O'Donnell, 2005; Maki and Kornberg, 1988). In *E. coli,* DnaN also associates with DNA Pols I, II, IV, and V (Indiani et al., 2005); the mismatch repair proteins MutS and MutL (López de Saro et al., 2006); and the replication regulator Hda (Kurz et al., 2004). The interaction of these proteins with DnaN is required for a number of processes, including translesion synthesis (Lenne-Samuel et al., 2002) and mismatch repair (Lenhart et al., 2013). These clamp-binding proteins contain variants of a short peptide motif, QL[SD]LF, that mediates binding to a conserved hydrophobic cleft on DnaN (Dalrymple et al., 2001). DnaN thus forms a central hub for DNA replication and repair in bacteria.

Here, we identify an atypical TA system in *Caulobacter crescentus,* SocAB. We find that the toxin SocB is normally unstable and constitutively degraded by the protease ClpXP. In contrast to canonical TA systems, in which the antitoxin neutralizes its toxin by sequestration, we find that the antitoxin SocA neutralizes SocB by acting as an adaptor for the degradation of SocB by ClpXP. The requirement of ClpXP for SocB degradation explains for the first time why ClpXP is

essential for viability in *Caulobacter*. Additionally, we provide evidence that SocB inhibits replication elongation through a direct interaction with DnaN. Mutations in DnaN or SocB that block their association occur in the hydrophobic cleft in DnaN or in a DnaN-binding motif in SocB, suggesting that SocB binds to DnaN in a similar manner as known clamp-binding proteins and competes for binding to DnaN during replication. In sum, our work elucidates novel mechanisms employed by TA systems to regulate toxin function and to inhibit cellular proliferation. Our results further suggest that protein interaction hubs such as DnaN may be ideal targets for the development of new protein- or small-molecule-based antimicrobials.

Results

Mutations in the toxin *socB* can bypass the essentiality of *clpXP*

ClpXP is a widely conserved AAA+ protease that uses the power of ATP hydrolysis to unfold and proteolyze substrates within cells (Sauer and Baker, 2011). ClpXP is comprised of two proteins: the unfoldase ClpX, which recognizes and unfolds substrates, and the peptidase ClpP, which degrades the unfolded substrates that it receives from ClpX. Unlike most bacteria, *clpX* and *clpP* are essential for viability in *Caulobacter* (Jenal and Fuchs, 1998). To identify factors responsible for *clpP* essentiality, we selected for transposon insertions that allow cells to grow in the absence of *clpP* expression. We identified multiple transposon insertions in a hypothetical gene (CCNA_03629) that we named *socB* for *s*uppressor *o*f *clpXP* (Figure 2.1A). A clean deletion of *socB* allowed cells to grow in the absence of either *clpX* or *clpP* expression (Figure 2.1B); however, growth was slower on medium that repressed *clpX* or *clpP*, consistent with the fact that ClpXP degrades a range of cellular proteins (Bhat et al., 2013).

To better assess the ability of *socB* mutants to suppress *clpX* essentiality, we used a *clpX* depletion strain and performed a time-course experiment following the switch to non-inducing medium. Using immunoblotting, we found that ClpX gradually declined to almost undetectable levels after 10 hours (Figure 2.1C) (Jenal and Fuchs, 1998). In $socB^+$ cells, the depletion of ClpX coincided with an increase in cellular filamentation (Figure 2.1D) and a more than 1000-fold decrease in colony-forming units (CFUs) (Figure 2.1E). However, in cells lacking *socB,* we observed only intermediate filamentation (Figure 2.1D) and no drop in CFUs (Figures 2.1E). These data support the conclusion that a *socB* deletion bypasses the essentiality of *clpXP.*

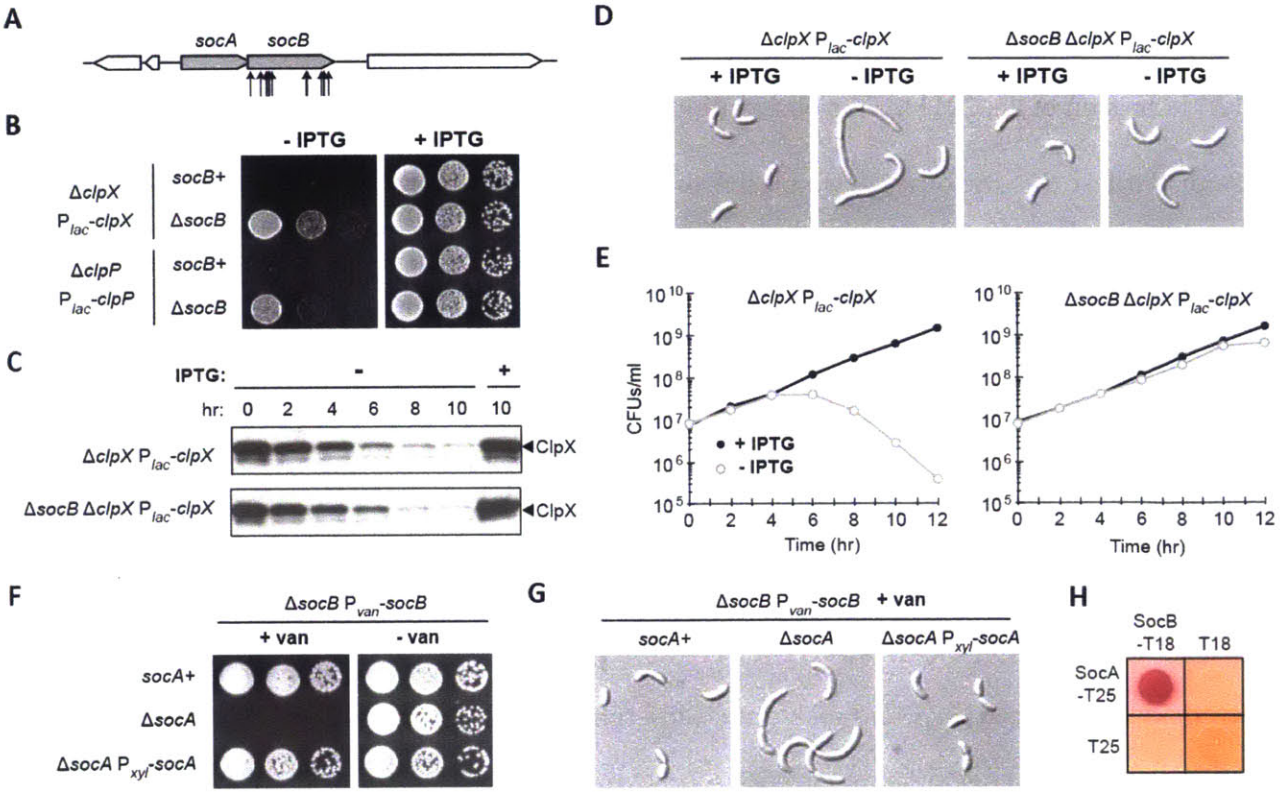


Figure 2.1. Mutations in the toxin *socB* bypass ClpXP essentiality

(**A**) Schematic of transposon insertions in *socB* (CCNA_03629) that suppressed the essentiality of *clpP*. (**B**) Growth of *clpX* and *clpP* depletion strains in *socB*+ and Δ*socB* backgrounds. Five-fold serial dilutions of the indicated strains were spotted onto media ± IPTG. (**C**) Kinetics of ClpX depletion. Indicated strains were shifted to media ± IPTG, and samples were subjected to immunoblotting. (**D**) Morphology of cells following ClpX depletion in *socB*+ and Δ*socB* backgrounds. Strains from (C) were imaged by DIC microscopy at 10 hr. (**E**) Viability of cells following ClpX depletion in *socB*+ and Δ*socB* backgrounds. Colony forming units (CFUs)/ml of the strains from (C) are shown; mean of two biological replicates. (**F**) Growth of strains expressing *socB* in the *socA*+ or Δ*socA* backgrounds. The indicated strains were five-fold serially diluted onto media that induces or represses *socB*. (**G**) Morphology of strains from (F). The indicated strains were grown for 4 hr in *socB* inducing conditions and then imaged by DIC microscopy. (**H**) Bacterial two-hybrid analysis of the interaction between SocA and SocB. T18/T25 were included as a negative control; red indicates a positive interaction. Cells were grown for 1 day at 30°C.

socB is present in an operon with an upstream gene, *socA*, that is predicted to be essential (Figure 2.1A) (Christen et al., 2011). This observation raised the possibility that *socAB* may encode a TA system. To test this hypothesis, we placed the putative toxin *socB* under an inducible promoter and tested whether its expression was toxic to cells in the presence or absence of its putative antitoxin *socA*. We found that inducing *socB* in *socA*+ cells had no significant effect on cell viability or morphology (Figures 2.1F and 2.1G). However, inducing *socB* in Δ*socA* cells inhibited colony formation and led to cellular filamentation (Figures 2.1F and 2.1G). These phenotypes could be rescued by expressing *socA* in *trans* from a plasmid behind a different inducible promoter (Figures 2.1F and 2.1G). To test whether *socB* expression is bacteriostatic or bactericidal, we placed *socA* and *socB* under different inducible promoters and measured viability following *socB* expression. *socB* expression caused a log-linear decrease in viability, even when *socA* was induced at later time points, suggesting that *socB* expression is primarily bactericidal (Figures 2.2A and 2.2B). Collectively, our data indicate that *socAB* behaves genetically like other TA systems.

For type II TA systems, the antitoxin functions by forming a complex with its cognate toxin and neutralizing its activity (Yamaguchi et al., 2011). To test whether SocA and SocB directly interact, we used the bacterial two-hybrid system based on complementation of the T18 and T25 fragments of adenylate cyclase (Karimova et al., 1998). We fused *socA* and *socB* to the T25 and T18 fragments, respectively, and co-expressed the gene fusions in *E. coli*. We observed a strong interaction between SocA and SocB, indicating that they likely form a complex (Figure 2.1H). Interestingly, *socB* expression did not kill *E. coli*, indicating that its toxicity may be

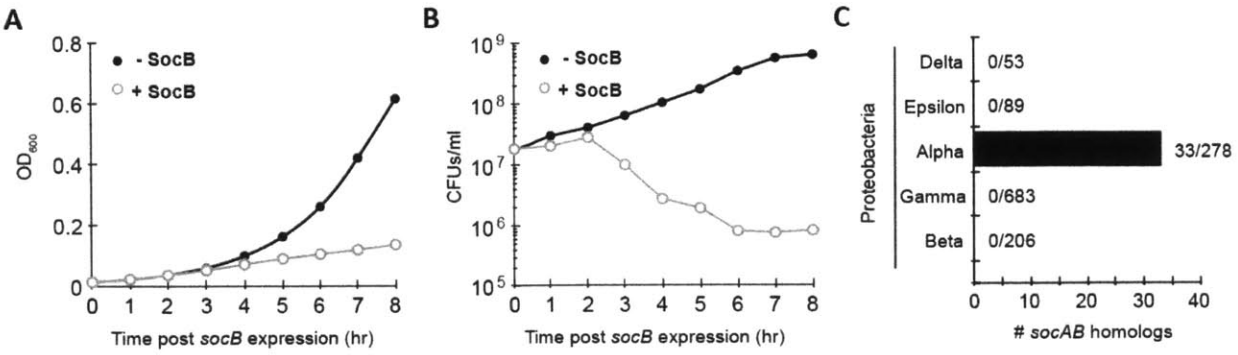


Figure 2.2. Effects of SocB production on viability; evolutionary conservation of *socAB*

(A) Growth curve of Δ*socAB* P_{xyl}-*socA* P_{van}-*socB* cells following vanillate addition at time zero to induce *socB* expression. **(B)** Measurement of colony-forming units (CFUs) from growth curve in (A). At each time point, cells were plated onto medium containing xylose to induce *socA* expression and neutralize any remaining SocB. **(C)** Number of *socAB* homologs in each class of proteobacteria. Homologs were identified by a modified version of reciprocal best hit against fully sequence genomes in the IMG database (September 2012). *socA* and *socB* homologs were first identified independently by reciprocal best hit in each bacterial genome. Afterward, *socA* homologs that also had a *socB* homolog directly downstream were called as *socAB* homologs. Numbers next to each bar indicate number of homologs / number of total species searched.

phylogenetically restricted. Indeed, homologs of *socAB* were identified only in the α-proteobacteria (Figure 2.2C).

SocA promotes SocB degradation by ClpXP

The observation that a deletion of *socB* can bypass the essentiality of ClpXP suggested that SocB may be a ClpXP substrate, and that accumulation of SocB in the absence of ClpXP inhibits growth. To test this possibility, we measured the accumulation of an M2-tagged variant of SocB in the presence or absence of ClpX. Whereas no M2-SocB was detected in the presence of ClpX, M2-SocB accumulation was observed when ClpX was first depleted for 12 hours (Figure 2.3A).

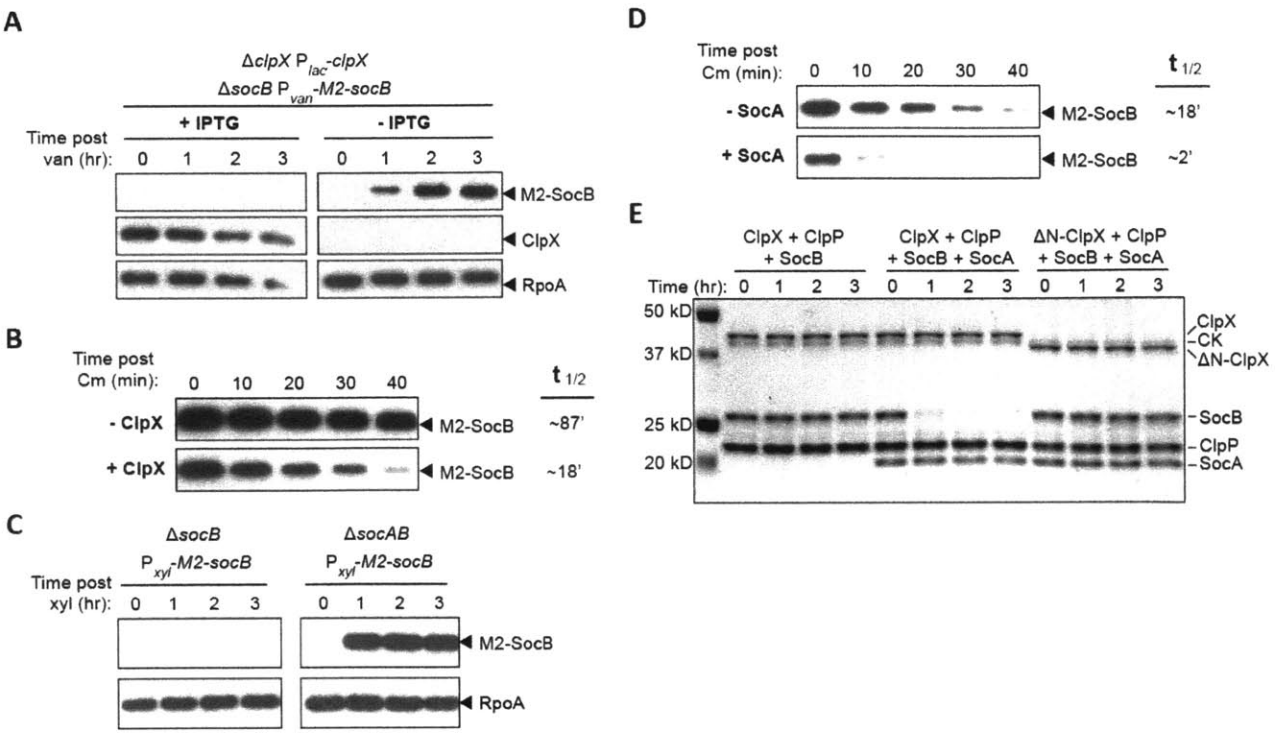


Figure 2.3. SocA promotes SocB degradation by ClpXP

(A) Abundance of M2-SocB ± ClpX assessed by immunoblotting. The indicated strain was grown in *clpX* inducing or repressing conditions for 12 hr and *M2-socB* expression was induced at time zero. RpoA is a loading control. **(B)** Stability of M2-SocB ± ClpX. *clpX* expression was repressed or induced for 12 hr, and then *M2-socB* expression was induced for 30 min prior to chloramphenicol (Cm) addition at time zero to shut off protein synthesis. Half-life quantified from two replicates (see Figure 2.4A). **(C)** Abundance of M2-SocB ± SocA assessed by immunoblotting. *M2-socB* expression was induced at time zero. **(D)** Stability of M2-SocB ± SocA. *M2-socB* expression was induced for 2 hr, and then *socA* expression was induced for an additional 1.5 hr prior to Cm addition at time zero. Half-life quantified from two replicates (see Figure 2.4B). **(E)** *In vitro* degradation of SocB by ClpXP ± SocA. Amounts are as follows: 0.5 μM ClpX or ΔN-ClpX, 1 μM ClpP, 5 μM SocB, 5 μM SocA, 32 μg/ml creatine kinase (CK), 16 mM creatine phosphate, and 4 mM ATP. Reaction performed at 4°C.

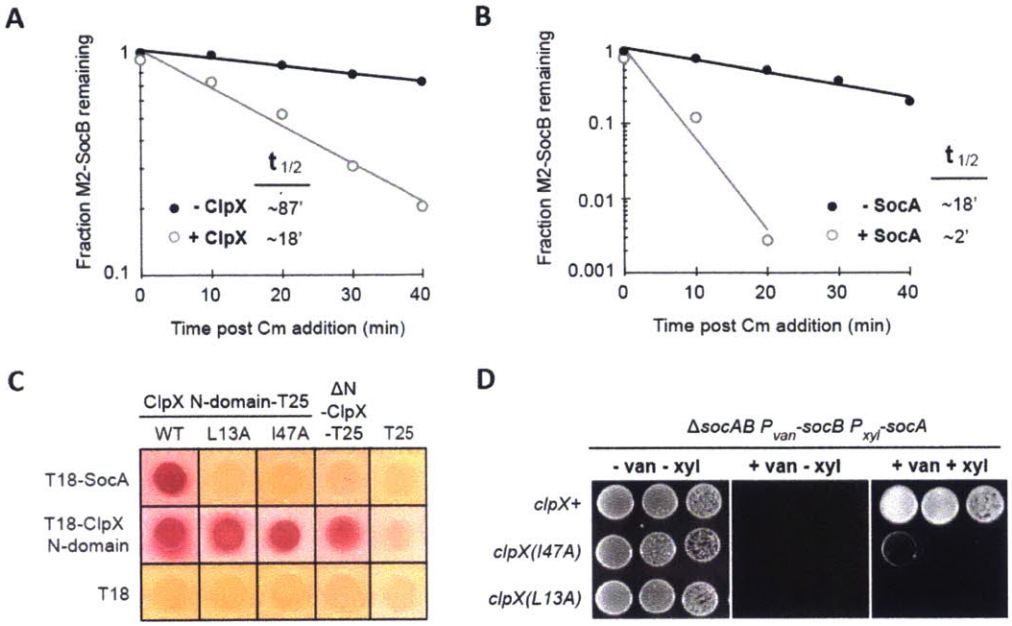


Figure 2.4. Quantification of M2-SocB stability; interaction of SocA with ClpX N-Domain

(A-B) Quantification of (A) M2-SocB stability ± ClpX (from Figure 2.3B) or (B) M2-SocB stability ± SocA (from Figure 2.3D). The average M2-SocB band intensity from two biological replicates was measured and plotted to calculate protein half-life. **(C)** SocA interacts with the ClpX N-domain, and mutations in the ClpX N-domain can abolish this interaction. Candidate residues in the ClpX N-domain were chosen based on the crystal structure of the E. coli ClpX N-domain in complex with the adaptor SspB (PDB ID: 2DS8) under the assumption that SocA and SspB may bind the ClpX N-domain in a similar fashion. We mutated a series of residues in the ClpX N-domain that directly contact SspB, and found that two mutations, L13A and I47A, were sufficient to abolish the ClpX N-domain-SocA interaction. Bacterial two-hybrid interaction is shown between the ClpX N-domain (residues 1-62), ClpX N-domain mutants, ΔN-ClpX (residues 63-420), and SocA. Dimerization between the ClpX N-domain is included as a control, to demonstrate that the L13A and I47A mutations specifically disrupted the SocA-N-domain interaction. Cells were grown for 2 days at 30°C. **(D)** Interaction of SocA with ClpX N-domain is required for SocA to function as an antitoxin in vivo. We generated a chromosomal replacement of *clpX* with *clpX(L13A)* or *clpX(I47A)*, which are mutations that abolish the ability of ClpX to interact with SocA (Figure 2.4C). We then spotted the indicated strains on media that represses both *socA* and *socB* (-van -xyl), induces only *socB* (+van -xyl), or induces both *socA* and *socB* (+van +xyl). Five-fold serial dilutions are shown.

To confirm that the decreased abundance of SocB in the presence of ClpX results from a change in protein stability, we produced M2-SocB from a plasmid and measured its half-life in the presence or absence of ClpX using a chloramphenicol shut-off assay. We found that the presence of ClpX reduced the half-life of M2-SocB from ~87 to ~18 min, indicating that SocB is likely a ClpXP substrate (Figures 2.3B and 2.4A).

Based on these results, we reasoned that SocB is normally present at low levels due to constitutive degradation by ClpXP. What role, then, does SocA play in the neutralization of SocB? Antitoxins of TA systems typically neutralize their cognate toxins by forming a stable complex (Yamaguchi et al., 2011). However, given that SocB is normally unstable, SocA may instead neutralize SocB by promoting its degradation. Indeed, we observed that M2-SocB accumulated in a strain lacking *socA*, but not in a *socA*+ strain (Figure 2.3C). To test whether SocA affects the stability of SocB, we measured the half-life of M2-SocB with and without *socA* expression from a low-copy plasmid. We found that SocA reduced the half-life of M2-SocB from ~18 to ~2 min, indicating that SocA promotes the degradation of SocB (Figures 2.3D and 2.4B). The half-life of M2-SocB in the presence of SocA (~2 min) was shorter than that measured above (~18 min), presumably due to differences in *socA* expression from a low-copy plasmid compared to its native chromosomal locus.

We hypothesized that SocA may be an adaptor for SocB degradation by ClpXP. Canonical adaptors, such as SspB, tether their substrates to the N-domain of ClpX. This tethering increases substrate concentration around the ClpX pore, which concomitantly increases the rate of substrate degradation (Dougan et al., 2003; Levchenko et al., 2000). To test whether SocA is a proteolytic adaptor, we purified SocA and SocB and performed an *in vitro* degradation reaction

with ClpXP. When SocB was combined with ClpXP alone, no degradation was observed (Figure 2.3E). However, when an equimolar amount of SocA was added to the reaction, we observed robust degradation of SocB, indicating that SocA promotes the degradation of SocB by ClpXP (Figure 2.3E).

We performed the same reaction with a variant of ClpX that lacks the N-domain (residues 1-62) and, consequently, is catalytically active but deficient in adaptor-mediated degradation (Dougan et al., 2003). We observed no detectable degradation of SocB in the presence of SocA, ΔN-ClpX, and ClpP, indicating that the N-domain of ClpX is required for SocA to promote SocB proteolysis (Figure 2.3E). Further, we observed a direct interaction between the ClpX N-domain and SocA (Figure 2.4C), and found that mutations in the ClpX N-domain that abolish the interaction between ClpX and SocA prevent SocA from functioning as an antitoxin *in vivo* (Figure 2.4D). Collectively, these data indicate that SocA is an adaptor for the degradation of SocB by ClpXP and they explain why *clpXP* and *socA* are essential for viability in *Caulobacter.*

Accumulation of SocB blocks replication elongation

Why is the accumulation of SocB toxic to *Caulobacter* cells? To better study SocB function, we sought to develop a stabilized variant that is toxic even in the presence of SocA and ClpXP. Because ClpX often recognizes the free C-terminus of its substrates (Flynn et al., 2003), we tested the effect of appending an M2 tag to SocB. We found that SocB-M2 was stabilized over 40-fold relative to M2-SocB (Figure 2.5A) and was toxic to *socA*+ *clpXP*+ cells (Figure 2.5B). This stabilized variant was subsequently used to assess the effects of SocB on cellular physiology.

