2016-11-01

Genetic and Biochemical Analysis of the Micrococcin Biosynthetic Pathway

Philip Ross Bennallack
Brigham Young University

Genetic and Biochemical Analysis of the Micrococcin Biosynthetic Pathway

Philip Ross Bennallack

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Richard A. Robison, Co-Chair
Joel S. Griffitts, Co-Chair
Julianne H. Grose
Steven M. Johnson
Bradley C. Bundy

Department of Microbiology and Molecular Biology

Brigham Young University

ABSTRACT

Genetic and Biochemical Analysis of the Micrococcin Biosynthetic Pathway

Philip Ross Bennallack
Department of Microbiology and Molecular Biology, BYU
Doctor of Philosophy

Declining antibiotic discovery and flourishing antibiotic resistance have led to a modern antibiotic crisis which threatens to compromise our ability to treat infectious disease. Consequently, there is significant interest in developing new antibiotics with novel modes of action and chemical properties. Ribosomally synthesized and post-translationally modified peptides (RiPPs) are natural compounds with the appealing attributes of being derived directly from a genetic template while possessing numerous exotic chemical features that contribute to stability and antimicrobial activity. Abundant in nature, their diverse range of biological activities makes them excellent prospects for antibiotic development. Thiopeptides, a RiPP family rich in chemical complexity, represent a particularly promising example. Characterized by post-translationally formed sulfur- and nitrogen-containing heterocycles, more than 100 different thiopeptides have been identified from various cultivable bacterial producers, and the mining of genomic and metagenomic data promises to uncover many more chemical species that have eluded discovery by conventional means. These peptides are potent inhibitors of bacterial protein synthesis and have been shown effective against many drug-resistant pathogens. Despite these attractive properties, therapeutic applications have been limited by the lack of an efficient synthetic route and poor aqueous solubility. Both of these challenges would be greatly alleviated by a more complete understanding of thiopeptide biosynthesis and improved systems for analysis and engineering.

Here we describe the characterization of a new thiopeptide gene cluster, which encodes the archetypal thiopeptide micrococcin P1. We describe the identification of the bioactive product and detail the mechanism of immunity in the producing strain. We also describe efforts to engineer this pathway for heterologous expression in *Bacillus subtilis*. Using this platform, we have been able to dissect this intricate biosynthetic pathway and parse the order and timing of the processing events involved in peptide maturation. The knowledge gained from these studies will inform future efforts to adapt thiopeptides for therapeutic use, and guide efforts to engineer unnatural compounds using the exotic enzymology employed by thiopeptide producing bacteria.

Keywords: RiPPs, thiopeptides, micrococcin, peptide engineering, *Bacillus subtilis*

ACKNOWLEDGEMENTS

This work would not have been possible without the tremendous support of my two incredible advisors: Joel Griffitts and Richard Robison. Joel has been a phenomenal mentor, example and friend. His curious mind, enthusiasm for conquering the unknown and sense of humor has truly made this experience a joy and a thrill. Richard Robison has also been a fantastic mentor and tremendous source of encouragement. Rich has been one of the biggest champions of my successes, and I have appreciated the intellectual freedoms he has granted me whilst always making himself available for professional and personal advice. I feel truly honored to have been mentored by both of these great men but more importantly, to call them my friends.

I also wish to acknowledge the many others who have contributed to this work. Members of my graduate committee — which also includes Steve Johnson, Julianne Grose and Brad Bundy — have provided invaluable insight and advice and been very influential in my academic development. Our collaborators at UCSF, Susan Miller and Kathryn Bewley have been amazing colleagues. Sue has provided a wealth of scientific insight and has been a true master of the literature. Kathryn has likewise been a great source of ideas and tremendous experimentalist. The assistance of these two remarkable scientists has been invaluable and has accelerated the pace and quality of my research.

Many others in the department have contributed to my academic and personal development. From the multitude of undergraduate students who have assisted on this project to my fellow graduate students for their ideas and friendship, I express appreciation. To Kim O'Neill, Brad Berges, Don Breakwell, and other lunchroom

companions I say “hello stupid!” Thanks for the laughs and friendly hallway greetings. These interactions have made this experience truly enjoyable and have made me feel at home in this department.

I am eternally indebted to my parents for their unconditional support. My dear parents, Paul and Andrea Bennallack, have been my biggest advocates and have always supported all of my goals and decisions, even when they take me across the world. Ken and Carrie Perucca have treated me like their own son and have likewise supported our family in all aspects of life. With such wonderful people behind me, it has often felt impossible to fail.

Finally, to my sweet girls Kaylee, Jayne and Elle, I dedicate this work. Kaylee has provided endless love, encouragement and (when needed) assistance at the bench. Never shy to embark on a new adventure, her faith in me and her optimism has provided clarity and motivation during good times and challenging times. Jayne and Elle, always so excited to visit Daddy at work, have helped me realize how blessed I am to work in such a friendly and intellectually stimulating environment. Knowing I have such a wonderful family to support is the best motivation to succeed. I will always cherish this incredible time in our lives.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

RiPP – ribosomally synthesized and post-translationally modified peptide

NRPS – nonribosomal peptide synthetases

Ser – serine

Thr – threonine

Cys – cysteine

LAPs – linear azol(in)e-containing peptides

MP1 – micrococcin P1

MP2 – micrococcin P2

tcl – thiocillin gene

Bs – *Bacillus subtilis*

Mc – *Macrococcus caseolyticus*

Bc – *Bacillus cereus*

wt – wild type

ORF – open reading frame

His_6 – hexa-histidine

GST – glutathione S-transferase

Ap – ampicillin

Cm – chloramphenicol

Sp - spectinomycin

Km – kanamycin

ATP - adenosine triphosphate

FMN – flavin mononucleotide

PCR – polymerase chain reaction

Chapter 1. Introduction

1.1 The need for new antibiotics

The antibiotic resistance crisis. The rise and spread of antibiotic resistance poses a significant threat to human health. This phenomenon, which occurs when the microbes we treat with antibiotics become insensitive to their effects, occurs naturally in response to exposure and selection, yet the widespread misuse and overuse of antibiotics greatly contributes to its emergence and spread. These bacteria can be particular problematic in a clinical setting, often resulting in delayed or even failed treatment regimens. In the United States alone, antibiotic resistant bacteria account for more than 2 million infections and 25,000 deaths annually (1). Alarmingly, resistance to virtually all major classes of antibiotics has been reported (Figure 1-1), including reports of multi-drug resistant (MDR), extensively-drug resistant (XDR) and pan-drug resistant (PDR) bacteria, the latter representing bacteria that are resistance to all approved antibiotic classes (2, 3). These observations alone are worrisome, but combined with the diminished rates of antibiotic discovery observed over the last 50+ years (4), present a global health crisis which threatens to compromise our ability to treat infectious disease altogether. The seriousness of this threat has been emphasized recently by many governments and health agencies, including the Centers for Disease Control (1), the World Health Organization (5) and most recently, the United Nations (6). While such forums have been successful in promoting policies that minimize the spread of antibiotic resistance, there is still a great need to discover and revitalize our dwindling antibiotic arsenal.

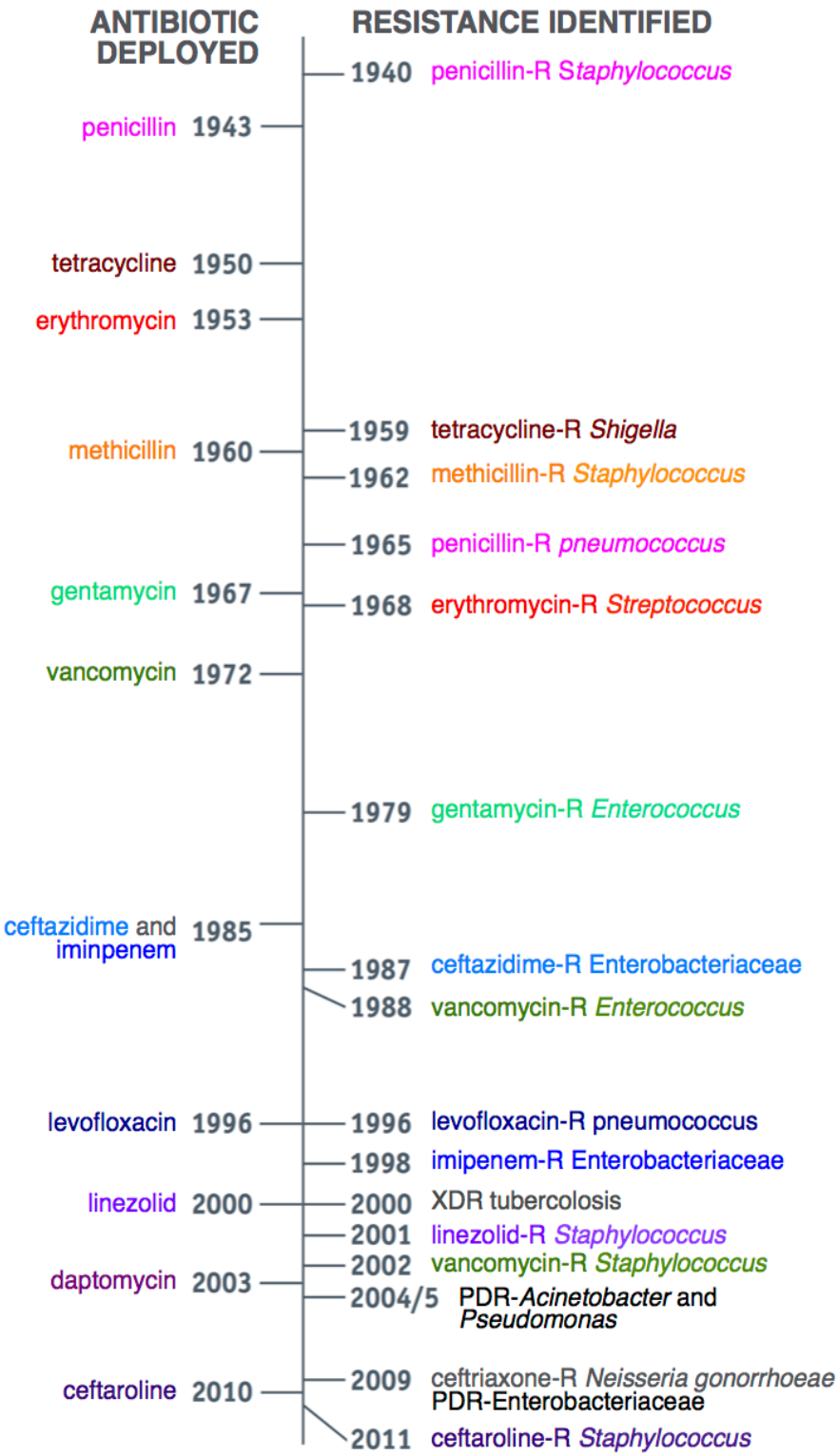


Figure 1-1. Timeline of antibiotic introduction and resistance for commonly used antibiotics. Year of resistance (R) is typically established from literature reports but in the case of PDR-*Acinetobacter* and *Pseudomonas*, data is derived from healthcare transmission reports. PDR, pan-drug resistance; XDR; extensive drug resistance. Adapted from (1).

A brief history of antibiotics. Traditionally, our best antibiotics have been those devised by nature. The serendipitous discovery of penicillin by Alexander Fleming in 1928 began what many consider to be the Golden Era of antibiotic discovery (7). His observation of the antibacterial properties of a contaminating penicillium mold led to the discovery of the first member of the β-lactam family of antibiotics, and established antibiotics as a legitimate means of controlling and treating microbial disease (8). Shortly thereafter, Selman Waksman proposed a new screening platform for antibiotic discovery, which became both highly successful and widely embraced (9). The platform was remarkably simple: Streptomyces isolated from the soil were grown on solid media and screened for antimicrobial activity against a sensitive indicator strain. The success of this screening approach is evidenced by its near universal implementation throughout the pharmaceutical industry at that time and the fact that Waksman was awarded the 1952 Nobel Prize for his discovery of streptomycin—the first antibiotic isolated using this approach (10). The remarkable acceptance of Waksman's antibiotic discovery platform accounts for its success as well as its eventual failure. So widely used was this platform, that the majority of our antibiotic scaffolds used to this day are products of it and were discovered during a short 20 year period after its introduction (11). However, the ubiquitous use of this platform also exhausted its effectiveness; high rates of rediscovery and a limited ability to culture diverse microbes in a laboratory environment at the time ultimately led to its abandonment in favor of alternative methods for antibiotic discovery and development.

What followed was the highly touted (yet largely unsuccessful) synthetic era of antibiotics. Technological advances made it possible to generate large chemical libraries and screen them for activity against specific targets. Equipped with an improved understanding of prokaryotic biology, combinatorial chemistry shifted the focus from nature to the bench (12).

Despite the tremendous investment in such approaches, the return was underwhelming. The fluoroquinilones represent the only synthetic antibiotics in mainstream use today (10) and an homage to an era of antibiotic discovery that never quite lived up to expectations. Combined with the financial and regulatory disincentives associated with antibiotic development, this lack of success contributed to a mass exodus from the industry and resulted in a serious innovation gap now referred to as the 'discovery void' (4).

A return to antibiotic discovery. As new technologies emerge, opportunities to discover, engineer and evolve new antibiotics have increased. During the last decade, we have observed a renewed interest in antibiotic discovery and development. This effort, driven largely by the academic community, has been spurred by an awareness of the antibiotic crisis and encouraged by new technological advances. Computational advances coupled with the reduced costs of high-throughput sequencing have fueled genome-mining efforts and allowed investigators to uncover new natural products and study their pathways in greater resolution than ever before. The development of new methods for culturing what we previously considered 'unculturable' bacteria has also expanded our capacity to survey new natural products. A remarkably simple yet innovative example is the iChip: this simple device brings the soil microenvironment into the laboratory and allows researchers to isolate and culture bacteria in a near-native environment (13). Doing so has allowed researchers to culture new bacterial species and identify novel antibiotic compounds that have not been accessible to previous generations (14-16).

Advances in chemistry and biochemistry have also opened doors to antibiotic development; improved synthetic capabilities make it possible to synthesize and derivatize new compounds with enhanced pharmacological properties. Enhanced analytical technologies also make the characterization of new lead compounds more efficient than ever. Together, these

advances give us great reason to be optimistic about the future of antimicrobial discovery and to expect that we will yet discover many new antimicrobials to replenish a declining supply.

Nature is the best chemist. Historically, the majority of our best antibiotics have come from natural products (12). Examples include such major classes as the β-lactams, aminoglycosides, tetracyclines and glycopeptides (10). The success of natural products likely stems from evolutionary selection; the fact that we observe these compounds in nature today is evidence of their utility to the organisms that produce them. Most antimicrobial compounds in use today are derived from secondary metabolites — organic compounds that are not essential for the growth and reproduction of the producing organism. Due to the non-essential nature of these compounds, natural selection would select against the production of compounds that do not enhance fitness. The production of antimicrobials, which antagonize competing microbes, certainly confers a selective advantage, allowing greater access to resources and space to the producing strain. In this manner, evolutionary selection favors the production of antimicrobials and nature maintains an exquisite pool of intriguing and potentially useful compounds.

Classes of natural products. Until recently, the majority of natural products could be classified into one of four main classes: terpenoids, alkaloids, polyketides and non-ribosomal peptides (Figure 1-2)(17). Terpenoids constitute a large and diverse class of compounds derived from 5-carbon isoprene units joined together and modified in countless ways. Examples include menthol, camphor, carotenoids and tetrahydrocannabinol (the psychoactive component of cannabis; see Figure 1-2A). Alkaloids are typically much simpler organic compounds consisting mainly of carbon, hydrogen and nitrogen atoms. In nature, alkaloids are found primarily in plants (18) but can also be found in various bacteria, fungi and animals (19). Well known alkaloids include morphine, caffeine, nicotine and cocaine (Figure 1-2B).

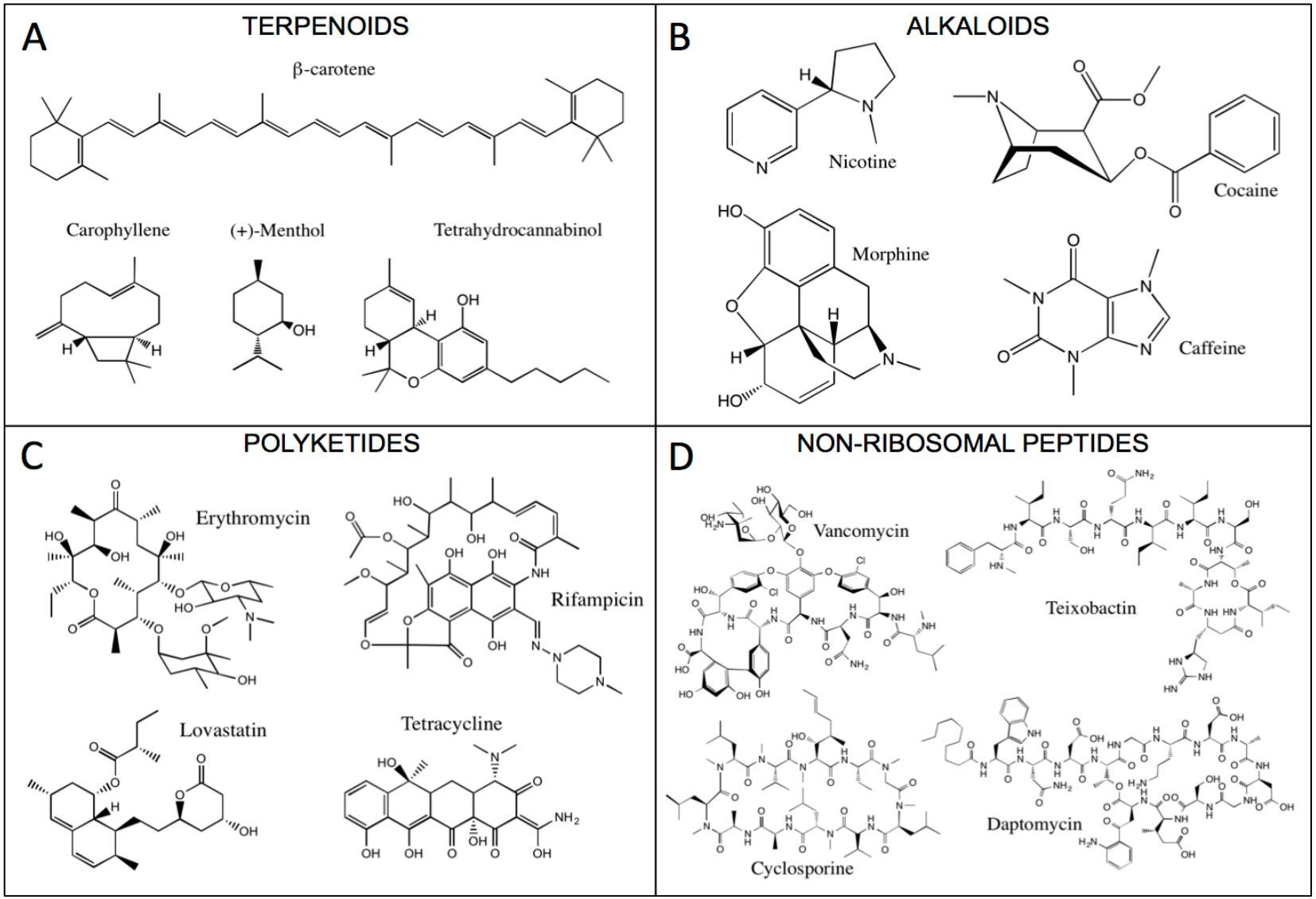


Figure 1-2. Examples of different natural products and their structures. (A) Terpenoids, (B) alkaloids (C) polyketides and (D) non-ribosomal

Polyketides and non-ribosomal peptides typically have more complex structures that arise from complex biosynthetic processes. In both cases, precursors are joined together by multifunctional enzyme modules to generate tremendous chemical diversity (20). In the case of polyketides, biosynthesis occurs from the iterative enzymatic joining of acyl-CoA monomers by specialized polyketide synthases (21). These natural products have been a rich source of many modern pharmaceuticals including the macrolide, ansamycin and tetracycline antibiotics, the antifungals nystatin, pimaricin and amphotericin, as well as the cholesterol lowering drug lovastatin (Figure 1-2C)(22). Non-ribosomal peptides are also complex secondary metabolites that rival polyketides in terms of chemical and biological diversity, but these arise from the concerted activities of non-ribosomal peptide-synthetases (NRPS). These are often large, multi-subunit modules capable of generating peptides with exotic modifications such as the presence of modified amino acids (glycosylation, methylation, hydroxylation and halogenation are common), thiazoles, oxazoles, and the incorporation of D-amino acids (23, 24). Many well-known antibiotics are the products of NRPS. These include vancomycin, bacitracin, daptomycin and the recently discovered teixobactin (Figure 1-2D). Because both polyketide synthases and NRPS act independent of the ribosome and possess overlapping biosynthetic pathways, some natural products are fusions of the two being acted on by both sets of machinery. For detailed reviews of polyketide and nonribosomal peptide biosynthesis, see (20-22, 25).

Represented among these four major classes of natural products is remarkable chemical and biological diversity, however the genome sequencing efforts of the 21st century uncovered another major class of natural products: ribosomally synthesized and post-translationally modified peptides (RiPPs) (26). This broad class of natural products display tremendous diversity in terms of structure and biological activity, and have become a focal point of antibiotic

research in recent years. In the following section, we outline our current understanding of RiPP biology and their potential as antibiotic compounds.

1.2 RiPPs: a new class of natural products

Overview of RiPPs. Ribosomally synthesized and post-translationally modified peptides (RiPPs) constitute a structurally diverse group of peptides found in all domains of life (17). As the name implies, RiPPs are synthesized at the ribosome and subsequently undergo significant post-translational modification (PTM) to produce the mature peptide. RiPP biosynthesis commences with a 7-110 residue precursor peptide encoded by a structural gene (27). In general, this precursor peptide is composed of an N-terminal leader fused to a C-terminal core peptide, the latter gives rise to the natural product (Figure 1-3). Modification of the core peptide is facilitated by dedicated biosynthetic enzymes that provide RiPPs with their characteristic structure, stability and (often) biological activity. The leader peptide is thought to contribute to substrate specificity and export; in many cases biosynthetic enzymes have been shown to interact directly with the leader and are likely directed to their cognate core peptide in this manner (28-30). In almost all cases, the leader peptide itself remains unmodified and is proteolytically cleaved during the final stages of processing (31). Notable exceptions to this general scheme of processing include the bottromycins and cyanobactins. Bottromycins lack the prototypical N-terminal leader sequence observed in most RiPPs but instead contain a C-terminal follower sequence that presumably serves a similar role (32-35). Cyanobactin precursor peptides contain highly conserved N-terminal leader sequences but additionally contain recognition sequences (orange boxes, Figure 1-3) that flank the core peptide and facilitate the head-to-tail macrocyclization that is characteristic of this RiPP family (36).