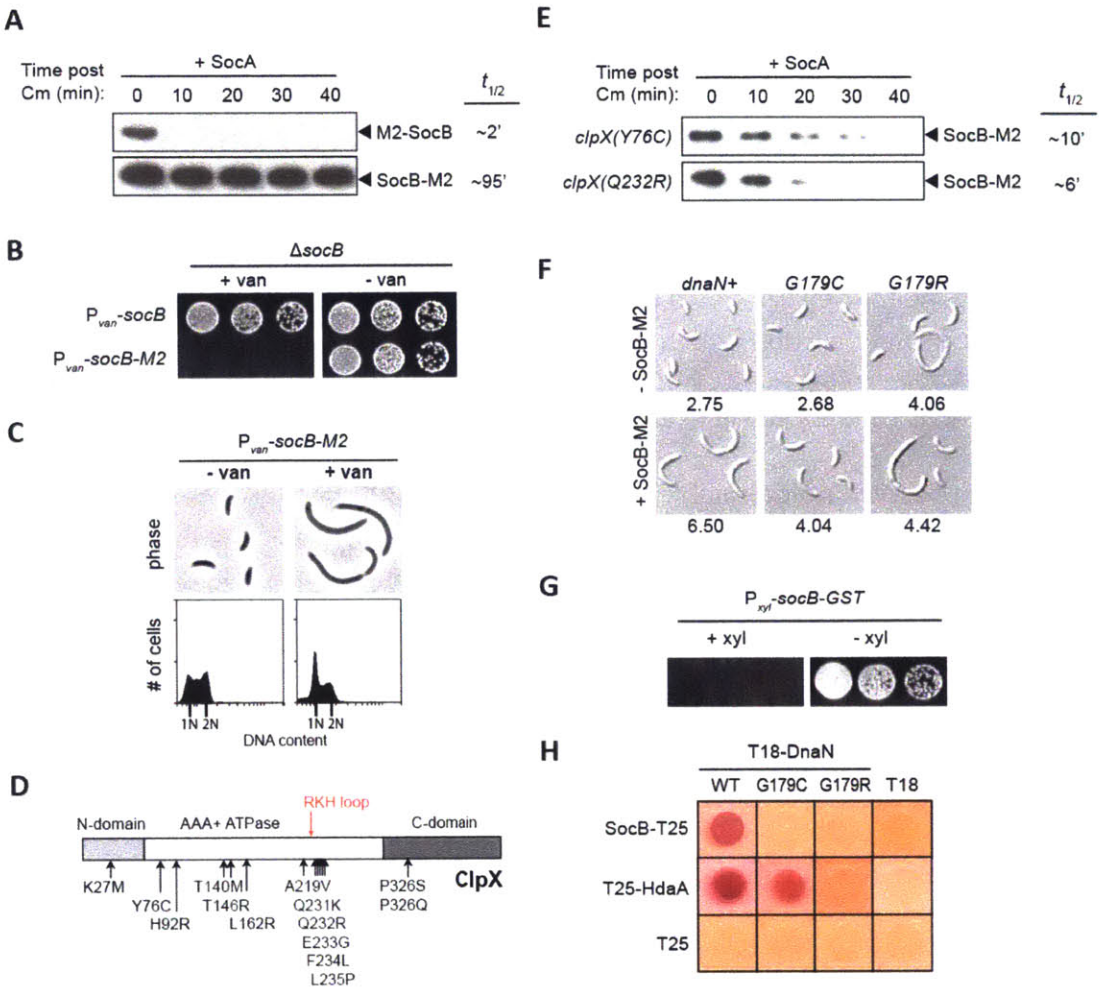


Figure 2.5. Suppressor mutations in clpX that destabilize SocB-M2; phenotype of suppressors in *dnaN*; interaction data between SocB, DnaN, and HdaA

(A) Stability of M2-SocB or SocB-M2 in the presence of SocA. *M2-socB* or *socB-M2* expression was induced for 2 hr, and then *socA* expression was induced for an additional 1.5 hr prior to Cm addition at time zero to shut off protein synthesis. Half-life was quantified from two replicates. **(B)** Growth of indicated strains on media that induces (+van) or represses (-van) expression of *socB* or *socB-M2*. Five-fold serial dilutions are shown. **(C)** Phase microscopy and DNA content of the indicated strain in the absence or presence of *socB-M2* expression for 6 hr. **(D)** Schematic of ClpX protein with mutations that bypass the toxicity of *socB-M2* expression indicated by arrows. Most mutations cluster near the RKH loop that is important for the recognition of ssrA-tagged substrates. **(E)** Stability of SocB-M2 in the presence of SocA in the *clpX(Y76C)* and *clpX(Q232R)* backgrounds. Performed as in Figure 2.5A. **(F)** Morphology, assessed by DIC microscopy, of indicated strains after 4 hr of growth in medium that induces or represses *socB-M2* expression. Average cell length (μm) is shown below each panel (n > 300 cells). **(G)** Growth of indicated strain on media that induces (+xyl) or represses (-xyl) expression of *socB-GST*. Five-fold serial

dilutions are shown. **(H)** Bacterial two-hybrid analysis of the interactions between DnaN, DnaN mutants, SocB, and HdaA. Cells were grown for 2 days at 30°C

Toxins of TA systems target a diverse range of targets within bacterial cells (Yamaguchi et al., 2011). Two observations suggested that SocB inhibits DNA replication at the level of elongation. First, induction of *socB-M2* caused cellular filamentation without chromosome accumulation, which may indicate that growth continues while replication elongation is blocked (Figure 2.5C). Second, global expression profiling indicated that *socB-M2* expression induced the SOS response (Figure 2.6A). The SOS response is often induced in response to replication perturbations, such as DNA damage, that disrupt replication fork progression (Little and Mount, 1982). To directly test whether replication elongation is inhibited by SocB, we measured DNA content in synchronized populations of cells harboring an inducible copy of *socB-M2*. In non-inducing conditions, DNA content increased linearly as a function of time post-synchrony (Figure 2.6B). In contrast, the induction of *socB-M2* caused a decrease in the rate of replication elongation, and cells eventually arrested with a DNA content between 1N and 2N (Figure 2.6B). As the completion of DNA replication is required for cell division, these cells also failed to divide (data not shown). These results indicate that SocB inhibits replication primarily at the level of elongation.

SocB blocks replication through an interaction with DnaN

To identify the putative target of SocB, we screened for mutants that can tolerate high levels of *socB-M2* expression. The first suppressors recovered were mutations in *clpX* that destabilized

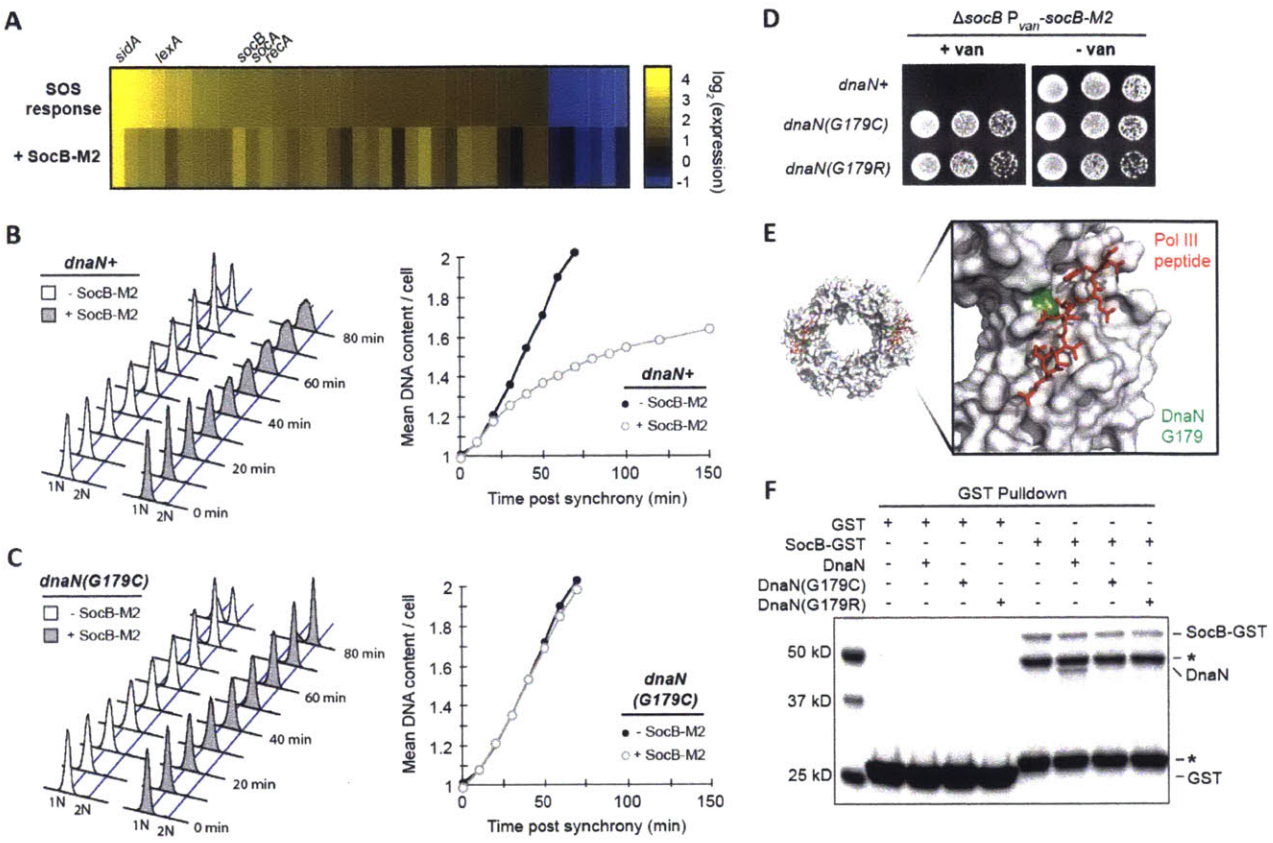


Figure 2.6. SocB blocks replication elongation through an interaction with DnaN

(A) Microarray analysis of gene expression changes following exposure to the DNA-damaging agent mitomycin C for 30 min (SOS response, top row) or following *socB-M2* expression for 2 hr (+SocB-M2, bottom row). For each treatment, the genes induced or repressed more than 2-fold following mitomycin C treatment are shown. **(B)** Flow cytometry of DNA content from synchronized cells grown ± *socB-M2* expression. For the +SocB-M2 condition, *socB-M2* was induced for 90 min prior to synchrony and release. Quantification of DNA content is shown on right. **(C)** Same as (B), except performed with the *dnaN(G179C)* strain. **(D)** Growth of indicated strains on *socB-M2* inducing or repressing medium. Five-fold serial dilutions are shown. **(E)** Structure of the *E. coli* sliding clamp in complex with a peptide derived from Pol III (PDB: 3D1F). Pol III peptide is in red, and the *E. coli* residue that corresponds to G179 in *Caulobacter* is colored in green. **(F)** Interaction between SocB-GST and DnaN. For each condition, the indicated proteins were mixed with glutathione sepharose beads, washed, eluted, and then loaded on an SDS-PAGE gel. SocB-GST protein is unstable; asterisks indicate truncated SocB-GST products that retain GST tag.

SocB-M2 (Figure 2.5D). The largest group of these mutations clustered near the RKH loop, which protrudes from the ClpX pore and is important for the recognition of SsrA-tagged substrates (Farrell et al., 2007; Martin et al., 2008). Additional mutations were found at sites distant from the ClpX pore: Y76 and P326, for example, are proximal to the ATP-binding site and roughly 40 Å away from the nearest RKH loop. In each tested case, the mutations resulted in an approximate 10-fold reduction in SocB-M2 stability (Figure 2.5E).

To identify suppressors outside of *clpX*, we continued our screen but focused on isolates that retained high levels of SocB-M2 and that did not harbor mutations in *clpX*. For two mutants, whole-genome sequencing revealed point mutations in *dnaN*, which encodes the β sliding clamp required for the processivity of DNA replication (Johnson and O'Donnell, 2005). Both mutations led to substitutions in glycine-179 of DnaN: G179C and G179R. We introduced these *dnaN* mutations into a clean genetic background and found that each was sufficient to bypass the replication block (Figure 2.6C) and growth inhibition (Figure 2.6D) normally observed following *socB-M2* expression. Furthermore, these mutations were able to partially suppress the filamentation observed following SocB-M2 accumulation (Figure 2.5F).

Glycine-179 resides within the hydrophobic groove on DnaN that is required for binding to Pol III and other replication proteins (Figure 2.6E) (Georgescu et al., 2008). The identification of suppressor mutations affecting this residue of DnaN raised the possibility that SocB blocks replication through a direct interaction with DnaN. To test this possibility, we purified SocB-GST and DnaN and measured their binding by affinity chromatography *in vitro*. Production of SocB-GST is toxic to *Caulobacter*, indicating that this fusion protein is functional (Figure 2.5G). We observed a strong interaction between DnaN and SocB-GST, but no interaction when DnaN was

incubated with GST alone (Figure 2.6F). Importantly, the suppressor mutations in DnaN, G179C and G179R, each disrupted the interaction with SocB-GST (Figure 2.6F). We obtained similar results using a bacterial two-hybrid system, confirming that DnaN and SocB can directly interact (Figure 2.5H).

The suppressor mutations isolated in DnaN reside within a hydrophobic groove required for its interaction with DnaN-binding proteins such as HdaA (Jonas et al., 2011), which must bind DnaA to regulate replication initiation (Kato and Katayama, 2001). To determine whether these mutations disrupt the interaction of DnaN with HdaA, we repeated our interaction assay and found that DnaN(G179C) retained the ability to interact with HdaA, whereas the G179R mutant did not (Figure 2.5H). Consistent with these binding data, cells producing DnaN(G179C) appeared similar to wild type in the absence of *socB* expression, suggesting that the mutant version of DnaN supported wild-type like growth. In contrast, cells producing DnaN(G179R) were often filamentous (Figure 2.5F), indicating that the interaction between DnaN and other proteins such as HdaA may be compromised. Taken together, our results support a model in which SocB blocks replication elongation through an interaction with DnaN, and that mutations in the hydrophobic cleft on DnaN can abrogate binding of SocB.

SocB induces loss of DnaN replication foci and replication fork collapse

During DNA replication, DnaN accumulates dynamically behind the lagging strand polymerase as a result of discontinuous DNA replication; consequently, YFP-tagged DnaN typically forms a discrete focus within cells during replication (Su'etsugu and Errington, 2011). Additionally, in *Caulobacter*, DnaN translocates along the major axis of the cell during replication

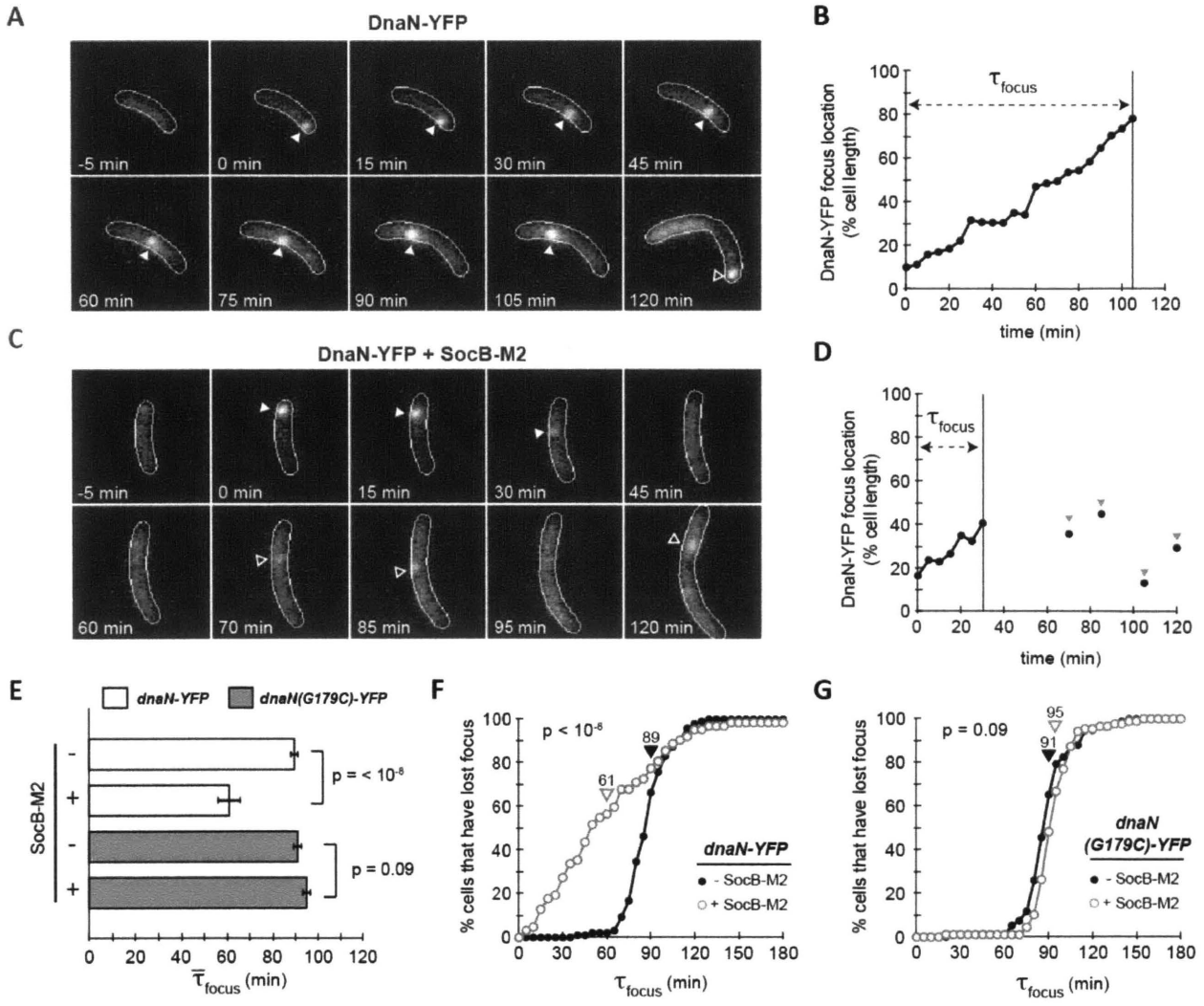


Figure 2.7. SocB induces replication fork collapse

(A) Fluorescence microscopy of strain bearing *dnaN-YFP* on a low-copy plasmid. *dnaN-YFP* was pre-induced for 2 h, and cells were synchronized and imaged every 5 min. Filled arrows indicates calculated position of DnaN-YFP focus; hollow arrows indicate a new round of replication following division. **(B)** Calculated DnaN-YFP focus position of the cell from (A). τ_{focus} is calculated as the time from focus formation to loss. **(C-D)** Same as (A-B), except *socB-M2* expression was induced for 90 min prior to synchrony and following release. Note the premature focus loss between 30 and 45 min. Hollow arrows indicate transient DnaN-YFP focus formation events that may be attempts at replication restart. **(E)** Average $\bar{\tau}_{focus}$ for *dnaN-YFP* strain (white bars) and *dnaN(G179C)-YFP* strain (grey bars) ± *socB-M2* expression. Error bars indicate ± S.E.M; n ≥ 60 cells per condition. **(F)** Percentage of cells that have lost their DnaN-YFP focus as a function of time post initiation in the absence (filled circles) or presence (hollow circles) of *socB-M2* expression. Average $\bar{\tau}_{focus}$ denoted by arrows. n ≥ 60 cells per condition. **(G)** Same as (F), but with *dnaN(G179C)-YFP* strain. n ≥ 60 cells per condition.

(Collier and Shapiro, 2009). To determine whether DnaN dynamics (and by proxy, ongoing replication) are affected by SocB, we imaged YFP-tagged DnaN in a strain harboring an inducible copy of *socB-M2*. In the absence of *socB-M2* expression, G1-phased cells typically showed diffuse DnaN-YFP followed by formation of a single focus after the initiation of DNA replication (Figure 2.7A). The focus moved along the major axis of the cell until its dispersal at the end of replication (Figure 2.7B). In the presence of *socB-M2* expression, we also observed formation of a single DnaN-YFP focus following initiation, but the focus often dispersed much earlier (Figures 2.7C and 2.7D). In cells that lost their DnaN-YFP focus earlier, transient focus formation was sometimes observed following initial focus loss (Figures 2.7C and 2.7D) although these foci often lasted only a single frame and appeared at varying points along the cell axis.

To quantify these effects, we measured the time from focus formation to dispersal (τ_{focus}), which reflects how long Pol III is engaged in replication before disengaging (either as a result of replication completion or pre-mature termination). We found that the average focus duration, $\bar{\tau}_{focus}$, dropped from 89 to 61 minutes in the presence of *socB-M2* expression (Figure 2.7E, $p < 10^{-8}$). Multiple observations indicated that the decrease in $\bar{\tau}_{focus}$ represented fork collapse prior to the completion of replication. First, measurements of τ_{focus} in the absence of *socB-M2* expression indicated that replication takes, on average, 89 minutes to complete, and that no cells finish prior to 40 minutes post initiation (Figure 2.7F); in the presence of *socB-M2* expression, however, 34% of cells terminated replication prior to 40 minutes, indicating that these cells most likely suffered collapsed forks and harbored incompletely replicated chromosomes. Second, measurements of DNA content in the DnaN-YFP strain expressing *socB-M2* indicated that these cells arrest with DNA content between 1N and 2N (data not shown). Third, transient DnaN-YFP focus formation

following initial dispersal is suggestive of abortive attempts at replication re-start. These findings, in addition to the observed induction of the SOS response upon expression *of socB-M2* (Figure 2.6A), are consistent with a model in which SocB induces replication fork collapse.

To test whether the mutations identified in *dnaN* bypass replication fork collapse, we imaged strains expressing *dnaN(G179C)-YFP* in the presence or absence of *socB-M2* expression and calculated τ_{focus}. In contrast to cells expressing *dnaN-YFP*, we no longer observed a significant difference in τ_{focus} following the production of SocB ($p = 0.09$) (Figures 2.7E and 2.7G). Whereas 34% of *dnaN-YFP* cells expressing *socB-M2* lost their foci within 40 minutes of replication initiation, only 1% of *dnaN(G179C)-YFP* cells did. These results indicate that mutations in DnaN that block binding to SocB also prevent replication fork collapse, further supporting the conclusion that SocB inhibits replication elongation through a direct interaction with DnaN.

SocB co-localizes with DnaN in a replication-dependent manner

The interaction between SocB and DnaN suggested that SocB may localize to the replisome. To examine the subcellular localization of SocB, we integrated an inducible *socB-YFP* fusion on the chromosome and imaged cells by fluorescence microscopy. Expression of *socB-YFP* inhibited colony formation, indicating that this translational fusion is functional (Figure 2.9A). After inducing *socB-YFP* for three hours, we observed the formation of SocB-YFP foci in a majority of cells (Figure 2.8A). The formation of these foci was dependent on the ability of SocB to bind DnaN, as foci were rarely seen in cells producing DnaN(G179C) or DnaN(G179R) (Figure 2.8A). SocB foci were also dependent on ongoing replication, as we saw a significant decrease in foci formation in cells depleted of DnaA, the replication initiator protein (Figure 2.8B).

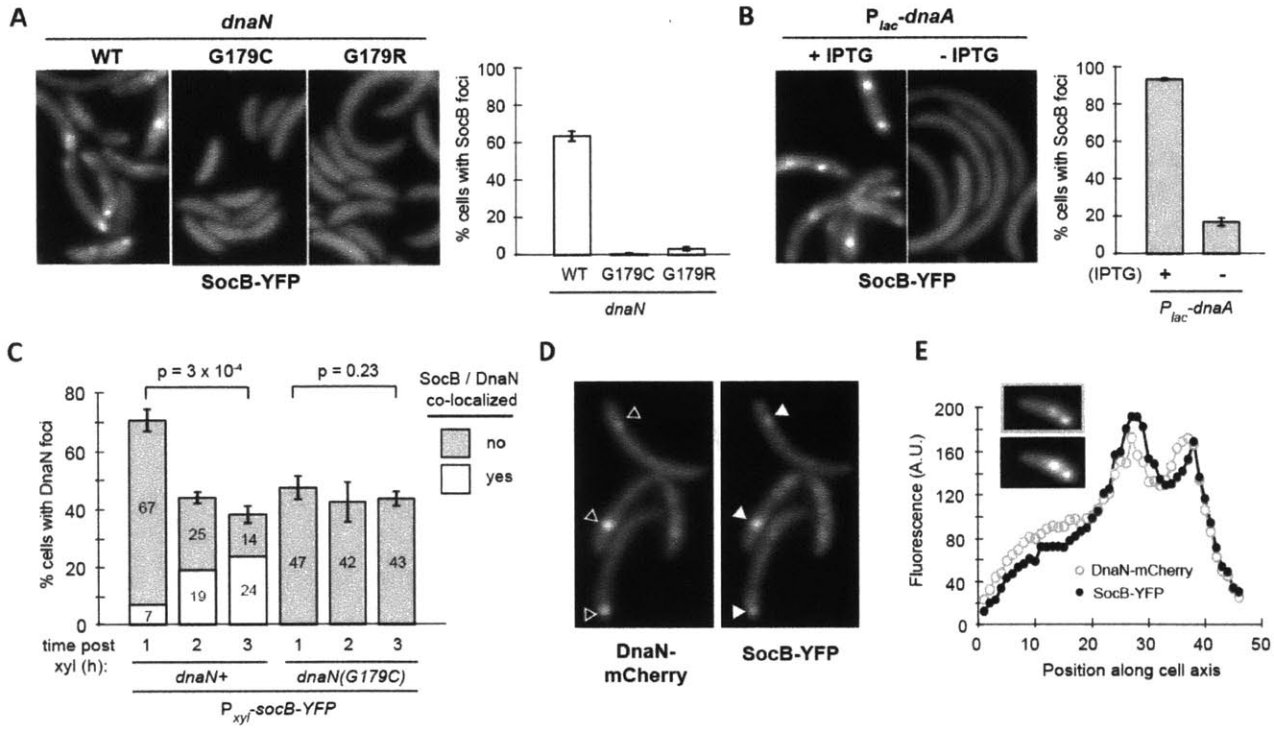


Figure 2.8. SocB forms foci that co-localize with DnaN

(A) Fluorescence microscopy of indicated strains at 3 hr post *socB-YFP* induction. Percentage of cells containing SocB-YFP foci is shown on the right. Errors bars indicate mean ± S.D. for three biological replicates (n > 400 cells per replicate). **(B)** Fluorescence microscopy of P_{lac}-*dnaA* cells grown in the presence or absence of IPTG for 2 hr prior to *socB-YFP* induction for 3 h. Percentage of cells containing SocB-YFP foci calculated as in (A). **(C)** Percentage of cells with DnaN-mCherry foci as a function of time post *socB-YFP* induction. The percentage of total cells with co-localized (white) or not co-localized (grey) DnaN-mCherry and SocB-YFP foci is shown within each bar. Error bars indicate mean ± S.D. for three biological replicates (n > 500 cells per replicate). **(D)** Co-localization of DnaN-mCherry foci (hollow arrows) and SocB-YFP foci (filled arrows) after induction of *socB-YFP* for 3 h. **(E)** Co-localization of multiple DnaN-mCherry and SocB-YFP foci in a single cell. Fluorescence profile for DnaN-mCherry (top inset, grey hollow circles) and SocB-YFP (bottom inset, black filled circles) is shown.

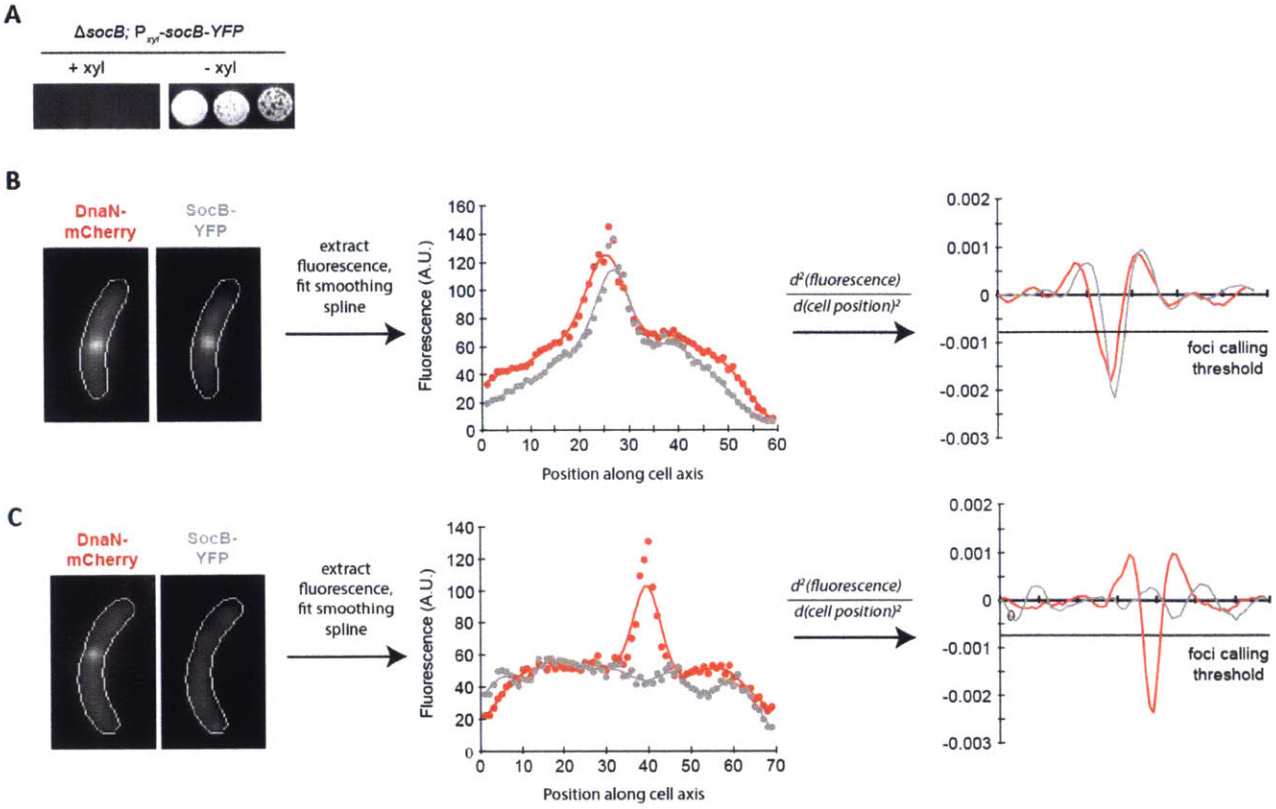


Figure 2.9. Toxicity of socB-YFP expression; workflow for automated co-localization calling

(**A**) Growth of indicated strain on media that induces (+xyl) or represses (-xyl) *socB-YFP* expression. Five-fold serial dilutions are shown. (**B**) Example cell exhibiting co-localization. Cells are first segmented and fluorescence signal extracted along the main cell axis. A smoothing spline is fit to the fluorescence signal and the second derivate of the fit is plotted. Local minima in the second derivative indicate "peaks" in the fluorescence signal; thus, the depth of the local minimum is a measure of the sharpness of the peak. Local minima that cross below a pre-determined threshold are called as foci, and cells are marked as co-localization positive if (1) the cell has foci for both DnaN-mCherry and SocB-YFP, and (2) the local minima in the second derivative for DnaN-mCherry and SocB-YFP overlap. (**C**) Example cell not exhibiting co-localization; methodology is the same as in (A). This cell is co-localization negative because no significant minima are detected in the second derivative of the SocB-YFP signal.

To test whether SocB and DnaN co-localize during replication, we fused *dnaN* to *mCherry* at the native chromosomal *dnaN* locus. We then integrated an inducible copy of *socB-YFP* on the chromosome and imaged cells by fluorescence microscopy at hour-long intervals post *socB-YFP* induction. Using an automated image analysis pipeline (Figures 2.9B-C and Experimental Procedures), we calculated the percentage of cells that have DnaN-mCherry foci, and of these cells, the percentage that also have co-localized SocB-YFP foci. As expected, we observed a decrease in the percentage of cells with DnaN-mCherry foci as a function of time post *socB-YFP* induction (Figure 2.8C, $p = 3 \times 10^{-4}$). The percentage of cells with DnaN-mCherry foci did not decrease to zero, presumably due to cells attempting to re-start replication following initial fork collapse (Figure 2.7C). Although the percentage of cells with DnaN-mCherry foci decreased over time, the percentage of cells with co-localized DnaN-mCherry and SocB-YFP increased significantly (Figure 2.8C). Most co-localization-positive cells had co-localized foci at a single point along the cell axis (Figure 2.8D). In a minority of cells, multiple co-localized foci could be observed, which may occur when the left- and right-arm replisomes are no longer overlapping (Figure 2.8E).

To test whether these localization effects were dependent on a direct interaction between SocB and DnaN, we repeated these microscopy experiments in a strain with *dnaN(G179C)-mCherry* integrated at its native chromosomal locus. In contrast to the *dnaN-mCherry* strain, we observed no significant decrease in the percentage of cells with DnaN(G179C)-mCherry foci as a function of time post *socB-YFP* induction (Figure 2.8C, $p = 0.23$). Further, fewer than 0.3% of cells exhibited co-localization between DnaN(G179C)-mCherry and SocB-YFP at any of the measured time points, in contrast to the 24% of cells that exhibit co-localization between DnaN-mCherry

and SocB-YFP at three hours post induction (Figure 2.8C). These results are consistent with a model in which SocB forms foci through its association with DnaN during active replication.

SocB interacts with DnaN through a DnaN-binding motif

In γ-proteobacteria, DnaN-binding proteins such as Hda and DnaE often contain a shared motif (QL[SD]LF) for binding the β-sliding clamp. Examination of HdaA (the *Caulobacter* ortholog of Hda) and DnaE orthologs from α-proteobacteria revealed a similar, putative DnaN binding motif (Figure 2.10A). SocB contained a short region similar to the DnaN-binding motif found in HdaA orthologs (Figure 2.10B). To test whether this region is required for the interaction of SocB with DnaN, we generated a Q52A mutant of *socB-YFP*. Mutation of this conserved glutamine in HdaA is sufficient to abolish its interaction with DnaN (Jonas et al., 2011). We found that this mutation abolished the toxicity of *socB-YFP* expression (Figure 2.10C), the formation of SocB-YFP foci (Figure 2.10D), and also the interaction between SocB and DnaN (Figures 2.10E-F). Importantly, this mutation did not affect the ability of SocB to interact with SocA, indicating that SocB(Q52A) is likely properly folded (Figure 2.10F). These results support a model in which SocB inhibits replication by binding to DnaN using a motif similar to that of HdaA.

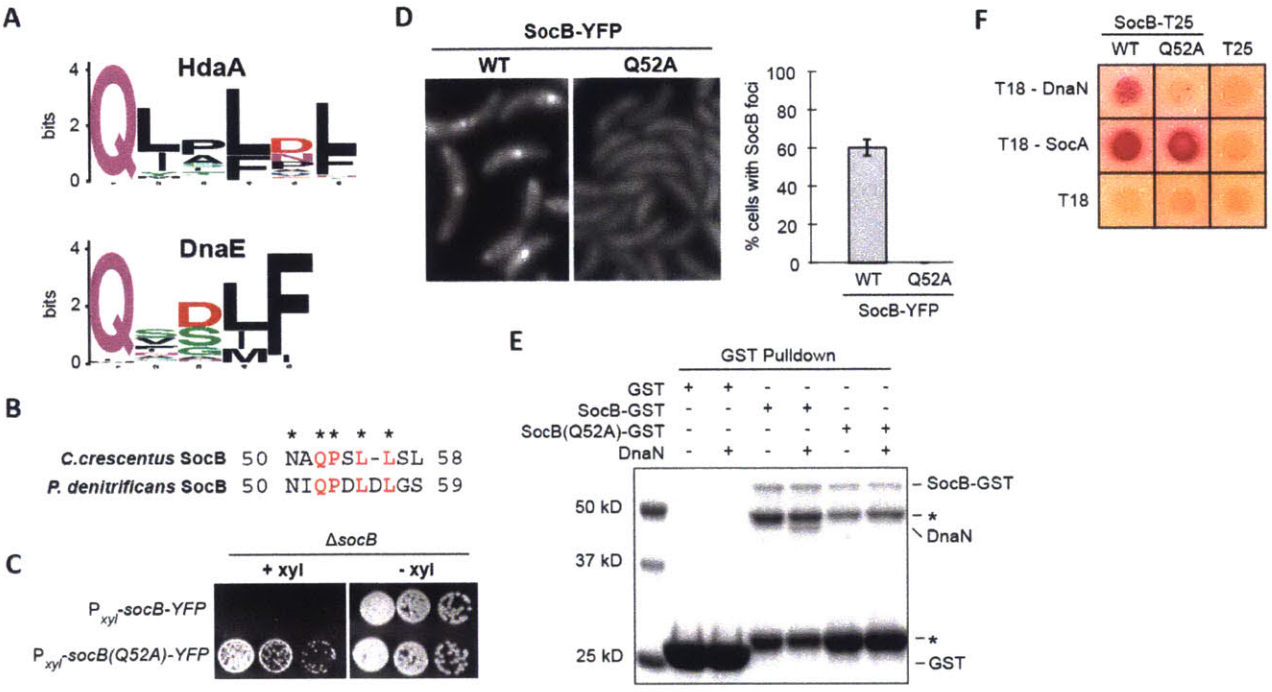


Figure 2.10. SocB interacts with DnaN through a conserved motif

(A) Sequence logo of the DnaN-binding motif in HdaA or DnaE from α-proteobacteria. **(B)** Putative DnaN-binding motif in SocB from *C. crescentus* and *P. denitrificans*. **(C)** Growth of indicated strains on media that induces (+xyl) or represses (-xyl) *socB-YFP* or *socB(Q52A)-YFP* expression. Five-fold serial dilutions are shown. **(D)** Fluorescence microscopy of *socB-YFP* or *socB(Q52A)-YFP* 3 hr post induction. Percentage of cells with SocB-YFP foci ± S.D. for three biological replicates is shown on right (n > 500 cells per replicate). **(E)** Interaction between SocB-GST, SocB(Q52A)-GST, and DnaN. Performed as in Figure 2.6H; as before, asterisk indicates SocB-GST N-terminal degradation products. **(F)** Bacterial two-hybrid analysis of the interaction between SocB, SocB(Q52A), DnaN, and SocA. Cells were grown for 2 days at 30°C.

Discussion

Essentiality of ClpXP and SocA mechanism of action

Unlike most bacteria, ClpXP is essential for the viability of *Caulobacter* cells (Jenal and Fuchs, 1998); however, the reason for this essentiality was not clear until now. Our work reveals that ClpXP is required for the degradation of a toxin, SocB, that is constitutively produced in cells (Figure 2.11, left). In the absence of ClpXP or SocA, SocB accumulates, leading to the collapse of replication forks, induction of the SOS response, and eventual cell death (Figure 2.11, right). Consequently, mutations in *socB* can bypass the essentiality of *clpX* or *clpP* (Figure 2.1B). We note, however, that cells lacking *socB* and either *clpX* or *clpP* do not grow as rapidly as wild-type, likely due to defects in the turnover of other ClpXP substrates such as the cell cycle regulator CtrA (Bhat et al., 2013).

The rapid turnover of SocB is unusual, given that, for most TA systems, the toxin is more stable than its cognate antitoxin. The reduced stability of SocB stems from the atypical mechanism of its antitoxin, SocA. Whereas most antitoxins inhibit their cognate toxins through sequestration, SocA is instead an adaptor for the degradation of SocB by ClpXP. SocA binds both SocB (Figure 2.1H) and the N-domain of ClpX (Figure 2.4C) and thereby promotes the degradation of SocB *in vivo* and *in vitro* (Figures 2.3D and 2.3E). SocA binding to the N-domain appears to be essential for this activity, as SocA was unable to promote SocB proteolysis when the N-domain was truncated from ClpX (Figure 2.3E). The observation that SocA decreases SocB stability does not rule out that SocA also blocks SocB toxicity through sequestration. However, multiple results

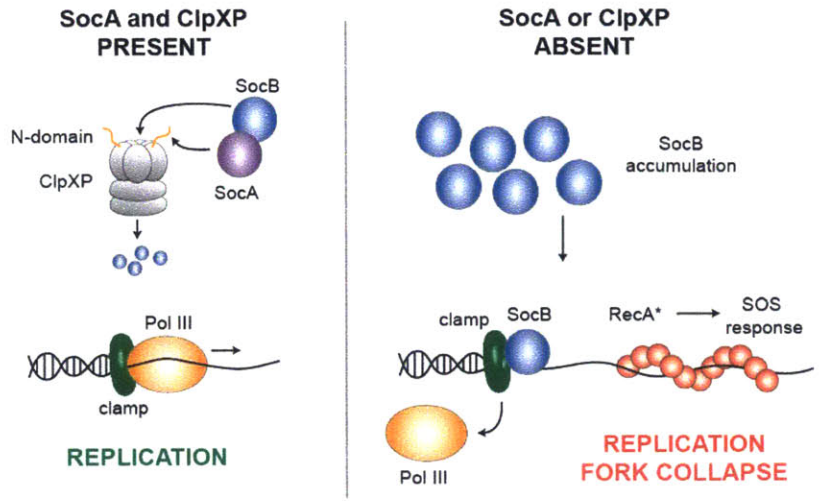


Figure 2.11. Model for SocAB function

Under normal growth conditions, the toxin SocB is delivered to ClpXP for degradation by its antitoxin SocA. Pol III thus remains in association with the clamp, and replication proceeds normally (left). However, in the absence of either ClpXP or SocA, SocB accumulates and competes for binding to the clamp with Pol III and other replication factors. This competition eventually results in the collapse of replication forks and induction of the RecA-mediated SOS response (right).

argue against this role for SocA: (1) targeted mutations in the N-domain of ClpX that abolish SocA binding also prevent SocA from functioning as an antitoxin *in vivo* (Figures 2.4C and 2.4D); and (2) production of the stable SocB-M2 variant in *socA*+ cells is lethal, even though SocB-M2 and SocA can still form a complex (Figure 2.5B and data not shown). Thus, we propose that the antitoxin activity of SocA results principally from its ability to promote SocB proteolysis through ClpXP, which makes SocA necessary, but not sufficient, for counteracting SocB toxicity.

In many bacteria, ClpXP is not formally essential for viability, although it is often required for other critical processes. In *E. coli*, deletions of *clpX* or *clpP* have no significant effect on viability or growth rate (Schweder et al., 1996), but ClpXP contributes to the proteolysis of a range of

cellular substrates (Flynn et al., 2003). In *B. subtilis*, *clpX* and *clpP* deletions are viable but have major defects in sporulation and competence (Nakano et al., 2001). These defects are mostly due to accumulation of a transcriptional regulator, Spx, that is normally degraded by ClpXP. ClpX or ClpP has, however, been found to be essential in several bacteria, including the pathogen *Streptococcus pneumoniae* (Piotrowski et al., 2009). The essentiality of *clpX* can be bypassed in *Streptococcus* by mutations in an uncharacterized gene, *spr1630*, suggesting that this gene product may accumulate in the absence of ClpXP and inhibit growth, similar to SocB. Characterization of the gene products that bypass ClpXP essentiality may reveal novel regulators of cell growth or physiology.

Mechanism of SocB inhibition of replication elongation

Our results indicate that, upon accumulating, SocB blocks replication elongation and triggers an SOS response (Figure 2.6). Multiple observations suggest that SocB mediates these effects through a direct interaction with DnaN: (1) SocB and DnaN interact *in vitro* and in a bacterial two-hybrid assay (Figures 2.6F and 2.5H); (2) mutations in DnaN that abolish its interaction with SocB also bypass the SocB-induced replication block (Figures 2.6C, 2.6F, and 2.5H); (3) SocB and DnaN co-localize, and this co-localization is dependent on the ability of SocB and DnaN to interact (Figure 2.8); and (4) SocB contains a DnaN-binding motif that is required for its interaction with DnaN and toxicity (Figure 2.10). In sum, these results are consistent with a model in which SocB binds to DnaN, and that this association leads to catastrophic replication fork collapse (Figure 2.7).

How, then, does the interaction between SocB and DnaN lead to the premature termination of replication? The simplest model is that SocB competes with Pol III for binding to DnaN,

potentially disrupting both lagging and leading strand synthesis. During lagging strand replication, the polymerase extends discontinuously and presumably must associate with a new clamp to produce each Okazaki fragment (Johnson and O'Donnell, 2005). Disruption of the Pol III-DnaN interaction could rapidly inhibit production of new Okazaki fragments and consequently block synthesis of the lagging strand. In contrast, during leading strand replication, the polymerase can extend continuously and may not need to load new clamps following initiation. However, depending on the stability of the Pol III-DnaN interaction, SocB may still be able to compete for binding to DnaN on the leading strand if Pol III and DnaN ever transiently dissociate, such as upon encountering a DNA lesion. Additionally, blocking synthesis on the lagging strand would also indirectly inhibit synthesis on the leading strand, given that the leading and lagging strand polymerases are physically tethered by the γ clamp loader complex. Such stalling of replication could eventually lead the replisome to disassemble.

Notably, the effects of SocB on DNA replication are not immediate; replication first slows down approximately 120 min post-induction (Figure 2.6B, note the 90 min pre-induction prior to time zero). This timing coincides with the first observable decrease in viability (Figure 2.2B) and an increase in the co-localization of SocB with DnaN (Figure 2.8C), but is later than may be expected if SocB immediately outcompetes Pol III for binding to DnaN. The delay may simply reflect a need for SocB to accumulate to sufficient levels before outcompeting Pol III and other DnaN-binding proteins. Alternatively, SocB may only disrupt new associations of DnaN and Pol III such that replication can initially continue following the accumulation of SocB by using DnaN molecules already in association with Pol III.

An additional question is why SocB, despite the presence of a DnaN-binding motif, is not toxic to *E. coli* which has a similar hydrophobic pocket on DnaN (Figure 2.1H and data not shown). It is possible that that SocB makes contacts with DnaN at a secondary site that is less well conserved between bacteria. Interestingly, a co-crystal structure of DnaN and the little finger domain of DNA Pol IV, a DnaN-interacting protein, revealed the presence of a secondary binding interface outside of the hydrophobic pocket (Bunting et al., 2003). The secondary interface comprised over 70 percent of the buried surface area, indicating that it likely contributes significantly to the affinity of interaction. The requirement of a secondary binding interface may explain why appending a DnaN-binding motif to GFP only results in weak DnaN-dependent foci formation in *B. subtilis* (Su'etsugu and Errington, 2011).

Given the ability of SocB to inhibit replication progression, an important remaining question is when SocB normally accumulates in wild-type cells. Genetically, we revealed that a loss of *clpX*, *clpP*, or *socA* is sufficient to allow SocB accumulation (Figure 2.1B and Figure 2.3). Whether the stability or activity of ClpXP, or SocA, is ever modulated to induce SocB accumulation is not yet clear. Intriguingly, the *socAB* operon is induced by the DNA-damaging agent MMC (Figure 2.6A), suggesting that it may play a regulatory role during the response to DNA damage. One attractive possibility is that sublethal amounts of SocB accumulate following DNA damage, or other stresses, to regulate the rate at which replication proceeds. Recent work has begun to reveal mechanisms by which cells regulate replication at the level of elongation. For example, the accumulation of (p)ppGpp during the stringent response in *B. subtilis* leads to an arrest of replication elongation through the inhibition of DNA primase (Wang et al., 2007). Although

replication is often controlled at the level of initiation, mechanisms for modulating elongation may be widespread.

Protein interaction hubs and antibiotic targets

In addition to blocking replication elongation, the binding of SocB to DnaN may disrupt other replication-associated processes. In addition to Pol III, the clamp interacts with proteins required for DNA repair and the regulation of replication initiation, including Pol IV and Pol V (translesion synthesis), MutS and MutL (mismatch repair), and Hda (regulation of initiation). These proteins all bind to DnaN within the same hydrophobic cleft on its surface that is likely bound by SocB (Robinson et al., 2010), suggesting that SocB may disrupt several cellular processes required for growth and genome maintenance.

Interaction hubs such as DnaN may be ideal targets for new antibiotic development because they coordinate multiple cellular processes and may be unable to mutate to prevent small molecule binding without significantly compromising their native functions. The ability of a protein toxin like SocB to arrest DNA replication and kill cells suggests that the hydrophobic cleft on DnaN may be a prime target for small molecule inhibitors. In fact, the small molecule RU-7 was recently shown to bind within this hydrophobic cleft on the clamp and to prevent its association with Pol III *in vitro* (Georgescu et al., 2008). This interaction was specific for the bacterial sliding clamp, as the interaction between the eukaryotic PCNA clamp and Pol δ was unaffected. A separate screen for inhibitors of an *in vitro* bacterial replication system identified six compounds that share the same core structure as RU-7, suggesting that they also function through binding to the clamp (Dallmann et al., 2010). The recent rise in multi-drug resistant strains of various pathogenic bacteria highlights the continued need to develop novel antibiotics to combat these

infections. The study of bacterially-encoded toxins and the mechanisms by which they inhibit cellular proliferation may prove valuable in the identification of promising new targets.

Experimental Procedures

Bacterial strains and media

Caulobacter strains used in this study are listed in Table 2.1. *Caulobacter* strains were grown in PYE broth at 30°C unless otherwise indicated. To induce expression from the P_{xyl}, P_{van}, or P_{lac} promoters, media was supplemented with xylose (0.3%), vanillate (500 μM), or IPTG (1 mM), respectively. For the P_{van}-*dnaN-YFP* and P_{van}-*dnaN(G179C)-YFP* strains, a lower concentration of vanillate (20 μM) was used for induction. All chromosomal deletions were generated using a two-step recombination method and *sacB* as a counterselection marker (Skerker et al., 2005). The first and last five codons were left intact to prevent polar effects on downstream genes. Chromosomal integrations were performed using the pVGFPN/pXGFPN vectors that integrate at the P_{van} or P_{xyl} locus (Thanbichler et al., 2007). The pMR20 and pJS14 vectors were used for low- and medium-copy plasmid expression, respectively. Synchronizations of swarmer/G1 cells were performed using Percoll (GE Healthcare) density gradient centrifugation at 10,000g for 20 min.

Suppressor screening and mapping

To screen for suppressors of *clpP*, a *clpP* depletion strain (UJ199; Δ*clpP;* P_{xyl}-*clpP*) was transposon mutagenized using the EZ-Tn5 kit (Epicentre) and plated onto PYE medium containing glucose (to repress *clpP* expression) and kanamycin (25 μg/ml, to select for transposon integration). Colonies that grew after 2-4 days were then isolated for genomic DNA extraction. To identify transposon insertion sites, genomic DNA from putative suppressor strains was digested, self-ligated, and transformed into *pir*-116 *E. coli*. Plasmids were prepped from the resulting transformants and sequenced using the Ez-Tn5 FP-1 forward primer.

To screen for suppressors of *socB-M2* expression, a copy of *socB-M2* was integrated at both the P_{van} and P_{xyl} loci to reduce the frequency of suppressor mutations that simply inactivate *socB-M2*. This strain was plated on PYE containing vanillate and xylose to induce *socB-M2* expression. Colonies that grew after 2-3 days were screened for mutations within *socB* or *clpX* by directed sequencing, and for high levels of SocB-M2 expression by Western blotting. Strains without mutations in *socB* and *clpX* that expressed SocB-M2 were sent for whole-genome Illumina sequencing (BioMicroCenter, MIT). Illumina reads were mapped to the *Caulobacter crescentus* NA1000 reference genome using the bwa package and then sorted, indexed, and mutations called using the samtools package. Mutations were verified by directed sequencing.

Bacterial two-hybrid assay

Protein interactions were assayed using the bacterial adenylate cyclase two-hybrid system (Karimova et al., 1998). Briefly, genes of interest were fused to the 3' or 5' end of the T18 or T25 fragments in the pUT18C pUT18C, pKT25, or pKNT25 vectors. The resulting T18- and T25-fusion plasmids were then co-transformed into BTH101 *E. coli*. Co-transformants were grown in M63 minimal media supplemented with maltose (0.2%), IPTG (1 mM, to induce the fusion proteins), and appropriate antibiotics. Saturated overnight cultures were then spotted onto MacConkey agar (40 g/L) plates supplemented with maltose (1%), IPTG (1 mM), and appropriate antibiotics. Plates were incubated at 30°C and pictures taken one or two days post-incubation.

Flow cytometry

Caulobacter samples were fixed in 70% ethanol, pelleted at 4000 RPM, and then re-suspended in 50 mM sodium citrate containing 2 μg/ml RNase. Samples were then incubated for 4 hours at

50°C to digest RNA. Samples were diluted to an OD_{600} of 0.001, stained with 2.5 μM SYTOX Green (Invitrogen), and then analyzed by flow cytometry on a BD Accuri C6 flow cytometer (BD Biosciences).

Protein purification

All proteins were cloned into pET-based vectors for expression in BL21 *E. coli*. For purification of His6-SocB, cells were grown to OD 0.4 and then induced with 0.1 mM IPTG at 18°C overnight. Cell pellets were re-suspended in 20 ml prep buffer-20 (20 mM Tris-HCl pH 7.4, 500 mM NaCl, 20 mM imidazole, 10% glycerol) per liter of starting culture and lysed by sonication. Lysates were cleared by centrifugation at 17,000g for 1 hr, and then the cleared lysates were bound to Ni-NTA agarose for 1 hr in batch format at 4°C. The Ni-NTA agarose was then applied to a 20 ml Econo-Pac chromatography column (Biorad), washed with five column volumes of prep buffer-20, and then eluted in 10 ml of prep buffer-250. His6-SocB was concentrated on an Amicon 10K Centrifugal Filter Unit (Millipore), buffer exchanged into PD-KCl-200 buffer (20 mM HEPES pH 7.6, 5 mM $MgCl_2$, 200 mM KCl, 0.032% NP-40, 10% glycerol) on PD-10 columns (GE Healthcare), and stored at -80°C. Due to low solubility at temperatures greater than 10°C, His6-SocB was always kept on ice after thawing. For purification of His6-SocA, expression was induced in BL21 *E. coli* that also contained the chaperone vectors pBB528 and pBB541 (de Marco, 2007), which we found greatly increased the yield of soluble His6-SocA protein. Otherwise, induction and purification conditions are identical as above, except that His6-SocA required further purification on a size exclusion column (S200, GE Healthcare) in order to remove high molecular weight contaminants. Fractions containing His6-SocA, as verified by

Coomassie staining, were pooled and stored at -80°C. ClpX, ΔN-ClpX, and ClpP were generous gifts of Peter Chien (University of Massachusetts Amherst).

For proteins used in affinity chromatography experiments, purification conditions are identical as His6-SocB except for the differences noted below. For GST-His6 and DnaN-His6 (wild-type and mutant proteins), cells were grown to OD 0.4 and then induced with 0.3 mM IPTG at 30°C for 4 hr prior to collecting pellets. Purified and concentrated GST-His6 or DnaN-His6 was buffer exchanged into low salt buffer (20 mM Tris-HCl pH 7.4, 200 mM NaCl, 10% glycerol) and stored at -80°C. For GST-SocB-His6 (wild-type and mutant proteins), cells were grown to OD 0.4 and then induced with 0.1 mM IPTG at 18°C overnight. Purified and concentrated GST-SocB-His6 was buffer exchanged into low salt buffer and stored at -80°C.

Degradation assays

Degradation reactions were performed in PD-KCl-200 buffer at 4°C. Reaction conditions were as follows: 0.5 μM ClpX, 0.5 μM ΔN-ClpX, 1 μM ClpP, 5 μM SocB, 5 μM SocA, 32 μg/ml creatine kinase, 16 mM creatine phosphate, and 4 mM ATP.