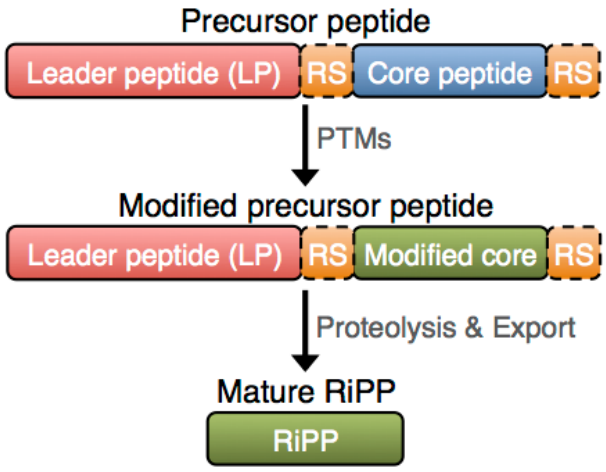


Figure 1-3. General schematic of RiPP processing. Post-translation modifications (PTMs) are installed on the core peptide to produce the modified precursor peptide. Subsequent leader peptide cleavage and export produce the fully processed natural product. In certain RiPP families, additional recognition sequences (RS) facilitate cyclization of the processed core peptide.

Structure and classification. Abundant in nature, more than 20 distinct RiPP families have been identified to date (17). These differ greatly in their source, structure and biological activity, but are commonly classified by the types of post-translational modifications they possess (Figure 1-4). One example is the lanthipeptides (or lanthionine-containing peptides), among which are some of the best studied RiPPs. This large group of polycyclic peptides is characterized by the presence of thioether (C-S-C) linkages in the form of lanthionine and/or methyllanthionine residues (37-39). These thioether bridges provide remarkable intramolecular stability and arise from the enzymatic dehydration of Ser/Thr residues to Dha/Dhb and subsequent reaction with Cys thiols (38). Antimicrobial lanthipeptides such as nisin (Figure 1-4, top) are referred to as lantibiotics (40), many of which have been shown to inhibit gram-positive bacteria (40, 41).

Linear azol(in)e-containing peptides (LAPs) constitute another example of a diverse RiPP family. These linear peptides contain thiazolines/thiazoles and oxazolines/oxazoles, which arise from the heterocyclization of Cys, Ser and Thr residues. This conversion occurs via a two-step process: in an initial reaction, five-membered thiazoline and oxazoline rings are formed from the

ATP-dependent cyclodehydration of Cys or Ser/Thr residues respectively (42). In a second step, azolines are oxidized to azoles by flavin-dependent dehydrogenases (43, 44). Conformational constraints introduced by these sulfur- and oxygen-containing heterocycles are thought to contribute to the biological activity of LAPs, which include the DNA-gyrase inhibitor microcin B17 (45, 46) and the narrow-spectrum bactericidal antibiotic plantazolicin (47).

Thiopeptides, which are discussed in greater detail in the following section, represent an intriguing family of peptides displaying both the dehydro-amino acids observed in lanthipeptides and the thiazole/oxazole modifications seen in the LAPs. Other fascinating examples of RiPP families include the lasso peptides (named for their characteristic shape), linaridins, proteusins (among which are the most highly modified peptides known), cyanobactins, bottromycins, amatoxins, microcins, microviridins, sactipeptides, cyclotides, and orbitides. Examples of structures from various families are shown in Figure 1-4. For a comprehensive review of RiPP classification, biosynthesis and biological activities, see (17, 31).

Therapeutic potential. Many (but not all) RiPPs have antimicrobial properties that make them attractive antibiotic candidates. Nisin for example, has broad antibacterial activity against gram-positive pathogens such as *Listeria monocytogenes*, *Clostridium botulinum* and *Staphylococcus aureus*, and is widely used as a food preservative (48-50). Another example, which was mentioned previously, is the LAP microcin B17. This highly modified linear peptide is bactericidal against *E. coli* and several related species, exerting its activity by inhibiting DNA gyrase (45, 46, 51). One final example is thiostrepton. This well-known thiopeptide has broad activity against gram-positive bacteria and is used widely in veterinary medicine (52). Many other RiPPs have well documented biological activities yet their properties are not fully appreciated due to a limited ability to efficiently synthesize, purify and/or derivatize them.

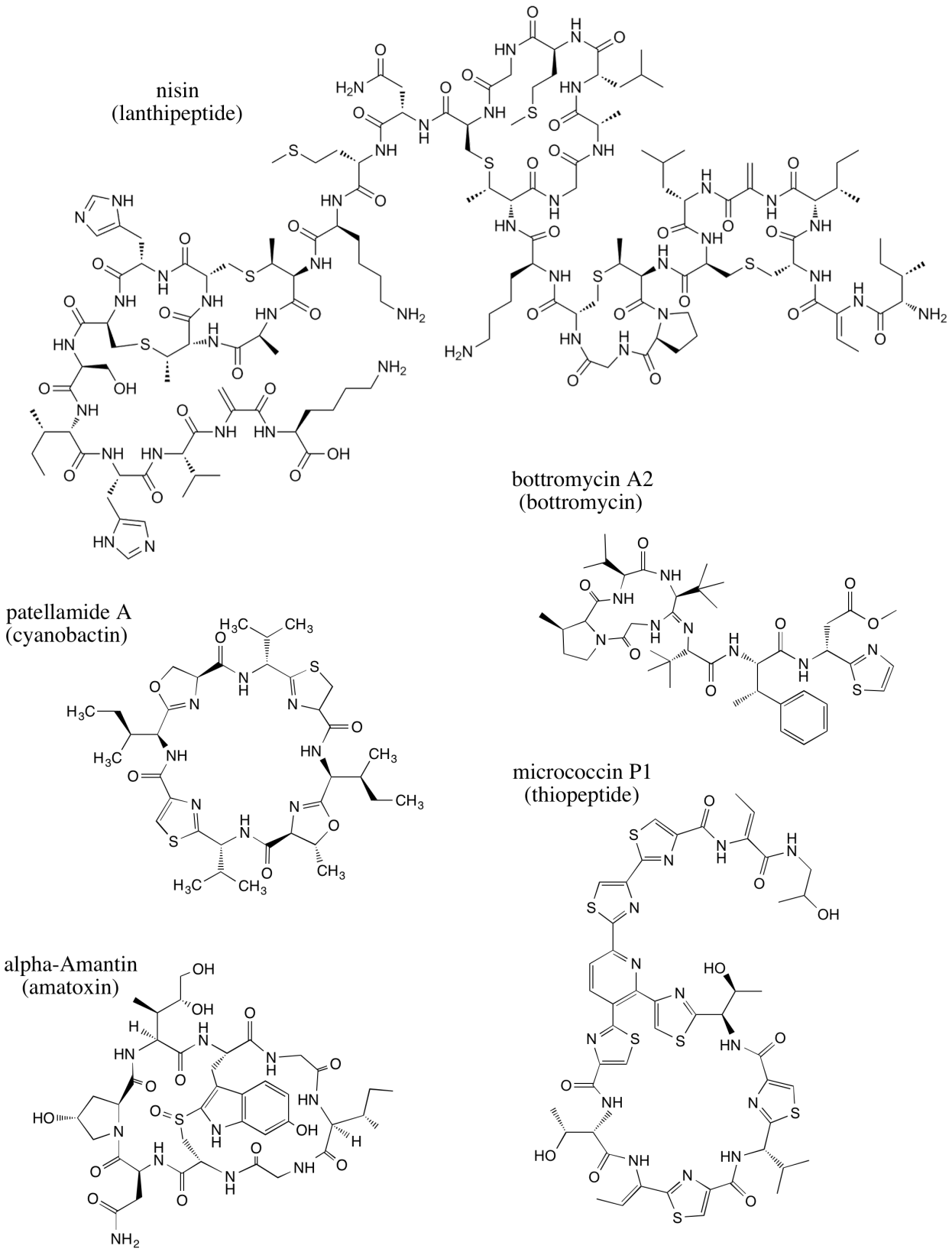


Figure 1-4. Examples of RiPP structures. RiPP family is indicated in parentheses.

Furthermore, many details of RiPP biosynthesis have remained elusive, complicating efforts to utilize heterologous expression as a viable route to large-scale production. The discovery of additional RiPP biosynthetic clusters has the potential to address many of these obstacles and is opening the door to further engineering studies, which may unlock even greater peptide diversity.

Peptide engineering and diversification. RiPPs possess many attributes that make them veritable candidates for engineering studies. For example, these peptides are readily detected by genome mining approaches as the final product can be traced to a genetically encoded precursor peptide, and this structural gene is typically located nearby or within a cluster of associated biosynthetic genes (38, 53, 54). In addition, the small size and conserved organization many biosynthetic clusters make them amenable to heterologous expression (55-57).

Another intriguing feature of RiPPs is that they can be rapidly diversified by simple genetic manipulation. Their ribosomal origin, combined with the plasticity of many RiPP enzymes and the permissiveness of core peptide sequences to variation, allow for tremendous chemical diversity to be generated from small changes in the structural gene. For example, several lanthipeptide studies have shown that unnatural core peptides fused to the nisin leader can still be dehydrated and cyclized by native biosynthetic machinery (58-60). Similar findings have been reported in cyanobactin (61-63) and lasso-peptide systems (64, 65). In each case, post-translational peptide cyclization can still be mediated by modifying enzymes despite changes to the core peptide sequence.

Several landmark thiopeptide studies also highlight the tolerance of processing enzymes to perturbations in the precursor peptide. Precursor replacement experiments using a thiocillin

producing strain of *B. cereus* strain (66) illustrate the promiscuity of the Tcl machinery. Whereas the wild type (*wt)* strain produces 8 detectable thiopeptide variants from a common precursor, simple mutations in core peptide expanded that diversity to over 100 compounds (67, 68). A number of these retained bioactivity and several displayed macrocyclic ring sizes not observed in nature (69). More recently, work conducted using an *in vitro* thiomuracin platform clarified the substrate specificity requirements of various pathway components (30). Results suggest that the early acting, thiazole-installing enzymes are most permissive to core peptide variation and that specificity during early stages of processing is governed principally by the leader peptide. Conversely, later acting Ser/Thr dehydratases are reportedly less dependent on the leader-peptide for recognition, instead recognizing specific core peptide structures incurred as a result of earlier modifications. Together, these examples highlight the remarkable plasticity of many biosynthetic enzymes and suggest that if equipped with a detailed grasp of their biosynthetic processes, thiopeptides scaffolds pose a tractable route to obtaining high chemical diversity at low genetic cost (17).

1.3 Thiopeptides: biosynthesis and engineering potential

Overview of thiopeptides. Thiopeptides represent an intriguing engineering opportunity because they display such broad chemical diversity and are reasonably well studied (70). This growing family of RiPPs brings together the Ser/Thr dehydratase activity of lanthipeptides with the thiazole/oxazole modifications observed in the LAPs, cyanobactins and bottromycins. In addition, thiopeptides contain a central six-membered ring and an assortment of other PTM-derived decorations, the significance of which we are only beginning to grasp. There have been over 100 thiopeptides discovered to date, the majority of which are produced by actinomycetes and the low GC firmicutes (17). Although thiopeptides have been reported since the late 1940s,

their ribosomal origin was disputed for a long time. It was only in 2009, when multiple groups independently reported the discovery of different thiopeptide gene clusters (66, 71-73), that thiopeptides were formally accepted as members of the RiPP family.

Structure and classification. The defining features of thiopeptides are the nitrogenous core, which gives them their macrocyclic shape. This central nitrogen heterocycle is found in various oxidation states (Figure 1-5A), the status of which is used to classify them (74). Series A thiopeptides (e.g. Sch 18640) contain a fully reduced piperidine ring. Series B thiopeptides (e.g. thiostrepton, siomycin) are further oxidized and contain a 1,2-dehydropiperidine ring. Series C peptides possess a piperidine ring fused with imidazole. At present, only one member of this family is known to exist — Sch 40832 produced by *Micromonospora carbonacea* (75). Series D thiopeptides are the most numerous group and contain a trisubstituted pyridine core. Some of the best-studied thiopeptides fall into this category, including the micrococcins, thiocillins and thiomuracins. Series E thiopeptides constitute the last class and include nosiheptide, the nocathiacins and others. These peptides possess a tetrasubstituted pyridine ring containing a hydroxyl group. Most series E thiopeptides also contain a second characteristic macrocycle (e.g. nosiheptide, below). Examples of various thiopeptide structures representing each of the classes discussed, are shown in Figure 1-5B.

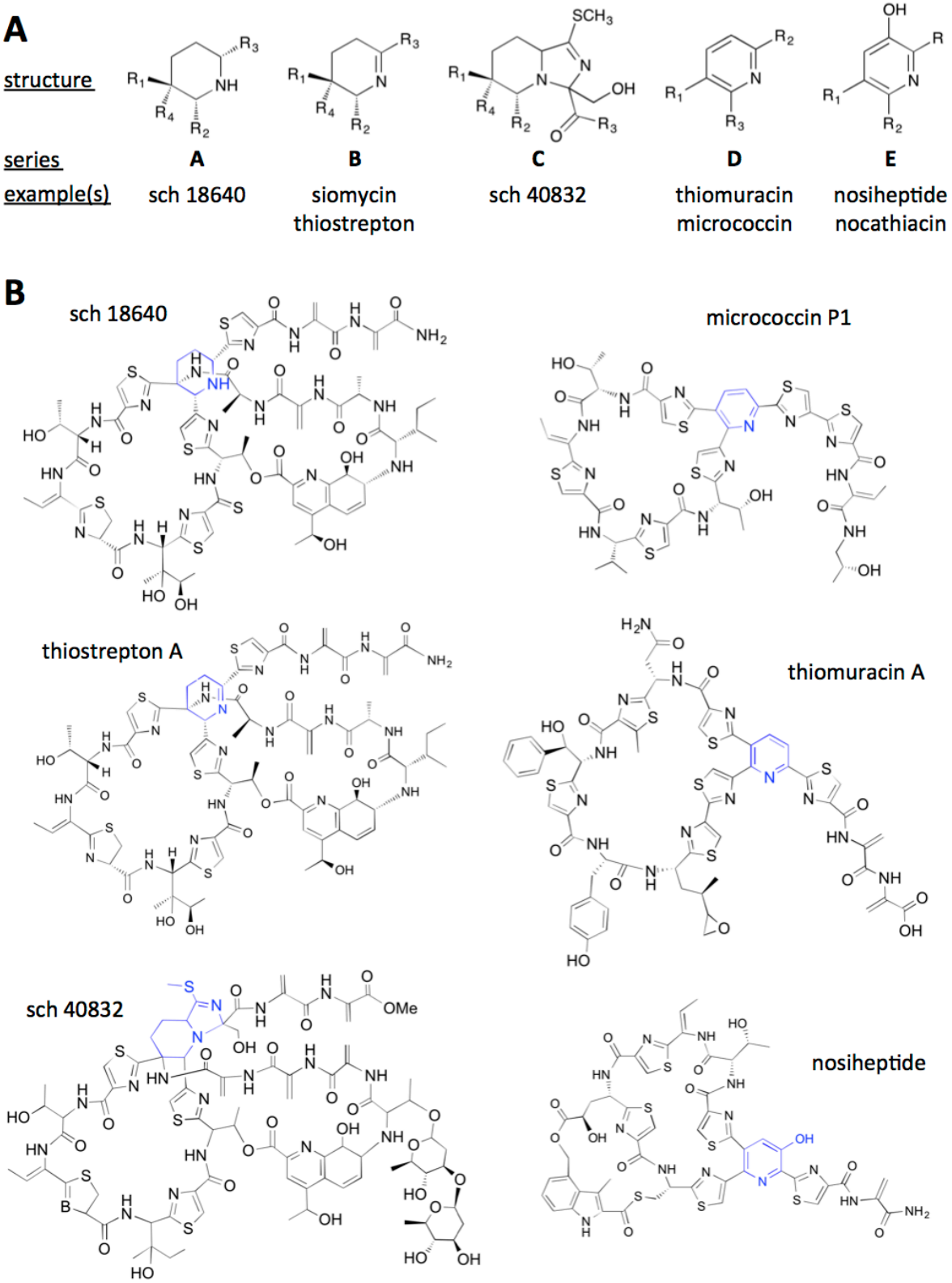


Figure 1-5. Thiopeptide core organization and structure. (A) Arrangement of the nitrogenous core differs between thiopeptide series. (B) Complete structures of various thiopeptides. In each structure the nitrogenous core is shown in blue.

Biological activity. Thiopeptides have well established antibacterial activity against gram-positive bacteria including many clinically relevant and drug resistant pathogens. In our laboratory, we have demonstrated the antibacterial action of micrococcin P1 against *Clostridium difficile*, vancomycin-resistant *Enterococcus faecalis*, methicillin-resistant *Staphylococcus aureus*, *Streptococcus pyogenes*, *Listeria monocytogenes*, *Bacillus anthracis*, *Mycobacterium tuberculosis*, and many other pathogens (76). Thiopeptides such as thiostrepton, thiocillin and thiomuracin display similar activity spectrums and are often active in the nanomolar range (68, 73, 77). Whilst ineffective against gram-negative bacteria in bioassays, thiostrepton has been shown to bind to the ribosomes of *E. coli* and *Streptomyces coelicolor in vitro* (78). Such studies suggest this lack of activity against gram-negatives can be attributed to poor penetration of cells as opposed to the lack of a suitable intracellular target. This is promising from an engineering perspective, as overcoming this penetration barrier (challenging but possible) would greatly expand the spectrum of thiopeptide activity.

In cases where the mechanisms of action have been empirically determined, there appears to be two main modes that correspond with the size of the central macrocycle. Those with smaller, 26-atom macrocycles (e.g. micrococcin, thiostrepton and nosiheptide) block translation in target cells by binding at a cleft between the 23S rRNA and the ribosomal protein L11 (79-81). When bound here, thiopeptides prevent the conformational changes required for ribosomal translocation along template mRNA and protein synthesis is halted (80-82). In contrast, thiopeptides with larger, 29-atom macrocycles (e.g. the thiomuracins, GE2270A) do not target the ribosome directly but instead abrogate the function of elongation factor Tu (EF-Tu) (83). These peptides have been shown to block the aminoacyl-tRNA binding site of EF-Tu and prevent these complexes from being delivered to the ribosome (84, 85). Thiopeptides with 35-

member macrocycles are known to exist (71, 86), but for these peptides, the precise mechanism of action remains unknown (52).

In addition to their well-documented antibacterial properties, various other biological activities of thiopeptides have been described. These include reports of antifungal (87), antiprotozoan (88, 89), antiviral (90) and anticancer properties (91, 92). While such reports are promising and illustrate the biological diversity that can arise from this peptide family, in many cases the precise molecular mechanisms are incompletely understood. Indeed further study in this area is warranted to fully appreciate the full range of biological functions that thiopeptides may serve.

Host-immunity and resistance. Until recently, specific details of host immunity mechanisms remained elusive (93). Methylation of the 23s rRNA had been shown to confer self-resistance in thiostrepton producing strains of *Streptomyces azureus* and in nosiheptide producing *Streptomyces actuosus* strains, but this method could not account for immunity in all thiopeptide producing strains (94). For example, in the *M. caseolyticus* MP1 producer (described in detail in subsequent chapters); cells are protected in presence of the plasmid encoded micrococcin-cluster, but become MP1-sensitive after spontaneous plasmid curing (76). Curiously, many thiopeptide producers antagonize members of the same, or closely related species. Taken together, these findings suggest that among thiopeptide producing bacteria, specific (but incompletely understood) immunity mechanisms must account for the observed host-resistance, as opposed to structural differences in the host target site.

While from an applied perspective, rapidly generated spontaneous resistance is a drawback, the mapping of resistance mutations has proven helpful in eliciting specific host-immunity mechanisms. For example, mutations in L11 from *Bacillus megaterium* have been

shown to confer resistance to micrococcin P1 (80, 95). Similarly, we have observed in our laboratory that selecting for spontaneous micrococcin resistance in *B. subtilis*, *S. aureus* and *M. caseolyticus* inevitably results in mutations in L11, especially in the N-terminal region proximal to the micrococcin binding site. In the context of these findings, the conservation of L11-encoding genes (TclQ and TclT) in both the *B. cereus* (66) and *M. caseolyticus* (76) gene clusters was suggestive of a role in host-immunity, which was later confirmed in this work.

An unrelated but fascinating observation is that while resistance to thiostrepton confers cross resistance to MP1, spontaneous MP1 resistance does not confer cross resistance to thiostrepton (78). These findings suggest that combination thiopeptide therapies, even by drugs that share a common target, may provide a viable strategy to delaying or circumventing the onset of resistance.

Biosynthesis. One of the most elusive aspects of thiopeptide biology relates to how such exquisite chemical complexity can be achieved from so few enzymes. Thiopeptides are among the most densely modified peptides known, yet at the onset of this dissertation project, very few of these details were well understood. While for some family members (including the prototypical thiopeptide, thiostrepton) this remains true, recent work from our group and several others has revealed important mechanistic details regarding the biosynthesis of micrococcin (96), thiomuracins (30, 77) and thiocillins (66). Here we summarize our current understanding of this intricate, multistep process and highlight further opportunities for study.

Current dogma asserts that the production of a bioactive thiopeptide requires expression of at least 7 genes: 1 structural gene that encodes the precursor peptide and 6 biosynthetic genes. Homologs of these 7 genes can be found in all confirmed thiopeptide-producing gene clusters

(Figure 1-6). Thiopeptide biosynthesis commences with an unmodified precursor peptide and concludes with a macrocyclic bioactive product. To initiate processing, an ocin-ThiF like protein binds to the leader peptide and recruits it for enzymatic modification (29, 30). A homolog of this protein family can be found in all known thiopeptide clusters and although they vary greatly at a sequence level, leader recognition elements are readily detected in these proteins (28). Once recruited, Cys/Ser-derived azoles are installed by two conserved enzymes: first, an ATP-dependent cyclodehydratrase aromatizes Cys/Ser residues to thiazolines/oxazolines respectively. Next, a flavin-dependent dehydrogenase oxidizes these to their corresponding thiazoles/oxazoles. *In vitro* thiomuracin studies have confirmed the essential nature of these three proteins and cofactors for heterocyclization (77). Furthermore, mounting evidence suggests that these azoles are installed in an ordered, but non-linear fashion (30).

Following heterocyclization, free Ser/Thr residues are selectively dehydrated by the coordinated activities of two conserved dehydratases. The first enzyme catalyzes the glutamylation of Ser/Thr residues via a tRNA-glu donor (96). The second eliminates glutamate side chains to produce the corresponding alkenes (Dha or Dhb). These two enzymes are homologous with the two domains of the multifunctional NisB lantibiotic dehydratase, which has been shown to dehydrate Ser/Thr residues in an analogous manner (97). Thiopeptide biosynthesis concludes when two of the Ser-derived Dha residues undergo formal [4 + 2] cycloaddition to produce the central six-membered ring. This process is catalyzed by a dedicated macrocyclase, which concomitantly cleaves the leader peptide (98), yielding the mature peptide.