Affinity chromatography

For binding assays, 100 μl of glutathione sepharose 4B beads (GE Healthcare) was pre-washed in binding buffer (50 mM HEPES-KOH pH 7.5, 100 mM KCl, 0.1 mM EDTA, 2.5 mM $MgCl_2$, 3% glycerol, 0.1% Triton X-100). Washed beads were then incubated with 1 mg of GST-tagged protein for 1.5 hr at 4°C with rotation. Beads were washed twice with 1 ml of binding buffer, and then 0.25 mg of bait protein (DnaN-His6 wild-type or mutant protein) was added and allowed to bind for an additional 1.5 hr. Beads were then transferred to a 10 ml Poly-Prep chromatography

column (BioRad) and then washed with six column volumes of binding buffer. Columns were capped and then 100 µl of elution buffer (50 mM Tris-HCl pH 8.0, 40 mM reduced glutathione) was added. After 30 min incubation, eluates were collected. An equal amount of protein was loaded on each lane of an SDS-PAGE gel and proteins were visualized by Coomassie blue staining.

Microscopy and image analysis

Differential interference contrast (DIC) and fluorescence images were taken using a Zeiss Axiovert 200 microscope with a 100x/1.45 oil immersion objective. Phase contrast images and time-lapse movies were taken on a Zeiss Observer Z1 microscope using a 100x/1.4 oil immersion objective and an LED-based Colibri illumination system. Cells were placed on a PYE + 1.5% low-melting agarose pad supplemented with xylose or vanillate as indicated, and imaged in a glass-bottomed petri dish wrapped with Parafilm to prevent desiccation. Pictures were taken every 5 minutes, and temperature was maintained at 30°C using the Zeiss Temp Module S1 and Heating Insert P S1. Automatic focusing was performed using the Zeiss Definite Focus module. To perform cell segmentation and tracking, images were processed using MicrobeTracker (Sliusarenko et al., 2011). Position of the DnaN-YFP focus along the cell axis was determined using custom-written MATLAB software. Briefly, fluorescence along the cell axis was extracted and fit to a smoothing spline. Presence of a focus was determined by looking for local minima in the second derivative of the smoothing spline that crossed below a pre-determined threshold. Cells were scored if (1) they had no focus in the first frame, and (2) a focus was observed to form at some point later in the movie. To calculate n_{focus}, the location of the focus was plotted as a function of time, and a smoothing spline was fit to its trajectory. The average of the absolute

value of the first derivative of this smoothing spline was taken as n_{focus}. To calculate t_{focus}, the time from focus appearance (defined as the first frame with a focus) to disappearance (defined as the first frame with a focus for which the following two frames have no focus) was calculated.

To calculate percentage of cells with SocB-YFP foci, the fluorescence along the cell axis was extracted and fit to a smoothing spline. Foci were defined as local minimum that crossed below a pre-determined threshold in the second derivative. To calculate co-localization between SocB-YFP and DnaN-mCherry, the same foci scoring function was used as above to determine the position of any SocB-YFP or DnaN-mCherry foci in the cell. If the position of the highest scoring DnaN-mCherry peak (as determined by lowest local minimum in the second derivative) and a SocB-YFP peak overlapped, the cell was classified as exhibiting co-localization.

Microarrays

Expression profiling of cells expressing *socB-M2* was performed as described previously (Gora et al., 2010). Expression data for cells exposed to the DNA-damaging agent mitomycin C was previously published (Modell et al., 2011).

Bacterial Strains

Table 2.1

Strain #	Genotype	Source
UJ199	Δ*clpP*; P_{xyl}-*clpP::tet*^{R} (*xyl* locus)	(Jenal and Fuchs, 1998)
ML2029	Δ*clpP*; pMR20: P_{lacI}-*lacI*-P_{lac}-*clpP::tet*^{R}	This study
ML2030	Δ*clpX*; pMR20: P_{lacI}-*lacI*-P_{lac}-*clpX::tet*^{R}	This study
ML2031	Δ*clpP*; Δ*socB*; pMR20: P_{lacI}-*lacI*-P_{lac}-*clpP::tet*^{R}	This study
ML2032	Δ*clpX*; Δ*socB*; pMR20: P_{lacI}-*lacI*-P_{lac}-*clpX::tet*^{R}	This study
ML2033	Δ*socB*; P_{van}-*socB::kan*^{R}	This study
ML2034	Δ*socAB*; P_{van}-*socB::kan*^{R}	This study
ML2035	Δ*socAB*; P_{van}-*socB::kan*^{R}; pMR20: P_{xyl}-*socA::tet*^{R}	This study
ML2036	Δ*clpX*; Δ*socB*; pMR20: P_{lacI}-*lacI*-P_{lac}-*clpX::tet*^{R}; P_{van}-*M2-socB::kan*^{R}	This study
ML2037	Δ*clpX*; Δ*socB*; pMR20: P_{lacI}-*lacI*-P_{lac}-*clpX::tet*^{R}; pJS14: P_{xyl}-*M2-socB::chlor*^{R}	This study
ML2038	Δ*socB*; pMR20: P_{xyl}-*M2-socB::tet*^{R}	This study
ML2039	Δ*socAB*; pMR20: P_{xyl}-*M2-socB::tet*^{R}	This study
ML2040	Δ*socAB*; P_{van}-*M2-socB::kan*^{R}; pMR20: P_{xyl}-*socA::tet*^{R}	This study
ML2041	*clpX(I47A)*; Δ*socAB*; P_{van}-*socB::kan*^{R}; pMR20: P_{xyl}-*socA::tet*^{R}	This study
ML2042	*clpX(L13A)*; Δ*socAB*; P_{van}-*socB::kan*^{R}; pMR20: P_{xyl}-*socA::tet*^{R}	This study
ML2044	Δ*socB*; pMR20: P_{xyl}-*socB-M2::tet*^{R}	This study
ML2045	Δ*socB*; P_{van}-*socB-M2::kan*^{R}	This study
ML2046	*dnaN(G179C)*; Δ*socB*; P_{van}-*socB-M2::kan*^{R}	This study
ML2047	*dnaN(G179R)*; Δ*socB*; P_{van}-*socB-M2::kan*^{R}	This study
ML2048	Δ*socB*; P_{xyl}-*socB-M2::gent*^{R}; pRVYFPC-5: P_{van}-*dnaN-YFP::tet*^{R}	This study
ML2049	*dnaN(G179C)*; Δ*socB*; P_{xyl}-*socB-M2::gent*^{R}; pRVYFPC-5: P_{van}-*dnaN(G179C)-YFP::tet*^{R}	This study
ML2050	Δ*socB*; P*xyl-socB-YFP::kan*^{R} (*xyl* locus)	This study
ML2051	*dnaN(G179C)*; Δ*socB*; P_{xyl}-*socB-YFP::kan*^{R} (*xyl* locus)	This study
ML2052	*dnaN(G179R)*; Δ*socB*; P_{xyl}-*socB-YFP::kan*^{R} (*xyl* locus)	This study
ML2053	P_{lacI}-*lacI* (*hfaA* locus); P_{lac}-*dnaA* (*dnaA* locus); P_{xyl}-*socB-YFP::kan*^{R} (*xyl* locus)	This study
ML2054	Δ*socB*; P_{xyl}-*socB-YFP::kan*^{R} (*xyl* locus); *dnaN-mCherry::gent*^{R} (native locus, promoter)	This study
ML2055	Δ*socB*; P_{xyl}-*socB-YFP::kan*^{R} (*xyl* locus); *dnaN(G179C)-mCherry::gent*^{R} (native locus, promoter)	This study
ML2056	Δ*socAB*; P_{van}-*socB-M2::kan*^{R}; pMR20: P_{xyl}-*socA::tet*^{R}	This study

ML2057	Δ*socB*; P_{van}-*M2-socB::kan*^{R}	This study
ML2058	Δ*socB*; *clpX(Y76C)*; P_{van}-*socB-M2::kan*^{R}	This study
ML2059	Δ*socB*; *clpX(Q232R)*; P_{van}-*socB-M2::kan*^{R}	This study
ML2060	Δ*socAB*; *clpX(Y76C)*; P_{van}-*socB-M2::kan*^{R}; pMR20: P_{xyl}-*socA::tet*^{R}	This study
ML2061	Δ*socAB*; *clpX(Q232R)*; P_{van}-*socB-M2::kan*^{R}; pMR20: P_{xyl}-*socA::tet*^{R}	This study
ML2115	Δ*socB*; P_{xyl}-*socB(Q52A)-YFP::kan*^{R} (*xyl* locus)	This study
ML2116	pMR20:P_{xyl}-*socB-GST-his6::tet*^{R}	This study

Author Contributions

Suppressor screening and bacterial two hybrid assays were performed by CDA and TNP. Affinity chromatography experiments were performed by TNP. Generation and testing of DnaN-binding motif mutants in SocB was performed by DH. All other experiments were conceived and performed by CDA. CDA and MTL analyzed the data and wrote the paper.

Acknowledgements

We thank members of the Laub laboratory for helpful discussions and comments on the manuscript. We also thank Peter Chien for generously providing purified ClpX, ΔN-ClpX, and ClpP. MTL is an Early Career Investigator at the Howard Hughes Medical Institute. This work was supported by an HHMI Summer Medical Fellowship to DH, an NSF Graduate Research Fellowship to CDA, and an NIH grant (R01GM082899) to MTL.

References

Aakre, C.D., Phung, T.N., Huang, D., and Laub, M.T. (2013). A bacterial toxin inhibits DNA replication elongation through a direct interaction with the β sliding clamp. Mol. Cell *52*, 617–628.

Bhat, N.H., Vass, R.H., Stoddard, P.R., Shin, D.K., and Chien, P. (2013). Identification of ClpP substrates in Caulobacter crescentus reveals a role for regulated proteolysis in bacterial development. Mol. Microbiol. *88*, 1083–1092.

Bunting, K.A., Roe, S.M., and Pearl, L.H. (2003). Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp. EMBO J. *22*, 5883–5892.

Bush, K., Courvalin, P., Dantas, G., Davies, J., Eisenstein, B., Huovinen, P., Jacoby, G.A., Kishony, R., Kreiswirth, B.N., Kutter, E., et al. (2011). Tackling antibiotic resistance. Nat. Rev. Microbiol. *9*, 894–896.

Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Coller, J.A., Fero, M.J., McAdams, H.H., and Shapiro, L. (2011). The essential genome of a bacterium. Mol. Syst. Biol. *7*, 528.

Coates, A.R.M., Halls, G., and Hu, Y. (2011). Novel classes of antibiotics or more of the same? Br. J. Pharmacol. *163*, 184–194.

Collier, J., and Shapiro, L. (2009). Feedback control of DnaA-mediated replication initiation by replisome-associated HdaA protein in Caulobacter. J. Bacteriol. *191*, 5706–5716.

Dallmann, H.G., Fackelmayer, O.J., Tomer, G., Chen, J., Wiktor-Becker, A., Ferrara, T., Pope, C., Oliveira, M.T., Burgers, P.M.J., Kaguni, L.S., et al. (2010). Parallel multiplicative target screening against divergent bacterial replicases: identification of specific inhibitors with broad spectrum potential. Biochemistry (Mosc.) *49*, 2551–2562.

Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E., and Jennings, P.A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proc. Natl. Acad. Sci. U. S. A. *98*, 11627–11632.

Dougan, D.A., Weber-Ban, E., and Bukau, B. (2003). Targeted delivery of an ssrA-tagged substrate by the adaptor protein SspB to its cognate AAA+ protein ClpX. Mol. Cell *12*, 373–380.

Farrell, C.M., Baker, T.A., and Sauer, R.T. (2007). Altered specificity of a AAA+ protease. Mol. Cell *25*, 161–166.

Flynn, J.M., Neher, S.B., Kim, Y.I., Sauer, R.T., and Baker, T.A. (2003). Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. Mol. Cell *11*, 671–683.

Georgescu, R.E., Yurieva, O., Kim, S.-S., Kuriyan, J., Kong, X.-P., and O'Donnell, M. (2008). Structure of a small-molecule inhibitor of a DNA polymerase sliding clamp. Proc. Natl. Acad. Sci. U. S. A. *105*, 11116–11121.

Gora, K.G., Tsokos, C.G., Chen, Y.E., Srinivasan, B.S., Perchuk, B.S., and Laub, M.T. (2010). A cell-type-specific protein-protein interaction modulates transcriptional activity of a master regulator in Caulobacter crescentus. Mol. Cell *39*, 455–467.

Indiani, C., McInerney, P., Georgescu, R., Goodman, M.F., and O'Donnell, M. (2005). A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously. Mol. Cell *19*, 805–815.

Jenal, U., and Fuchs, T. (1998). An essential protease involved in bacterial cell-cycle control. EMBO J. *17*, 5658–5669.

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. Annu. Rev. Biochem. *74*, 283–315.

Jonas, K., Chen, Y.E., and Laub, M.T. (2011). Modularity of the bacterial cell cycle enables independent spatial and temporal control of DNA replication. Curr. Biol. CB *21*, 1092–1101.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc. Natl. Acad. Sci. U. S. A. *95*, 5752–5756.

Kato, J., and Katayama, T. (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in Escherichia coli. EMBO J. *20*, 4253–4262.

Kurz, M., Dalrymple, B., Wijffels, G., and Kongsuwan, K. (2004). Interaction of the sliding clamp beta-subunit and Hda, a DnaA-related protein. J. Bacteriol. *186*, 3508–3515.

Lenhart, J.S., Sharma, A., Hingorani, M.M., and Simmons, L.A. (2013). DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. Mol. Microbiol. *87*, 553–568.

Lenne-Samuel, N., Wagner, J., Etienne, H., and Fuchs, R.P.P. (2002). The processivity factor beta controls DNA polymerase IV traffic during spontaneous mutagenesis and translesion synthesis in vivo. EMBO Rep. *3*, 45–49.

Levchenko, I., Seidel, M., Sauer, R.T., and Baker, T.A. (2000). A specificity-enhancing factor for the ClpXP degradation machine. Science *289*, 2354–2356.

Little, J.W., and Mount, D.W. (1982). The SOS regulatory system of Escherichia coli. Cell *29*, 11–22.

Liu, M., Zhang, Y., Inouye, M., and Woychik, N.A. (2008). Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. Proc. Natl. Acad. Sci. U. S. A. *105*, 5885–5890.

López de Saro, F.J., Marinus, M.G., Modrich, P., and O'Donnell, M. (2006). The beta sliding clamp binds to multiple sites within MutL and MutS. J. Biol. Chem. *281*, 14340–14349.

Maisonneuve, E., Shakespeare, L.J., Jørgensen, M.G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. Proc. Natl. Acad. Sci. *108*, 13206–13211.

Maki, S., and Kornberg, A. (1988). DNA polymerase III holoenzyme of Escherichia coli. II. A novel complex including the gamma subunit essential for processive synthesis. J. Biol. Chem. *263*, 6555–6560.

De Marco, A. (2007). Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in Escherichia coli. Nat. Protoc. *2*, 2632–2639.

Martin, A., Baker, T.A., and Sauer, R.T. (2008). Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. Mol. Cell *29*, 441–450.

Modell, J.W., Hopkins, A.C., and Laub, M.T. (2011). A DNA damage checkpoint in Caulobacter crescentus inhibits cell division through a direct interaction with FtsW. Genes Dev. *25*, 1328–1343.

Mutschler, H., Gebhardt, M., Shoeman, R.L., and Meinhart, A. (2011). A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. PLoS Biol. *9*, e1001033.

Nakano, M.M., Hajarizadeh, F., Zhu, Y., and Zuber, P. (2001). Loss-of-function mutations in yjbD result in ClpX- and ClpP-independent competence development of Bacillus subtilis. Mol. Microbiol. *42*, 383–394.

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

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

Piotrowski, A., Burghout, P., and Morrison, D.A. (2009). spr1630 is responsible for the lethality of clpX mutations in Streptococcus pneumoniae. J. Bacteriol. *191*, 4888–4895.

Robinson, A., Brzoska, A.J., Turner, K.M., Withers, R., Harry, E.J., Lewis, P.J., and Dixon, N.E. (2010). Essential biological processes of an emerging pathogen: DNA replication, transcription, and cell division in Acinetobacter spp. Microbiol. Mol. Biol. Rev. MMBR *74*, 273–297.

Robinson, A., Causer, R.J., and Dixon, N.E. (2012). Architecture and conservation of the bacterial DNA replication machinery, an underexploited drug target. Curr. Drug Targets *13*, 352–372.

Sauer, R.T., and Baker, T.A. (2011). AAA+ proteases: ATP-fueled machines of protein destruction. Annu. Rev. Biochem. *80*, 587–612.

Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U., and Sorek, R. (2013). Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. Mol. Cell *50*, 136–148.

Schweder, T., Lee, K.H., Lomovskaya, O., and Matin, A. (1996). Regulation of Escherichia coli starvation sigma factor (sigma s) by ClpXP protease. J. Bacteriol. *178*, 470–476.

Skerker, J.M., Prasol, M.S., Perchuk, B.S., Biondi, E.G., and Laub, M.T. (2005). Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. PLoS Biol. *3*, e334.

Sliusarenko, O., Heinritz, J., Emonet, T., and Jacobs-Wagner, C. (2011). High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. Mol. Microbiol. *80*, 612–627.

Su'etsugu, M., and Errington, J. (2011). The replicase sliding clamp dynamically accumulates behind progressing replication forks in Bacillus subtilis cells. Mol. Cell *41*, 720–732.

Thanbichler, M., Iniesta, A.A., and Shapiro, L. (2007). A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in Caulobacter crescentus. Nucleic Acids Res. *35*, e137.

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H., and Moll, I. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli. Cell *147*, 147–157.

Walsh, C. (2003). Antibiotics : actions, origins, resistance (ASM Press).

Wang, J.D., Sanders, G.M., and Grossman, A.D. (2007). Nutritional control of elongation of DNA replication by (p)ppGpp. Cell *128*, 865–875.

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

Yamaguchi, Y., Park, J.-H., and Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. Annu. Rev. Genet. *45*, 61–79.

Yuan, J., Sterckx, Y., Mitchenall, L.A., Maxwell, A., Loris, R., and Waldor, M.K. (2010). Vibrio cholerae ParE2 poisons DNA gyrase via a mechanism distinct from other gyrase inhibitors. J. Biol. Chem. *285*, 40397–40408.

Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G., and Inouye, M. (2003). MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. Mol. Cell *12*, 913–923.

Chapter 3

Coevolving residues in toxin-antitoxin systems enforce interaction specificity

This work is in preparation for submission as Aakre, C.D., Herrou, J., Phung T.N., Perchuk, B., Crosson, S., and Laub M.T.

Summary

Toxin-antitoxin (TA) systems are important for bacterial responses to stress and are often present in multiple copies on bacterial chromosomes. However, the interaction specificity of these paralogous TA systems is unclear. Here, we show that members of the ParD-ParE TA family typically interact with only a single, cognate partner. To identify determinants of this interaction specificity, we measured amino-acid coevolution in large sets of cognate ParD-ParE pairs using a pseudo-likelihood-based approach. This method identified a set of specificity residues that are sufficient, when mutated, to reprogram a ParD antitoxin to interact with non-cognate ParE toxins. Further, we generated a library of ~10^4 ParD variants harboring different specificity residues and selected those that can antagonize a cognate toxin, a non-cognate toxin, or both. This analysis revealed an abundance of promiscuous ParD variants that are densely connected in sequence space to more specific variants. We demonstrate how such promiscuous states facilitate changes in TA specificity and promote the expansion of these systems, and likely other signaling proteins, by duplication and divergence.

Introduction

Toxin-antitoxin (TA) systems are genetic modules composed of co-operonic toxin and antitoxin genes. In general, the toxin encodes a stable globular protein and the antitoxin encodes an unstable protein that wraps around the toxin to inactivate it. Originally identified on plasmids (Ogura and Hiraga, 1983), TA systems were later found to exist widely on bacterial chromosomes and can constitute up to 2.6% of open reading frames in certain species (Leplae et al., 2011; Pandey and Gerdes, 2005). Toxins and antitoxins belong to diverse structural families, and cognate pairs often comprise a "mix and match" of toxins and antitoxins of different structural folds (Arbing et al., 2010). The biological function of TA systems is unclear, but they have been implicated in stress responses (Vesper et al., 2011), resistance to phage (Sberro et al., 2013), formation of persister cells (Maisonneuve et al., 2011), and bacterial pathogenicity (Helaine et al., 2014), among other functions (Van Melderen, 2010).

Given their prevalence in bacterial genomes, an unresolved question is whether TA systems interact in an exclusive one-to-one manner – that is, whether antitoxins only interact with their co-transcribed toxin and vice versa. Genetic data suggests that these interactions may be specific, as deleting an antitoxin is generally lethal unless it is supplied in *trans* (Fiebig et al., 2010). Additionally, certain antitoxins have been observed to only neutralize their co-operonic toxins, although these analyses have been limited in the number of cognate pairs tested (Hallez et al., 2010; Ramage et al., 2009). In contrast, other groups have reported that toxins and antitoxins encoded in different operons are capable of interacting *in vivo* and *in vitro* (Yang et al., 2010; Zhu et al., 2010). These interactions often occur between toxin families, suggesting that toxins and antitoxins could, in principle, form large, promiscuous networks that alter cellular physiology in

response to environmental cues, although the functional relevance of these non-cognate interactions remains uncertain (Zhu et al., 2010).

Protein interaction specificity can be encoded at multiple levels, but for many protein families, specificity is encoded primarily at the level of molecular recognition (Newman and Keating, 2003; Stiffler et al., 2007). Identifying the sequence determinants of this specificity is often difficult and relies on a combination of mutagenesis, computational, and library screening methods (Chen and Keating, 2012). One approach that has been used successfully is the analysis of amino-acid coevolution in large, multiple sequence alignments of cognate protein pairs. For example, a mutual information metric has been used to identify the specificity determining residues in the kinase-substrate pairs that comprise bacterial two-component signaling pathways (Capra et al., 2010; Skerker et al., 2008) and the residues controlling histidine kinase dimerization specificity (Ashenberg et al., 2011). More recent methods for analyzing coevolution, such as the pseudo-likelihood-based algorithm GREMLIN, have been introduced that are more accurate than mutual information in predicting residue-residue contacts (Kamisetty et al., 2013; Mao et al., 2015; Ovchinnikov et al., 2014). However, residues identified using these newer methods have not been experimentally demonstrated to control interaction specificity.

Here, we systematically measure the binding preferences for twenty ParD-ParE family members and find that toxins and antitoxins display a remarkable capacity to distinguish between cognate and non-cognate partners. This specificity appears to be encoded by a small set of coevolving residues at the toxin-antitoxin interface, as mutations in these residues are sufficient to reprogram the ParD antitoxin to interact with non-cognate toxins. Additionally, we generate a library of ~10^4 variants in these coevolving residues in the ParD antitoxin, and we select those

mutants that are able to antagonize a cognate toxin, non-cognate toxin, or both. We find that promiscuous variants, or those that can antagonize multiple toxins, are readily obtained, and that these promiscuous variants are densely connected in sequence space to specific variants. These results suggest that mutational paths that reprogram specificity tend to pass through promiscuous intermediates, and that the abundance of promiscuous states facilitates the expansion of paralogous protein families through duplication and divergence.

Results

Toxins and antitoxins from the ParDE family exhibit high interaction specificity

To more systematically measure interaction specificity of TA systems, we chose to focus on the ParD-ParE family, which has been experimentally characterized and is often found in multiple copies on bacterial chromosomes (Fiebig et al., 2010) (Figure 3.2A). To measure interaction specificity in this family, we cloned three chromosomal ParD-ParE pairs from the α-proteobacterium *Mesorhizobium opportunistum* into vectors that allow for separate and inducible expression of the ParE toxin and ParD antitoxin. We then co-transformed all possible toxin and antitoxin plasmid combinations into *E. coli*, and assessed whether the induced expression of each ParD antitoxin rescues the growth arrest resulting from inducing each ParE. As a control, we first confirmed that inducing each ParE toxin inhibited growth of *E. coli* (Figure 3.1A). Then, plating on a medium that induces both ParD and ParE, we observed growth for each of the three cognate ParD-ParE pairings (Figure 3.1A). No growth was observed for the six non-cognate pairs, indicating that ParD antitoxins from *M. opportunistum* can only neutralize their cognate ParE toxins.