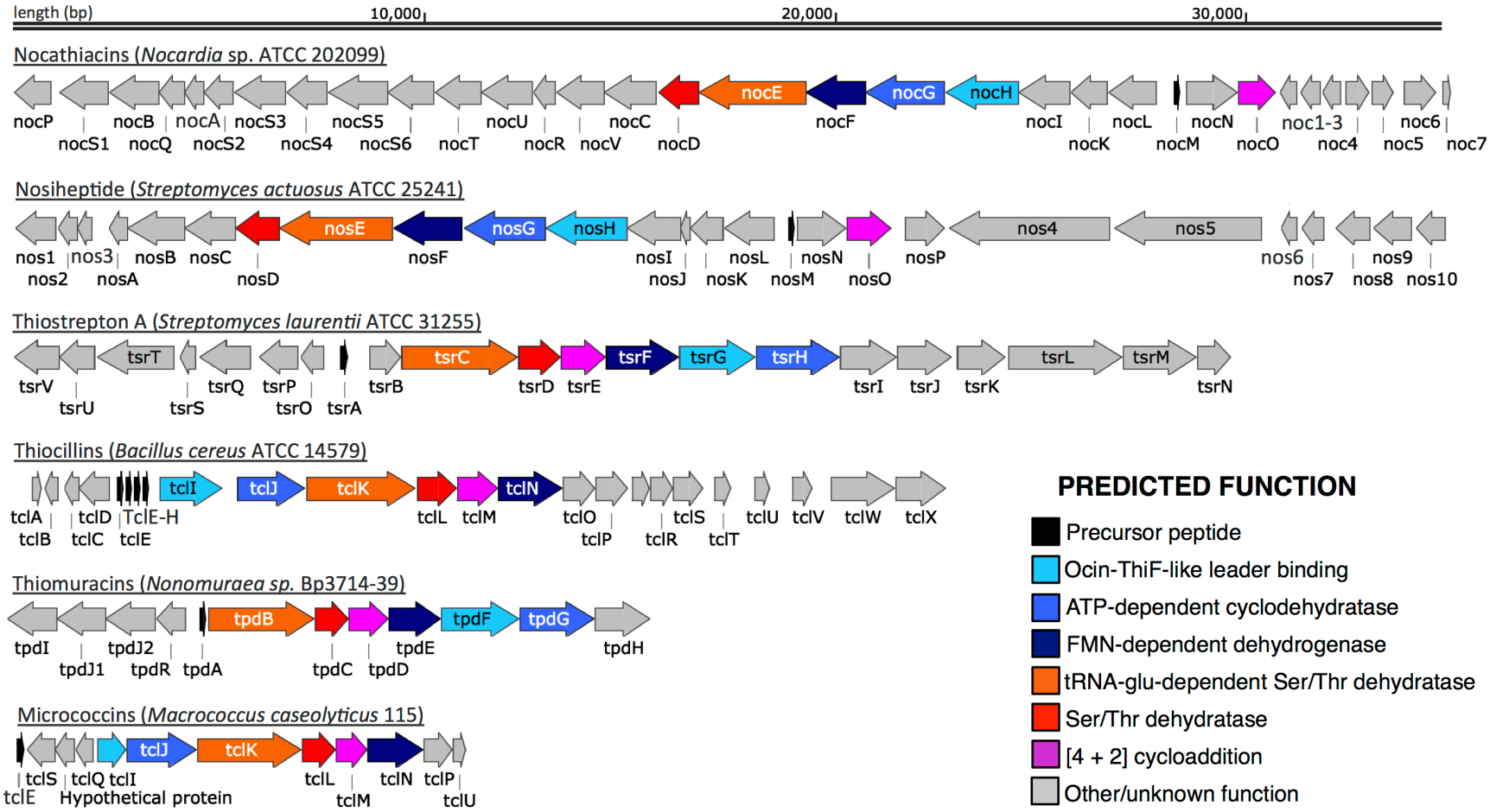


Figure 1-6. Comparison of well characterized thiopeptide gene clusters. The six conserved biosynthetic enzymes in each cluster are colored based on their alignment with the nosiheptide (*nos*) cluster. Predicted functions are based on *in vitro* thiomuracin studies and *in vivo* studies using micrococcin- and thiocillin-producing strains

These core modifications are common to all thiopeptides in which the biosynthetic details have been elucidated, however additional tailoring is often observed. While in most cases these are thought to represent accessory modifications that are not required for biological activity, some interesting exceptions exist. For example, in the micrococcin/thiocillin pathway, we have additionally shown that TclP, a short-chain dehydrogenase, is essential for full peptide maturation and bioactivity (57). This enzyme oxidatively decarboxylates the core peptide C-terminus; and serves as a prerequisite for Ser/Thr dehydration and subsequent macrocyclization steps (96). Chapters 3 and 4 outline these findings in greater detail.

While these recent reports have greatly illuminated this our understanding of these complicated processing schemes, many biosynthetic details are still unknown. For example, the biological significance of the diverse C-terminal modifications observed throughout the thiopeptide family and the exact mechanistic details of their installation remain to be deciphered. An enhanced understanding of the substrate specificities of each pathway component and a better grasp of the molecular transitions that account for progress through this multi-step process would also improve our knowledge and aid efforts to engineer these pathways. In each case, the availability of genetically tractable platforms and effective analytical tools will greatly fuel these studies.

1.4 Micrococcin: an ideal platform for biosynthetic and therapeutic exploration.

An ideal platform for studying thiopeptide biosynthesis would retain all of the chemical and biological properties common to most members of this RiPP family, but afford researchers the opportunity to easily and rapidly manipulate the compound, its precursors and associated processing enzymes. Micrococcin, the first thiopeptide ever discovered (99), meets all of these

criteria. Specifically, it retains all of the properties that make thiopeptides attractive for study (potency, broad spectrum activity, chemical stability), yet it is structurally relatively simple compared with other family members, and multiple biosynthetic gene clusters have been identified and validated (66, 76, 93).

A brief history of micrococcin. Micrococcin exists in two stable forms in nature that differ only in the redox state of the C-terminus. Micrococcin P1 (MP1) contains a C-terminal alcohol group, whereas its sister compound Micrococcin P2 (MP2) contains a ketone at the same position. Both forms are biologically active and are often jointly produced by the same organism (66, 100, 101). The first reported micrococcin producer was isolated from a sewage sample in Oxford, UK in 1948 (99). Likely producing a mixture of both compounds (100), the strain was identified as a *Micrococcus sp*. and its remarkable antibacterial properties were noted. Several other groups have subsequently discovered micrococcin producing strains of bacteria, often from diverse genera. One of the most notable examples is the *Bacillus cereus* strain reported by the Walsh lab in 2009 (66). In addition to discovering a new thiopeptide producing strain, by identifying the cognate gene cluster, they firm established thiopeptides as legitimate members of the RiPP class of antibiotics. Their thiocillin (*tcl* for short) gene cluster was reported to contain as many as 24 genes and was shown to give rise to a set of 8 related compounds from the same core scaffold. This cluster would later serve as the basis of pioneering studies probing the basic principles of thiopeptides biosynthesis, however it lacked the genetic versatility we desired. In addition, the lack of a homogenous product clouded efforts to manipulate this system and determine the precise molecular consequences of such changes.

The discovery of a new micrococcin cluster in *Macrococcus caseolyticus* str. 115 provided an opportunity to overcome both of these challenges. Originally discovered in the early

1980s (102), this avian isolate was first noted for its antibacterial properties and was initially characterized as a member of the species *Staphylococcus epidermidis* (103). Initial attempts to identify the antimicrobial compound were unsuccessful, but important physical characteristics were noted and purification methodologies were developed, ultimately aiding future efforts to characterize this compound. Although successful identification of this thiopeptide would not occur until 2012 (76), the therapeutic potential of this strain was vigorously evaluated shortly after its discovery. From 1984-1985 a large-scale control program was enacted among the Utah Turkey population in which exposure to aerosolized *M. caseolyticus* 115 was shown to significantly reduce the incidence of staphylococcosis in turkeys (104). In 1984 ~3 million turkeys in Utah were exposed at 1-10 days of life and again at 4-6 weeks. Staphylococcosis among this population was still observed but at a lower rate than in previous years. In 1985, a more controlled experiment was conducted in which 1 million turkeys were treated, and 2 million control birds were untreated. Incidence among treated turkeys was 2.7% lower ($P<0.001$) than in the untreated turkeys that were monitored. These preliminary findings led to the continued use and preservation of this strain, which would become the subject of intense investigation decades later.

A simpler gene cluster. In 2012, the identity of this antimicrobial compound would finally be realized. A combination of biochemical and genetic methods resulted in the characterization of the peptide produced by *M. caseolyticus* as MP1 (76). More importantly, a new gene cluster was identified, which unlike all other known thiopeptide clusters at the time was encoded on a small, replicative plasmid. The overall architecture of this cluster was reminiscent of the *tcl* cluster identified just a few years earlier in *B. cereus,* but it was much simpler and smaller in size (66). This strain also afforded us the luxury of producing a much more homogenous product;

relatively pure (>97%) MP1 preparations were readily obtained by organic extraction. This provided an exciting opportunity to parse the finer details of micrococcin biosynthesis and although significant engineering would be required to develop a suitable genetic system, this new micrococcin platform would become a powerful tool through which the biosynthetic pathway could be dissected in greater resolution than had previously been possible.

1.5 Summary of research chapters

This dissertation represents a rigorous analysis of the micrococcin biosynthetic pathway identified in *Macrococcus caseolyticus* strain 115. In chapter 2 we describe the discovery of this plasmid encoded gene cluster, predict gene functions via *in silico* analysis and characterize the mechanism of self-immunity. The small size of this new micrococcin cluster, and the simple yet effective nature of this immunity mechanism are striking and set the stage for future strain manipulation.

Chapter 3 outlines our efforts to engineer this pathway for heterologous expression in *Bacillus subtilis.* This includes promoter engineering, genetic refactoring and modularization of the gene cluster to produce a highly versatile genetic platform allowing for rapid gene manipulation and product detection. We demonstrate the versatility of this platform by testing various biochemical tags on all pathway components and establish methods for the isolation and characterization of pathway products. Finally, we establish a role for TclS in reducing the decarboxylated C-terminus to MP1 and demonstrate that deleting of *tclS* provides a simple route to homogenous MP2 production.

Chapter 4 represents the most complete analysis of the micrococcin pathway to date and leverages our *B. subtilis* platform to genetically block the pathway at different stages of

processing and characterize partially processed peptides by mass spectrometry. We outline the order of processing events and describe two distinct phases of processing punctuated by the accumulation of stable intermediates. Most importantly, we uncover a key role for micrococcin C-terminal modification in mediating the transition between these two phases of processing and demonstrate the effectiveness of this platform in detecting all pathway products, however completely or incompletely processed they are.

In the final chapter, we discuss the current trajectory of the field and propose future research aims and methods designed to expound upon our understanding of thiopeptide biosynthesis. We highlight opportunities to use this platform for the generation of micrococcin derivatives with improved chemical properties as well as for the generation and screening of unnatural peptide libraries to uncover new antibiotic lead compounds and new antimicrobial modes of action.

Chapter 2. Characterization of a novel plasmid-borne thiopeptide gene cluster in *Macrococcus caseolyticus* strain 115

The following chapter is an adaptation from an article entitled "Characterization of a novel plasmid-borne thiopeptide gene cluster in *Staphylococcus epidermidis* strain 115" published in the Journal of Bacteriology in December 2014. This chapter mirrors the structure and content of the article, however all references to the species of the producing strain have been updated to "*Macrococcus caseolyticus*" to reflect the more precise species identification that took place after the manuscript was published.

2.1 Summary

Thiopeptides are small (12-17 amino acid), heavily modified peptides of bacterial origin. This antibiotic family, with more than 100 known members, is characterized by the presence of sulfur-containing heterocyclic rings and dehydrated residues within a macrocyclic peptide structure. Thiopeptides, including micrococcin P1, have garnered significant attention in recent years for their potent antimicrobial activity against bacteria, fungi and even protozoa. Micrococcin P1 is known to target the ribosome; however, like other thiopeptides, its biosynthesis and mechanisms of self-immunity are poorly characterized. We have discovered an isolate of *Macrococcus caseolyticus* harboring the genes for thiopeptide production and self-protection on a 24-kb plasmid. Here we report the characterization of this plasmid, identify the antimicrobial peptide that it encodes, and provide evidence of a target replacement-mediated mechanism of self-immunity.

2.2 Introduction

Thiopeptides are an intriguing class of peptide antibiotics that have been investigated for more than six decades for their antimicrobial activity. Widely produced by bacteria in nature, thiopeptides possess complex structures that arise from extensive post-translational tailoring (Figure 2-1). A hallmark of thiopeptide structure is the presence of numerous thiazolyl groups arranged around a macrocyclic structure with a pyridine, piperidine or dehydropiperidine core (17). It is also common for thiopeptides to possess dehydro-amino acids (Dha and Dhb) and other modifications directed by specialized biosynthetic enzymes. Detailed reviews of thiopeptide biosynthesis can be found elsewhere (72-74, 93). These alterations make thiopeptides some of the most densely post-translationally modified peptides known (66).

Thiopeptides are potent inhibitors of protein synthesis, particularly among gram-positive bacteria. In general, thiopeptides have been shown to act by two distinct mechanisms, depending on the size of the macrocycle: 26-member macrocycles (e.g. micrococcin P1 and thiostrepton; Figure 2-1) bind at a cleft between the ribosomal protein L11 and the 23S rRNA known as the GTPase associated center (80-82). This region is a critical binding site for elongation factor G (EF-G), which is necessary for ribosome translocation along template mRNA. When thiopeptides are bound here, EF-G is prevented from binding and protein synthesis is halted (105-108). Thiopeptides with 29-membered rings (e.g. GE2270A; Figure 2-1) bind to elongation factor Tu (EF-Tu), blocking its aminoacyl-tRNA binding site, thus preventing peptide elongation (83-85). Beyond their intrinsic antibacterial properties, thiopeptides (especially thiostrepton and micrococcin P1) have been investigated for a variety of other applications including antifungal (87), antiplasmodial (88, 89, 109, 110) and anti-cancer properties (91, 92, 111).

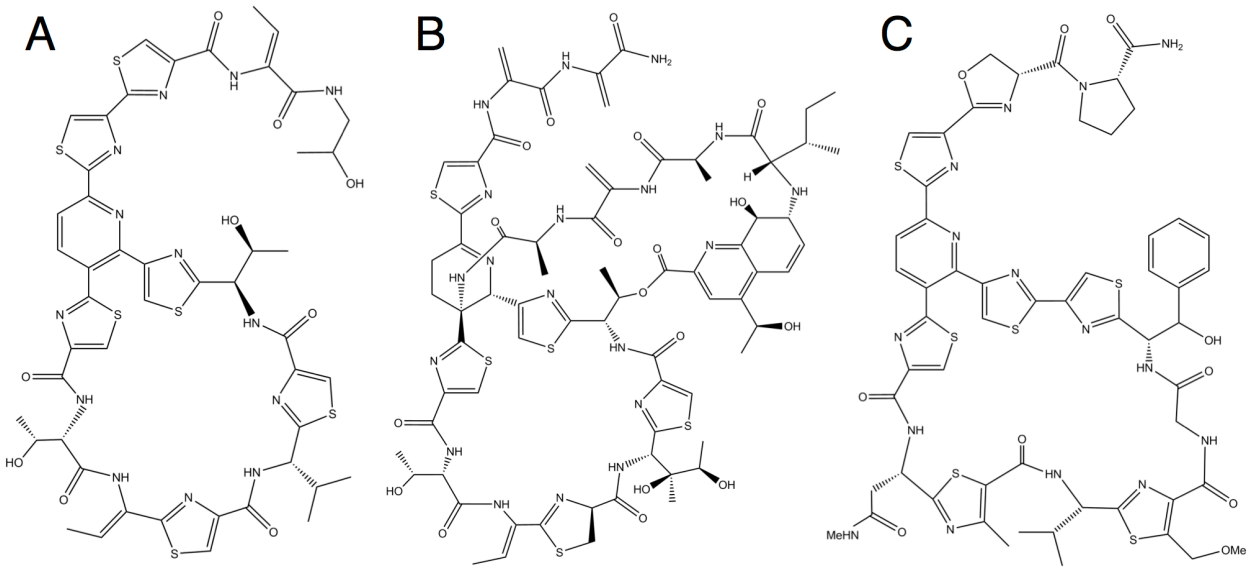


Figure 2-1. Structures of various thiopeptide antibiotics. (A) micrococcin P1, (B) thiostrepton, and (C) GE2270A

Mechanisms of thiopeptide self-immunity are poorly understood. The majority of thiopeptide-producing bacteria have been shown to antagonize closely related species or even members of the same species, implying that mechanisms of immunity are encoded alongside the genes for biosynthesis. Methylation of 23S rRNA has been shown to confer self-immunity in thiostrepton-producing strains of *Streptomyces azureus* and nosiheptide-producing strains of *Streptomyces actuosus* (94). However, for many other thiopeptide producers, the precise self-immunity mechanisms remain unknown or untested.

Here we report the characterization of a novel thiopeptide gene cluster and mechanism of self-immunity in a micrococcin P1 (MP1)-producing isolate of *Macrococcus caseolyticus*, strain 115. First discovered in 1983, this avian isolate was reported to produce an interfering compound which inhibited virulent strains of *S. aureus* (102). In subsequent studies, researchers devised means to extract the compound in order to evaluate certain physical characteristics (103). However, they were unable to identify the compound or map the genes required for antimicrobial

production. Using this strain, we have identified the compound as MP1, defined the MP1 biosynthetic gene cluster, and addressed the mechanism by which strain 115 protects itself from MP1 toxicity.

2.3 Materials and Methods

Target range assays. The *M. caseolyticus* 115 target spectrum was assessed using either the flanking-patch assay or the spot-on-lawn assay. Precise media types varied according to the strain being tested, and these details are given in Table 2-S1. Flanking patch assays were performed by growing *M. caseolyticus* 115 (source patch) for 18 h at 37°C before patching indicator strains 5 mm from the source patch (see Figure 2-2A). Strains were incubated for a further 12-24 h before assessing resistance or sensitivity. For assays using purified MP1, spot-on-lawn assays were performed. These assays were conducted by spreading 150 µl of a 1:50 diluted overnight culture of the test strain on a compatible growth medium. Plates were allowed to dry before applying 5-µl spots of MP1 dilutions. Spots were allowed to dry before incubating plates for 18-24 h at 37°C. Strains were deemed sensitive when a clear zone of inhibition was present surrounding MP1 spots, but absent around vehicle-only spots.

MP1 purification. *M. caseolyticus* strain 115 was cultured on 2xYT agar then harvested by scraping plates into 0.9% NaCl. Harvested cells were pelleted in 50-ml Oak Ridge centrifuge tubes (Nalgene), weighed, then resuspended in 7 M urea (4 ml per 1-g cell pellet). The suspension was allowed to incubate at 25°C for 5 minutes, with intermittent vortexing, followed by centrifugation at 12,000 x *g* for 10 min. MP1-containing supernatant was transferred to a new tube and re-centrifuged to remove remaining particulate material. The final supernatant was heat-treated to kill residual cells (65°C, 20 min) and then tested for activity by spot-on-lawn assay.

Urea-extracted MP1 was further purified using a Sep-Pak® C18 Plus Long cartridge (Waters). The cartridge was equilibrated with 50% then 20% acetonitrile (ACN). Approximately 30 ml of extract was applied to the cartridge before washing with 30% ACN. Elution was accomplished with 50% ACN. All solvents were acidified with 0.1% formic acid (FA).

Final purification was performed using reversed-phase high performance liquid chromatography (RP-HPLC) on an Agilent 300SB-C_8 column. MP1 was eluted isocratically using 35% ACN-0.1% FA (v/v) over 16 min at a flow rate of 0.8 ml/min. The active fraction (retention time = 10-12.5 min) was collected for further analysis. All purification steps were performed at room temperature. In parallel, the same steps were followed for pBac115-cured *M. caseolyticus* strain 115C. This isogenic control strain is incapable of antimicrobial compound production (see below).

Plasmid curing. To generate plasmid-cured derivatives, overnight cultures of *M. caseolyticus* 115 were diluted 4 x 10^4-fold into fresh tryptic soy broth (TSB) containing 0.004% SDS, and cultured with shaking (225 rpm) at 45°C overnight. Cultures were plated on tryptic soy agar (TSA) and incubated for 18-24 h at 37°C. Heat/SDS-surviving (HS) strains were tested by flanking patch assay for MP1 production and sensitivity. The plasmid profiles of HS strains were evaluated by standard alkaline lysis miniprep after treatment of cells with lysostaphin (Sigma) at a concentration of 120 μg/ml for 30 min at 37°C. The resulting DNA was analyzed on a 1% agarose gel.

Plasmid sequencing and annotation. Plasmid minipreps of strain 115 were generated as described above and submitted for 454 shotgun sequencing (Roche) at the Brigham Young University DNA Sequencing Center. Sequence assembly was accomplished using GS *De Novo*

Assembler (Roche) with localized PCR and Sanger sequencing implemented to resolve ambiguous areas. Contigs representing portions of pBac115 were identified based on successful PCR amplification from strain 115 and no amplification from pBac115-cured strain 115C. These tests are summarized in Table 2-S2. Annotation of pBac115 was accomplished using GeneMark.hmm prokaryotic gene prediction software (112) to find open reading frames (ORFs), and blast analysis (http://www.ncbi.nlm.nih.gov/blast) to assign predicted functions. We confirmed the sequence of the MP1 biosynthetic gene cluster by re-sequencing the entire region using Sanger sequencing.

MS and NMR Analysis. The identity of purified MP1 was determined by a combination of high-resolution electrospray ionization mass spectrometry (ESI-MS) and high-resolution nuclear magnetic resonance (NMR) spectroscopy. ESI-MS was performed at the MS Core Facility (Brigham Young University) using an Agilent MSD-ToF instrument with electrospray ionization. Theoretical MP1 masses were calculated using mMass (113). 1D and 2D NMR spectra were collected on a Varian NMR-System 500 MHz spectrometer equipped with a OneNMR probe and ProTune (NMR Facility, Brigham Young University). All NMR experiments are summarized in Table 2-S3. The sample used for NMR analysis was composed of approximately 14 mg MP1 dissolved in 0.8 ml of DMSO-d6 in a 5 mm NMR tube (~10 mM). Both the ^{13}C and ^{15}N spectra were acquired at natural abundance. The spectra were analyzed using VNMRJ 3.2 software.

Genetic manipulation of *Bacillus subtilis.* Plasmid pRB014 (Figure 2-S1A) was used to generate the *B. subtilis* strain expressing *tclQ* in *trans* at the *amyE* locus. A similarly constructed plasmid (pRB015, Figure 2-S1B), was used for introducing the *tclWX* cassette to the *amyE* locus. These preliminary experiments were performed in *B. subtilis* strain SCK6, which is modified for

xylose-inducible competence. The more definitive allelic exchange strains depicted in Figure 2-6 were created in *B. subtilis* strain 168. Strains PB213, PB230, PB214, and PB215 resulted from *rplK* allelic replacement using plasmids pRB029, pRB031, pRB030, and pRB024, respectively. Full plasmid sequences are provided in Figure 2-S1. The allelic replacements were designed to maintain the native overall transcriptional unit so that the expression of introduced alleles and downstream genes would be left unperturbed.