We extended this analysis to 20 chromosomal ParDE pairs from eight different bacteria, including the 3 pairs from *M. opportunistum* (Figure 3.2B). For this 20x20 matrix of ParD and ParE pairs we observed strong interactions between all 20 cognate ParDE pairs, but only 11 of the 380 (or 3%) possible non-cognate pairings (Figure 3.1B). Importantly, these cross-reactions were only observed between ParDE pairs not encoded in the same species, indicating that the ParDE pairs within a given organism are typically specific and insulated from one another. To determine whether sequence similarity between ParD antitoxins is a predictor of cross-reactivity,

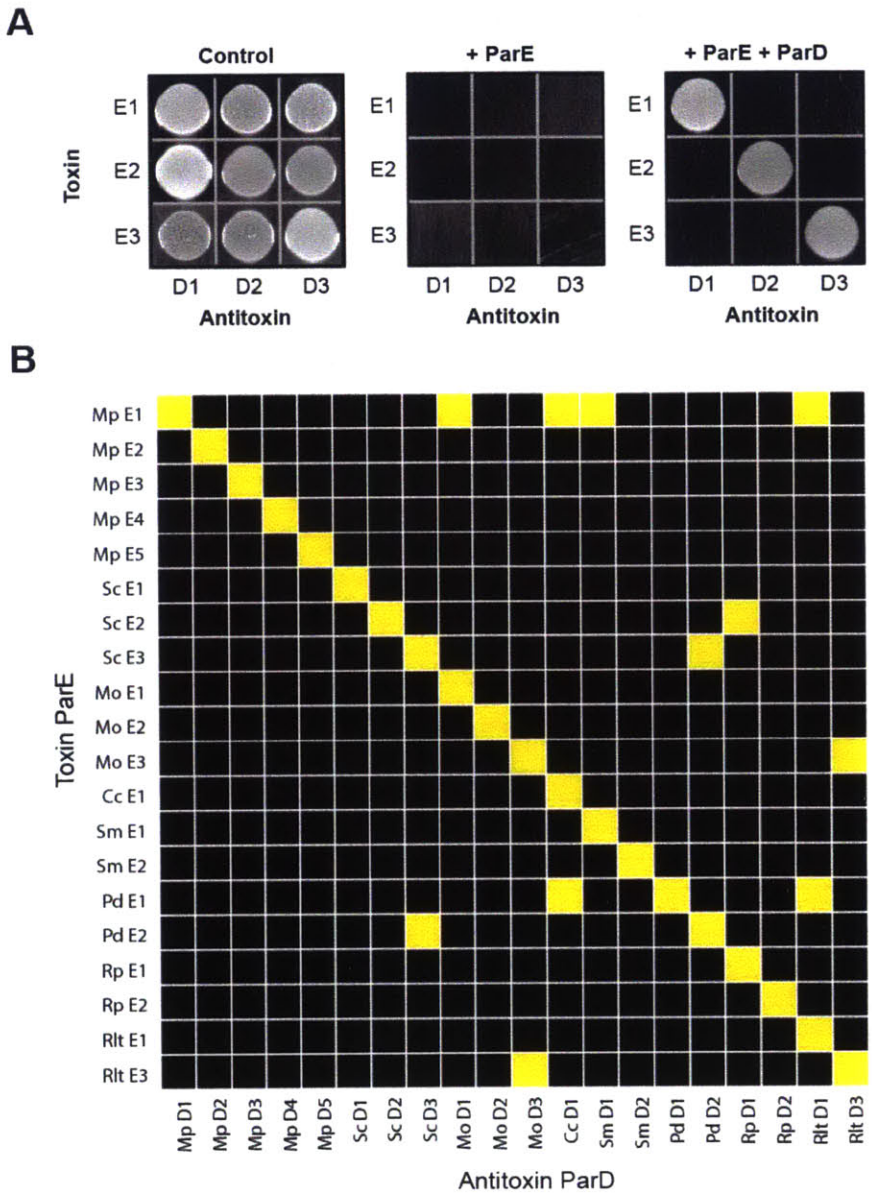
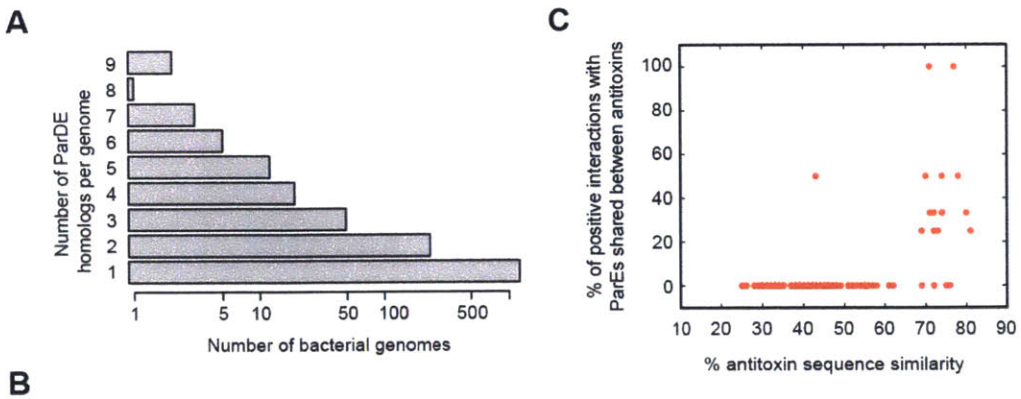


Figure 3.1. Toxins and antitoxins from ParD-ParE family exhibit binding specificity.

(A) Individual testing of interaction specificity for ParD antitoxins and ParE toxins from *Mesorhizobium opportunistum*. Control, plated on non-inducing conditions; + ParE, plated on medium containing 0.2% arabinose; + ParE + ParD, plated on medium containing 0.2% arabinose and 100 μM IPTG. **(B)** Comprehensive testing of interaction specificity for 20 ParD and ParE pairs from eight different bacteria. Cells containing each toxin and antitoxin plasmid combination were plated onto ParD and ParE inducing conditions (0.2% arabinose, 100 μM IPTG) and grown overnight at 37°C. Yellow, visible colonies observed following serial dilution; black, no visible colonies observed.



B

	Species	Name	ParD NCBI GeneID	ParE NCBI GeneID
1	Methylobacterium populi BJ001	Mp parDE1	188584325	188584324
2	Methylobacterium populi BJ001	Mp parDE2	188579862	188579861
3	Methylobacterium populi BJ001	Mp parDE3	188584614	188584613
4	Methylobacterium populi BJ001	Mp parDE4	188584366	188584365
5	Methylobacterium populi BJ001	Mp parDE5	188583979	188583978
6	Sphingobium chlorophenolicum L-1	Sc parDE1	334344591	334344592
7	Sphingobium chlorophenolicum L-1	Sc parDE2	334346364	334346365
8	Sphingobium chlorophenolicum L-1	Sc parDE3	334346363	334346362
9	Mesorhizobium opportunistum WSM2075	Mo parDE1	337266996	337266995
10	Mesorhizobium opportunistum WSM2075	Mo parDE2	337269626	337269627
11	Mesorhizobium opportunistum WSM2075	Mo parDE3	337270049	337270048
12	C. crescentus NA1000	Cc parDE1	221233854	221233853
13	Sinorhizobium meliloti 1021	Sm parDE1	16263810	16263811
14	Sinorhizobium meliloti 1021	Sm parDE2	15966432	15966431
15	Paracoccus denitrificans PD1222	Pd parDE1	119386537	119386538
16	Paracoccus denitrificans PD1222	Pd parDE2	119385788	119385787
17	Rhodopseudomonas palustris BisB18	Rp parDE1	90423718	90423719
18	Rhodopseudomonas palustris BisB18	Rp parDE2	90421684	90421683
19	Rhizobium leguminosarum bv. Trifolii WSM2304	Rlt parDE1	209552210	209552209
20	Rhizobium leguminosarum bv. Trifolii WSM2304	Rtl parDE3	209552404	209552403

Figure 3.2. TA systems are widely present on bacterial genomes; sequence similarity is a poor predictor of interaction affinity.

(A) Homologs of *Mesorhizobium opportunistum* ParD3 and ParE3 were identified independently using an iterative jackhmmer search (Experimental Procedures). Pairs of ParD and ParE homologs were then paired if they were genome neighbors. The number of bacterial genomes that have the indicated number of ParD-ParE homologs is shown. **(B)** Name and genome accession numbers for the twenty cloned ParD-ParE systems. **(C)** The percentage sequence similarity between antitoxins is a poor predictor of shared positive interactions with ParE toxins. For each possible antitoxin pairing from Figure 3.1, we calculated the percentage sequence similarity and the percentage of shared positive ParE interactions (number of shared positive ParE interactions divided by the total positive interactions between both antitoxins).

we compared the percentage of shared positive interactions to the average sequence similarity for all ParD antitoxin pairings. Although the percentage of shared positive interactions increased with an increase in sequence identity, there were many highly similar antitoxins (>75% sequence similarity) that shared none of the same positive interactions (Figure 3.2C). These results indicate that ParD antitoxins are highly specific for their cognate ParE toxins and that global sequence similarity is a poor predictor of interaction specificity.

Computational identification of covarying residues in ParD and ParE

As a first step to understanding the molecular basis of specificity in ParD-ParE complexes, we solved a crystal structure of *M. opportunistum* ParD3 in complex with its cognate toxin ParE3 at a resolution of 1.59 Å. This structure revealed a heterotetrameric asymmetric unit composed of ParD3 and ParE3 dimers (Figures 3.4A). The crystal packing suggested that the biological assembly may be composed of two tetramers, which was supported by an estimated mass of ~87 kDa in solution (Figure 3.4B). We found that ParD3 makes extensive contacts with ParE3 through its second and third alpha helices, for a total buried surface area of 1624 $Å^2$ between the ParD3 and ParE3 monomers (Figure 3.3A).

Previous work with bacterial two-component signaling systems has revealed that interaction specificity is controlled by a subset of residues at the protein-protein interface formed by a histidine kinase and response regulator (Capra et al., 2010; Skerker et al., 2008). These specificity-determining residues coevolve to maintain the interaction between cognate signaling proteins. Thus, to pinpoint the residues that contribute to the specificity of ParD-ParE interactions, we searched for residues that strongly covary in a multiple sequence alignment of concatenated, co-operonic ParD and ParE proteins. We used the GREMLIN pseudo-likelihood-based approach,

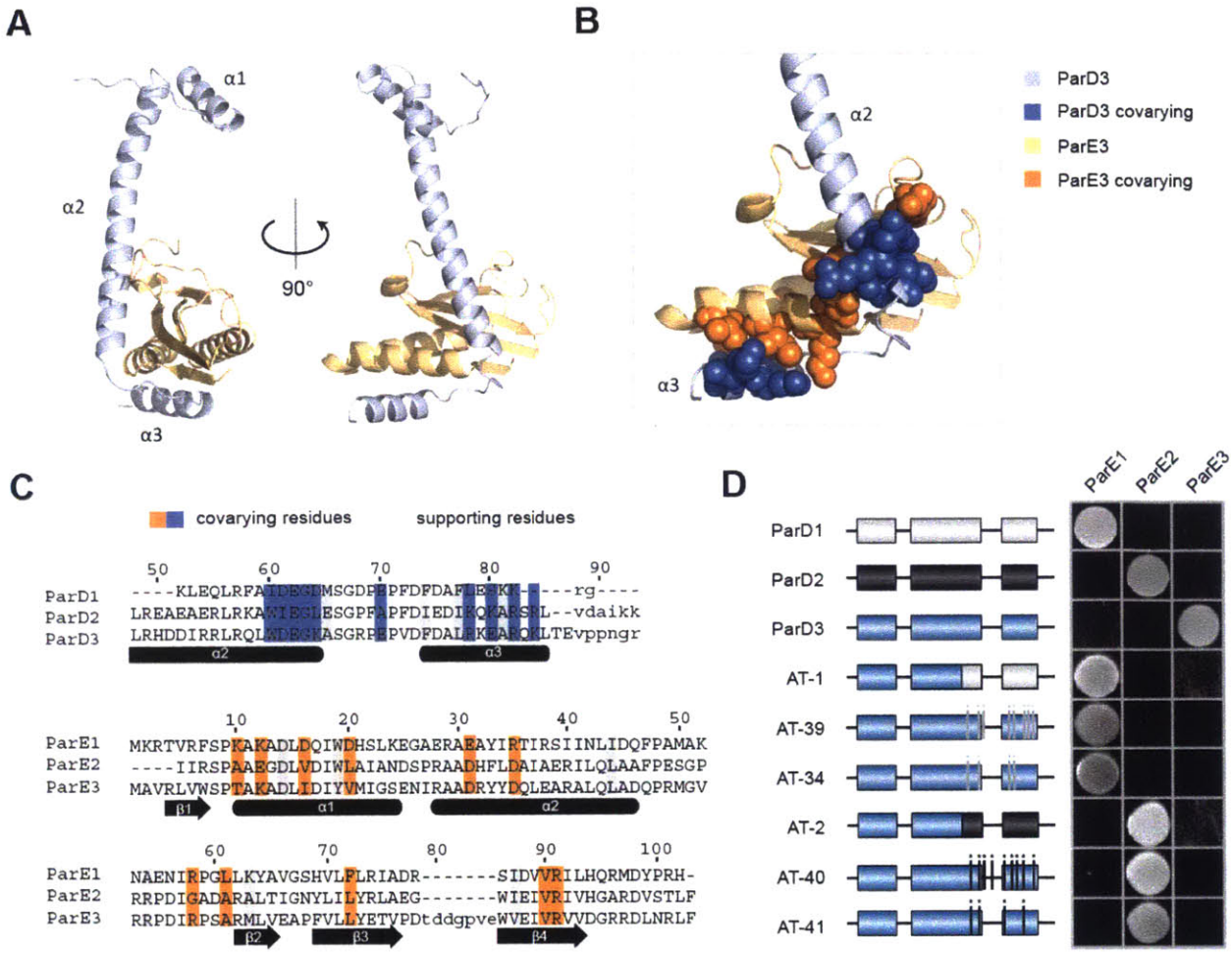


Figure 3.3. Covarying residues dictate interaction specificity in ParD-ParE family.

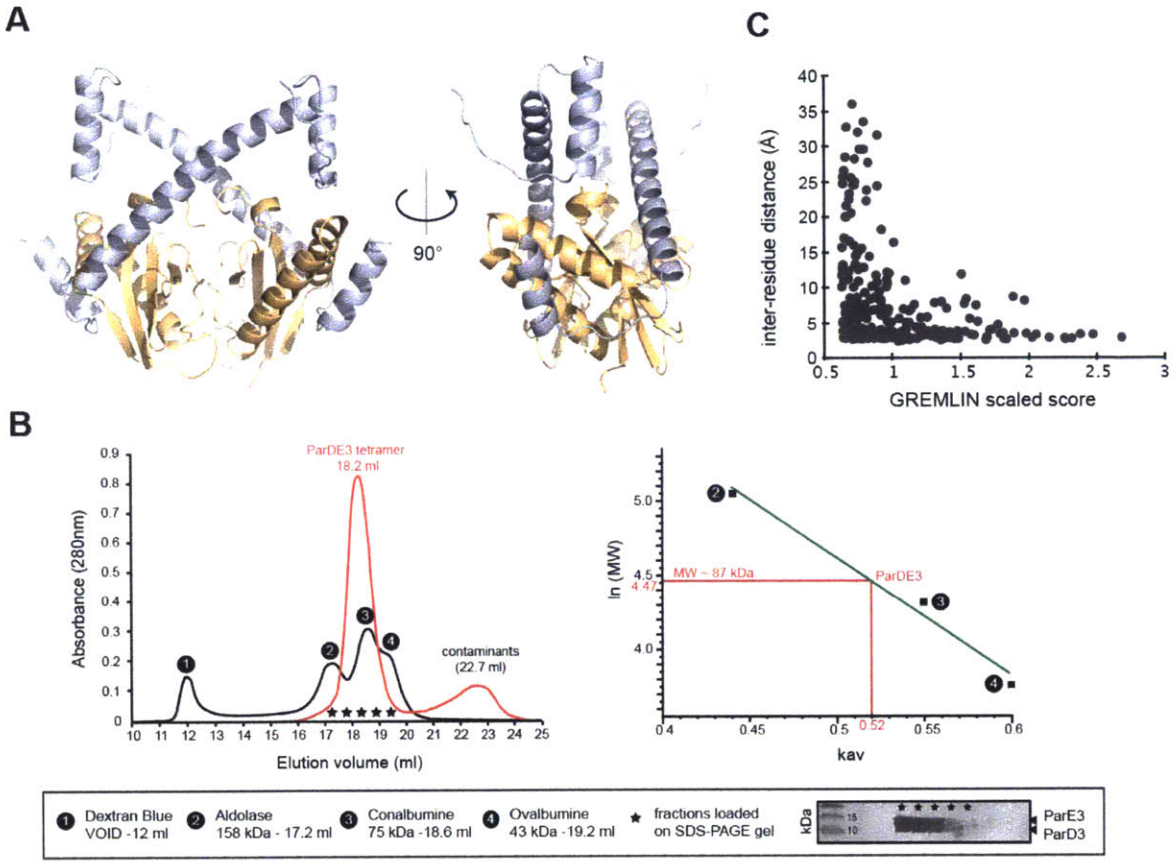
(A) Structure of the *M. opportunistum* ParD3-ParE3 complex. Light orange, ParE3 monomer; light blue, ParD3 monomer. The three alpha helices of ParD3 are labeled. **(B)** Coevolving residues cluster into two groups. ParD3-ParE3 structure from (A) is shown magnified; covarying residues are shown in space-filling representation. **(C)** Alignment of ParD and ParE homologs from *M. opportunistum* with coevolving residues highlighted in blue or orange for ParD or ParE, respectively. Supporting residues, which are coevolving with the interfacial coevolving residues, are shown in grey. **(D)** Mutations in the C-terminus of ParD3 can re-program interaction specificity. The indicated ParD3 mutants were tested against each ParE homolog in *M. opportunistum* using the toxicity-rescue assay. AT-1 and AT-2 have the entire C-terminus of ParD3 substituted for the corresponding region in ParD1 or ParD2. AT-39, AT-34, AT-40, and AT-41 have more targeted mutations in the ParD3 C-terminus. For a full list of substitutions in these mutants, see Figure 3.4D.

which has improved accuracy in identifying residue-residue contacts compared to previous methods (Kamisetty et al., 2013; Ovchinnikov et al., 2014). We found that the average inter-residue distance in our ParD3-ParE3 structure decreased as a function of increasing GREMLIN scaled coupling score (Figure 3.4C), indicating that GREMLIN was able to identify residues at the molecular interface of this complex.

At a scaled coupling score threshold of 1.25, we identified a total of 10 residues in ParD and 11 residues in ParE that are coevolving most strongly, hereafter referred to as 'specificity' residues as our work below indicates that they play the dominant role in determining the partner interaction specificity. Mapping these specificity residues onto the ParD3-ParE3 crystal structure indicated that they cluster into two groups (Figure 3.3B). The first group sits at the base of the second alpha helix in ParD3 and covaries with residues in the four-stranded beta sheet in ParE3 (Figures 3.3C). The second group clusters in the third alpha helix in ParD3 and covaries with residues in the first and second alpha helices of ParE3 (Figures 3.3C). We also identified residues within each protein that coevolve with the specificity residues. These 'supporting' residues may indirectly contribute to ParD-ParE interaction specificity by influencing the orientation or packing of interfacial specificity residues. At a scaled coupling score threshold of 1.25, we identified four and six supporting residues in ParD and ParE, respectively (Figures 3.3C, 3.4F).

Covarying residues dictate interaction specificity in the ParD-ParE family

To determine whether the coevolving residues identified are sufficient to dictate interaction specificity in the ParD-ParE family, we constructed a series of chimeric proteins in which different regions of the *M. opportunistum* ParD3 were replaced with the corresponding regions of ParD1 or ParD2 (Figure 3.3D). Replacing the entire C-terminal region of ParD3 with the



Please see figure legend on next page

Figure 3.4. ParD3-ParE3 biological assembly is a dimer of tetramers; supplemental information on mutants in ParD3 C-terminus.

(A) Structure of the tetrameric *M. opportunistum* ParD3-ParE3 complex. Light orange, ParE3 dimer; light blue, ParD3 dimer. The full complex is a dimer of tetramers (not shown). **(B)** Estimation of the oligomeric state of the ParD3-ParE3 complex using size exclusion chromatography. Comparison of ParD3-ParE3 elution profile (left panel) to molecular weight standards (right panel) yields an estimated molecular weight of ~87 kDa, which is consistent with a dimer of tetramers model. **(C)** Inter-residue distance decreases as the GREMLIN scaled score increases. All residue-residue pairs (within and between proteins) were scored. **(D)** List of ParD3 mutants tested in Figures 3.3D and 3.4E. Substitutions from the ParD3 wild-type sequence are shown in red. **(E)** Mutants in the ParD3 C-terminus were tested against each ParE homolog in *M. opportunistum* as in Figure 3.3D. **(F)** Structure of *M. opportunistum* ParD3-ParE3 with the supporting residues (Figure 3.3C) shown in grey space-filling representation.

corresponding region of ParD1 or ParD2 produced a chimera that had lost its ability to interact with ParE3 and gained the ability to interact with ParE1 or ParE2, respectively (Figure 3.3D, AT-1 and AT-2). These chimeras involved both clusters of residues identified above as coevolving between the ParD and ParE proteins. Replacing only one of these clusters of residues gave different results depending on the antitoxin tested. For example, replacing the first or second cluster of ParD3 with the corresponding region of ParD1 resulted in a non-functional antitoxin that does not neutralize any tested toxin (Figure 3.4E, AT-3 and AT-5). In contrast, replacing the first or second cluster of ParD3 with the corresponding region of ParD2 resulted in a chimera that interacts with ParE2, although the chimera in which only the second cluster was replaced still interacted with ParE3 (Figure 3.4E, AT-4 and AT-6). These results indicate that the C-terminal region containing the specificity and supporting residues is sufficient to dictate interaction specificity.

To pinpoint the residues required for interaction specificity, we focused our mutagenesis on the coevolving residues identified computationally using GREMLIN. We generated variants of ParD3 in which all of the specificity and supporting residues were replaced with the corresponding residues in ParD1 or ParD2, for a total of eight or nine substitutions, respectively. We found that in each case, these mutations were sufficient to reprogram ParD3 to interact with ParE1 or ParE2 and lose its ability to interact with ParE3 (Figures 3.3D, AT-39 and AT-40). Interestingly, ParD3 could be reprogrammed to interact with ParE1 or ParE2 with fewer substitutions. For example, we found sets of four substitutions that were sufficient to reprogram ParD3 to interact with ParE1 or ParE2 (Figures 3.3D, AT-34 and AT-41). Taken together, our results indicate that mutating the most highly coevolving residues in an antitoxin can be sufficient to reprogram its interaction specificity, and that in some cases mutating only a subset of these residues is required for a complete switch in its partner specificity.

High-throughput mapping of mutant fitness at coevolving interface

Previous work has identified extensive degeneracy of the coevolving specificity residues in a two-component signaling protein interface (Podgornaia and Laub, 2015). To determine the extent of degeneracy in the ParD-ParE interface, we generated a library of mutants at four interfacial positions in the ParD3 antitoxin, Leu-59, Trp-60, Asp-61, and Lys-64 (LWDK), that include three strongly coevolving residues (positions 60, 61, and 64 have scaled scores > 1.8) and one more weakly coevolving residue (position 59 has a scaled score of 0.9). To reduce the complexity of our library, we only allowed residues at each library position that are commonly found in naturally occurring ParD homologs (see Experimental Procedures). The resulting library has a theoretical diversity of 9,360 variants, with 12, 6, 13, and 10 residues encoded at the four

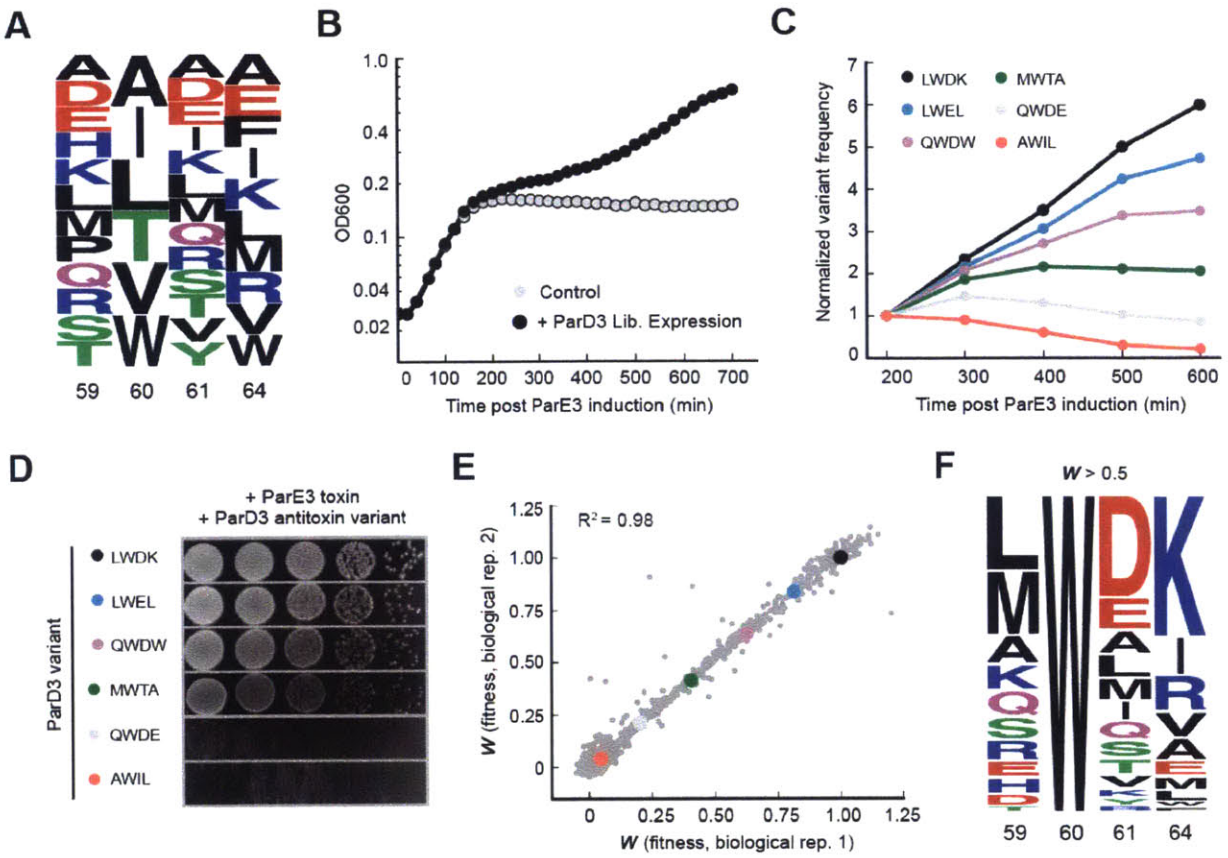


Figure 3.5. High-throughput mapping of mutant fitness at co-evolving interface.

(A) Amino acid composition of ParD3 antitoxin library at each randomized position. **(B)** Liquid growth assay following ParE3 toxin induction. Control, absence of ParD3 library expression; + ParD3 Lib. Expression, ParD3 library expression induced at time 0 with the addition of 100 μM IPTG. **(C)** Time-resolved frequency changes for the indicated ParD3 variants following ParE3 induction. **(D)** Individual testing of ParD3 library variants from (C) using the toxicity rescue assay. Cells containing the indicated antitoxin and toxin plasmids were serially diluted and plated onto medium that induces toxin and antitoxin expression (0.2% arabinose, 100 μM IPTG). Cells were grown overnight at 37°C and imaged. **(E)** Fitness measurements of the ParD3 library against the ParE3 toxin are highly reproducible between biological replicates. ParD3 library variants are as follows: black, LWDK; blue, LWEL; purple, QWDW; green, MWTA; grey, QWDE; red, AWIL. **(F)** Amino acid composition of ParD3 library variants with high fitness ($W > 0.5$) against ParE3.

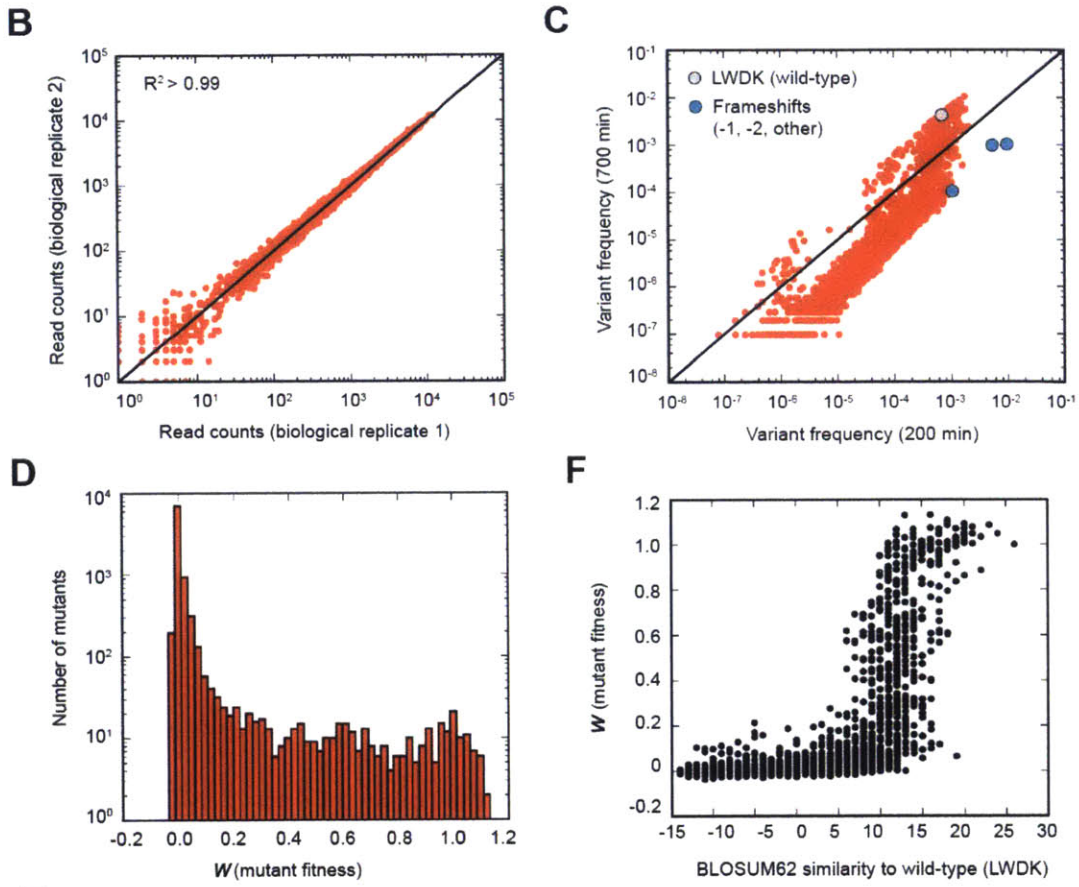
respective positions of the library (Figure 3.5A). Deep-sequencing of the relevant region in *parD3* in the initial library revealed that >98% of the predicted variants were represented by at least 10 reads and >94% had at least 100 reads (Figure 3.6A). Measurements of read numbers were highly reproducible between replicates ($R^2 > 0.99$, Figure 3.6B).

To assess the ability of each ParD3 variant to bind and antagonize ParE3, we co-transformed *E. coli* with the ParD3 library and a vector for inducing the expression of wild-type ParE3. When cultured in conditions that do not induce ParD3, cell growth arrested within 200 minutes after inducing the ParE3 toxin (Figure 3.5B). In contrast, when the ParD3 library was expressed, growth slowed after inducing the toxin but eventually resumed, suggesting that some fraction of the population could neutralize ParE3 toxicity (Figure 3.5B). To determine which mutants could neutralize ParE3 and hence were enriched during the course of this experiment, we harvested samples every 100 minutes starting 200 minutes after inducing ParE3 and deep-sequenced the relevant region of *parD3*. We observed large changes in the frequency of individual variants over this 400 minute time course (Figure 3.6C). For example, the wild-type sequence (LWDK) was enriched ~6-fold, whereas sequences with frameshift mutations in *parD3* that are presumably non-functional were depleted ~7-fold (Figure 3.6C).

To validate the functionality of variants identified in this competitive growth assay, we isolated six mutants that exhibited different frequency dynamics following toxin induction (Figure 3.5C). We tested these six mutants individually using our toxicity-rescue assay, and found clear agreement between the change in the frequency of each variant and plating efficiency (Figure 3.5D). To quantify the differences in variant behavior during competitive growth, we generated a linear fit to the frequencies of each mutant as a function of time, and then calculated the log-fold

A

	>3 reads	>10 reads	>100 reads
replicate 1	9290 (99.3%)	9215 (98.5%)	8876 (94.8%)
replicate 2	9292 (99.3%)	9211 (98.4%)	8860 (94.7%)
both	9267 (99.0%)	9200 (98.3%)	8835 (94.4%)



E

mutations	scored	functional (W_{raw} > 0.5)
0	1	1 (100%)
1	37	31 (84%)
2	499	189 (38%)
3	2798	31 (1%)
4	5859	0 (0%)

Figure 3.6. Statistics on high-throughput sequencing of ParD3 library against ParE3 toxin.

(A) Number of library variants present with at least the indicated number of reads. **(B)** Read counts are highly reproducible between replicates. The ParD3 library was independently transformed into *E. coli* and grown to saturation overnight. This library was then subjected to deep sequencing and read counts compared. **(C)** Library variant frequencies change following induction of the ParE3 toxin. Shown is a scatter plot of variant frequency pre-toxin induction (200 min) and after 500 minutes of toxin induction (500 min). Grey point, wild-type LWDK sequence; blue points; frameshift mutations that are predicted to be non-functional. **(D)**

Histogram of library variants by fitness against ParE3. Most variants scored against ParE3 are low fitness (W < 0.05). **(E)** Number of functional library variants with the indicated number of mutations from the wild-type sequence, LWDK. **(F)** Chemical similarity of library residues to the wild-type sequence, LWDK, is a modest predictor of fitness. Shown is a scatter plot of BLOSUM62 similarity to LWDK versus mutant fitness.

expansion of each mutant relative to the rest of the population. These calculations yielded W_{raw}, the fitness value for each mutant (Experimental Methods). We then transformed these relative fitness values such that the *W* value for frameshift variants was 0 and the *W* value for the wild-type (LWDK) sequence was 1; the resulting distribution of *W* values ranges from -0.04 to 1.13, with >91% of values falling below 0.05 (Figure 3.6D). These fitness values were highly reproducible between biological replicates (Figure 3.5E, $R^2 = 0.98$).

To determine the degeneracy of the ParD3 interface residues varied in our library, we selected those variants with *W* values > 0.5. There were 252 such variants, representing 2.7% of the total (Figure 3.5F). This set included the wild-type combination of residues (LWDK) and, relative to the wild-type sequence, 31 single, 189 double, and 31 triple mutants (Figure 3.6E). There were no quadruple mutants as position 60 was invariantly tryptophan. The most common residues in this set as a whole were the wild-type residues: Leu-59, Trp-60, Asp-61, and Lys-64. Additionally, we detected a correlation between the chemical similarity of each variant to the wild-type sequence, as measured by BLOSUM62, and our fitness values (Figure 3.6F). Together, our results indicate that degeneracy varied by library position and that most high-scoring variants tended to be similar to the wild-type sequence.

High-throughput mapping of mutants against non-cognate toxin

We showed earlier that a small number of substitutions can reprogram ParD3 to interact with ParE2, a non-cognate toxin also present in the *M. opportunistum* genome (Figure 3.3D). We were thus interested in whether our library contained variants that could interact with the non-cognate ParE2 toxin. To identify ParE2-reactive variants, we co-transformed our ParD3 library with a plasmid carrying an inducible copy of ParE2 and performed a competitive growth experiment. As before, we observed growth rescue following ParD3 library expression and large changes in the frequency of individual variants over time (Figures 3.7A, 3.8A). However, the frequency changes observed here differed from those observed against the cognate toxin ParE3. For example, a variant containing the specificity residues found in the native ParD2 antitoxin, AWIL, was enriched in the ParD3 library screened against ParE2 but was depleted when screened against ParE3 (Figure 3.7B). Some variants, such as LWEL, were enriched against both ParE2 and ParE3, demonstrating the existence of promiscuous mutants that can bind to more than one toxin (Figures 3.5C, 3.7B).

To quantify the observed differences in enrichment rates, we calculated the fitness of each variant against ParE2 in biological replicates (Figure 3.8B, $R^2 = 0.94$). We found a total of 151 variants, or 1.6% of the total, with W values > 0.5 (Figure 3.7C). The most common residues were Ala-59, Trp-60, Leu-61, and Leu-64, with the tryptophan at position 60 again being invariant (Figure 3.7C). However, we noted important differences between variants reactive against ParE2 and ParE3. ParE2-specific variants tended to have small hydrophobic or positively-charged residues at position 61, whereas ParE3-specific variants favored negatively-charged residues. Additionally, ParE2-specific variants were more likely to contain small hydrophobic residues at position 64,

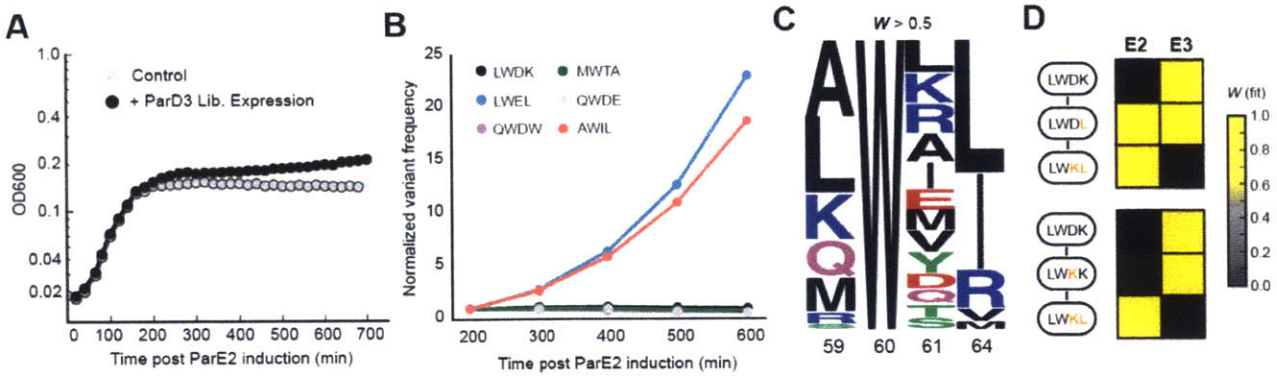


Figure 3.7. High-throughput mapping of mutant fitness against non-cognate toxin.

(A) Liquid growth assay following induction of the non-cognate ParE2 toxin. Control, absence of ParD3 library expression; + ParD3 Lib. Expression, ParD3 library expression induced at time 0 with the addition of 100 μM IPTG. **(B)** Time-resolved frequency changes for the indicated ParD3 library variants. **(C)** Amino acid composition of ParD3 library variants with high fitness (W > 0.5) against ParE2. **(D)** Contribution of positions 61 and 64 to insulation of ParD3-ParE3 interaction. Mutational paths from ParD3(LWDK) to ParD3(LWKL) are shown. Fitnesses for each ParD3 variant against ParE2 or ParE3 are shown on right as a heatmap; yellow is high fitness, black is low fitness.

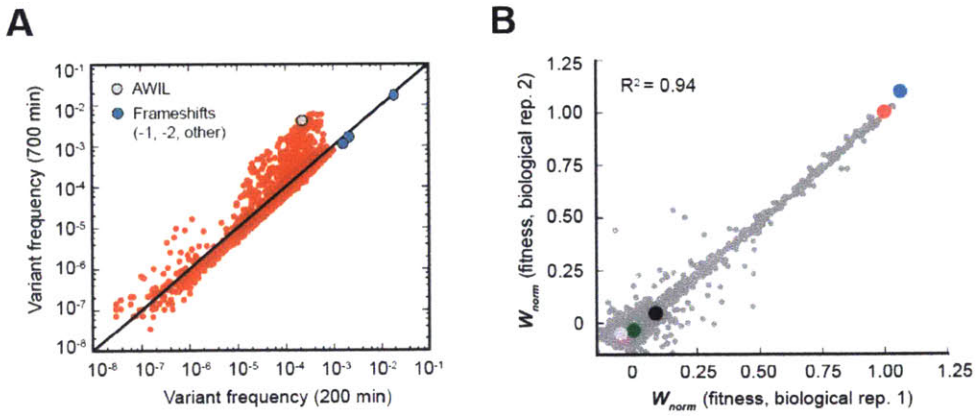


Figure 3.8. Statistics on high-throughput sequence of library against non-cognate ParE2 toxin.

(A) Change in library variant frequencies post induction of the non-cognate ParE2 toxin (as in Figure 3.6C). Grey point, AWIL sequence; blue points, frameshift mutations that are predicted to be non-functional. **(B)** Fitness measurements of ParD3 library against non-cognate ParE2 toxin in biological replicates. ParD3 library variants are as follows: black, LWDK; blue, LWEL; purple, QWDW; green, MWTA; grey, QWDE; red, AWIL.

whereas ParE3-specific variants tended to have positively-charged residues (Figures 3.5F, 3.7C). Further, we found that positions 61 and 64 contribute significantly to the insulation and specificity of the ParD-ParE system, as a single substitution (K64L) in ParD3, producing variant LWDL, resulted in promiscuous binding to ParE2 and ParE3, and a second substitution (D61K), resulting in variant LWKL, conferred specificity for ParE2 (Figure 3.7D). Interestingly, incorporating these substitutions in the reverse order, D61K and then K64L, results in a "switch-like" change in specificity that is potentiated by the first mutation, rather than a promiscuous intermediate (Figure 3.7D). These results underscore how a small number of mutations can fully reprogram interaction specificity and demonstrate that the order of mutations can strongly affect whether the path to a new specificity state involves a promiscuous intermediate or immediate switch.

Mutational paths that reprogram specificity tend to involve promiscuous variants

To more systematically probe the sequence space governing the specificity of ParD3, we generated a scatterplot of ParD3 variant fitness against the ParE2 and ParE3 toxins (Figure 3.9A). This analysis revealed variants spanning all ranges of fitness, including those capable of antagonizing ParE2, ParE3, or both ParE2 and ParE3 simultaneously (Figure 3.9A). We identified a total of 31 variants that we classified as promiscuous ($W > 0.5$ for both toxins), which represents a subset of the 252 ParE3-reactive and 151 ParE2-reactive variants noted above (Figure 3.9B). We grouped variants by specificity class and found that promiscuous variants tend to harbor sequence elements from both ParD3 and ParD2, such as negatively charged residues at position 61 (ParD3-like) and small hydrophobic residues at position 64 (ParD2-like) (Figure 3.10A).

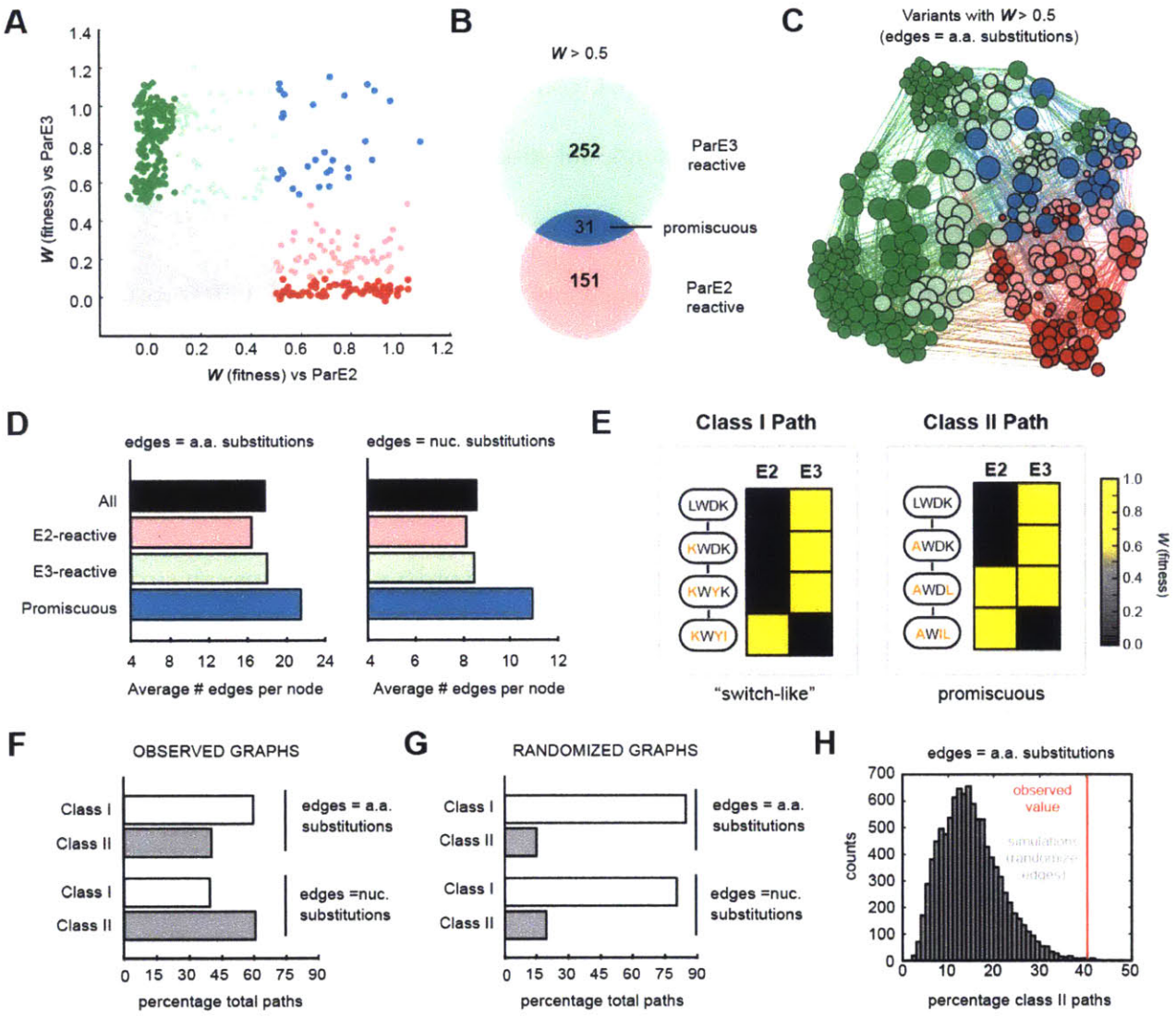


Figure 3.9. Mutational paths through sequence and specificity space.

(A) Fitness space of ParD3 variants against ParE2 and ParE3. Green, specific for ParE3; blue, capable of antagonizing both ParE2 and ParE3; red, specific for ParE2. **(B)** Venn diagram of ParD3 variants that are reactive against ParE3, ParE2 or both. **(C)** Force-directed graph of all ParD3 variants with $W > 0.5$ against ParE3 or ParE2. Nodes represent individual variants and edges represent single amino acid substitutions. Node size scales with increasing degree and color corresponds to the specificity classes in (A). **(D)** Average number of edges per node for each specificity class. Left, calculations for amino acid substitution graph; right, calculations for nucleotide substitution graph. **(E)** Class I and class II path examples with corresponding fitness data for each trajectory. Yellow, high fitness; black, low fitness. Class I paths do not cross any promiscuous intermediates, whereas class II paths cross at least one. **(F)** Percentage class I and class II paths from the wild-type ParD3 sequence (LWDK) to each of the 66 ParE2-specific variants ($W_{E2} > 0.5$, $W_{E3} < 0.1$). Top, calculations for amino acid substitution graph; bottom,

calculations for nucleotide substitution graph. **(G)** Same as (F) except calculated on 10,000 simulations in which the graph edges were randomly shuffled while keeping the total edge count constant. Shown are the average class I and class II path percentages for these simulations. **(H)** Enrichment of class II paths is highly significant. Histogram represents the percentage of class II paths for 10,000 simulations in which the edges were randomly shuffled, red line represents percentage of class II paths for the observed graph.

To visualize the connectivity of functional variants, we created a force-directed graph where nodes represent functional sequences, node size increases with greater connectivity, node color represents specificity class, and edges connect variants that differ by one amino acid. The resulting graph is densely connected and groups variants based on their specificity (Figure 3.9C). The average number of edges per node, or degree, was 17.8 and ranged from 7 to 31. However, we noted that the average number of edges per node was 23% higher for promiscuous variants than for variants specific for ParE2 or ParE3 (Figure 3.9D). We also generated a force-directed graph in which edges represent variants that differ by a single nucleotide substitution, following the standard genetic code (Figure 3.10B). For this graph, promiscuous variants were, on average, 31% more connected to other functional nodes than their ParE2- or ParE3-specific counterparts (Figure 3.9D). This increased connectivity of promiscuous variants was highly significant as it was lost when the graph's edges were randomly shuffled ($p < 10^{-4}$, Figures 3.10C-D). These results indicate that functional sequences are densely connected in sequence space and that promiscuous variants tend to be more highly connected than other variants.

The high degree of connectivity of promiscuous variants suggested that mutational paths that change ParD3 specificity (from ParE3-specific to ParE2-specific, or vice versa) may tend to travel through promiscuous intermediates. To test this hypothesis, we first defined two types of

specificity-reprogramming paths. Note that for the following analysis, we exclude paths in which ParD3 does not interact with either ParE3 or ParE2 (also, see Discussion). Class I paths are "switch-like" and only involve intermediates that are specific for ParE2 or ParE3, whereas class II paths travel through at least one promiscuous intermediate (Figure 3.9E). To determine whether paths that change the interaction specificity of ParD3 tend to be class I or class II, we identified the shortest mutational paths from the wild-type ParD3 variant (LWDK) to each of the 66 variants that strongly prefer ParE2 over ParE3 ($W_{E2} > 0.5$, $W_{E3} < 0.1$); for this analysis, each mutational step involved a single amino-acid substitution. We found a total of 370 shortest paths, of which 40% belonged to the class II category involving a promiscuous intermediate (Figure 3.9F). The percentage of class II paths increased to 61% when considering only paths that involve single nucleotide substitutions (Figure 3.9F).

To determine whether the number of paths that involve promiscuous variants is greater than would be expected by chance, we generated graphs in which the edges were randomly shuffled and again calculated the percentage of class I and class II paths from ParD3 (LWDK) to the ParE2-specific variants. For such graphs with randomized edges, the percentage of class II paths dropped to 15% for the amino acid neighbor graph and 20% for the nucleotide neighbor graph (Figure 3.9G). Thus, the enrichment of class II paths in the observed graphs is highly significant ($p = 0.002$, amino acid neighbor graph; $p = 0.003$, nucleotide neighbor graph) (Figures 3.9H, 3.10E). Collectively, our results demonstrate the dense connectivity of the sequence space governing specificity of the ParD-ParE family and reveal that specificity-reprogramming paths are enriched for those that involve promiscuous variants, which likely facilitates the evolution of ParD-ParE systems with new specificities.

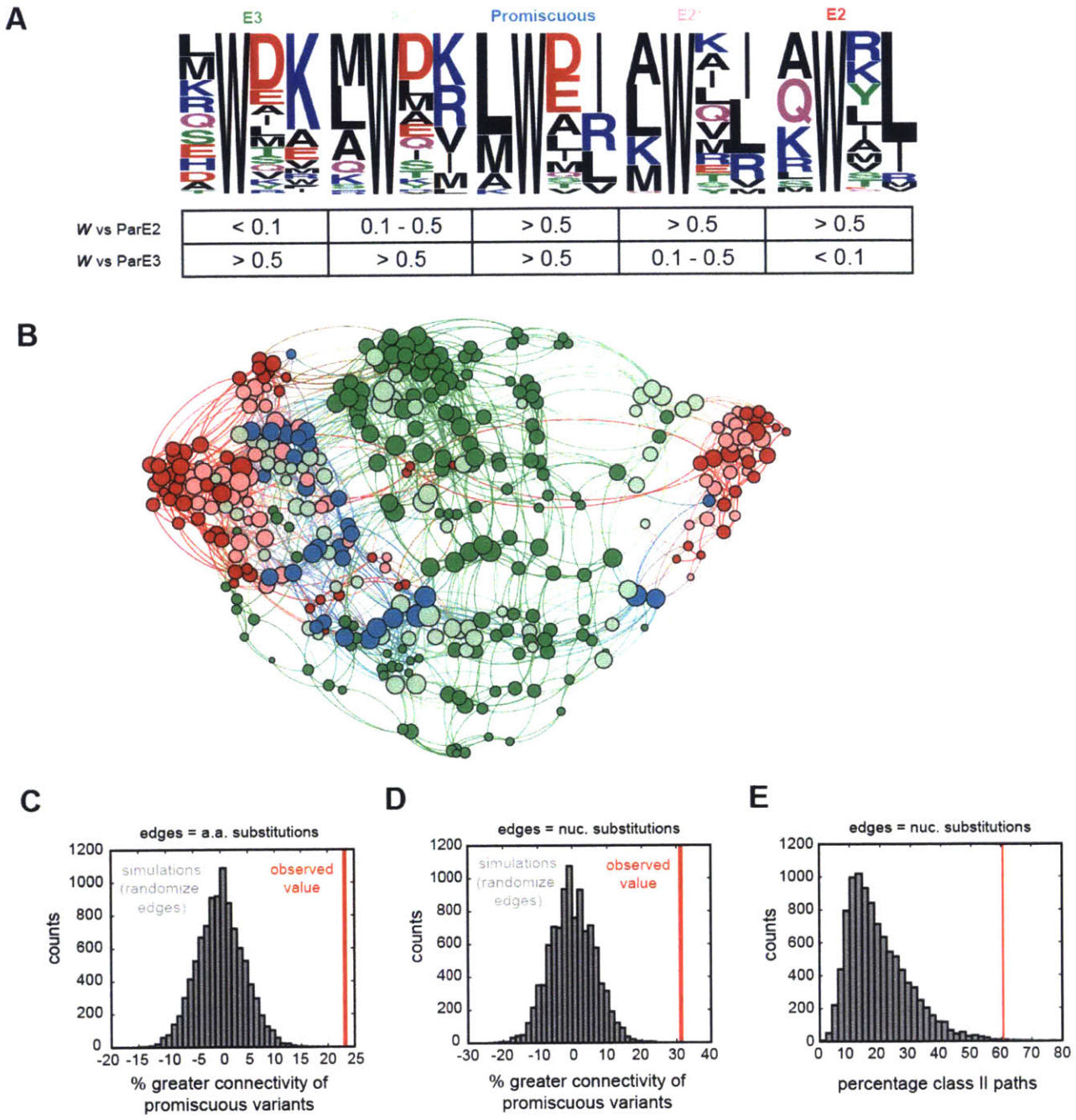


Figure 3.10. Sequence composition by specificity class; connectivity of nucleotide neighbors

(A) Amino acid composition of library variants in each specificity class. **(B)** Force-directed graph of all ParD3 variants with $W > 0.5$ against ParE3 or ParE2. Nodes represent individual variants and edges represent single nucleotide substitutions. Node size scales with increasing degree and color corresponds to specificity classes in Figure 3.9A. **(C)** The greater connectivity of promiscuous variants is highly significant. To determine whether the increased connectivity could be observed by chance, we randomized the edge connectivity in the amino acid

substitution graph while keeping the number of edges constant. We then calculated the percentage greater connectivity of promiscuous variants versus non-promiscuous variants. Grey, result of 10,000 simulations; red line, observed value. **(D)** Same as (C), except calculated on the nucleotide substitution graph. **(E)** Enrichment of class II paths is highly significant. Same calculations as Figure 3.9H, except for nucleotide substitution graph.