Plasmids were constructed using standard techniques with enzymes purchased from New England Biolabs. The high-fidelity polymerases Q5 (New England Biolabs) and *Pfx50* (Invitrogen) were used for insert amplification. All constructs and genetic insertions/replacements were confirmed by Sanger sequencing. Custom oligonucleotides were purchased from Invitrogen.

Spot dilution assays. Overnight liquid cultures of *B. subtilis* mutants were normalized to an OD_{600} of 0.1 before performing 10-fold serial dilutions. 5-µl spots were made on LB agar supplemented with chloramphenicol (5 µg/ml), with or without MP1 (18 µg/ml). Spots were allowed to dry before incubating plates at 37°C for 24 h.

Sequence information. The complete sequence of pBac115 has been deposited in the Genbank database (http://www.ncbi.nlm.nih.gov/genbank/) under the accession number KM613043.

2.4 Results

***M. caseolyticus* 115 produces a potent, broad-spectrum antimicrobial compound.** In order to test the target range of strain 115, we devised a flanking patch assay (Figure 2-2A), which we used on a broad panel of bacterial isolates in our possession. We observed that strain 115 secretes a diffusible compound that inhibits the growth of diverse gram-positive bacteria (Table 2-1). Of particular interest was the observation that many clinically relevant pathogens were inhibited, including vancomycin-resistant *Enterococcus faecalis* (VRE), methicillin-resistant *S. aureus* (MRSA), *Clostridium difficile, Streptococcus agalactiae, Bacillus anthracis, Listeria monocytogenes* and *S. pyogenes*. Interestingly, none of the gram-negative bacteria tested were sensitive. This activity profile is consistent with that of many known thiopeptides (74).

The antimicrobial activity of *M. caseolyticus* 115 is plasmid-encoded. Earlier speculation suggested that antimicrobial production by strain 115 may be plasmid-mediated (103). We investigated this hypothesis by screening for derivatives of *M. caseolyticus* 115 that lose antimicrobial production under conditions that promote plasmid curing. Following growth at elevated temperatures and in the presence of low concentrations of SDS, we successfully screened for heat/SDS surviving derivatives (HS strains) that had lost the ability to inhibit the growth of *B. subtilis* 168. The loss of antimicrobial production always coincided with loss of resistance to the compound (see strain 115C in Figure 2-2). Plasmid preparations from HS strains revealed that derivatives deficient in both the production and immunity phenotypes were devoid of a high molecular weight plasmid which we have named pBac115 (Figure 2-2B). We were able to replicate this phenomenon on repeated occasions; in every instance the loss of antimicrobial production and immunity was perfectly correlated with the absence of pBac115.

Table 2-1. Inhibitory spectrum of strain 115

Species	No. of Strains	
	Sensitive	**Resistant**
Gram-positive bacteria		
Staphylococcus spp.		
S. aureus	2	0
S. aureus (MRSA)	6	0
S. epidermidis	2	0
Enterococcus spp.		
E. faecalis	1	0
E. faecalis (VRE)	1	0
Streptococcus spp.		
S. agalactiae	1	0
S. pyogenes	2	0
S. pneumoniae	1	0
Bacillus spp.		
B. anthracis	1	0
B. brevis	1	0
B. cereus	0	1
B. licheniformis	1	0
B. megaterium	1	0
B. mycoides	1	0
B. sphaericus	1	0
B. subtilis	4	1
B. thuringiensis	2	0
Clostridium spp.		
C. difficile	1	0
C. perfringens	1	0
Listeria monocytogenes	2	0
Paenibacillus polymyxa	1	0
Gram-negative bacteria		
Escherichia coli	0	2
Yersinia pseudotuberculosis	0	1
Burkholderia thailandensis	0	1
Acid fast bacteria		
Mycobacterium smegmatis	1	0

[a]Detailed strain information is given in Table 2-S1

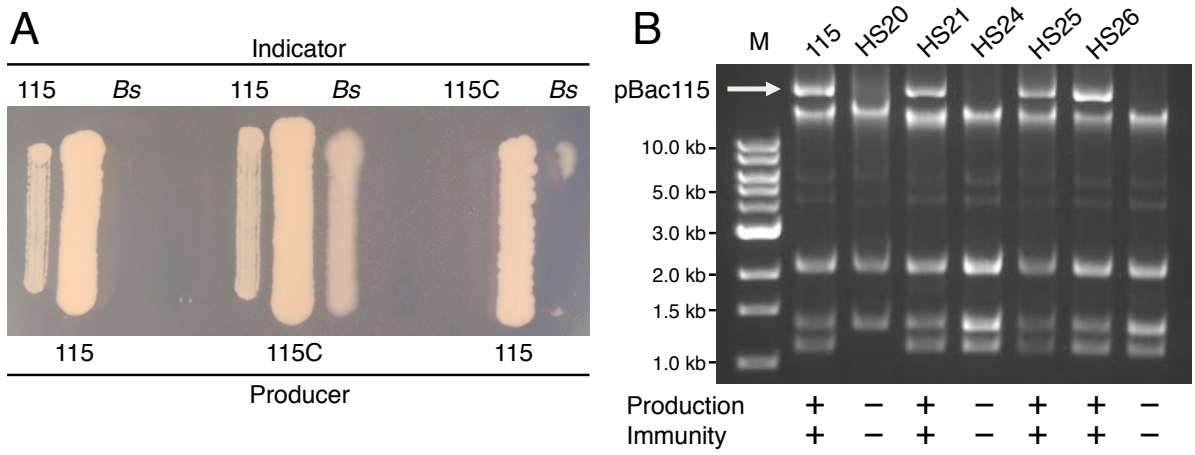


Figure 2-2. *M. caseolyticus* strain 115 antimicrobial activity is plasmid-encoded. (A) Examples of flanking patch assays for production and immunity. In each triplet the central patch was applied to the agar 18 hours prior to the flanking patches. 115C is a pBac115-cured derivative. Bs corresponds to B. subtilis 168, a susceptible indicator strain. (B) Plasmid preparations from M. caseolyticus 115 and heat/SDS (HS)-surviving derivatives.

pBac115 contains thiopeptide biosynthetic genes. In order to identify the plasmid-encoded biosynthetic genes, we sequenced DNA from an alkaline-lysis plasmid preparation derived from *M. caseolyticus* 115. This preparation appeared to contain at least three low molecular weight plasmids, as well as pBac115. A second high molecular weight band is attributed to chromosomal DNA fragments (see Figure 2-2B). The resulting 454 pyrosequencing data allowed us to discern a circular 24-kb plasmid consistent with the properties of pBac115. First, in our panel of HS strains (see Figure 2-2B), strains harboring the pBac115 band also test positive for the 24-kb molecule by PCR, whereas strains lacking the pBac115 band test negative for the 24-kb molecule. Second, annotation of the 24-kb plasmid reveals 27 putative open reading frames (ORFs), including an 11 gene cluster predicted to encode thiopeptide biosynthetic functions (*tcl* genes; see Figure 2-3A).The *tcl* designations are according to (66). A detailed report of predicted pBac115 gene functions is given in Table 2-S4.

Thiopeptide biosynthesis consists of extensive post-translational tailoring of a structural peptide by multiple biosynthetic enzymes. Although the biosynthesis of thiopeptides is incompletely understood, thiopeptide maturation is thought to occur on a precursor peptide scaffold from which the mature modified peptide is cleaved by a final macrocyclization reaction (114). Peptide modification steps include dehydration of Ser and Thr residues, Cys cyclization to give thiazolyl groups, and the joining of Dha pairs to form the central macrocycle, consequently cleaving the leader peptide (17, 52). On pBac115, the short ORF *tclE* drew our attention as potentially encoding a precursor peptide (Figure 2-3B) because the C-terminal 14 amino acids are identical to the core peptide from which a certain family of thiopeptides (thiocillins) are derived in other organisms.

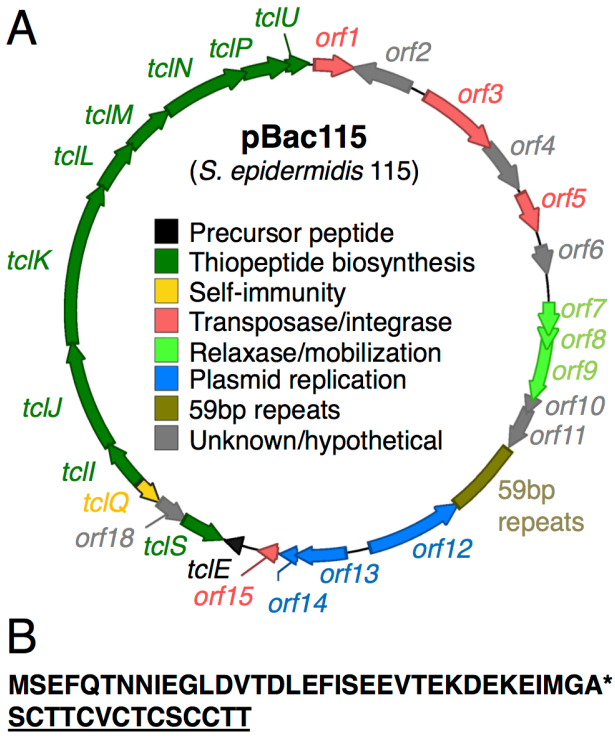
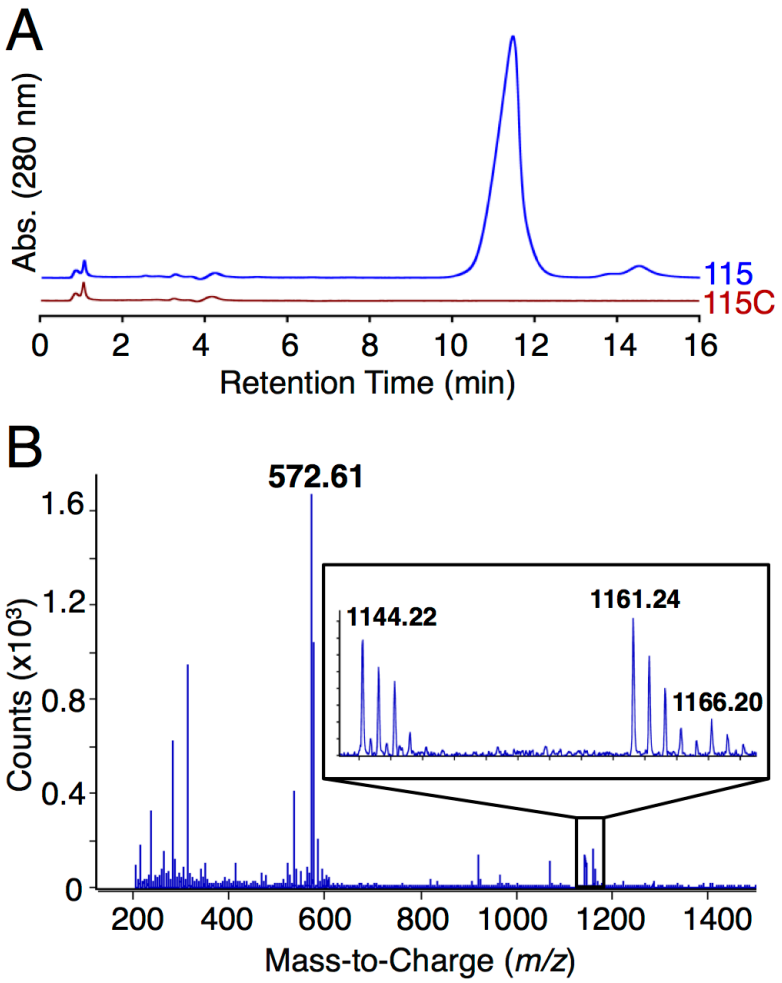


Figure 2-3. pBac115 contains thiopeptide biosynthetic (*tcl*) genes. (A) Plasmid map of pBac115 with annotation based upon results from blastx and direct alignments with previously identified *tcl* genes. (B) Amino acid sequence of the TclE precursor peptide. Asterisk (*) indicates the predicted cleavage site; the core peptide, which undergoes modification to produce the mature thiopeptide, is underlined.

Purification of the antimicrobial activity of *M. caseolyticus* 115. According to previous reports, the antimicrobial compound produced by strain 115 can be extracted by suspending cell pellets in 7M urea (103). We used this approach to obtain high yields of urea-extracted compound with high biological activity. We further enriched this extract by reversed-phase cartridge chromatography. We hypothesized that this purified activity would be found in strain 115 but not the pBac115-cured strain, 115C. To test this, we performed parallel purification procedures using both strains. As expected, the 115 extract retained antimicrobial activity in a spot-on-lawn assay even when diluted 1000-fold, whereas the 115C sample exhibited no detectable activity. We analyzed the 115 and 115C samples by HPLC and observed a very prominent peak present in the 115 sample that was absent in the 115C sample (Figure 2-4A). In spot-on-lawn assays, fractions corresponding to this peak exhibited antimicrobial activity, while all other fractions were devoid of activity.

Structural analysis identifies the unknown compound as Micrococcin P1. The antimicrobial compound was analyzed by mass spectrometry (MS) and nuclear magnetic resonance (NMR). Using high-resolution electrospray-ionization (ESI)-MS, we observed dominant peaks at m/z = 572.61, 1144.22, 1161.24 and 1166.20 (Figure 2-4B). These peaks are remarkably consistent with the thiopeptide micrococcin P1 (MP1) in various ionization states: the dually protonated form (572.61), the singly protonated form (1144.22), the ammonium adduct (1161.24) and the sodium adduct (1166.20; Figure 2-4C). MP1, also produced by *Bacillus cereus* strain ATCC 14579 (66), is known to be synthesized on the peptide scaffold SCTTCVCTCSCCTT, which we noted above is encoded by *tclE* on pBac115. The pBac115-dependent compound was finally analyzed by NMR, which confirmed its identity as MP1 (Tables 2-S3 and 2-S5, Figure 2-S2). The structure of MP1 is shown in Figure 2-1A.



C

MP1 Predicted Mass	Observed Mass
1144.2173 [MP1 + H]	1144.2175
572.6123 [MP1 + 2H]	572.6100
1161.2439 [MP1 + NH_4]	1161.2409
1166.1993 [MP1 + Na]	1166.1966

Figure 2-4. Biochemical analysis of M. caseolyticus 115 antimicrobial activity. (A) RP-HPLC analysis of purified material from strain 115 and material similarly prepared from pBac115-cured strain 115C. (B) ESI-MS fragmentation of the major HPLC peak from strain 115. Inset: close-up of the region from *m/z* = 1142-1168. (C) Comparison of observed *m/z* values with those predicted for MP1.

The plasmid pBac115 harbors only a single immunity gene. Having identified the antimicrobial compound known as MP1, we compared the putative MP1-encoding gene cluster on pBac115 with the only other empirically tested MP1 biosynthetic gene cluster; the *tcl* cluster from *B. cereus* ATCC 14579 (66). Comprised of 24 *tcl* genes within a 22-kb region of the *B. cereus* chromosome, the *tcl* cluster reportedly gives rise to 8 different thiopeptides (including MP1) from the SCTTCVCTCSCCTT scaffold. Genetic comparisons between pBac115 and the *B. cereus* gene cluster identified 11 genes clustered within an 11-kb region of pBac115 with homology to *tcl* genes (Figure 2-5, see also Table 2-S4). Of these 11, only one copy of the precursor peptide was identified (*tclE*) contrasted with the four tandem copies (*tclE-H*) in *B. cereus*. Additionally, of the four genes presumed to encode immunity functions in *B. cereus* (*tclQ, tclT, tclW* and *tclX*), the pBac115 cluster contains only a single gene; a homolog of *tclQ* and *tclT* (*tclQ* and *tclT* are identical copies of the same gene in *B. cereus*). This observation led us to hypothesize that a single gene (hereafter referred to as *tclQ*) may be sufficient to confer self-immunity in *M. caseolyticus* 115.

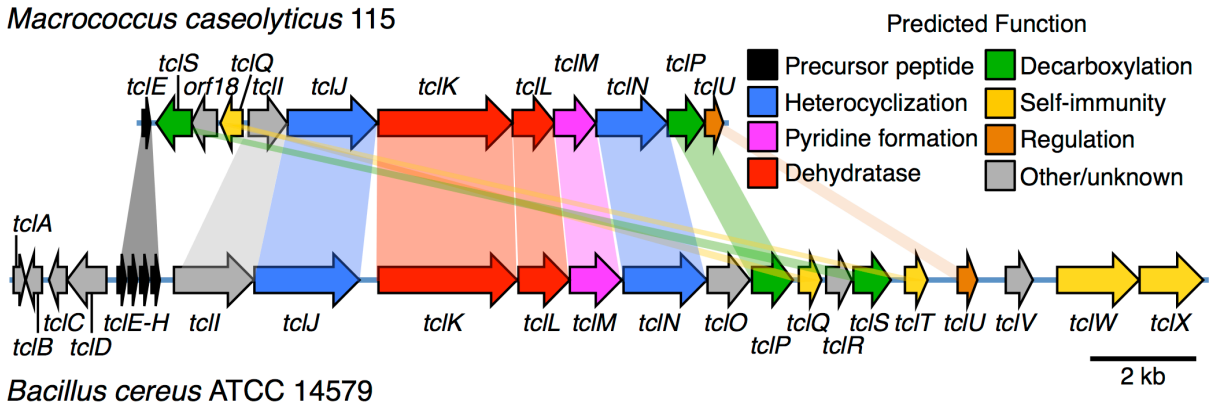


Figure 2-5. Comparisons of the *tcl* gene clusters from *M. caseolyticus* 115 and *B. cereus* ATCC 14579. Connectors indicate homologous genes based on blastp comparisons of their predicted protein products. Predicted functions are given in the color-coded key.

TclQ confers resistance to MP1 in *Bacillus subtilis.* TclQ is a homolog of the 50S ribosomal protein L11, which is located near the ribosomal GTPase center (79, 115) and is known to be involved in the MP1 mechanism of action (80). As such, we hypothesized that TclQ confers immunity by replacing the native L11 to protect ribosomes from MP1. To test our "target replacement" hypothesis, we evaluated whether TclQ could confer MP1 resistance in the model *Bacillus subtilis* strain 168. This strain is sensitive to MP1 and we have determined that the ribosomal protein L11 is a major target, since spontaneous resistance is invariably associated with mutations in the L11-encoding gene *rplK* (data not shown). In multiple independent MP1-resistant isolates, we see the same *rplK* mutation that results in deletion of Pro-22 (referred to hereafter as the Δ*P22* allele).

As an initial test of our target replacement hypothesis, the *tclQ* gene from pBac115 was expressed in *trans* from its native promoter at the *B. subtilis amyE* locus. This strain was modestly (10-fold) more resistant to MP1 than the wild type. In a similar experiment, *tclWX*, the putative efflux system from *B. cereus*, did not confer detectable resistance (data not shown). We reasoned that TclQ expression in *trans* may not confer very high resistance to *B. subtilis* due to competition for ribosomal incorporation with the native L11 protein. According to this logic, replacement of *rplK* (the L11-encoding gene) with *tclQ* would lead to much higher levels of resistance, as competition for ribosomal incorporation is eliminated. As shown in Figure 2-6, replacement of *rplK* with *tclQ* allows near-normal growth and high-level resistance to MP1. MP1 dilution experiments show that the replacement strain was ≥200-fold resistant relative to the wild type. This increased resistance suggests that TclQ substitutes for L11 on the ribosome, but we could not rule out the possibility that removal of L11 would lead to high resistance regardless of what replaces it. To address this, we constructed an isogenic strain in which the *rplK* ORF was

simply removed from the chromosome and not replaced with *tclQ*. This strain was viable, but very slow growing. It was, however, noticeably resistant to MP1 (see Figure 2-6B). The robust growth observed for the *tclQ* replacement strain shows that *tclQ* functionally complements the *rplK* mutant for normal protein synthesis and growth. This strongly suggests that the TclQ protein incorporates into the ribosome in place of L11, while also conferring resistance to MP1.

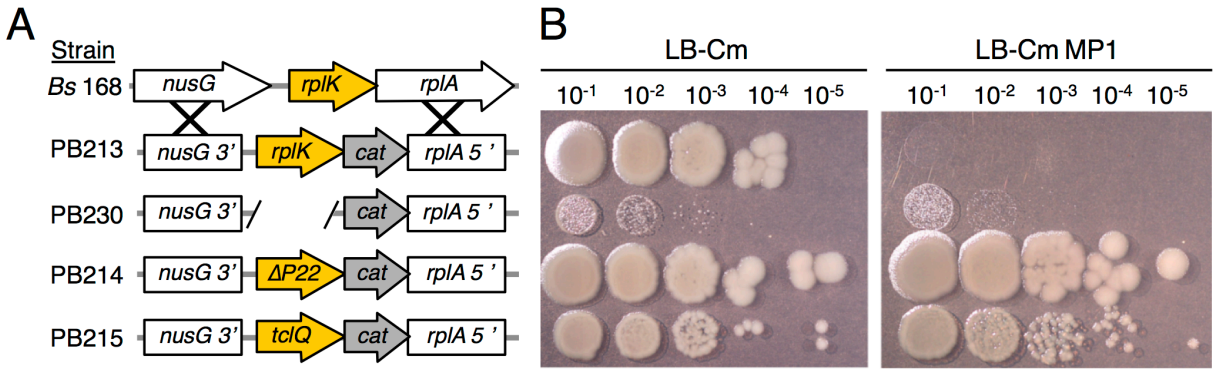


Figure 2-6. Allelic exchange experiments at the L11-encoding *rplK* locus. (A) Map of *rplK* alterations on the *B. subtilis* 168 chromosome, with strain names indicated. Δ*P22* is an abbreviation for *B. subtilis rplK*Δ*P22* allele that is known to give resistance to MP1. The *tclQ* allele was amplified from *M. caseolyticus* 115. (B) Spot dilution assays showing growth at 24 hrs on LB-agar with or without MP1 (18 µg/ml); rows correspond to genotypes indicated in part A.

2.5 Discussion

We have shown that *M. caseolyticus* 115 produces the broad-spectrum antibiotic MP1. This is the first reported instance of MP1 being produced by a member of this species and the second reported example of thiopeptide production among members of the family Staphylococcaceae (116). *M. caseolyticus* is an animal-derived bacterial isolate, whereas the majority of thiopeptide producing bacteria isolated to date were derived from environmental samples (52). In their early work, Jensen and colleagues (102-104) demonstrated that strain 115 confers protective benefits to host animals, suggesting that thiopeptide-producing microbiota may play an important role in shaping microbial communities in animal hosts.

We have also provided evidence for a remarkably simple yet efficient mechanism of self-immunity in *M. caseolyticus* 115: TclQ replaces L11 at the ribosome and prevents the action of MP1. In the annotation of the thiopeptide gene cluster from *B. cereus*, the authors predicted that *tclQ* as well as *tclWX* might contribute to self-immunity (66). In our *B. subtilis* bioassay, we did not observe *tclWX*-based immunity. With this observation, and the fact that *tclWX* is not found in the *M. caseolyticus* MP1 gene cluster, it appears that *tclQ* may provide the dominant mechanism of self-immunity in both organisms.

The plasmid-encoded nature of the *M. caseolyticus* MP1 biosynthetic gene cluster is also of interest. Beyond pBac115, only one other plasmid encoded thiopeptide gene cluster has been reported to date (53). pBac115 shares strong sequence similarity with plasmid pMCCL1 from *Macrococcus caseolyticus* JCSC5402 (117). Both plasmids share more than 94% sequence identity through an 8.7-kb stretch of DNA that includes replication- and mobilization-associated genes. However, all *tcl* genes on pBac115 are absent from pMCCL1, suggesting they are gained and lost as a single functional unit.

With few exceptions, the biosynthetic functions of many thiopeptide-modifying enzymes remain largely undemonstrated. The *B. cereus* TclM enzyme has been implicated in what is thought to be the culminating macrocyclization step of thiopeptide maturation (118). In a few other cases, researchers have modified the gene encoding the precursor peptide and detected resultant biochemical changes in the mature thiopeptide, including expansion of macrocyclic ring size (67-69). The *tcl* gene cluster on pBac115 provides a unique platform for continuing these kinds of studies because it is half the size of the homologous gene cluster from *B. cereus*, and it controls the biosynthesis of a single thiopeptide product.

Access to a simpler thiopeptide gene cluster may also facilitate pharmaceutical development of these potent antibiotics. At present, poor pharmacokinetics and poor aqueous solubility have limited the translation of thiopeptide technology from laboratory to clinic. Thiostrepton and nosiheptide remain the only successfully commercialized thiopeptides to date, both used in veterinary applications (52). Some attempts have been made to improve thiopeptide solubility and generate synthetic derivatives with improved pharmacokinetics (119-121), but this has been difficult. The discovery of this streamlined MP1 gene cluster opens the door to effectively address these issues and improve synthetic strategies.

MP1 is considered the optimal platform to explore the medicinal chemistry of thiopeptides for a number of reasons: its mechanism of action has been well addressed, it is produced by a variety of bacterial species, its structure is relatively simple compared with other thiopeptides, yet it retains all of the attractive biochemical properties (potency, stability etc.) of this chemical family (74). Furthermore, the availability of a simple system of self-immunity, which we have demonstrated here, enables further development using model bacterial hosts such as *B. subtilis*.

2.6 Acknowledgements

We thank Marcus Jensen for providing *M. caseolyticus* strain 115 and for helpful insights into earlier work. We also wish to acknowledge technical assistance provided by Jared Balaich, Joseph Thiriot, Cody Ashcroft, Joann Diray-Arce, Brendan Coutu, Andrew Mathis, Jordon March and Kaylee Bennallack. We sincerely appreciate Petra Levin for her generous assistance with *Bacillus* genetics as well as Jiping Zou for his assistance with HPLC. This work was supported by a grant from BYU College of Life Sciences Vaccine Royalties Fund.

2.7 Supplemental Material

Supplemental Material for:

Characterization of a novel plasmid-borne thiopeptide gene cluster in *Macrococcus caseolyticus* strain 115

Philip R. Bennallack[1], Scott R. Burt[2], Michael J. Heder[1], Richard A. Robison[1*], and Joel S. Griffitts[1*]

[1]Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT 84602, USA

[2]Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT 84602, USA

Contents:

SUPPLEMENTAL METHODS

I. Transformation of *Bacillus subtilis* 168 by starvation-induced competence

- modified from (122)

a) Media preparation

Starvation medium 1 (SM1) contains 0.2% ammonium sulfate, 1.2% dipostassium hydrogen phosphate, 0.6% potassium dihydrogen phosphate, 0.1% sodium citrate dihydrate, 0.2% magnesium sulfate heptahydrate, 0.2% yeast extract, 0.025% casamino acids, 0.01% L-tryptophan and 0.5% glucose.

Starvation medium 2 (SM2) differs slightly from SM1, containing 0.2% ammonium sulfate, 1.2% dipostassium hydrogen phosphate, 0.6% potassium dihydrogen phosphate, 0.1% sodium citrate dihydrate, 0.8% magnesium sulfate heptahydrate, 0.1% yeast extract, 0.0125% casamino acids, 0.01% L-tryptophan, 0.5% glucose and 0.022% calcium chloride dihydrate.

Both solutions were mixed fresh on the day of transformation from stock solutions prepared ahead of time. The ammonium sulfate, dipostassium hydrogen phosphate, potassium dihydrogen phosphate and sodium citrate dihydrate can be combined and autoclaved to provide the base (ST Base) for both media. All other solutions were maintained as separate stocks, sterilized by autoclaving (magnesium sulfate heptahydrate and calcium chloride) or filter sterilization (yeast extract, casamino acids, tryptophan and glucose).

b) Preparation of competent cells

Pure cultures of *B. subtilis* 168 cells were grown on LB agar without antibiotics, then used to inoculate a 4-ml liquid culture of LB. Cells were grown overnight at 37°C, 225rpm for 14-18 h. 1 ml of overnight culture was combined with 15 ml of warm, freshly prepared SM1 (final OD600 ≈ 0.4-0.6) and incubated at 37°C, 225rpm for 5 h. Following incubation in SM1, an equal volume of pre-warmed SM2 was added and cells were further incubated for 2 h under the same conditions. At this stage and for at least 60-90 min after, cells are competent and are ready for transformation. Cells may be combined with glycerol (final conc. 10%) and frozen at -80^{o}C at this stage, however subsequent transformations may be at least two-fold less efficient than those performed with fresh competent cells.

c) Transformation of competent cells

500 µl of competent *B. subtilis* cells were combined with 100-500 ng plasmid DNA (<5 µl volume) and incubated at 37°C (with rotation) for 30 min. 200 µl of fresh LB was then added to cells and incubated for an additional 60 min. Following this final incubation step, the desired volume of cells was plated on antibiotic-containing LB (LB-Cm_5) and grown overnight at 37°C. Transformation efficiencies in the range of 2 x 10^4-2 x 10^5 transformants per µg DNA are routinely observed using this procedure.

I. Transformation of *Bacillus subtilis* SCK6 by xylose induced competence

B. subtilis strain SCK6 (123) harbors a Pxyl-*comK* insertion for rapid induction of competence in the presence of xylose. Strains were transformed by diluting overnight cultures of strain SCK6 into LB-xyl (1% w/v) for 2 hours at 37°C to induce competence. At this stage cells were mixed with 100-500 ng plasmid DNA and incubated for 60 min prior to selection. Following this final incubation step, the desired volume of cells was plated on antibiotic-containing LB (LB-Cm_5) and grown overnight at 37°C. High efficiency transformations are readily obtained in this manner.

LIST OF SUPPLEMENTAL TABLES AND FIGURES

SUPPLEMENTAL TABLES 2-S1 – S5

Table 2-S1. Strains and assay conditions used in inhibition studies

Indicator (Test) Strain	Gram Stain	Assay Method[a]	Sensitive	Culture Media[b]
B. anthracis Sterne strain 1043	Positive	Patch	+	TSA
B. brevis ATCC 8246	Positive	Spot	+	NB
B. cereus ATCC 14579	Positive	Patch, spot	-	LB, TSA
B. licheniformis ATCC 14580	Positive	Spot	+	NB
B. megaterium ATCC 15127	Positive	Patch	+	LB, TSA
B. mycoides ATCC 6462	Positive	Spot	+	NB
B. sphaericus ATCC 14577	Positive	Spot	+	NB
B. subtilis ATCC 19659	Positive	Spot	+	LB,TSA
B. subtilis ATCC 6051	Positive	Spot	+	LB
B. subtilis SCK6	Positive	Spot	+	LB
B. subtilis subsp. *spizizenii* ATCC 6633	Positive	Patch, spot	-	NB
B. subtilis 168	Positive	Spot	+	LB
B. thuringiensis ATCC 19269	Positive	Spot	+	NB
B. thuringiensis Al Hakam	Positive	Spot	+	NB
C. difficile ATCC 43598	Positive	Spot	+	LB
C. perfringens ATCC 13124	Positive	Spot	+	TSA w/ blood
E. faecalis ATCC 19433	Positive	Spot	+	LB
E. faecalis clinical isolate RL112105 (VRE)	Positive	Patch	+	TSA
L. *monocytogenes* ATCC 13932	Positive	Spot	+	LB,TSA
L. monocytogenes ATCC 15313	Positive	Spot	+	LB,TSA
P. polymyxa ATCC 842	Positive	Spot	+	NB
S. agalactiae ATCC 12386	Positive	Spot	+	TSA w/ blood
S. aureus ATCC 29213	Positive	Spot	+	LB
S. aureus ATCC 6538	Positive	Spot	+	LB
S. aureus ATCC 43300 (MRSA)	Positive	Spot	+	LB
S. aureus environmental isolate RL1 (MRSA)	Positive	Spot	+	LB
S. aureus environmental isolate RL2 (MRSA)	Positive	Spot	+	LB
S. aureus environmental isolate RL3 (MRSA)	Positive	Spot	+	LB
S. aureus environmental isolate RL4 (MRSA)	Positive	Spot	+	TSA
M. caseolyticus PB023	Positive	Spot	+	LB
S. pneumoniae ATCC 6303	Positive	Patch	+	LB
S. pyogenes RL421682	Positive	Patch	+	LB
S. pyogenes ATCC 51339	Positive	Patch	+	LB
B. thailandensis E135	Negative	Spot	-	LB
E. *coli* 0157:H7	Negative	Patch	-	TSA
Y. pseudotuberculosis RLPI399	Negative	Patch	-	LB
M. smegmatis ATCC 14468	Acid fast	Spot	+	7H11

[a] Patch, flanking patch assay; Spot, spot-on-lawn assay
[b] LB, Luria broth; NB, nutrient broth; TSA, tryptic soy agar; 7H11, mycobacteria 7H11 agar

Table 2-S2. Primer pairs used in assembly of pBac115[a]

Primer pair	Primer name	Direction	Sequence (5' - 3')	Expected product (bp)	Target
1	oPB009	Forward	GGAGTATTATACGCTGCCAGC	2068	Contig 1
	oPB010	Reverse	CCACAAGGAAATGCATCCGCA		
2	oPB011	Forward	GCTGACACAACTTCTCCTGG	2493	Contig 1
	oPB012	Reverse	ATTGGACCAGGAATGGGCAG		
3	oPB015	Forward	CTTGTGTGATCTCCGGCATTCC	2397	Contig 2
	oPB016	Reverse	TGCCTGGTGGCTCATCTTTG		
4	oPB019	Forward	TGAACGTGGAGATGTTGCTGG	1895	Contig 3
	oPB020	Reverse	CAGAATGGAGGGACGAACAAGTG		
5	oPB023	Forward	TTACAAGCAAGCGTAGCGAGC	659	Contig 4
	oPB024	Reverse	ATCGCTTGAGCCCTACTCTCC		
6	oPB027	Forward	CGAATGATGTGGTTATCGCAGG	776	Contig 5
	oPB028	Reverse	TCACTTCCATATCCAGCGTCG		

[a]Primer pairs 1, 2 and 4 amplified the expected product from strain 115 but not the pBac115-deficient strain 115C. All other primer pairs amplified the expected products from both strains.

Table 2-S3. NMR protocols

Protocol	**Key Parameters**
^{1}H	64 scans
^{13}C	20000 scans
gradient-filtered DQ-COSY	16 scans, 400 increments
z-filtered TOCSY	32 scans, 256 increments, 90 ms DIPSI-2 spinlock
adiabatic HSQC	16 scans, 96 increments, J1=146 Hz
band-selective adiabatic HSQC	32 scans, 128 increments, 5-80 ppm
band-selective adiabatic HSQC	32 scans, 128 increments, 105-150 ppm
adiabatic HSQC-TOXY	64 scans, 96 increments, 90 ms MLEV-17 spinlock
gradient-filtered adiabatic HMBC	16 scans, 96 increments, Jn=8 Hz
band-selective grad.-filtered adiab. HMBC	64 scans, 128 increments, Jn=12 Hz, 110-180 ppm
band-selective grad.-filtered adiab. HMBC	256 scans, 128 increments, Jn=3 Hz, 110-180 ppm
adiabatic ^{15}N-HSQC	128 scans, 96 increments, J1= 90 Hz
adiabatic ^{15}N-HSQC-TOXY	128 scans, 96 increments, 80 ms MLEV-17 spinlock

Table 2-S4. Summary of pBac115 blastp results

pBac115 Gene Product	Size (aa)	Thiocillin Protein Homolog[a]	Size (aa)	Identity/similarity /coverage %[b]	Top blastp Results	Predicted function[c]
Orf16 (TclE)	50	TclE	53	52/65/98 (entire product); 100/100/100 (core peptide)	none	Structural (precursor) peptide
Orf17 (TclS)	224	TclS	237	23/44/81	Short-chain dehydrogenase/reductase SDR [*Pseudomonas sp*. CFII64], oxidoreductase [*Pseudomonas syringae*], putative uncharacterized protein [*Megamonas funiformis* CAG:377]	Stochastic changes (decarboxylation) at residue 14(118)
Orf18	160	-	-	-	Hypothetical cyanophage protein [*Synechococcus* phage S-RSM4], putative regulator of cell autolysis [*Solitalea canadensis*], serine phosphatase [uncultured bacterium]	Unknown
Orf19 (TclQ)	141	TclQ/TclT	142	60/79/99	50S ribosomal protein L11 [*Leptolyngbya sp*. PCC 7376], 50S ribosomal protein L11 [*Nocardioides sp*. JS614], 50S ribosomal protein L11 [*Geitlerinema sp*. PCC 7105]	Self-immunity
Orf20 (TclI)	243	TclI	504	41/53/13	Bacterioferritin comigratory protein [*Lactobacillus gasseri*], thiazole-containing bacteriocin maturation protein [*Geobacillus stearothermophilus* NUB3621], bacteriocin maturation protein [*Bacillus sp*. UNC41MFS5]	Other modification
Orf21 (TclJ)	564	TclJ	661	35/50/87	Conserved hypothetical protein [*Lactobacillus gasseri*], SagD family biosynthesis docking scaffold protein [*Bacillus cereus*], bacteriocin biosynthesis protein SagD [*Alicyclobacillus pomorum*], cyclodehydratase [*Streptomyces sp*. NRRL 30471]	Heterocyclization (94), YcaO (DUF181) homolog(66)
Orf22 (TclK)	844	TclK	872	28/47/92	Lanthionine biosynthesis protein [*Bacillus cereus*], lantibiotic dehydratase, superfamily protein [*Lactobacillus gasseri*], NosE [*Streptomyces actuosus*], NocE [*Nocardia sp*. ATCC 202099]	Dehydratase (full length)(94)

Table 2-S4. (continued) Summary of pBac115 Blastp Results

pBac115 Gene Product	Size (aa)	Thiocillin Protein Homolog[a]	Size (aa)	Identity/similarity/coverage %[b]	Top blastp Results	Predicted function[c]
Orf23 (TclL)	268	TclL	324	28/45/95	Lantibiotic biosynthesis protein [*Lactobacillus gasseri*], lantibiotic biosynthesis protein [*Bacillus cereus*], thiopeptide-type bacteriocin biosynthesis domain-containing protein [*Enterococcus faecalis*]	Dehydratase (truncated)(94)
Orf24 (TclM)	265	TclM	326	32/52/79	Hypothetical protein [*Bacillus cereus*], hypothetical protein [*Lactobacillus gasseri*], hypothetical protein [*Actinoalloteichus spitiensis*]	Cycloaddition of Dha1 and Dha10 to produce pyridine(118)
Orf25 (TclN)	448	TclN	523	25/43/95	NADH oxidase [*Bacillus cereus*], conserved hypothetical protein [*Lactobacillus gasseri*], SagB-type dehydrogenase domain-containing protein [*B. cereus*]	Heterocyclization(94)McbC homolog(66)
Orf26 (TclP)	233	TclP	257	38/57/99	Short-chain dehydrogenase [*Bacillus cereus*], 3-oxoacyl-[acyl-carrier-protein] reductase FabG [*Clostridium saccharobutylicum*], hypothetical protein [*Saccharibacillus kuerlensis*]	Stochastic changes (decarboxylation) at residue 14(118)
Orf27 (TclU)	114	TclU	124	29/57/58	MULTISPECIES: transcriptional activator tipA [*Lactobacillus casei* group], transcriptional activator tipA [*Lactobacillus casei*], MerR family transcriptional regulator [*Lactobacillus casei*]	Regulation(94)

[a] based upon comparisons with *B. cereus* ATCC 14579
[b] results of directed blastp between translated products of pBac115 ORFS and *B. cereus* ATCC 14579 tcl genes
[c] numbers in parentheses correspond to references which can be found in the Supplemental References section

Table 2-S5. Chemical shifts of micrococcin P1 in DMSO-d6

Position	Residue	Type	δ ^{13}C (ppm)	δ ^{15}N (ppm)	δ ^{1}H (ppm)	Mult.	J (Hz)
1	Thr 1, γ	CH3	21.51	-	1.00	d	6.2
2	Thr 1, β	CH	65.55	-	3.69	m	-
2	Thr 1, OH	OH	-	-	4.63	d	br
3	[a]Thr 1, α	CH2	47.35	-	3.06	m	-
4	Thr 1, NH	NH	-	107.30	7.90	t	5.8
5	Thr 2, CO	C=O	164.79	-	-	-	-
6	Thr 2, α	C	131.07	-	-	-	-
7	[b]Thr 2, β	CH	128.31	-	6.50	q	7.0
8	[b]Thr 2, γ	CH3	13.99	-	1.68	d	7.0
9	Thr 2, NH	NH	-	117.20	9.51	s	-
10	[c]Cys 1, CO	C=O	159.44	-	-	-	-
11	Thiazole 1	C	150.90	-	-	-	-
12	Thiazole 1	CH	125.85	-	8.44	s	-
13	Thiazole 1	C	161.87	-	-	-	-
14	Thiazole 2	C	149.90	-	-	-	-
15	Thiazole 2	CH	122.14	-	8.57	s	-
16	Thiazole 2	C	168.81	-	-	-	-
17	Pyridine	C	150.21	-	-	-	-
18	Pyridine	CH	119.12	-	8.32	d	8.1
19	Pyridine	CH	141.14	-	8.43	d	8.1
20	Pyridine	C	129.60	-	-	-	-
21	Pyridine	C	151.53	-	-	-	-
22	Thiazole 3	C	152.95	-	-	-	-
23	Thiazole 3	CH	121.90	-	8.10	s	-
24	Thiazole 3	C	170.46	-	-	-	-
25	Thr 3, α	CH	56.59	-	5.06	dd	4.6, 8.8
26	Thr 3, β	CH	67.60	-	3.99	m	-
26	Thr 3, OH	OH	-	-	4.76	d	br
27	Thr 3, γ	CH3	21.07	-	1.02	d	6.4
28	Thr 3, NH	NH	-	115.40	8.21	d	8.9
29	[c]Cys 4, CO	C=O	160.91	-	-	-	-
30	Thiazole 4	C	149.56	-	-	-	-
31	Thiazole 4	CH	125.15	-	8.28	s	-
32	Thiazole 4	C	170.32	-	-	-	-
33	Val α	CH	55.85	-	5.12	dd	9.0, 9.0
34	Val β	CH	32.68	-	2.50	m	-

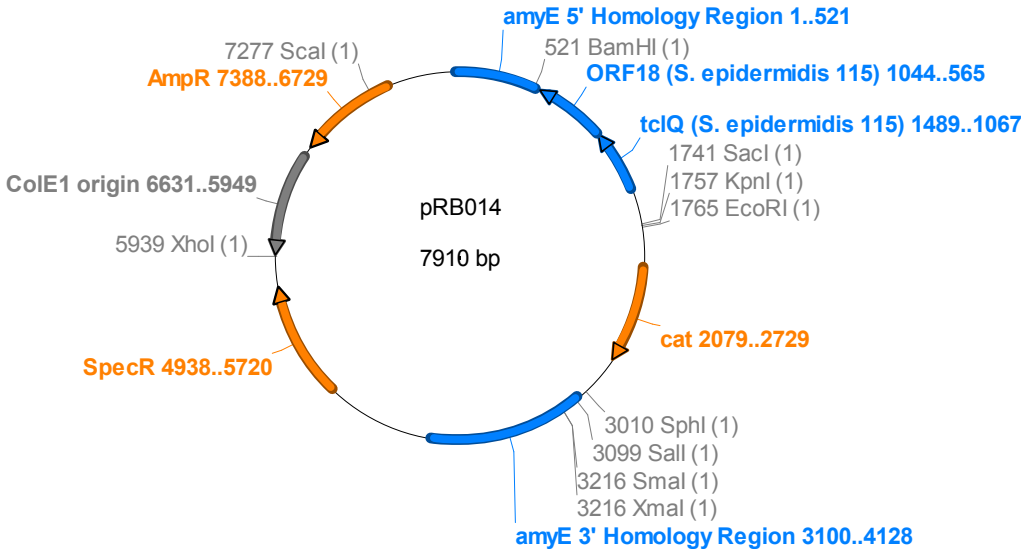
35	[d]Val γ	CH3	18.94	-	0.96	d	6.7
36	[d]Val γ	CH3	20.03	-	0.85	d	6.6
37	Val NH	NH	-	120.10	8.38	d	9.0
38	[c,e]Cys 5, CO	C=O	160.24	-	-	-	-
39	Thiazole 5	C	148.71	-	-	-	-
40	Thiazole 5	CH	124.92	-	8.19	s	-
41	Thiazole 5	C	166.81	-	-	-	-
42	Thr 4, α	C	130.09	-	-	-	-
43	[b]Thr 4, β	CH	129.07	-	6.45	q	6.9
44	[b]Thr 4, γ	CH3	14.24	-	1.74	d	6.8
45	Thr 4, NH	NH	-	120.40	9.51	s	-
46	Thr 5, CO	C=O	168.98	-	-	-	-
47	Thr 5, α	CH	58.09	-	4.67	dd	3.4, 8.0
48	Thr 5, β	CH	67.98	-	4.37	m	-
48	Thr 5, OH	OH	-	-	5.42	d	6.0
49	Thr 5, γ	CH3	20.31	-	1.36	d	6.3
50	Thr 5, NH	NH	-	109.30	7.85	d	7.9
51	[c,e]Cys 6, CO	C=O	160.27	-	-	-	-
52	Thiazole 6	C	150.12	-	-	-	-
53	Thiazole 6	CH	126.31	-	8.36	s	-
54	Thiazole 6	C	164.54	-	-	-	-

^{1}H and ^{13}C chemical shifts obtained from both 1D and 2D spectra; ^{15}N chemical shifts obtained from 2D spectra. Multiplicities and J values obtained from 1D ^{1}H spectrum. [a]Thr residue is decarboxylated. [b]Thr residue is dehydrated. [c]Remainder of the Cys is in a thiazole ring. [d]The assignment of carbons 35 and 36 is ambiguous. [e]The assignment of carbons 38 and 51 is ambiguous. *br* denotes a signal that was too broad to accurately measure the J value.