Mutational trajectories to an orthogonal ParD3-ParE3 pair

To further probe how the interaction specificity of the ParD-ParE system can evolve, we sought to first generate a variant of the toxin ParE3 containing a combination of specificity residues that does not allow for an interaction with ParD3, and then select ParD3 variants from our library that can neutralize the novel toxin. To this end, we first generated a mutational variant of the toxin, ParE3*, that retains toxicity but is incapable of binding to its cognate antitoxin, ParD3. We focused on five positions in ParE3 that strongly covary with the specificity residues Trp-60 and Asp-61 in ParD3. These residues in ParE3 are Arg-54, Arg-58, Ala-61, Met-63, and Leu-72, or RRAML. We chose to incorporate the residue combination VEIRF into ParE3, creating the putative orthogonal toxin ParE3*. The individual residues VEIRF were both frequently observed in ParE3 homologs and the most chemically distant from the wild-type combination RRAML (Figure 3.12A). To confirm that ParE3* is insulated from ParD3 we used our toxicity-rescue assay, and we found that ParE3* retained its toxicity but was no longer neutralized by ParD3 (Figure 3.11).

Next, to determine whether there are variants in the ParD3 library capable of neutralizing ParE3*, we co-transformed our ParD3 library with the ParE3* toxin and performed a competitive growth experiment. Similar to before, we noticed enrichment of particular variants over time, and these changes in frequency were converted to reproducible fitness measurements ($R^2 = 0.96$,

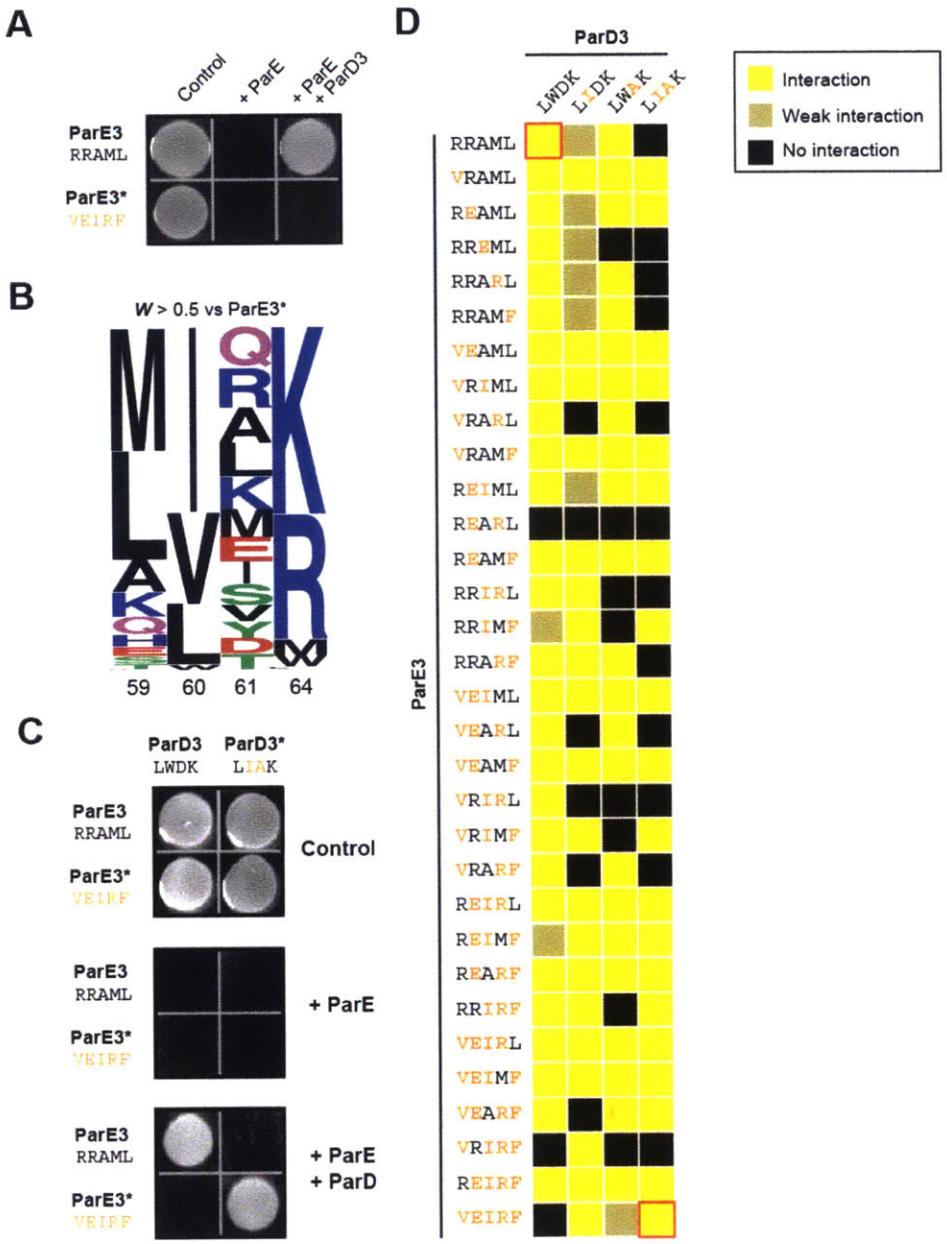


Figure 3.11. Mutational trajectories to an orthogonal ParD3*-ParE3* pair

(A) ParE3* is insulated from its cognate antitoxin ParD3. Plasmids containing ParE3 or ParE3* were co-transformed with ParD3 into *E. coli*, and cells were plated on medium that induces or represses expression of the toxin and antitoxin. **(B)** Amino acid composition of library variants with high fitness ($W > 0.5$) against ParE3*. **(C)** ParE3*-ParD3* are insulated from the wild-type ParD3-ParE3 pair. **(D)** Toxicity-rescue interaction assays for all ParD3 and ParE3 mutant combinations. Top left, wild-type ParD3-ParE3 pair; bottom right, orthogonal ParD3*-ParE3* pair.

Figure 3.12B). Sequence analysis of the high-fitness mutants ($W > 0.5$) revealed large differences in amino-acid preferences at positions 60 and 61 relative to those shown above to be reactive against ParE3 (Figures 3.5F, 3.11B). In particular, for the ParD3 variants that neutralize ParE3*, the invariant Trp at position 60 is replaced by Ile, Val, or Leu; and the strong preference for a negatively charged residue at position 61 is replaced by positively-charged or neutral residues (Figure 3.11B). One of the high-fitness variants identified had specificity residues LIAK, which we renamed ParD3*. We found that ParD3* no longer neutralized ParE3, but robustly interacted with ParE3* (Figure 3.11C). Taken together, our results indicate that mutations in the specificity residues of ParD3 and ParE3 are sufficient to create an orthogonal, interacting pair.

Our previous results indicated that mutational paths that change ParD specificity tend to pass through promiscuous intermediates (Figure 3.9). Given that we generated an orthogonal ParD3*-ParE3* pair, we wanted to determine whether mutational paths between the wild-type and orthogonal system also can pass through promiscuous intermediates. We therefore generated variants of ParE3 containing all possible subsets of the substitutions in ParE3* (32 mutants) and variants of ParD3 containing all possible subsets of the substitutions in ParD3* (4 mutants). We then co-transformed each possible pairing of ParD3 and ParE3 variants (128 pairs total) into *E. coli* and measured interaction strength using the toxicity-rescue assay (Figure 3.11D). Interestingly, we found that most (90 of 128) ParD3-ParE3 pairings were functional. We also found that most (17 out of 32 mutants) intermediate ParE3 states were promiscuous, which we defined as capable of interacting strongly with both ParD3 and ParD3* (Figure 3.11D).

To analyze mutational paths between the wild-type and insulated pair, we first enumerated the total number of trajectories between these systems. Assuming that one residue is mutated per

step and no reversions are considered, there are a total of 5,040 paths from ParD3-ParE3 to the orthogonal ParD3*-ParE3* pair. Only 1,030 (or 20%) of these paths retain functionality at each step. To determine whether these functional paths tend to involve promiscuous intermediates, we tallied the number of promiscuous intermediates that appear in each of these 1,030 paths. On average, we found that functional paths contain 4.8 promiscuous ParE3 variants, with all paths involving at least one promiscuous intermediate (Figure 3.12C). These promiscuous states may be necessary for the ParD-ParE system to evolve a new interaction specificity, otherwise the mutational path would require a intermediate state in which a substitution in one component (either ParD or ParE) yields a "switch-like" change in specificity that breaks the interaction until a second substitution restores the interaction. The prevalence of mutational paths involving promiscuous intermediates may help to explain the apparent ease with which ParD-ParE systems, and perhaps other protein-protein interaction systems, have expanded during the course of evolution.

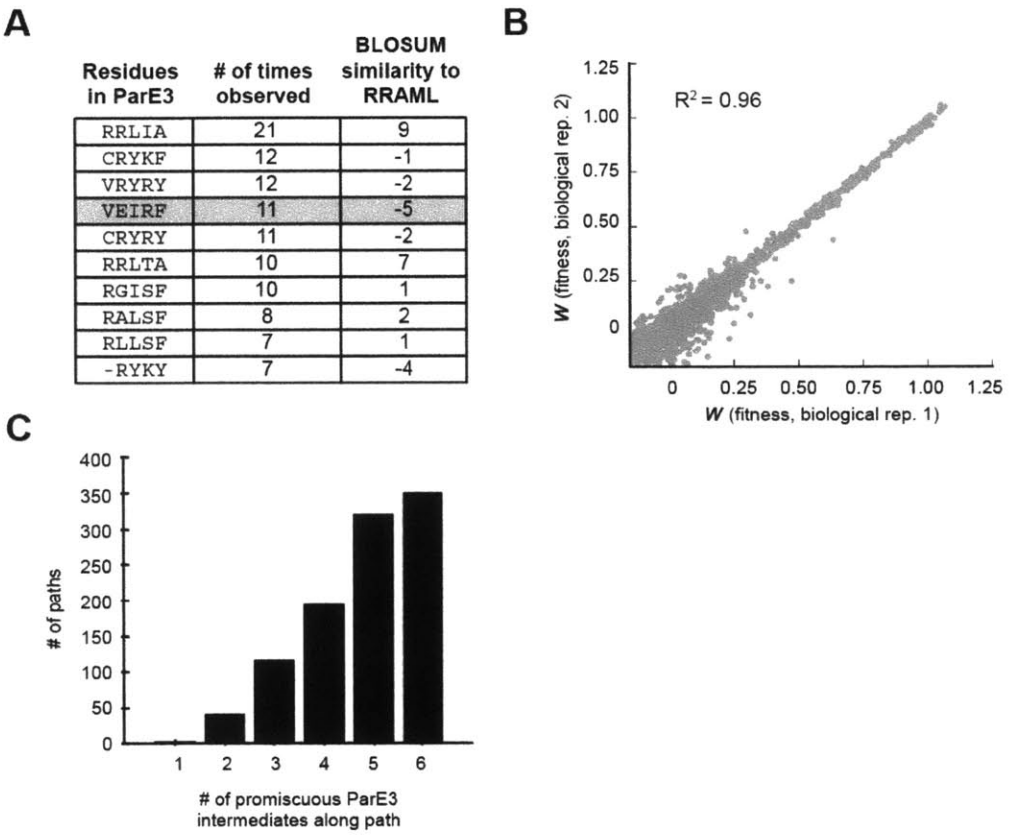


Figure 3.12. Generation and testing of ParE3* variant against ParD3 library; mutational paths between ParD3-ParE3 and ParD3*-ParE3*

(A) An alignment of ParE homologs was analyzed for the most commonly occurring residues at positions 54, 58, 61, 63, and 72 in ParE3 (middle column). The residue combinations were then scored based on their chemical similarity to the residues in ParE3, RRAML, using BLOSUM62 (right column). We chose to incorporate the residue combination VEIRF given that it was the most distant from RRAML by BLOSUM62. **(B)** Fitness measurements of the ParD3 library against the ParE3* toxin are reproducible between biological replicates. **(C)** Most mutational paths between ParD3-ParE3 and ParD3*-ParE3* pass through multiple promiscuous ParE3 intermediates. Promiscuous ParE3 intermediates are defined as those capable of interacting with both ParD3 and ParD3*. Only functional paths are scored.

Discussion

Specificity of protein-protein interactions in the ParD-ParE family

The evidence for specificity in toxin-antitoxin systems has previously been mixed, with some reports indicating that these protein-protein interactions are specific (Fiebig et al., 2010) and others suggesting that TA systems form large, promiscuous interaction networks (Yang et al., 2010; Zhu et al., 2010). In this work, we perform the first systematic assessment of interaction specificity in a TA family and find that antitoxins exhibit an exquisite preference for binding to their co-transcribed, cognate toxins. For the 180 non-cognate pairings that we tested in the ParD-ParE family, we found cross-talk between only 11 of them (Figure 3.1). Importantly, no cross-talk was observed for non-cognate pairs present in the same species. This high degree of interaction specificity is remarkable given that the average pairwise similarity between the tested ParD antitoxins is 43%.

How is this specificity achieved? Similar to other paralogous protein families (Capra et al., 2010; Skerker et al., 2008), this specificity appears to be encoded by a small subset of coevolving residues at the protein-protein interface (Figure 3.3). We found that as few as four mutations in these coevolving residues were sufficient to reprogram the antitoxin ParD3 to interact with the non-cognate toxin ParE1 or ParE2 and lose its ability to interact with ParE3 (Figure 3.3D). Interestingly, fewer mutations are required to make ParD3 binding promiscuous, as the single K64L mutation allows ParD3 to antagonize both the ParE2 and ParE3 toxins (Figure 3.7D). The smaller number of mutations required to make ParD3 promiscuous is consistent with our understanding of how specificity evolves. Specificity requires not only mutations that enable binding to a desired partner but also mutations that block binding to non-cognate proteins

(Schreiber and Keating, 2011). Insulation from undesired binding partners may be achieved by a single frustrated contact between two surfaces, as is the case for the non-cognate interaction between ColE9 and Im2 (Meenan et al., 2010).

The level of specificity observed for the ParD-ParE family is similar to that observed for other large, paralogous protein families. For example, sensor kinases from two-component systems exhibit a kinetic preference for their cognate response regulator over all possible non-cognate partners from the same species (Skerker et al., 2005). Additionally, a comprehensive analysis of interactions between 157 mouse PDZ domains and 217 genome-encoded peptides revealed interactions for only 4% of the tested pairs (Stiffler et al., 2007). A similar study that looked at interaction specificity for 59 coiled-coil family bZIP transcription factors detected binding for 14% of the pairs (Newman and Keating, 2003). The low rate of crosstalk for many of these paralogous families is attributed to selection for specificity (Capra et al., 2012; Zarrinpar et al., 2003), which raises the possibility that the ParD-ParE family may be under similar selective pressures.

An alternative explanation for the observed specificity of ParD-ParE interactions is neutral drift. However, we do not currently favor the neutral hypothesis for two reasons. First, purifying selection appears to be necessary in other paralogous protein families to prevent crosstalk. An example is the two-component system PhoR-PhoB, which underwent selection in the α-proteobacterial lineage to avoid crosstalk with a paralogous system, NtrY-NtrX (Capra et al., 2012). Second, we never observed crosstalk for ParD-ParE pairs present in the same genome, which is consistent with selection acting to promote specificity within species (Figure 3.1). Interaction specificity within species has been observed for other systems, such as SH3-peptide

interactions in yeast (Zarrinpar et al., 2003). Interestingly, this specificity is lost when presenting yeast peptides to SH3 domains from metazoans – an analogous situation to the loss of specificity for ParD-ParE pairs between species. The biological rationale for specificity in TA systems is unclear, and will require a deeper understanding of the function of these systems in bacterial physiology.

High-throughput mapping of mutant fitness at the ParD-ParE interface

Libraries have been used previously to interrogate the relationship between protein sequence and function with great success. These studies have found that the mutational tolerance of a protein depends strongly on the queried position (Melamed et al., 2013) and also demonstrated that epistasis severely constrains paths between functional variants (Podgornaia and Laub, 2015). Libraries have also helped illuminate how mutations affect ligand binding specificity. For example, single point mutants in a PDZ domain were identified that restored binding to a non-canonical peptide ligand (McLaughlin Jr et al., 2012). However, in general, these studies have been limited to looking at how mutations alter the function of a single protein (Hietpas et al., 2011), or how mutations affect binding between cognate proteins (Fowler et al., 2010; Podgornaia and Laub, 2015).

Our work expands on this foundation by querying a library of ParD3 variants against two separate proteins: the cognate toxin ParE3 and the non-cognate toxin ParE2. This approach yielded a high-density mapping of sequence space against an orthogonal pair of proteins (Figure 3.9A). From these data, we arrived at three main insights. First, we were able to uncover the residues in ParD3 responsible for the selective antagonism of one ParE toxin over another (Figures 3.5F, 3.7C). We found that this selectivity was due primarily to charge switching and the

exchange of charged for small hydrophobic residues at positions 61 and 64 in ParD3. Second, these data demonstrated the prevalence of specificity residue combinations that promote a promiscuous state of ParD3. We found 252 variants of ParD3 that can antagonize ParE3, and of these, 12% also neutralized the non-cognate toxin ParE2 (Figure 3.9B). These results suggest that ParD antitoxins possess a latent ability to interact with multiple ParE toxins, perhaps by relying on a core set of conserved, interacting residues. This situation is analogous to the binding between colicin and immunity proteins, which is mediated by a common anchoring point (conserved between all proteins) and a second variable region that confers affinity and selectivity for its cognate protein (Levin et al., 2009).

Third, and perhaps most interestingly, our data allowed us to map mutational trajectories that reprogram the specificity of ParD3. We found that these paths were strongly shaped by the high connectivity of promiscuous variants in sequence space (Figure 3.9C-D). In particular, we found that paths leading to a change in specificity are highly enriched for promiscuous intermediates (Figure 3.9F-G). This enrichment of promiscuous intermediates was more pronounced for graphs of functional variants connected by single nucleotide substitutions, suggesting that the genetic code may be biased toward mutations that permit promiscuity. This phenomenon was even more pronounced when considering mutations on both side of the interface. Mapping of the mutational trajectories between the wild-type ParD3-ParE3 and orthogonal ParD3*-ParE3* pair revealed that all functional paths crossed through at least one promiscuous intermediate (Figure 3.11).

How does the presence of so many promiscuous intermediates shape our view of protein evolution? In one model, these promiscuous intermediates provide evolution with functional

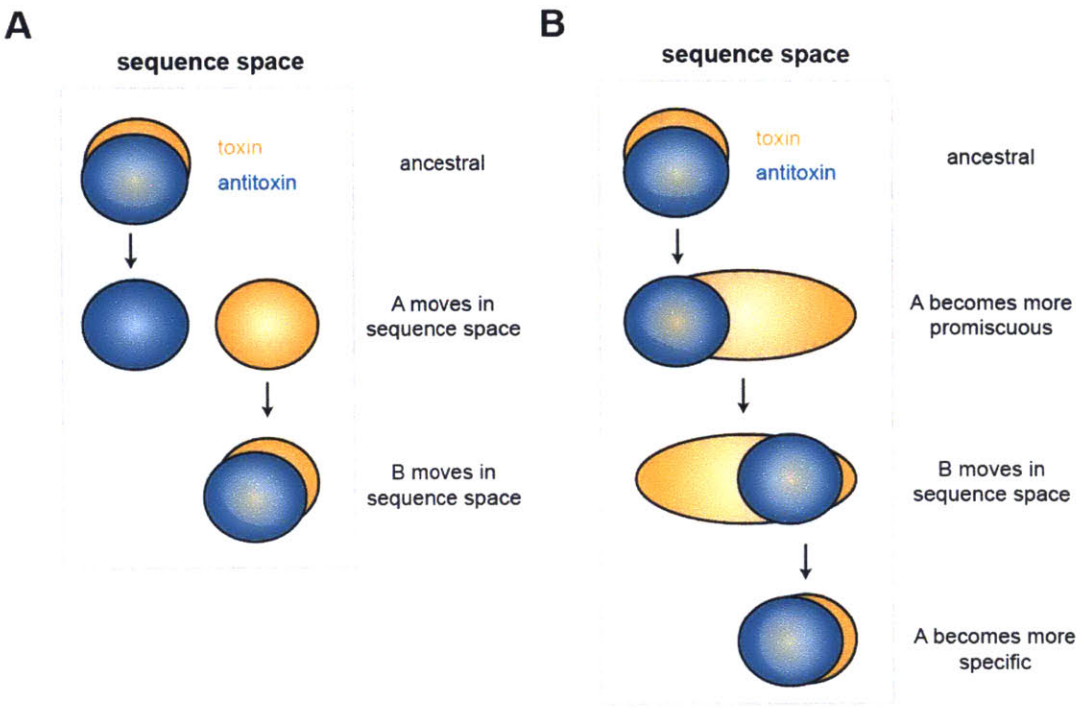


Figure 3.13. Model for divergence of proteins through promiscuous intermediates.

Cognate pairs of toxins and antitoxin normally overlap in specificity. A movement in sequence space away from the antitoxin would normally be lethal (A). However, if the toxin first becomes more promiscuous, then the antitoxin is free to move in sequence space within a region defined by the toxin (B). A subsequent narrowing of specificity is then sufficient to create a pair that is orthogonal to its ancestral state.

trajectories to states that would be otherwise inaccessible due to non-functional intermediates. For example, if we consider the sequence space defined by the specificity-determining residues of the toxin and antitoxin, cognate pairs are normally overlapping (Figure 3.13A). If the toxin moves in sequence space away from its cognate antitoxin, this intermediate state is lethal to the cell. However, an alternate possibility is that the toxin first travels through a promiscuous, intermediate state (Figure 3.13B). The antitoxin is then free to move in sequence space within the region dictated by the toxin. A final narrowing of specificity for the toxin then results in a new,

orthogonal protein pair. The prevalence of promiscuous intermediates may thus facilitate the expansion of paralogous protein families by increasing the number of mutational paths to new, insulated states. Interestingly, enzyme evolution is also thought to involve intermediate states with broadened substrate specificity (Aharoni et al., 2005; Matsumura and Ellington, 2001). As such, passing through promiscuous intermediate states may be a broadly conserved mechanism of divergence in evolution.

Experimental Procedures

Bacterial strains and media

Escherichia coli strains were grown in M9L medium (M9 minimal medium supplemented with 5% LB (v/v) and 0.4% glycerol) at 37°C, unless otherwise indicated. To induce expression from the P_{BAD} and P_{lac} promoters, media was supplemented with 0.2% arabinose or 100 μM IPTG, respectively. All toxins were cloned into the SacI and HindIII sites of the arabinose-inducible pBAD33 vector, and all antitoxins were cloned into the SacI and HindIII sites of the IPTG-inducible pEXT20 vector.

Toxicity rescue assay

To test whether a toxin and antitoxin interact, we co-transformed toxin and antitoxin plasmids into *E. coli* TOP10 cells (Life Technologies) and plated on LB medium with 0.4% glucose and appropriate antibiotics. Single colonies were then grown to saturation overnight in M9L medium with 0.4% glucose and antibiotics. The following morning, cultures were serially diluted and spotted onto M9L plates supplemented with antibiotics and 0.4% glucose, 0.2% arabinose, or 0.2% arabinose and 100 μM IPTG. Plates were then incubated at 37°C for 24 hours. Positive interactions yielded single colonies on M9L with 0.2% arabinose and 100 μM IPTG after 24 hours of growth. Intermediate interactions yielded modest growth on plates but no visible single colonies. No intermediate growth phenotypes were observed for the 20x20 matrix (Figure 3.1).

Identification of coevolving residues

Coevolving residues in the ParDE family were identified using GREMLIN (Kamisetty et al., 2013; Ovchinnikov et al., 2014). We used the GREMLIN webserver (gremlin.bakerlab.org) which is maintained by the David Baker lab at the University of Washington. Input sequences were ParD3 and ParE3 from *Mesorhizobium opportunistum*, and we set the number of iterations to four and the E-value cutoff to 1E-04. To identify specificity residues, we isolated all residue pairings that had a scaled coupling score greater than 1.25. To identify supporting residues, we performed the following iterative procedure using a score cutoff of 1.25: (1) identify residues within ParD or ParE that covary with the specificity residues; (2) identify residues within ParD or ParE that covary with either the specificity residues or the supporting residues identified in step (1); (3) repeat step (2) until no new supporting residues are identified.

Creation of the orthogonal ParE3* toxin

To create a ParE3 toxin with a novel specificity profile, we focused on residues in ParE3 that covary with W60/D61 from ParD3. We found that residues R58/A61/L72 in ParE3 covary with W60/D61 from ParD3 with a GREMLIN scaled score greater than 1. We then searched for residues within ParE3 that covary with R58/A61/L72 (termed "supporting specificity residues") with a GREMLIN scaled score greater than 1. Repeating this search process iteratively produced two more supporting specificity residues in ParE3, M63/R54, for a total of five specificity and supporting specificity residues in ParE3: R54/R58/A61/M63/L72.

To identify which mutations to make in these five residues, we searched naturally existing ParE sequences for combinations of residues that often occur at these positions. Clustering analysis revealed groups of residues that often covary.

ParDE3 expression and purification

Recombinant *Mesorhizobium opportunistum* ParDE3 protein complex was expressed in *E. coli* Rosetta(DE3)pLysS (Novagen). A 50 ml overnight culture in LB medium supplemented with 50 μg/ml kanamycin (LB-Kan_{50}) was used to inoculate 2 liters of LB-Kan_{50}; this culture was incubated at 37°C in a rotary shaker at 220 rpm. Transcription of recombinant *parDE3* was induced at an OD_{660} of 0.8 by adding 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After 4 h of induction, the cells were harvested by centrifugation at 12,000g for 20 min at 4°C. Cell pellets were resuspended in 30 ml of lysing/binding buffer (10 mM Tris (pH 7.4), 150 mM NaCl, 10 mM imidazole with 5 μg/ml of DNase I (Sigma-Aldrich) and half a tablet of cOmplet EDTA free protease inhibitor cocktail (Roche Life Science).