SUPPLEMENTAL FIGURES 2-S1 – S2

Figure 2-S1. Complete maps and sequences of plasmids used to generate *B. subtilis* mutants.

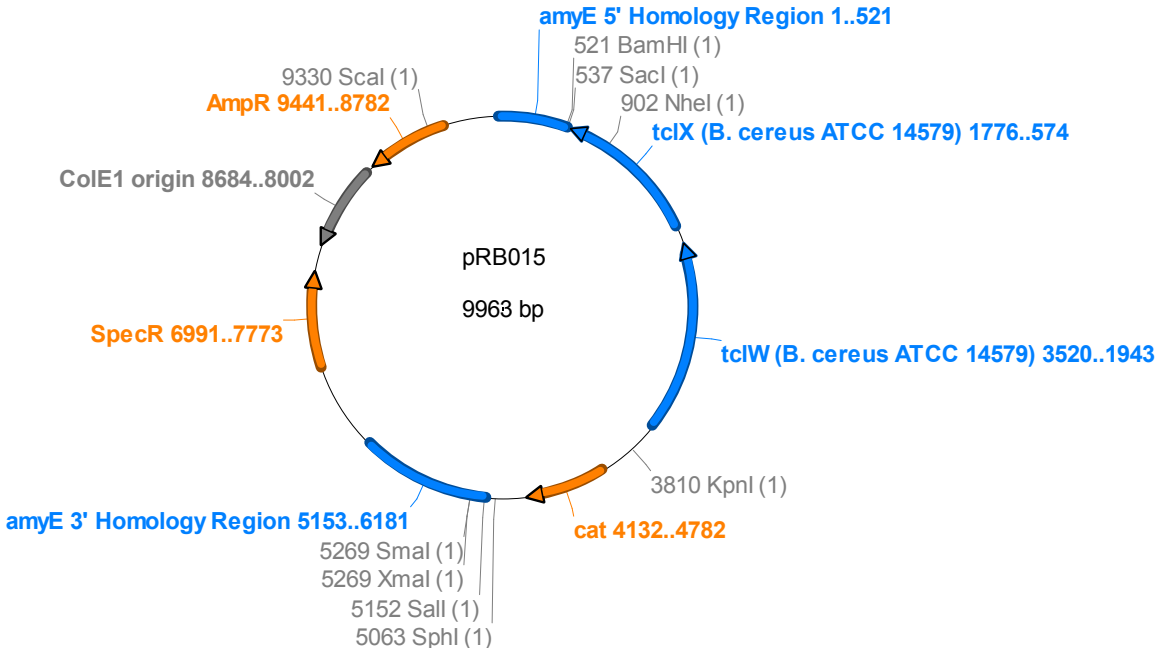
A Plasmid pRB014 used to generate *B. subtilis* SCK6 *amyE*::*tclQ-orf18-cat* mutant. The plasmid map shows critical features and unique restriction sites, where relevant. The complete nucleotide sequence is provided below with *tclQ* and *cat* shown in lowercase font.



ATGTTTGCAAAACGATTCAAAACCTCTTTACTGCCGTTATTCGCTGGATTTTTATTGCTGTTTCATTTGGTTCTGGCAGGACCGG
CGGCTGCGAGTGCTGAAACGGCGAACAAATCGAATGAGCTTACAGCACCGTCGATCAAAAGCGGAACCATTCTTCATGCATGGAA
TTGGTCGTTCAATACGTTAAAACACAATATGAAGGATATTCATGATGCAGGATATACAGCCATTCAGACATCTCCGATTAACCAA
GTAAAGGAAGGGAATCAAGGAGATAAAAGCATGTCGAACTGGTACTGGCTGTATCAGCCGACATCGTATCAAATTGGCAACCGTT
ACTTAGGTACTGAACAAGAATTTAAAGAAATGTGTGCAGCCGCTGAAGAATATGGCATAAAGGTCATTGTTGACGCGGTCATCAA
TCATACCACCAGTGATTATGCCGCGATTTCCAATGAGGTTAAGAGTATTCCAAACTGGACACATGGAAACACACAAATTAAAAAC
TGGTCTGATCGGATCCGTAAAGCAATAGGACTTGTTCCTCCATATATTATACAATTATCCATGATTATTATCCTCATAAATTGGA
TAGTTGATTTTACCATATCTAGAGTAGGTAGTAATTAAGTAATCTTCAATAAAAATGATATCTTTAAATAACTCCATCTGTGGAT
TGGTAATATTACATTCACTTAACAAATAATATACTAGTTCTGATATTTTTAAAATTATGTATCTGTGATGCCTAAAGGTAGATTG
TTTAAAATTTCCTATAATATTAAAAGCTATTTCAAGTATAAGTTCTTCGTAACTTTCACCTTGATAGTTAATTAAAAATTCGTTA
CAATTATTGGAATTTTTAACAAAAATAACTATTTTAATATCTGTCGATTCTTTTCTAATAGTCATCACTTCTACATCAAAATTCT
CGAACAAAGGGTTGCCATATATACATTCTGAGAAATCTAACGTTTCTACTTTATTCGACGAATCACACATACATTGATATTTAAT
CTCGTCAAAAATTTTTTTATACATAGGATATAAAGAAAGTAGAAGActaatcttctactttgactcccatatttctagcagtacc
ttcaatcattgacattgctgtttcaatagaacttgcattcagatcaggcatttttatttctgcaatctcttgtaaatctttcttt
gaaatacttcctacaatttctttgtttggaacaccagatcctttattgatatgtgctttctcttttattaattcagctgtgggag
gagtttttagtataaagttaaagcttcgatcttcatataccgtaatttctgcaggaataaccccagacttttctttagtctgagc
attatattgtgttacaaactccataatattaactcctgctgggcctaatgcagtacctaccggaggtgctggagaagcttttcca
gcttctaattgaatattaataatttgctttatagcttttgccatTATTCGTCACTCACTTTCAATAATTTTCTATTTAAATATAT
CCTAAAGAACTATACACTAGTGTAACTAAAAAATCAAGAAAAAGAAAGAGGGCTAATATGTATTTACTATTTAAAAAAGATACAT
TTTACATAAAAACACATAATGAAGGTATTTTATTCAAAAATAATTTTACAAATTTAGAAGTAAAAACTCATAAAAGTTACTATGT
CTTTGAAAACTTAATTGAGTATCTCAATGGTTCCTATACAGAGCTCAGTCTAGACTGGTACCTCGAATTCTCATGTTTGACAGCT
TATCATCGGCAATAGTTACCCTTATTATCAAGATAAGAAAGAAAAGGATTTTTCGCTACGCTCAAATCCTTTAAAAAAACACAAA
AGACCACATTTTTTAATGTGGTCTTTTATTCTTCAACTAAAGCACCCATTAGTTCAACAAACGAAAATTGGATAAAGTGGGATAT
TTTTAAAATATATATTTATGTTACAGTAATATTGACTTTTAAAAAAGGATTGATTCTAATGAAGAAAGCAGACAAGTAAGCCTCC
TAAATTCACTTTAGATAAAAATTTAGGAGGCATATCAAatgaactttaataaaattgatttagacaattggaagagaaaagagat
atttaatcattatttgaaccaacaaacgacttttagtataaccacagaaattgatattagtgttttataccgaaacataaaacaa
gaaggatataaattttaccctgcatttatttcttagtgacaagggtgataaactcaaatacagcttttagaactggttacaata
gcgacggagagttaggttattgggataagttagagccactttatacaatttttgatggtgtatctaaaacattctctggtatttg
gactcctgtaaagaatgacttcaaagagttttatgatttatacctttctgatgtagagaaatataatggttcggggaaattgttt
cccaaaacacctatacctgaaaatgctttttctctttctattattccatggacttcatttactgggtttaacttaaatatcaata
ataatagtaattaccttctacccattattacagcaggaaaattcattaataaaggtaattcaatatatttaccgctatctttaca
ggtacatcattctgtttgtgatggttatcatgcaggattgtttatgaactctattcaggaattgtcagataggcctaatgactgg

cttttataaTATGAGATAATGCCGACTGTACTTTTTACAGTCGGTTTTCTAATGTCACTAACCTGCCCCGTTAGTTGAAGAAGGT
TTTTATATTACAGCTCCAGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTAT
ATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCC
CCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACT
GGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACATGGATGAGCGATGATGATATCCGTTTAGGCTGGGCGGTGA
TAGCTTCTCGTTCAGGCAGTACGCCTCTTTTCTTTTCCAGACCTGAGGGAGGCGGAAATGGTGTGAGGTTCCCGGGGAAAAGCCA
AATAGGCGATCGCGGGAGTGCTTTATTTGAAGATCAGGCTATCACTGCGGTCAATAGATTTCACAATGTGATGGCTGGACAGCCT
GAGGAACTCTCGAACCCGAATGGAAACAACCAGATATTTATGAATCAGCGCGGCTCACATGGCGTTGTGCTGGCAAATGCAGGTT
CATCCTCTGTCTCTATCAATACGGCAACAAAATTGCCTGATGGCAGGTATGACAATAAAGCTGGAGCGGGTTCATTTCAAGTGAA
CGATGGTAAACTGACAGGCACGATCAATGCCAGGTCTGTAGCTGTGCTTTATCCTGATGATATTGCAAAAGCGCCTCATGTTTTC
CTTGAGAATTACAAAACAGGTGTAACACATTCTTTCAATGATCAACTGACGATTACCTTGCGTGCAGATGCGAATACAACAAAAG
CCGTTTATCAAATCAATAATGGACCAGACGACAGGCGTTTAAGGATGGAGATCAATTCACAATCGGAAAAGGAGATCCAATTTGG
CAAAACATACACCATCATGTTAAAAGGAACGAACAGTGATGGTGTAACGAGGACCGAGAAATACAGTTTTGTTAAAAGAGATCCA
GCGTCGGCCAAAACCATCGGCTATCAAAATCCGAATCATTGGAGCCAGGTAAATGCTTATATCTATAAACATGATGGGAGCCGAG
TAATTGAATTGACCGGATCTTGGCCTGGAAAACCAATGACTAAAAATGCAGACGGAATTTACACGCTGACGCTGCCTGCGGACAC
GGATACAACCAACGCAAAAGTGATTTTTAATAATGGCAGCGCCCAAGTGCCCGGTCAGAATCAGCCTGGCTTTGATTACGTGCTA
AATGGTTTATATAATGACTCGGGCTTAAGCGGTTCTCTTCCCCATTGAGGGCAAGGCTAGACGGGACTTACCGAAAGAAACCATC
AATGATGGTTTCTTTTTTGTTCATAAATCAGACAAAACTTTTCTCTTGCAAAAGTTTGTGAAGTGTTGCACAATATAAATGTGAA
ATACTTCACAAACAAAAAGACATCAAAGAGAAACATACCCTGCAAGGATGCTGATATTGTCTGCATTTGCGCCGGAGCAAACCAA
AAACCTGGTGAGACACGCCTTGAATTAGTAGAAAAGAACTTGAAGATTTTCAAAGGCATCGTTAGTGAAGTCATGGCGAGCGGAT
TTGACGGCATTTTCTTAGTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC
CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTA
ACTTCGATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCG
CCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTACTGAAGTGGATTTCTTTAAGCTAG
AGGATCGATCTGTATAATAAAGAATAATTATTAATCTGTAGACAAATTGTGAAAGGATGTACTTAAACGCTAACGGTCAGCTTTA
TTGAACAGTAATTTAAGTATATGTCCAATCTAGGGTAAGTAAATTGAGTATCAATATAAACTTTATATGAACATAATCAACGAGG
TGAAATCATGAGCAATTTGATTAACGGAAAAATACCAAATCAAGCGATTCAAACATTAAAAATCGTAAAAGATTTATTTGGAAGT
TCAATAGTTGGAGTATATCTATTTGGTTCAGCAGTAAATGGTGGTTTACGCATTAACAGCGATGTAGATGTTCTAGTCGTCGTGA
ATCATAGTTTACCTCAATTAACTCGAAAAAAACTAACAGAAAGACTAATGACTATATCAGGAAAGATTGGAAATACGGATTCTGT
TAGACCACTTGAAGTTACGGTTATAAATAGGAGTGAAGTTGTCCCTTGGCAATATCCTCCAAAAAGAGAATTTATATACGGTGAG
TGGCTCAGGGGTGAATTTGAGAATGGACAAATTCAGGAACCAAGCTATGATCCTGATTTGGCTATTGTTTTAGCACAAGCAAGAA
AGAATAGTATTTCTCTATTTGGTCCTGATTCTTCAAGTATACTTGTCTCCGTACCTTTGACAGATATTCGAAGAGCAATTAAGGA
TTCTTTGCCAGAACTAATTGAGGGGATAAAAGGTGATGAGCGTAATGTAATTTTAACCCTAGCTCGAATGTGGCAAACAGTGACT
ACTGGTGAAATTACCTCGAAAGATGTCGCTGCAGAATGGGCTATACCTCTTTTACCTAAAGAGCATGTAACTTTACTGGATATAG
CTAGAAAAGGCTATCGGGGAGAGTGTGATGATAAGTGGGAAGGACTATATTCAAAGGTGAAAGCACTCGTTAAGTATATGAAAAA
TTCTATAGAAACTTCTCTCAATTAGGCTAATTTTATTGCAATAACAGGTGCTTACTTTTCTGGAGTTCTTTAGCAAATTTTTTTA
TTAGCTGAACTTAGTATTAGTGGCCATACTCCTCCAATCCAAAGCTATTTAGAAAGATTACTATATCCTCAAACAGGCGGTAACC
GGCCTCTTCATCGGGAATGCGCGCGACCTTCAGCATCGCCGGCATGTCCCCCTGGCGGACGGGAAGTATCCAGCTCGAGGTCGGG
CCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCC
GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATAC
CTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCT
CCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT
AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAG
TTGATAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAA
AGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATG
AGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCC
CGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCT
CCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTA
TTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCGTGGT
GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAA
AAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGC
ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTAT
GCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGA
AAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT
CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGAC
ACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATA
TTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTA
TTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTAACAAAATTCTCCAGTCTTCACAT
CGGTTTGAAAGGAGGAAGCGGAAGAATGAAGTAAGAGGGATTTTTGACTCCGAAGTAAGTCTTCAAAAAATCAAATAAGGAGTGT
CAAGA

B Plasmid pRB015 used to generate *B. subtilis* SCK6 *amyE*::*tclWX-cat* mutant. The plasmid map shows critical features and unique restriction sites, where relevant. The complete nucleotide sequence is provided below with *tclW, tclX* and *cat* indicated in lowercase font.

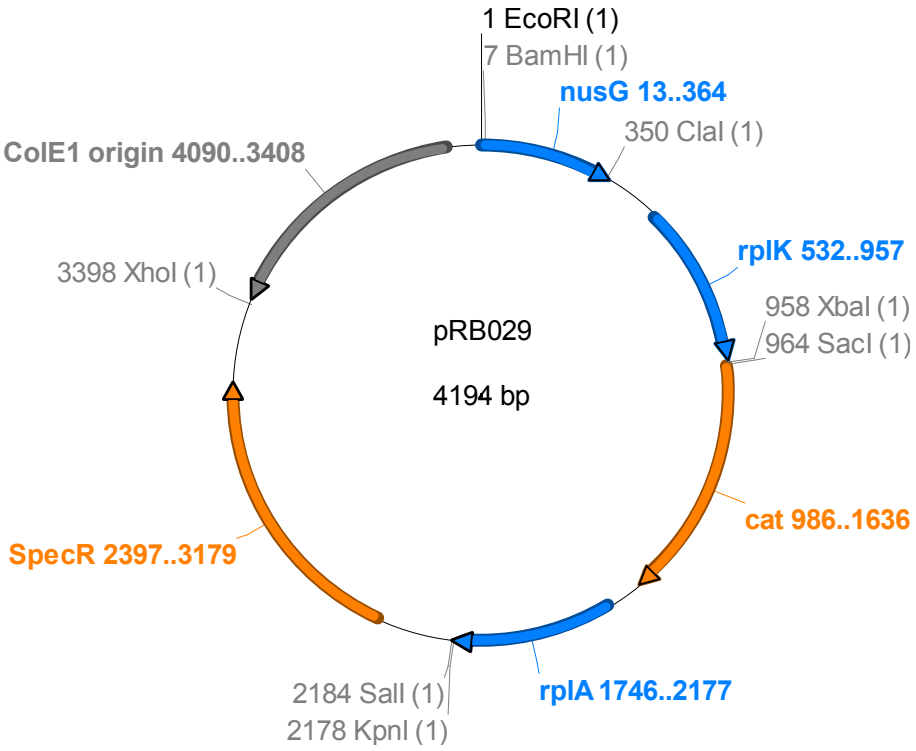


ATGTTTGCAAAACGATTCAAAACCTCTTTACTGCCGTTATTCGCTGGATTTTTATTGCTGTTTCATTTGGTTCTGGCAGGACCGGC
GGCTGCGAGTGCTGAAACGGCGAACAAATCGAATGAGCTTACAGCACCGTCGATCAAAAGCGGAACCATTCTTCATGCATGGAATT
GGTCGTTCAATACGTTAAAACACAATATGAAGGATATTCATGATGCAGGATATACAGCCATTCAGACATCTCCGATTAACCAAGTA
AAGGAAGGGAATCAAGGAGATAAAAGCATGTCGAACTGGTACTGGCTGTATCAGCCGACATCGTATCAAATTGGCAACCGTTACTT
AGGTACTGAACAAGAATTTAAAGAAATGTGTGCAGCCGCTGAAGAATATGGCATAAAGGTCATTGTTGACGCGGTCATCAATCATA
CCACCAGTGATTATGCCGCGATTTCCAATGAGGTTAAGAGTATTCCAAACTGGACACATGGAAACACACAAATTAAAAACTGGTCT
GATCGGATCCACAAGCTTCAGAGCTCTCACTTCGCAAACCCTCTGTATTATAAATATttacacttcaaacctcttgcttaatacaa
tatacattgcgcaaaatagcacgaatatcccagcataaattaacaaaatctctacactaaacatatccgacaatggcccaaatatt
atcattcctagaggaagcgcgcatgaattcgcaatttgtacaagactgaacatccgtccgtgcattgatggttctactttctcctg
caataacacagtaactggtgtattgaagcacggcatcgttataccaattaaaaaattaaatactaaatacataataaatatcggtg
ctattccgagtccaatcataagcagtccatataatgcacaagctagcgctgttgtatgcatacgattacgaaaaccaccccacgcg
gcaatcaatacgccaccagctgccgcaccagcactaaaggtcatctcactagcagtaagtctccacatttcggccccgaaagaacg
agttaccattaaaggtgttaaaaaagcagctggacttattaaaatcatgatgataatcaataaaacgattagccttttaataaata
cattttcttttaaataagagaatccttctttaattccttgtagattagattttgcttattctctgcattcacatgtccctctact
tgaatgataagcataatactaattccaatcatagccgtaataacatcaataaaaaatgttgtttcaatgctggcaatagaaagaat
cgccgcacttgctgctggagaaagaaacataattaaagatgtaatactgctatttattccatttactttcataagatactctttcg
gaacgatttgcggaattaatgcattcacagctggcgtttgtactcctgttccagctgaacgaatgagtaacacaataaaaagtaac
caaatacttttatatcctgttaaaaataaaattgctaaaactaacgtcgcaatcgctataattccatccgataacataatcatctt
tttacggtcataacgatcaatccaaacacctgcaaacaaagatatcgcaatttgtggtagaaaaccgcaaacagttgaaatagtca
acatcactcctgatgatgttgtaagggtgatataccatataatcgcgtattgcacgagagaggaacctaacaacgaaatagtttga
gcggtcaaaaataatataattttcgttttccaattgttttcactcatatgtttcacACCCCCTTAACTAATCCCGTCTCTATATGA
AAGATACTATCTGCCCAATCAAGATAAAAACTCGCTTCGTGAGATACGAGTATAATGCTGCCCTTAAACTTAATTAACGCCTTTTG
CAAAGCTTCCTTCGCTTCTGCATCTAAATGATTTGTCGGCTCATCTAAAAttaataagttacaaggtgataataataatccgcaac
ccccttaactaatcccgtctctatatgaaagatactatctgcccaatcaagataaaaactcgcttcgtgagatacgagtataatgc
tgcccttaaacttaattaacgccttttgcaaagcttccttcgcttctgcatctaaatgatttgtcggctcatctaaaattaataag
ttacaaggtgataataataatccgcacagtttcaccttcgattgttcaccaccacttaacgtacgaatttcttgcgaaacgtgcgt
atccttcactccacacgcagctaaatgatggcgtatttccttcatattcagtttaggaaactgttctgaaataatttgaagcggcg
ttaaagaatcattgctccaatttaaatcttgctcataatatccaatctttatttgctctgaaaagtgaaactctccagaaatacta
tcaatatttccaacgattgtttttaataaagtagattttcctactccattgaaacctgtaataacaatttttttgcctacccattac
tgaaaaattcaactttggtaagagagcgaagtcgtaaccaacttcaaagttatttacttcaagaactgtctgaactgaaatgggaa
gttcctgaaaatgaatagacggtttatgagtaaaacttggtggcgcaattctttccatacgctgtaattgttttcgacgaccttgc
gctatttttgaattcaccccagcaatatttctacgaatatactcttccgtttttcaattttcttttgctgtgcataatattggcg

aatataatcttttcgcaaatgctccttttgtcttacaaaatctgagtatttgccgtaatacttcttaactgttccgaactctacat
cacatatgcaagaagttacctttttccagaaaatcaaaatcatgtgatacgacaataaacgctccttcaaaattcattaaataatta
gcaagccactccacatgctccttatcgagaaaattcgtcggttcatctagcaataaaatattaggctcttctaataagagcttcgc
taaaattacctttgcacgctgccctccgcttaattcaccaatcacacgattcatcccaatcacatcaatacctaaaccggctgcca
ccttattcgtaatactatctactgaataaaaatcatgagcctcgagcagtccttgaatttctgccgctttcaatagttgattttca
tctccagtcatcgcactttcttcatataacttgttcattctcttttccatttcatacaaatcataaaaggctctctttaaatattg
agaaatggtataaccctcatcaatttctgcgtattgatcaagatgtccaatttgaatatttggttgccactttatgtctcccttat
cagaaataatttcccccaccaatattttcattaatgtgctttttcctgcaccattctgcccaacaattcccatatgctcccctta
tgcaaatcaaaagaagcattttcgtataatactttatctataaaactatgtgataaactttcaactttaagtaaactcatTTTTAT
TCCTCTTTTCCTGAAATTAAAAAAAGCAGAGAGATAGCGTCATTCGACACTAGCCCTCTGCTTCATGGTTTTTCCTATACTACTCC
AATTTATAGTAAAGACTCATCGGAATAATATACACAAAGGGCCATAGACAAAGCATTGTCTATGGCCCCTTAATCTATTTTATACA
CCAGCTGAAAAACCATAACGAAATAAGCATAGATAAACTATACTTTTTTAGAAAAGTTATTCAGCCGATGAATATAGGGGATTTAT
ACCCCTTGATTAAGCTCCGTACAACGGTACCTCGAATTCTCATGTTTGACAGCTTATCATCGGCAATAGTTACCCTTATTATCAAG
ATAAGAAAGAAAAGGATTTTTCGCTACGCTCAAATCCTTTAAAAAAACACAAAAGACCACATTTTTTAATGTGGTCTTTTATTCTT
CAACTAAAGCACCCATTAGTTCAACAAACGAAAATTGGATAAAGTGGGATATTTTTAAAATATATATTTATGTTACAGTAATATTG
ACTTTTAAAAAAGGATTGATTCTAATGAAGAAAGCAGACAAGTAAGCCTCCTAAATTCACTTTAGATAAAAATTTAGGAGGCATAT
CAAatgaactttaataaaattgatttagacaattggaagagaaaagagatatttaatcattatttgaaccaacaaacgacttttag
tataaccacagaaattgatattagtgttttataccgaaacataaaacaagaaggatataaattttaccctgcatttatttcttag
tgacaagggtgataaactcaaatacagcttttagaactggttacaatagcgacggagagttaggttattgggataagttagagcca
ctttatacaatttttgatggtgtatctaaaacattctctggtatttggactcctgtaaagaatgacttcaaagagttttatgattt
atacctttctgatgtagagaaatataatggttcggggaaattgtttcccaaaacacctatacctgaaaatgctttttctctttcta
ttattccatggacttcatttactgggtttaacttaaatatcaataataatagtaattaccttctacccattattacagcaggaaaa
ttcattaataaaggtaattcaatatatttaccgctatctttacaggtacatcattctgtttgtgatggttatcatgcaggattgtt
tatgaactctattcaggaattgtcagataggcctaatgactggcttttataaTATGAGATAATGCCGACTGTACTTTTTACAGTCG
GTTTTCTAATGTCACTAACCTGCCCCGTTAGTTGAAGAAGGTTTTTATATTACAGCTCCAGATCCTCTACGCCGGACGCATCGTGG
CCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGG
CTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATT
CCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACATG
GATGAGCGATGATGATATCCGTTTAGGCTGGGCGGTGATAGCTTCTCGTTCAGGCAGTACGCCTCTTTTCTTTTCCAGACCTGAGG
GAGGCGGAAATGGTGTGAGGTTCCCGGGGAAAAGCCAAATAGGCGATCGCGGGAGTGCTTTATTTGAAGATCAGGCTATCACTGCG
GTCAATAGATTTCACAATGTGATGGCTGGACAGCCTGAGGAACTCTCGAACCCGAATGGAAACAACCAGATATTTATGAATCAGCG
CGGCTCACATGGCGTTGTGCTGGCAAATGCAGGTTCATCCTCTGTCTCTATCAATACGGCAACAAAATTGCCTGATGGCAGGTATG
ACAATAAAGCTGGAGCGGGTTCATTTCAAGTGAACGATGGTAAACTGACAGGCACGATCAATGCCAGGTCTGTAGCTGTGCTTTAT
CCTGATGATATTGCAAAAGCGCCTCATGTTTTCCTTGAGAATTACAAAACAGGTGTAACACATTCTTTCAATGATCAACTGACGAT
TACCTTGCGTGCAGATGCGAATACAACAAAAGCCGTTTATCAAATCAATAATGGACCAGACGACAGGCGTTTAAGGATGGAGATCA
ATTCACAATCGGAAAAGGAGATCCAATTTGGCAAAACATACACCATCATGTTAAAAGGAACGAACAGTGATGGTGTAACGAGGACC
GAGAAATACAGTTTTGTTAAAAGAGATCCAGCGTCGGCCAAAACCATCGGCTATCAAAATCCGAATCATTGGAGCCAGGTAAATGC
TTATATCTATAAACATGATGGGAGCCGAGTAATTGAATTGACCGGATCTTGGCCTGGAAAACCAATGACTAAAAATGCAGACGGAA
TTTACACGCTGACGCTGCCTGCGGACACGGATACAACCAACGCAAAAGTGATTTTTAATAATGGCAGCGCCCAAGTGCCCGGTCAG
AATCAGCCTGGCTTTGATTACGTGCTAAATGGTTTATATAATGACTCGGGCTTAAGCGGTTCTCTTCCCCATTGAGGGCAAGGCTA
GACGGGACTTACCGAAAGAAACCATCAATGATGGTTTCTTTTTTGTTCATAAATCAGACAAAACTTTTCTCTTGCAAAAGTTTGTG
AAGTGTTGCACAATATAAATGTGAAATACTTCACAAACAAAAAGACATCAAAGAGAAACATACCCTGCAAGGATGCTGATATTGTC
TGCATTTGCGCCGGAGCAAACCAAAAACCTGGTGAGACACGCCTTGAATTAGTAGAAAAGAACTTGAAGATTTTCAAAGGCATCGT
TAGTGAAGTCATGGCGAGCGGATTTGACGGCATTTTCTTAGTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCG
CTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCG
CTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGG
GTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTACTGAA
GTGGATTTCTTTAAGCTAGAGGATCGATCTGTATAATAAAGAATAATTATTAATCTGTAGACAAATTGTGAAAGGATGTACTTAAA
CGCTAACGGTCAGCTTTATTGAACAGTAATTTAAGTATATGTCCAATCTAGGGTAAGTAAATTGAGTATCAATATAAACTTTATAT
GAACATAATCAACGAGGTGAAATCATGAGCAATTTGATTAACGGAAAAATACCAAATCAAGCGATTCAAACATTAAAAATCGTAAA
AGATTTATTTGGAAGTTCAATAGTTGGAGTATATCTATTTGGTTCAGCAGTAAATGGTGGTTTACGCATTAACAGCGATGTAGATG
TTCTAGTCGTCGTGAATCATAGTTTACCTCAATTAACTCGAAAAAAACTAACAGAAAGACTAATGACTATATCAGGAAAGATTGGA
AATACGGATTCTGTTAGACCACTTGAAGTTACGGTTATAAATAGGAGTGAAGTTGTCCCTTGGCAATATCCTCCAAAAAGAGAATT
TATATACGGTGAGTGGCTCAGGGGTGAATTTGAGAATGGACAAATTCAGGAACCAAGCTATGATCCTGATTTGGCTATTGTTTTAG
CACAAGCAAGAAAGAATAGTATTTCTCTATTTGGTCCTGATTCTTCAAGTATACTTGTCTCCGTACCTTTGACAGATATTCGAAGA
GCAATTAAGGATTCTTTGCCAGAACTAATTGAGGGGATAAAAGGTGATGAGCGTAATGTAATTTTAACCCTAGCTCGAATGTGGCA
AACAGTGACTACTGGTGAAATTACCTCGAAAGATGTCGCTGCAGAATGGGCTATACCTCTTTTACCTAAAGAGCATGTAACTTTAC
TGGATATAGCTAGAAAAGGCTATCGGGGAGAGTGTGATGATAAGTGGGAAGGACTATATTCAAAGGTGAAAGCACTCGTTAAGTAT
ATGAAAAATTCTATAGAAACTTCTCTCAATTAGGCTAATTTTATTGCAATAACAGGTGCTTACTTTTCTGGAGTTCTTTAGCAAAT
TTTTTTATTAGCTGAACTTAGTATTAGTGGCCATACTCCTCCAATCCAAAGCTATTTAGAAAGATTACTATATCCTCAAACAGGCG
GTAACCGGCCTCTTCATCGGGAATGCGCGCGACCTTCAGCATCGCCGGCATGTCCCCCTGGCGGACGGGAAGTATCCAGCTCGAGG
TCGGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA
ACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA

TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCG
CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGT
TGATAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAG
GATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGA
TTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTC
TGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCG
TGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGAT
TTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTG
TTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCGTGGTGTCACGCT
CGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTT
AGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCT
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGA
GTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCG
GGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTT
TACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAA
TACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAAC
CTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTAACAAAATTCTCCAGTCTTCACATCGGTTTGAAAGGAGGAA
GCGGAAGAATGAAGTAAGAGGGATTTTTGACTCCGAAGTAAGTCTTCAAAAAATCAAATAAGGAGTGTCAAGA

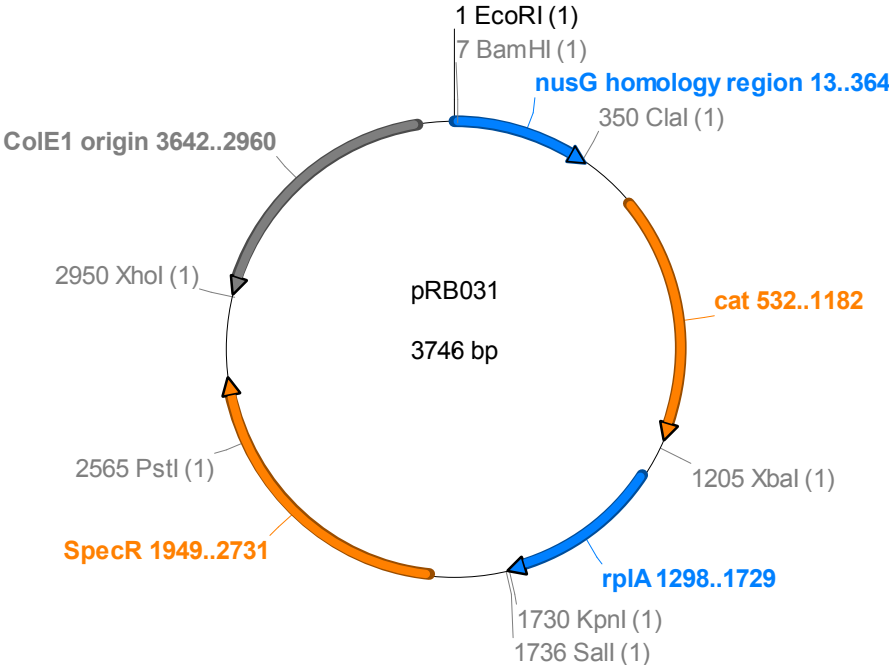
C Plasmid pRB029 used to generate *B. subtilis* 168 Δ*rplK*::*rplK-cat* mutant (PB213). The plasmid map shows critical features and unique restriction sites, where relevant. The complete nucleotide sequence is provided below with *rplK* and *cat* shown in lowercase font.



GAATTCGGATCCCCCTGGTTATGTGCTTGTTGAAATTGTAATGACAGACGACTCTTGGTATGTCGTCCGCAACACGCCGGGCGTTA
CTGGATTCGTAGGATCTGCCGGGTCAGGTTCAAAACCGACGCCGCTTCTTCCGGGGGAAGCAGAAACCATTCTGAAGAGAATGGGC
ATGGATGAACGAAAAACTGATATTGACTTTGAACTGAAAGAGACAGTGAAAGTAATAGACGGACCTTTTGCTAACTTTACAGGATC
AATTGAAGAGATTGATTATGATAAAAGCAAGGTCAAAGTTTTCGTTAATATGTTTGGCCGTGAAACGCCGGTTGAGCTGGAATTTA
CCCAAATCGATAAATTGTAATGTGAAAAAACTTGAAATTGCTATATATAAATGATAATATAGCAAAGGTACGTCTTGGACTTATCC
AAGGCAACTAGCATGATATTTCGTCATTCATATAAAGAATGAAACCTTGAGTGGGAGGGTTTACCCTATTACCACATCACGGACTT

AAGGAGGTGTGTCTCgtggctaaaaaagtagttaaagttgtaaaattgcaaattcctgctggaaaagctaacccagcaccaccagt
tggacctgcacttggtcaagccggtgttaacatcatgggattctgtaaggagtttaacgctcgtacagctgaccaagctggtttaa
tcattcctgttgaaatttcggtttacgaagaccgttcatttacatttattacaaaaactccacctgctgcagtattgcttaaaaaa
gcagctggaattgagtctggttctggtgaaccaaaccgtaataaagtggcaaccgttaagcgcgataaagtacgcgaaatcgctga
aacgaaaatgcctgacttaaacgcagcagacgttgaagcggcaatgcgcatggttgaaggtactgcccgcagtatgggtattgtaa
tcgaggattaaTCTAGAGAGCTCTTAGGAGGCATATCAAatgaactttaataaaattgatttagacaattggaagagaaaagagat
atttaatcattatttgaaccaacaaacgacttttagtataaccacagaaattgatattagtgttttataccgaaacataaaacaag
aaggatataaattttaccctgcatttattttcttagtgacaagggtgataaactcaaatacagcttttagaactggttacaatagc
gacggagagttaggttattgggataagttagagccactttatacaatttttgatggtgtatctaaaacattctctggtatttggac
tcctgtaaagaatgacttcaaagagttttatgatttatacctttctgatgtagagaaatataatggttcggggaaattgtttccca
aaacacctatacctgaaaatgctttttctctttctattattccatggacttcatttactgggtttaacttaaatatcaataataat
agtaattaccttctacccattattacagcaggaaaattcattaataaaggtaattcaatatatttaccgctatctttacaggtaca
tcattctgtttgtgatggttatcatgcaggattgtttatgaactctattcaggaattgtcagataggcctaatgactggctttttat
aaTATGAGATAATGCCGACTGTACCTTGTCGGGTTGCGAGTTTTAACAAGTTCGCAACCCTTATTCGTGGGAGGTTATTCCGCTAT
AACCACATAAGGAGGAAATTTTAAAATGGCTAAAAAAGGTAAAAAGTACGTTGAAGCTGCTAAGCTTGTAGACCGTTCTAAAGCTT
ACGACGTCTCTGAAGCAGTAGCTCTCGTTAAAAAAACAAACACAGCTAAATTCGACGCTACAGTTGAAGTGGCTTTCCGTTTAGGG
GTTGACCCTCGTAAAAACGACCAGCAAATCCGTGGAGCAGTTGTGCTTCCAAACGGAACTGGTAAAACTCAGCGCGTTCTCGTTTT
CGCAAAAGGCGAAAAAGCGAAAGAAGCTGAAGCTGCTGGTGCAGATTTCGTAGGCGATACTGACTACATCAACAAAATTCAACAAG
GCTGGTTCGATTTCGATGTTATCGTAGCTACACCTGACATGATGGGTGAAGTTGGTAAAATCGGTCGTGTACTTGGACCAAAAGGT
TTAATGCCGAACCCTAAAACTGGTACAGGTACCGTCGACCACCTACTGAAGTGGATTTCTTTAAGCTAGAGGATCGATCTGTATAA
TAAAGAATAATTATTAATCTGTAGACAAATTGTGAAAGGATGTACTTAAACGCTAACGGTCAGCTTTATTGAACAGTAATTTAAGT
ATATGTCCAATCTAGGGTAAGTAAATTGAGTATCAATATAAACTTTATATGAACATAATCAACGAGGTGAAATCATGAGCAATTTG
ATTAACGGAAAAATACCAAATCAAGCGATTCAAACATTAAAAATCGTAAAAGATTTATTTGGAAGTTCAATAGTTGGAGTATATCT
ATTTGGTTCAGCAGTAAATGGTGGTTTACGCATTAACAGCGATGTAGATGTTCTAGTCGTCGTGAATCATAGTTTACCTCAATTAA
CTCGAAAAAAACTAACAGAAAGACTAATGACTATATCAGGAAAGATTGGAAATACGGATTCTGTTAGACCACTTGAAGTTACGGTT
ATAAATAGGAGTGAAGTTGTCCCTTGGCAATATCCTCCAAAAAGAGAATTTATATACGGTGAGTGGCTCAGGGGTGAATTTGAGAA
TGGACAAATTCAGGAACCAAGCTATGATCCTGATTTGGCTATTGTTTTAGCACAAGCAAGAAAGAATAGTATTTCTCTATTTGGTC
CTGATTCTTCAAGTATACTTGTCTCCGTACCTTTGACAGATATTCGAAGAGCAATTAAGGATTCTTTGCCAGAACTAATTGAGGGG
ATAAAAGGTGATGAGCGTAATGTAATTTTAACCCTAGCTCGAATGTGGCAAACAGTGACTACTGGTGAAATTACCTCGAAAGATGT
CGCTGCAGAATGGGCTATACCTCTTTTACCTAAAGAGCATGTAACTTTACTGGATATAGCTAGAAAAGGCTATCGGGGAGAGTGTG
ATGATAAGTGGGAAGGACTATATTCAAAGGTGAAAGCACTCGTTAAGTATATGAAAAATTCTATAGAAACTTCTCTCAATTAGGCT
AATTTTATTGCAATAACAGGTGCTTACTTTTCTGGAGTTCTTTAGCAAATTTTTTTATTAGCTGAACTTAGTATTAGTGGCCATAC
TCCTCCAATCCAAAGCTATTTAGAAAGATTACTATATCCTCAAACAGGCGGTAACCGGCCTCTTCATCGGGAATGCGCGCGACCTT
CAGCATCGCCGGCATGTCCCCCTGGCGGACGGGAAGTATCCAGCTCGAGGTCGGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG
CCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCC
CTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCG
CTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCA
GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACA
GTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGATAGCTCTTGATCCGGCAAACAAACCACCGCTGG
TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAAT
TAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGC

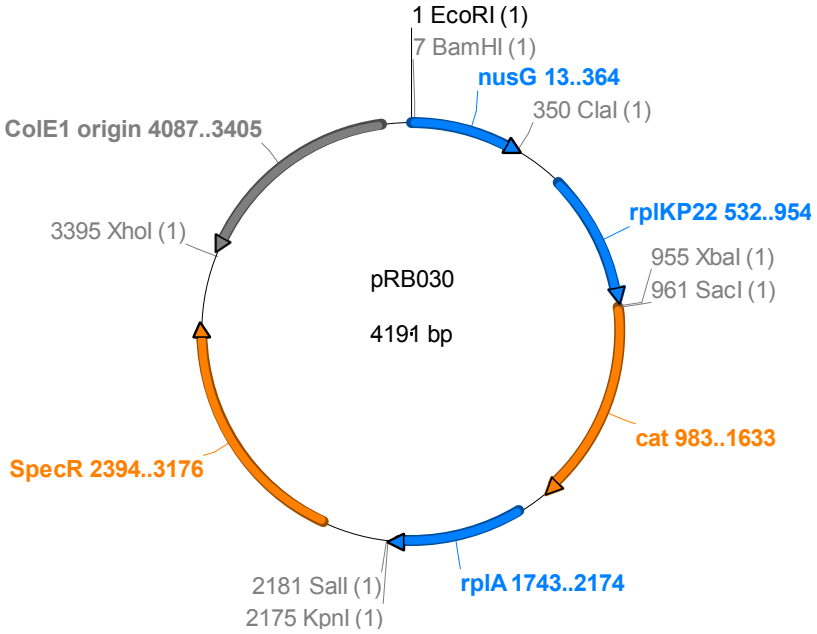
D Plasmid pRB031 used to generate *B. subtilis* 168 Δ*rplK*::*cat* mutant (PB230). The plasmid map shows critical features and unique restriction sites, where relevant. The complete nucleotide sequence is provided below with *cat* shown in lowercase font.



GAATTCGGATCCCCCTGGTTATGTGCTTGTTGAAATTGTAATGACAGACGACTCTTGGTATGTCGTCCGCAACACGCCGGGCGTTA
CTGGATTCGTAGGATCTGCCGGGTCAGGTTCAAAACCGACGCCGCTTCTTCCGGGGGAAGCAGAAACCATTCTGAAGAGAATGGGC
ATGGATGAACGAAAAACTGATATTGACTTTGAACTGAAAGAGACAGTGAAAGTAATAGACGGACCTTTTGCTAACTTTACAGGATC
AATTGAAGAGATTGATTATGATAAAAGCAAGGTCAAAGTTTTCGTTAATATGTTTGGCCGTGAAACGCCGGTTGAGCTGGAATTTA
CCCAAATCGATAAATTGTAATGTGAAAAAACTTGAAATTGCTATATATAAATGATAATATAGCAAAGGTACGTCTTGGACTTATCC
AAGGCAACTAGCATGATATTTCGTCATTCATATAAAGAATGAAACCTTGAGTGGGAGGGTTTACCCTATTACCACATCACGGACTT
AAGGAGGTGTGTCTCatgaactttaataaaattgatttagacaattggaagagaaaagagatatttaatcattatttgaaccaaca
aacgacttttagtataaccacagaaattgatattagtgttttataccgaaacataaaacaagaaggatataaattttaccctgcat
ttatttttcttagtgacaagggtgataaactcaaatacagcttttagaactggttacaatagcgacggagagttaggttattgggat
aagttagagccactttatacaatttttgatggtgtatctaaaacattctctggtatttggactcctgtaaagaatgacttcaaaga
gttttatgatttatacctttctgatgtagagaaatataatggttcggggaaattgtttcccaaaacacctatacctgaaaatgctt
tttctctttctattattccatggacttcatttactgggtttaacttaaatatcaataataatagtaattaccttctacccattatt
acagcaggaaaattcattaataaaggtaattcaatatatttaccgctatctttacaggtacatcattctgtttgtgatggttatca
tgcaggattgtttatgaactctattcaggaattgtcagataggcctaatgactggcttttataaTATGAGATAATGCCGACTGTAC
TCTAGACTTGTCGGGTTGCGAGTTTTAACAAGTTCGCAACCCTTATTCGTGGGAGGTTATTCCGCTATAACCACATAAGGAGGAAA
TTTTAAAATGGCTAAAAAAGGTAAAAAGTACGTTGAAGCTGCTAAGCTTGTAGACCGTTCTAAAGCTTACGACGTCTCTGAAGCAG
TAGCTCTCGTTAAAAAAACAAACACAGCTAAATTCGACGCTACAGTTGAAGTGGCTTTCCGTTTAGGGGTTGACCCTCGTAAAAAC
GACCAGCAAATCCGTGGAGCAGTTGTGCTTCCAAACGGAACTGGTAAAACTCAGCGCGTTCTCGTTTTCGCAAAAGGCGAAAAAGC
GAAAGAAGCTGAAGCTGCTGGTGCAGATTTCGTAGGCGATACTGACTACATCAACAAAATTCAACAAGGCTGGTTCGATTTCGATG
TTATCGTAGCTACACCTGACATGATGGGTGAAGTTGGTAAAATCGGTCGTGTACTTGGACCAAAAGGTTTAATGCCGAACCCTAAA
ACTGGTACAGGTACCGTCGACCACCTACTGAAGTGGATTTCTTTAAGCTAGAGGATCGATCTGTATAATAAAGAATAATTATTAAT
CTGTAGACAAATTGTGAAAGGATGTACTTAAACGCTAACGGTCAGCTTTATTGAACAGTAATTTAAGTATATGTCCAATCTAGGGT
AAGTAAATTGAGTATCAATATAAACTTTATATGAACATAATCAACGAGGTGAAATCATGAGCAATTTGATTAACGGAAAAATACCA
AATCAAGCGATTCAAACATTAAAAATCGTAAAAGATTTATTTGGAAGTTCAATAGTTGGAGTATATCTATTTGGTTCAGCAGTAAA
TGGTGGTTTACGCATTAACAGCGATGTAGATGTTCTAGTCGTCGTGAATCATAGTTTACCTCAATTAACTCGAAAAAAACTAACAG
AAAGACTAATGACTATATCAGGAAAGATTGGAAATACGGATTCTGTTAGACCACTTGAAGTTACGGTTATAAATAGGAGTGAAGTT
GTCCCTTGGCAATATCCTCCAAAAAGAGAATTTATATACGGTGAGTGGCTCAGGGGTGAATTTGAGAATGGACAAATTCAGGAACC
AAGCTATGATCCTGATTTGGCTATTGTTTTAGCACAAGCAAGAAAGAATAGTATTTCTCTATTTGGTCCTGATTCTTCAAGTATAC
TTGTCTCCGTACCTTTGACAGATATTCGAAGAGCAATTAAGGATTCTTTGCCAGAACTAATTGAGGGGATAAAAGGTGATGAGCGT