Cells were disrupted by one passage through an LV1 microfluidizer (Microfluidics, Westwood, MA) and the cell debris was removed by centrifugation for 20 min at 25,000g. The supernatant was loaded onto a Ni^{2+} Sepharose affinity column (GE Life Sciences) pre-equilibrated with the binding buffer. Two washing steps were performed using 10 mM and 75 mM of imidazole followed by two elution steps with 200 mM and 1 M imidazole in the binding buffer. After purity of the different fractions was assessed by SDS-PAGE, the protein solution was dialyzed against 10 mM Tris (pH 7.4), 150 mM NaCl, 200 mM imidazole buffer

Crystallization of ParDE3

Purified ParDE3 was purified and concentrated using a centrifugal filter (3 kDa MWCO, Amicon-Millipore). Protein purity was estimated to be 95% as assessed by 14% SDS-PAGE stained with Coomassie brilliant blue. Initial crystallization screening was carried out using the

sitting-drop, vapor-diffusion technique in 96-well microplates (Nunc). Trays were set up using a Mosquito robot (TTP LabTech) and commercial crystallization kits (Nextal-Qiagen). The drops were set up by mixing equal volumes (0.1 μl) of the protein and the precipitant solutions equilibrated against 75 μl of the precipitant solution. In all trials, the protein concentration was ~ 40 mg/ml. In approximately five days, needle-like crystals appeared in condition 15 of the Pro-complex Suite crystallization kit (Qiagen). After manual refinement of the crystallization condition, the best crystals were obtained at 19°C with the following crystallization solution: 400 mM Sodium Acetate, 100 mM Sodium Citrate pH5.5, 20% PEG 4000, 20 % glycerol. All manual crystallization attempts were carried out using the hanging-drop, vapor-diffusion technique in 24-well plates (Hampton). Prior to flash freezing in liquid nitrogen, drops containing the crystals were mixed with 1 μl of a crystallization solution containing 100 mM sodium iodide and incubated for 4 hours. Crystals were then cryo-protected by soaking them in the crystallization solution containing 25 % glycerol and 100 mM sodium iodide.

Crystallographic data collection and data processing

Crystal diffraction was measured at a temperature of 100 K using a 1 degree oscillation range on beamline 21-ID-D (LS-CAT, Advanced Photon Source, Argonne, Illinois); diffraction images were collected on a MAR Mosaic 300 detector. Diffraction images were processed using the HKL 2000 suite . Geometric refinement and examination of the scaled amplitudes revealed that the ParDE3 crystals belong to orthorhombic space group I222, with cell dimensions a=43.18, b=118.84, c=211.42 ($\alpha=\beta=\gamma=90^{\circ}$).

Diffraction from a single ParDE3 protein crystal was measured to 1.53 Å at an energy of 12.66 keV (0.979 Å). The anomalous signal in the data was used to locate iodide atoms in the lattice,

and the structure was phased by single wavelength anomalous dispersion using the Autosol SAD routine in Phenix. Two ParDE3 complexes are present in the asymmetric unit. Six iodine sites were located within the asymmetric unit. A preliminary ParDE3 structural model was built *de novo* from the initial experimental, solvent-flattened maps using the AutoBuild routine and phenix.refine. This initial model was then manually examined and corrected; solvent addition and refinement of the structure was conducted iteratively using Coot and phenix.refine. The final structural model was refined to an R_{work} of 16.6 % and R_{free} of 18.8 %. Coordinates of ParDE3 will e deposited in the Protein Data Bank.

Size exclusion chromatography

A purified sample of ParDE3 (10 mg/ml-300 µl) was injected on a GE Healthcare Superdex 200 10/300 GL column (flow rate 0.5 ml/min) and fractions of 500 µl were collected. 10 mM Tris pH 7.4, 150 mM NaCl, 200 mM imidazole was used as a running buffer. Collected fractions were resolved on 14% SDS-PAGE gels and compared to the elution profile. To estimate the molecular weight and, hence, oligomeric state of the ParDE3 complex in solution, its elution volume was compared to molecular weight standards (blue dextan, aldolase, conalbumin and ovalbumin) resolved on the same column using the same buffer and flow protocol.

Author Contributions

Crystallization experiments were performed by JH. Protein chimeras for Figure 3.3 were generated by TNP. Specificity assays for Figure 3.11 performed by BP. All other experiments were conceived and performed by CDA. CDA and MTL analyzed the data and wrote the paper.

Acknowledgements

We thank members of the Laub laboratory for helpful discussions and comments on the manuscript. We would also like to acknowledge Sergey Ovchinnikov and Christopher Bahl for their valuable discussions on GREMLIN.

References

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C., and Tawfik, D.S. (2005). The "evolvability" of promiscuous protein functions. Nat. Genet. *37*, 73–76.

Arbing, M.A., Handelman, S.K., Kuzin, A.P., Verdon, G., Wang, C., Su, M., Rothenbacher, F.P., Abashidze, M., Liu, M., Hurley, J.M., et al. (2010). Crystal structures of Phd-Doc, HigA, and YeeU establish multiple evolutionary links between microbial growth-regulating toxin-antitoxin systems. Struct. Lond. Engl. 1993 *18*, 996–1010.

Ashenberg, O., Rozen-Gagnon, K., Laub, M.T., and Keating, A.E. (2011). Determinants of homodimerization specificity in histidine kinases. J. Mol. Biol. *413*, 222–235.

Capra, E.J., Perchuk, B.S., Lubin, E.A., Ashenberg, O., Skerker, J.M., and Laub, M.T. (2010). Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. PLoS Genet. *6*, e1001220.

Capra, E.J., Perchuk, B.S., Skerker, J.M., and Laub, M.T. (2012). Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families. Cell *150*, 222–232.

Chen, T.S., and Keating, A.E. (2012). Designing specific protein-protein interactions using computation, experimental library screening, or integrated methods. Protein Sci. Publ. Protein Soc. *21*, 949–963.

Fiebig, A., Castro Rojas, C.M., Siegal-Gaskins, D., and Crosson, S. (2010). Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin-antitoxin systems. Mol. Microbiol. *77*, 236–251.

Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., and Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. Nat. Methods *7*, 741–746.

Hallez, R., Geeraerts, D., Sterckx, Y., Mine, N., Loris, R., and Van Melderen, L. (2010). New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in Escherichia coli O157:H7. Mol. Microbiol. *76*, 719–732.

Helaine, S., Cheverton, A.M., Watson, K.G., Faure, L.M., Matthews, S.A., and Holden, D.W. (2014). Internalization of Salmonella by macrophages induces formation of nonreplicating persisters. Science *343*, 204–208.

Hietpas, R.T., Jensen, J.D., and Bolon, D.N.A. (2011). Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. *108*, 7896–7901.

Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013). Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc. Natl. Acad. Sci. U. S. A. *110*, 15674–15679.

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

Levin, K.B., Dym, O., Albeck, S., Magdassi, S., Keeble, A.H., Kleanthous, C., and Tawfik, D.S. (2009). Following evolutionary paths to protein-protein interactions with high affinity and selectivity. Nat. Struct. Mol. Biol. *16*, 1049–1055.

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

Mao, W., Kaya, C., Dutta, A., Horovitz, A., and Bahar, I. (2015). Comparative study of the effectiveness and limitations of current methods for detecting sequence coevolution. Bioinforma. Oxf. Engl.

Matsumura, I., and Ellington, A.D. (2001). In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. J. Mol. Biol. *305*, 331–339.

McLaughlin Jr, R.N., Poelwijk, F.J., Raman, A., Gosal, W.S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. Nature *491*, 138–142.

Meenan, N.A.G., Sharma, A., Fleishman, S.J., MacDonald, C.J., Morel, B., Boetzel, R., Moore, G.R., Baker, D., and Kleanthous, C. (2010). The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. Proc. Natl. Acad. Sci. *107*, 10080–10085.

Melamed, D., Young, D.L., Gamble, C.E., Miller, C.R., and Fields, S. (2013). Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA *19*, 1537–1551.

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

Newman, J.R.S., and Keating, A.E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science *300*, 2097–2101.

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

Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife *3*, e02030.

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

Podgornaia, A.I., and Laub, M.T. (2015). Protein evolution. Pervasive degeneracy and epistasis in a protein-protein interface. Science *347*, 673–677.

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

Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U., and Sorek, R. (2013). Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. Mol. Cell *50*, 136–148.

Schreiber, G., and Keating, A.E. (2011). Protein binding specificity versus promiscuity. Curr. Opin. Struct. Biol. *21*, 50–61.

Skerker, J.M., Prasol, M.S., Perchuk, B.S., Biondi, E.G., and Laub, M.T. (2005). Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis. PLoS Biol *3*, e334.

Skerker, J.M., Perchuk, B.S., Siryaporn, A., Lubin, E.A., Ashenberg, O., Goulian, M., and Laub, M.T. (2008). Rewiring the specificity of two-component signal transduction systems. Cell *133*, 1043–1054.

Stiffler, M.A., Chen, J.R., Grantcharova, V.P., Lei, Y., Fuchs, D., Allen, J.E., Zaslavskaia, L.A., and MacBeath, G. (2007). PDZ domain binding selectivity is optimized across the mouse proteome. Science *317*, 364–369.

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H., and Moll, I. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli. Cell *147*, 147–157.

Yang, M., Gao, C., Wang, Y., Zhang, H., and He, Z.-G. (2010). Characterization of the interaction and cross-regulation of three Mycobacterium tuberculosis RelBE modules. PloS One *5*, e10672.

Zarrinpar, A., Park, S.-H., and Lim, W.A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. Nature *426*, 676–680.

Zhu, L., Sharp, J.D., Kobayashi, H., Woychik, N.A., and Inouye, M. (2010). Noncognate Mycobacterium tuberculosis toxin-antitoxins can physically and functionally interact. J. Biol. Chem. *285*, 39732–39738.

Chapter 4

Conclusions and Future Directions

These data are unpublished.

Conclusions

My graduate work has focused on understanding the mechanisms, targets, and specificity of toxin-antitoxin systems from bacteria. These systems are widely present on bacterial chromosomes, yet our understanding of their biology is incomplete. How do antitoxins neutralize their toxins? How do toxins inhibit growth of the cell, and how are these mechanisms analogous to the action of antibiotics? Do toxins and antitoxins only interact with their cognate pairs? If so, how is this specificity encoded? What evolutionary paths are available to toxins and antitoxins for divergence in sequence space following duplication? These questions are broad, and my thesis work touches on each of them in turn.

To understand better the mechanisms and targets of TA systems, I first characterized the SocAB system from *Caulobacter crescentus* (Chapter 2). I found that SocAB deviates from the canonical TA system in two interesting ways. First, the typical stability logic is switched – instead of the antitoxin being unstable, the toxin SocB is unstable and constitutively degraded by the protease ClpXP. The instability of SocB is promoted by its antitoxin, SocA, which acts as a proteolytic adaptor. This mechanism of action does not seem to fit within the five current mechanistic TA types (Chapter 1.III), leading other groups to propose a new class of "SocAB-like" TA systems (Markovski and Wickner, 2013). Second, the toxin SocB inhibits replication, but not through poisoning gyrase such as the ParE and CcdB toxins. Rather, SocB blocks replication through an interaction with the sliding clamp, making SocB the first toxin to inhibit replication by targeting a component of the replisome. Given the overlap between the targets of toxins and antibiotics, I suggested that the sliding clamp may make an attractive target for the development of antimicrobials.

Following my work on SocAB, I became interested in whether TA systems interact exclusively with their cognate partners (Chapter 3). This question was motivated by the observation that a single species can encode more than 40 copies of the same paralogous TA family (Ramage et al., 2009). Using the ParD-ParE system as a model, this work had two key takeaways. First, I found that that toxins and antitoxins display a remarkable capacity to distinguish between cognate and non-cognate partners. This specificity appears to be encoded by a small set of coevolving residues at the toxin-antitoxin interface, as mutations in these residues are sufficient to reprogram ParD to interact with non-cognate toxins. Second, I systematically mapped the sequence space of ~10^4 ParD variants in these specificity residues, and found that promiscuous variants, or those that are capable of binding to multiple toxins, are readily obtained. These promiscuous mutants are densely connected to specific variants in sequence space, and as such, mutational paths that cross specificity thresholds tend to cross through these promiscuous intermediates. These results suggest that promiscuous states may facilitate changes in TA specificity and promote the expansion of these systems by duplication and divergence.

These results thus build on our current understanding of targets, mechanisms, and specificity in toxin-antitoxin systems, but they also raise additional questions. For example, what is the physiological function of the SocA-SocB systems? We have hints about the function of particular TA systems (Fineran et al., 2009; Helaine et al., 2014; Maisonneuve et al., 2011; Vesper et al., 2011), but a comprehensive model for TA system function is still lacking. Furthermore, how does evolutionary information and the mapping of so many mutants at a protein-protein interface inform our understanding of protein design? In the following sections, I discuss these questions in more detail, with particular focus on avenues for future research.

Future Directions

Physiological function of the SocA-SocB System

My work on the SocA-SocB system in Chapter 2 revealed a novel antitoxin mechanism and toxin target, yet left unresolved a central question: what is the physiological function of the SocA-SocB system in *Caulobacter crescentus*? My current hypothesis is that SocA-SocB acts as an abortive infection system (Samson et al., 2013). In this model, SocA-SocB acts as a monitoring system for the continued health and activity of the ClpXP protease. If ClpXP is inhibited during phage infection, then SocB would accumulate and kill the cell, thus restricting the spread of phage in the population. There are two lines of evidence in support of this hypothesis. First, phages are known to encode inhibitors of cellular proteases. For example, the phage proteins PinA and Gp4.5 are inhibitors of Lon protease (Sberro et al., 2013; Skorupski et al., 1988), and the λ phage protein RexB inhibits the ClpP protease (Engelberg-Kulka et al., 1998). Thus, monitoring ClpXP activity may be sensitive way to detect phage infection. Second, the SocA antitoxin bears homology to phage-associated proteins.[5] This association suggests that SocA-homologous proteins have been co-opted by phages in other evolutionary contexts, perhaps to direct the degradation of particular phage proteins to host-encoded proteases. Thus, the use of SocA homologs to monitor the activity of host proteases is a possibility.

One way to test this hypothesis is to screen libraries of phage against both wild-type and Δ*socAB* *Caulobacter* cells. The expectation is that if SocA-SocB functions as an abortive infection system, then there will be phage isolates that can grow on Δ*socAB* but not wild-type cells. This approach may require the identification of new phage isolates, as most *Caulobacter* phages have been

[5] NCBI BLASTP search on non-redundant protein database

isolated on the wild-type NA1000 strain (Gill et al., 2012). However, given that *Caulobacter* can be found in most lakes and stream, the identification of new phages is straightforward. Phages capable of infecting *Caulobacter* can be isolated from most surface water samples in a matter of days (Gill et al., 2012). Thus, this project is quite feasible and has the potential to broaden our understanding of how TA systems and phages interact during evolution.

Physiological function of other TA systems

As reviewed in Chapter 1.V, toxin-antitoxin systems are hypothesized to function in a number of different cellular processes, including abortive infection, persister formation, stress responses, and virulence (Fineran et al., 2009; Helaine et al., 2014; Maisonneuve et al., 2011; Vesper et al., 2011). However, the number of chromosomal TA systems with an ascribed function are few and far between, and the majority of TA systems have no phenotype upon deletion. There are thus a number of unanswered questions. When are these TA systems activated? How does their activation alter cellular physiology? And most importantly, how does their deletion impact bacterial fitness?

There is general agreement that the inhibition of translation results in the activation of many TA systems in the cell (Christensen and Gerdes, 2003; Christensen et al., 2001; Christensen-Dalsgaard et al., 2010). This activation is based on the observation that antitoxins are unstable, and as such, blocking translation results in the rapid clearance of antitoxins and the activation of their cognate toxins. However, we are currently lacking a comprehensive understanding of which TA systems are activated during different stresses. Furthermore, we are also lacking mutants that are deleted for all TA systems activated during a particular stress. These multiple deletion

mutants are required to control for the fact that TA systems may be redundantly activated (Christensen-Dalsgaard et al., 2010; Maisonneuve et al., 2011).

These gaps in our current understanding can be addressed by a series of straightforward experiments. To determine the entire panel of TA systems that are activated during translational stress, I recommend performing RNA-Seq on cells exposed to translational inhibitors such as chloramphenicol or serine hydroxamate. Additional stresses, such as nutrient starvation, temperature shock, or exposure to heavy metals can also be tested. Given that most TA complexes transcriptionally repress their own promoter, we can then use transcriptional activation of TA systems as an accurate readout for activation (Li et al., 2008). This experiment will provide the first comprehensive map of TA activation across the genome in response to different stresses. To determine how the activation of these TA systems is affecting bacterial fitness, I recommend generating mutants deleted for the TA systems activated by a particular stress. These multiple deletions can be made using traditional techniques (Murphy and Campellone, 2003) or through next-generation genome modification methods such as MAGE (Wang et al., 2009). The response of these TA deletion mutants to particular stresses can then be compared to the isogenic wild-type strain. One prediction is that the TA deletion mutant may exhibit heightened sensitivity to the stress, or may recover more slowly after the stress has been removed.

Using evolutionary information to predict mutant fitness

My work on the ParD-ParE system revealed that protein libraries can be used to measure the fitness of ~10^4 different ParD3 mutants simultaneously (Chapter 3). However, an unresolved question is to what extent we can use evolutionary information – such as the statistical couplings

between positions in an alignment of ParD-ParE orthologs – to predict the fitness of these mutants *in silico*. For example, we used GREMLIN to measure coevolution in the ParD-ParE family, and GREMLIN explicitly models statistical coupling between different residues combinations at each pair of positions in the alignment (Kamisetty et al., 2013). The GREMLIN coupling score can be split into residue-residue couplings that occur within proteins ("intra GREMLIN coupling") and between proteins ("inter GREMLIN coupling"). Intra coupling can be thought of as a proxy for protein stability, and inter coupling can be considered a proxy for strength of binding between the two proteins. By combining the intra and inter coupling scores, the GREMLIN coupling score can thus be used to score the predicted stability and binding strength for different mutants.

To test whether GREMLIN coupling scores can accurately predict the effect of mutations at the ParD-ParE interface, I compared the experimentally measured fitness values for ~10^4 ParD3

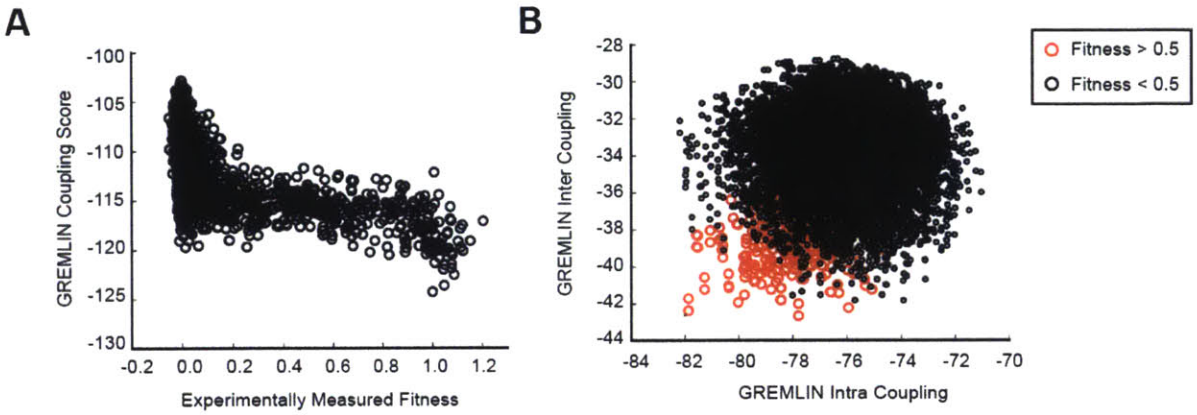


Figure 4.1. Correlation between fitness measurements and GREMLIN coupling score.

(A) Scatter plot of experimentally measured fitness for 9,194 mutants in the ParD3 specificity residues (against ParE3 toxin) versus GREMLIN coupling score. **(B)** Scatter plot of GREMLIN intra and inter coupling score for the mutants from (A). Fit variants, red; unfit variants, black. Note that fit variants cluster in the bottom left portion of the graph.

variants with the GREMLIN coupling score. This analysis revealed a modest correlation between fitness and GREMLIN coupling scores (Figure 4.1A). To determine whether this correlation could be improved be considering intra and inter coupling separately, I generated a scatter plot of GREMLIN intra and inter coupling scores for the ~10^4 ParD3 variants and color-coded by fitness (Figure 4.2B). Interestingly, functional ParD3 variants (W > 0.5) clustered in the bottom left corner of the graph, suggesting that variants need to be both stable (low intra coupling) and strong binders (low inter coupling) in order to be functional.

Another way to score the accuracy of our predictions is to rank all mutants by their GREMLIN coupling score (from best to worst), and then calculate the fraction of high-fitness variants as a function of the number of GREMLIN predictions that are scored. Using this approach, I found that ranking variants by the sum of their inter and intra coupling scores produces the most accurate predictions (Figure 4.2). In particular, this method yields an accuracy of 81% when considering the top 100 scoring variants. This accuracy drops to 62% when considering only the inter coupling score, and 20% when considering only the intra coupling score. Recently, my collaborators have tried to take structural information into account by calculating the Rosetta free energy of binding ($\Delta\Delta G$) for the different ParD3 mutants, using the crystal structure that we solved previously (Chapter 3).[6] Surprisingly, we found that the accuracy of Rosetta was worse than GREMLIN – only 48% of the top 100 scoring Rosetta variants were found to be functional. We are currently working on improving the accuracy of our Rosetta model, either in isolation or in combination with GREMLIN.

[6] This work was done by Sergey Ovchinnikov and Christopher Bahl from the Baker lab at the University of Washington. Rosetta modeling was done with minimal backbone movement.

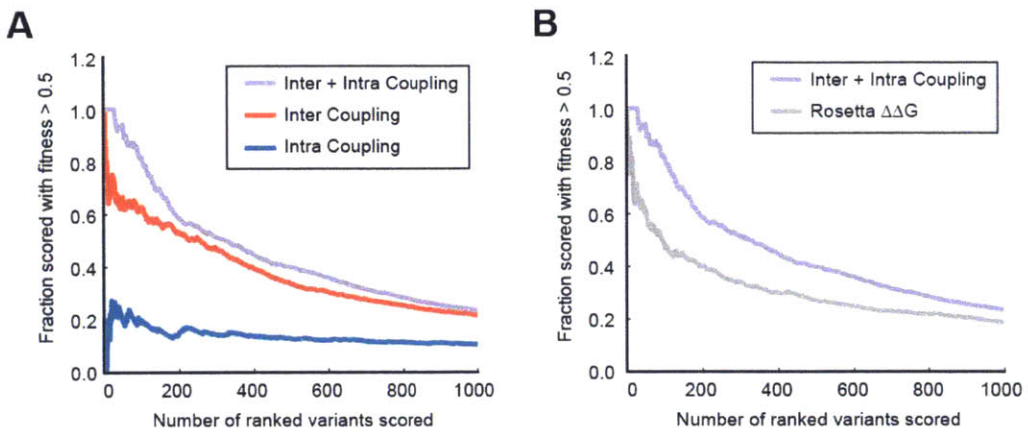


Figure 4.2. Accuracy of GREMLIN and Rosetta-based methods for predicting the fitness of mutants at the ParD-ParE interface.

(A) A combination of GREMLIN inter and intra coupling scores most accurately predicts the fitness of ParD3 mutants. GREMLIN scores for the 9,194 ParD3 mutants were sorted into a ranked list, and the fraction of variants that are functional (W > 0.5) was determined as a function of the number of ranked variants evaluated. As expected, the accuracy of the method decreases as more low-ranking variants are scored. Red, inter coupling scores only; blue, intra coupling scores only; purple, sum of intra and inter coupling scores. **(B)** GREMLIN predicts high-fitness variants more accurately than Rosetta free energy of binding (ΔΔG). Starting with the *M. opportunistum* ParDE3 crystal structure (Chapter 3), we introduced mutations from each of the 9,194 ParD3 variants and then modeled change in ΔΔG using Rosetta. The ΔΔG values were then sorted into a ranked list and scored as in (A). Purple, sum of GREMLIN inter and intra coupling scores; grey, Rosetta ΔΔG.

These results indicate that GREMLIN can be used to predict, with moderate to high accuracy, the effect of mutations at the ParD-ParE interface. It will be interesting to see whether these results can be extended to other protein families for which there is sufficient evolutionary information. For example, GREMLIN has been used to help predict which parts of a protein make contact with each other in a complex (Ovchinnikov et al., 2014). The next step is to see whether GREMLIN can be used to predict mutations that will block particular protein interactions within a complex, with the goal of assessing the effect of these mutations on cellular physiology.

References

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

Christensen-Dalsgaard, M., Jørgensen, M.G., and Gerdes, K. (2010). Three new RelE-homologous mRNA interferases of Escherichia coli differentially induced by environmental stresses. Mol. Microbiol. *75*, 333–348.

Engelberg-Kulka, H., Reches, M., Narasimhan, S., Schoulaker-Schwarz, R., Klemes, Y., Aizenman, E., and Glaser, G. (1998). rexB of bacteriophage lambda is an anti-cell death gene. Proc. Natl. Acad. Sci. U. S. A. *95*, 15481–15486.

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., and Salmond, G.P.C. (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.

Gill, J.J., Berry, J.D., Russell, W.K., Lessor, L., Escobar-Garcia, D.A., Hernandez, D., Kane, A., Keene, J., Maddox, M., Martin, R., et al. (2012). The Caulobacter crescentus phage phiCbK: genomics of a canonical phage. BMC Genomics *13*, 542.

Helaine, S., Cheverton, A.M., Watson, K.G., Faure, L.M., Matthews, S.A., and Holden, D.W. (2014). Internalization of Salmonella by Macrophages Induces Formation of Nonreplicating Persisters. Science *343*, 204–208.

Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013). Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc. Natl. Acad. Sci. U. S. A. *110*, 15674–15679.

Li, G.-Y., Zhang, Y., Inouye, M., and Ikura, M. (2008). Structural mechanism of transcriptional autorepression of the Escherichia coli RelB/RelE antitoxin/toxin module. J. Mol. Biol. *380*, 107–119.

Maisonneuve, E., Shakespeare, L.J., Jørgensen, M.G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. Proc. Natl. Acad. Sci. *108*, 13206–13211.

Markovski, M., and Wickner, S. (2013). Preventing bacterial suicide: a novel toxin-antitoxin strategy. Mol. Cell *52*, 611–612.

Murphy, K.C., and Campellone, K.G. (2003). Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic E. coli. BMC Mol. Biol. *4*, 11.

Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife *3*, e02030.

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

Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. *11*, 675–687.

Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U., and Sorek, R. (2013). Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. Mol. Cell *50*, 136–148.

Skorupski, K., Tomaschewski, J., Rüger, W., and Simon, L.D. (1988). A bacteriophage T4 gene which functions to inhibit Escherichia coli Lon protease. J. Bacteriol. *170*, 3016–3024.

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H., and Moll, I. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli. Cell *147*, 147–157.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894–898.