AATGTAATTTTAACCCTAGCTCGAATGTGGCAAACAGTGACTACTGGTGAAATTACCTCGAAAGATGTCGCTGCAGAATGGGCTAT
ACCTCTTTTACCTAAAGAGCATGTAACTTTACTGGATATAGCTAGAAAAGGCTATCGGGGAGAGTGTGATGATAAGTGGGAAGGAC
TATATTCAAAGGTGAAAGCACTCGTTAAGTATATGAAAAATTCTATAGAAACTTCTCTCAATTAGGCTAATTTTATTGCAATAACA
GGTGCTTACTTTTCTGGAGTTCTTTAGCAAATTTTTTATTAGCTGAACTTAGTATTAGTGGCCATACTCCTCCAATCCAAAGCTA
TTTAGAAAGATTACTATATCCTCAAACAGGCGGTAACCGGCCTCTTCATCGGGAATGCGCGCGACCTTCAGCATCGCCGGCATGTC
CCCCTGGCGGACGGGAAGTATCCAGCTCGAGGTCGGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCA
CAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGC
TGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGATAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGT
TTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACG
AAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAA
TCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGC

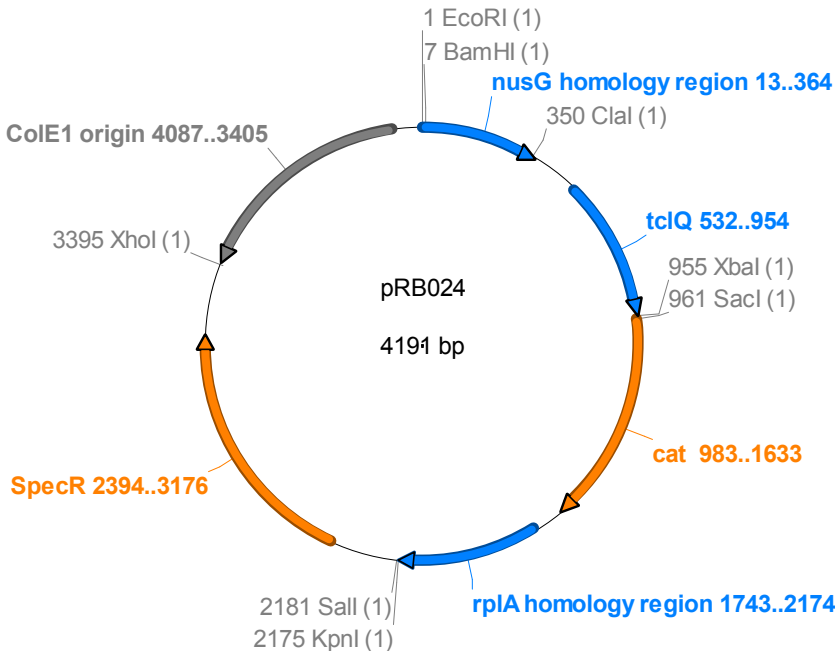
E Plasmid pRB030 used to generate *B. subtilis* 168 Δ*rplK*::*rplK*Δ*P22-cat* mutant (PB214). The plasmid map shows critical features and unique restriction sites, where relevant. The complete nucleotide sequence is provided below with *rplK*Δ*P22* and *cat* shown in lowercase font.



GAATTCGGATCCCCCTGGTTATGTGCTTGTTGAAATTGTAATGACAGACGACTCTTGGTATGTCGTCCGCAACACGCCGGGCGTTA
CTGGATTCGTAGGATCTGCCGGGTCAGGTTCAAAACCGACGCCGCTTCTTCCGGGGGAAGCAGAAACCATTCTGAAGAGAATGGGC
ATGGATGAACGAAAAACTGATATTGACTTTGAACTGAAAGAGACAGTGAAAGTAATAGACGGACCTTTTGCTAACTTTACAGGATC
AATTGAAGAGATTGATTATGATAAAAGCAAGGTCAAAGTTTTCGTTAATATGTTTGGCCGTGAAACGCCGGTTGAGCTGGAATTTA
CCCAAATCGATAAATTGTAATGTGAAAAAACTTGAAATTGCTATATATAAATGATAATATAGCAAAGGTACGTCTTGGACTTATCC
AAGGCAACTAGCATGATATTTCGTCATTCATATAAAGAATGAAACCTTGAGTGGGAGGGTTTACCCTATTACCACATCACGGACTT
AAGGAGGTGTGTCTCgtggctaaaaaagtagttaaagttgtaaaattgcaaattcctgctggaaaagctaacccagcaccagttgg
acctgcacttggtcaagccggtgttaacatcatgggattctgtaaggagtttaacgctcgtacagctgaccaagctggtttaatca
ttcctgttgaaatttcggtttacgaagaccgttcatttacatttattacaaaaactccacctgctgcagtattgcttaaaaaagca
gctggaattgagtctggttctggtgaaccaaaccgtaataaagtggcaaccgttaagcgcgataaagtacgcgaaatcgctgaaac
gaaaatgcctgacttaaacgcagcagacgttgaagcggcaatgcgcatggttgaaggtactgcccgcagtatgggtattgtaatcg
aggattaaTCTAGAGAGCTCTTAGGAGGCATATCAAatgaactttaataaaattgatttagacaattggaagagaaaagagatatt
taatcattatttgaaccaacaaacgacttttagtataaccacagaaattgatattagtgttttataccgaaacataaaacaagaag
gatataaattttaccctgcatttatttttcttagtgacaagggtgataaactcaaatacagcttttagaactggttacaatagcgac
ggagagttaggttattgggataagttagagccactttatacaatttttgatggtgtatctaaaacattctctggtatttggactcc
tgtaaagaatgacttcaaagagttttatgatttatacctttctgatgtagagaaatataatggttcggggaaattgtttcccaaaa
cacctatacctgaaaatgctttttctctttctattattccatggacttcatttactgggtttaacttaaatatcaataataatagt

aattaccttctacccattattacagcaggaaaattcattaataaaggtaattcaatatatttaccgctatctttacaggtacatca
ttctgtttgtgatggttatcatgcaggattgtttatgaactctattcaggaattgtcagataggcctaatgactggctttataaT
ATGAGATAATGCCGACTGTACCTTGTCGGGTTGCGAGTTTTAACAAGTTCGCAACCCTTATTCGTGGGAGGTTATTCCGCTATAAC
CACATAAGGAGGAAATTTTAAAATGGCTAAAAAAGGTAAAAAGTACGTTGAAGCTGCTAAGCTTGTAGACCGTTCTAAAGCTTACG
ACGTCTCTGAAGCAGTAGCTCTCGTTAAAAAAACAAACACAGCTAAATTCGACGCTACAGTTGAAGTGGCTTTCCGTTTAGGGGTT
GACCCTCGTAAAAACGACCAGCAAATCCGTGGAGCAGTTGTGCTTCCAAACGGAACTGGTAAAACTCAGCGCGTTCTCGTTTTCGC
AAAAGGCGAAAAAGCGAAAGAAGCTGAAGCTGCTGGTGCAGATTTCGTAGGCGATACTGACTACATCAACAAAATTCAACAAGGCT
GGTTCGATTTCGATGTTATCGTAGCTACACCTGACATGATGGGTGAAGTTGGTAAAATCGGTCGTGTACTTGGACCAAAAGGTTTA
ATGCCGAACCCTAAAACTGGTACAGGTACCGTCGACCACCTACTGAAGTGGATTTCTTTAAGCTAGAGGATCGATCTGTATAATAA
AGAATAATTATTAATCTGTAGACAAATTGTGAAAGGATGTACTTAAACGCTAACGGTCAGCTTTATTGAACAGTAATTTAAGTATA
TGTCCAATCTAGGGTAAGTAAATTGAGTATCAATATAAACTTTATATGAACATAATCAACGAGGTGAAATCATGAGCAATTTGATT
AACGGAAAAATACCAAATCAAGCGATTCAAACATTAAAAATCGTAAAAGATTTATTTGGAAGTTCAATAGTTGGAGTATATCTATT
TGGTTCAGCAGTAAATGGTGGTTTACGCATTAACAGCGATGTAGATGTTCTAGTCGTCGTGAATCATAGTTTACCTCAATTAACTC
GAAAAAAACTAACAGAAAGACTAATGACTATATCAGGAAAGATTGGAAATACGGATTCTGTTAGACCACTTGAAGTTACGGTTATA
AATAGGAGTGAAGTTGTCCCTTGGCAATATCCTCCAAAAAGAGAATTTATATACGGTGAGTGGCTCAGGGGTGAATTTGAGAATGG
ACAAATTCAGGAACCAAGCTATGATCCTGATTTGGCTATTGTTTTAGCACAAGCAAGAAAGAATAGTATTTCTCTATTTGGTCCTG
ATTCTTCAAGTATACTTGTCTCCGTACCTTTGACAGATATTCGAAGAGCAATTAAGGATTCTTTGCCAGAACTAATTGAGGGGATA
AAAGGTGATGAGCGTAATGTAATTTTAACCCTAGCTCGAATGTGGCAAACAGTGACTACTGGTGAAATTACCTCGAAAGATGTCGC
TGCAGAATGGGCTATACCTCTTTTACCTAAAGAGCATGTAACTTTACTGGATATAGCTAGAAAAGGCTATCGGGGAGAGTGTGATG
ATAAGTGGGAAGGACTATATTCAAAGGTGAAAGCACTCGTTAAGTATATGAAAAATTCTATAGAAACTTCTCTCAATTAGGCTAAT
TTTATTGCAATAACAGGTGCTTACTTTTCTGGAGTTCTTTAGCAAATTTTTTTATTAGCTGAACTTAGTATTAGTGGCCATACTCC
TCCAATCCAAAGCTATTTAGAAAGATTACTATATCCTCAAACAGGCGGTAACCGGCCTCTTCATCGGGAATGCGCGCGACCTTCAG
CATCGCCGGCATGTCCCCCTGGCGGACGGGAAGTATCCAGCTCGAGGTCGGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT
TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCC
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGATAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAG
CGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAA
AAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGC

F Plasmid pRB024 used to generate *B. subtilis* 168 Δ*rplK*::*tclQ-cat* mutant (PB215). A plasmid map shows critical features and unique restriction sites, where relevant. The complete nucleotide sequence is provided below with *tclQ* and *cat* shown in lowercase font.



GAATTCGGATCCCCCTGGTTATGTGCTTGTTGAAATTGTAATGACAGACGACTCTTGGTATGTCGTCCGCAACACGCCGGGCGTTA
CTGGATTCGTAGGATCTGCCGGGTCAGGTTCAAAACCGACGCCGCTTCTTCCGGGGGAAGCAGAAACCATTCTGAAGAGAATGGGC
ATGGATGAACGAAAAACTGATATTGACTTTGAACTGAAAGAGACAGTGAAAGTAATAGACGGACCTTTTGCTAACTTTACAGGATC
AATTGAAGAGATTGATTATGATAAAAGCAAGGTCAAAGTTTTCGTTAATATGTTTGGCCGTGAAACGCCGGTTGAGCTGGAATTTA
CCCAAATCGATAAATTGTAATGTGAAAAAACTTGAAATTGCTATATATAAATGATAATATAGCAAAGGTACGTCTTGGACTTATCC
AAGGCAACTAGCATGATATTTCGTCATTCATATAAAGAATGAAACCTTGAGTGGGAGGGTTTACCCTATTACCACATCACGGACTT
AAGGAGGTGTGTCTCatggcaaaagctataaagcaaattattaatattcaattagaagctggaaaagcttctccagcacctccggt
aggtactgcattaggcccagcaggagttaatattatggagtttgtaacacaatataatgctcagactaaagaaaagtctggggtta
ttcctgcagaaattacggtatatgaagatcgaagctttaactttatactaaaaactcctcccacagctgaattaataaaagagaaa
gcacatatcaataaaggatctggtgttccaaacaaagaaattgtaggaagtatttcaaagaaagatttacaagagattgcagaaat
aaaaatgcctgatctgaatgcaagttctattgaaacagcaatgtcaatgattgaaggtactgctagaaatatgggagtcaaagtag
aagattagTCTAGAGAGCTCTTAGGAGGCATATCAAatgaactttaataaaattgatttagacaattggaagagaaaagagatatt
taatcattatttgaaccaacaaacgacttttagtataaccacagaaattgatattagtgttttataccgaaacataaaacaagaag
gatataaattttaccctgcatttatttttcttagtgacaagggtgataaactcaaatacagcttttagaactggttacaatagcgac
ggagagttaggttattgggataagttagagccactttatacaatttttgatggtgtatctaaaacattctctggtatttggactcc
tgtaaagaatgacttcaaagagttttatgatttatacctttctgatgtagagaaatataatggttcggggaaattgtttcccaaaa
cacctatacctgaaaatgctttttctctttctattattccatggacttcatttactgggtttaacttaaatatcaataataatagt
aattaccttctacccattattacagcaggaaaattcattaataaaggtaattcaatatatttaccgctatctttacaggtacatca
ttctgtttgtgatggttatcatgcaggattgtttatgaactctattcaggaattgtcagataggcctaatgactggcttttataaT
ATGAGATAATGCCGACTGTACCTTGTCGGGTTGCGAGTTTTAACAAGTTCGCAACCCTTATTCGTGGGAGGTTATTCCGCTATAAC
CACATAAGGAGGAAATTTTAAAATGGCTAAAAAAGGTAAAAAGTACGTTGAAGCTGCTAAGCTTGTAGACCGTTCTAAAGCTTACG
ACGTCTCTGAAGCAGTAGCTCTCGTTAAAAAAACAAACACAGCTAAATTCGACGCTACAGTTGAAGTGGCTTTCCGTTTAGGGGTT
GACCCTCGTAAAAACGACCAGCAAATCCGTGGAGCAGTTGTGCTTCCAAACGGAACTGGTAAAACTCAGCGCGTTCTCGTTTTCGC
AAAAGGCGAAAAAGCGAAAGAAGCTGAAGCTGCTGGTGCAGATTTCGTAGGCGATACTGACTACATCAACAAAATTCAACAAGGCT
GGTTCGATTTCGATGTTATCGTAGCTACACCTGACATGATGGGTGAAGTTGGTAAAATCGGTCGTGTACTTGGACCAAAAGGTTTA
ATGCCGAACCCTAAAACTGGTACAGGTACCGTCGACCACCTACTGAAGTGGATTTCTTTAAGCTAGAGGATCGATCTGTATAATAA
AGAATAATTATTAATCTGTAGACAAATTGTGAAAGGATGTACTTAAACGCTAACGGTCAGCTTTATTGAACAGTAATTTAAGTATA
TGTCCAATCTAGGGTAAGTAAATTGAGTATCAATATAAACTTTATATGAACATAATCAACGAGGTGAAATCATGAGCAATTTGATT
AACGGAAAAATACCAAATCAAGCGATTCAAACATTAAAAATCGTAAAAGATTTATTTGGAAGTTCAATAGTTGGAGTATATCTATT
TGGTTCAGCAGTAAATGGTGGTTTACGCATTAACAGCGATGTAGATGTTCTAGTCGTCGTGAATCATAGTTTACCTCAATTAACTC
GAAAAAAACTAACAGAAAGACTAATGACTATATCAGGAAAGATTGGAAATACGGATTCTGTTAGACCACTTGAAGTTACGGTTATA
AATAGGAGTGAAGTTGTCCCTTGGCAATATCCTCCAAAAAGAGAATTTATATACGGTGAGTGGCTCAGGGGTGAATTTGAGAATGG
ACAAATTCAGGAACCAAGCTATGATCCTGATTTGGCTATTGTTTTAGCACAAGCAAGAAAGAATAGTATTTCTCTATTTGGTCCTG
ATTCTTCAAGTATACTTGTCTCCGTACCTTTGACAGATATTCGAAGAGCAATTAAGGATTCTTTGCCAGAACTAATTGAGGGGATA
AAAGGTGATGAGCGTAATGTAATTTTAACCCTAGCTCGAATGTGGCAAACAGTGACTACTGGTGAAATTACCTCGAAAGATGTCGC
TGCAGAATGGGCTATACCTCTTTTACCTAAAGAGCATGTAACTTTACTGGATATAGCTAGAAAAGGCTATCGGGGAGAGTGTGATG
ATAAGTGGGAAGGACTATATTCAAAGGTGAAAGCACTCGTTAAGTATATGAAAAATTCTATAGAAACTTCTCTCAATTAGGCTAAT
TTTATTGCAATAACAGGTGCTTACTTTTCTGGAGTTCTTTAGCAAATTTTTTTATTAGCTGAACTTAGTATTAGTGGCCATACTCC
TCCAATCCAAAGCTATTTAGAAAGATTACTATATCCTCAAACAGGCGGTAACCGGCCTCTTCATCGGGAATGCGCGCGACCTTCAG
CATCGCCGGCATGTCCCCCTGGCGGACGGGAAGTATCCAGCTCGAGGTCGGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT
TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCC
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGATAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAG
CGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAA
AAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGC

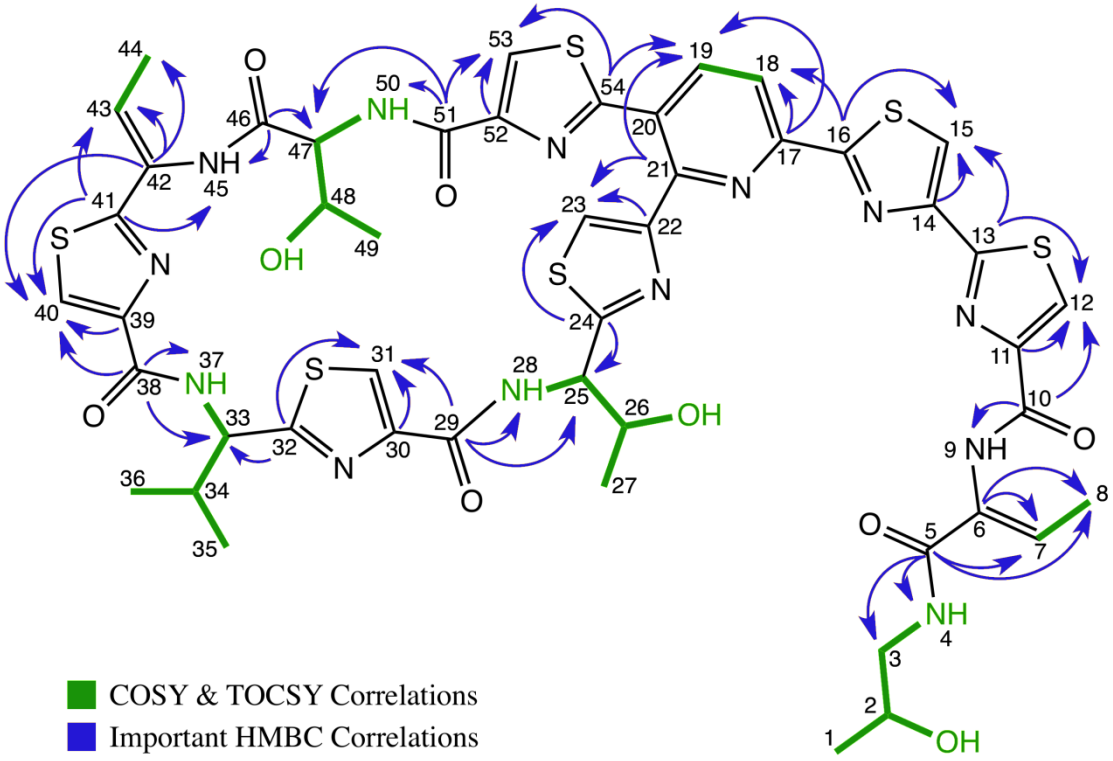


Figure 2-S2. Summary of NMR 2D correlations. NMR analysis analysis identified a peptide-like backbone with a series of highly modified substituents, no N-terminus, and with a decarboxylated threonine at the C-terminus. The 2D NMR correlations implied a macrocyclic ring structure as well as several smaller aromatic rings, consistent with the structure of Micrococcin P1. Also, two of the threonine residues were dehydrated, leaving a double bond between the α and β carbons; the six cysteine residues were each converted into a thiazole ring; and the two serine residues were condensed into a pyridine ring, making the macrocycle.

Chapter 3. Reconstitution and minimization of a micrococcin biosynthetic pathway in *Bacillus subtilis*

The following chapter is taken from an article with the same title, published in the Journal of Bacteriology in September 2016. All content, including supplemental material has been formatted appropriately for this dissertation but is otherwise unchanged.

3.1 Summary

Thiopeptides represent one of several families of highly modified peptide antibiotics that hold great promise for natural product engineering. These macrocyclic peptides are produced by a combination of ribosomal synthesis and extensive post-translational modification by dedicated processing enzymes. We previously identified a compact, plasmid-encoded gene cluster for the biosynthesis of micrococcin P1 (MP1), an archetypal thiopeptide antibiotic. In an effort to genetically dissect this pathway, we have reconstituted it in *Bacillus subtilis*. Successful MP1 production required promoter engineering and reassembly of essential biosynthetic genes in a modular plasmid. The resulting system allows for rapid pathway manipulation, including protein tagging and gene deletion. We find that 8 processing proteins are sufficient for production of MP1, and that the tailoring enzyme TclS catalyzes a C-terminal reduction step that distinguishes MP1 from its sister compound micrococcin P2.

3.2 Introduction

Maintaining an arsenal of effective antibiotics is a tremendous biomedical challenge as antibiotic discovery is outpaced by the evolution of resistance (10, 124, 125). Ribosomally synthesized and Post-translationally modified Peptides (RiPPs) are natural compounds with the appealing attributes of being derived directly from a genetic template while possessing numerous exotic chemical features that contribute to stability and antimicrobial activity (17). Thiopeptides, characterized by post-translationally formed sulfur- and nitrogen-containing heterocycles, constitute a rapidly expanding class of RiPP antibiotics (52). Over 100 different thiopeptides have been chemically identified from numerous culturable bacterial producers, and the mining of genomic and metagenomic data promises to uncover many more chemical species that have eluded discovery by conventional means (53, 126, 127).

Thiopeptides are generally encoded by chromosomal or plasmid-localized gene clusters (see example in Figure 3-1A) consisting of approximately 6 to 25 protein-coding genes (126). These include a gene encoding the precursor peptide that is modified to become the thiopeptide product, as well as an assemblage of post-translational modification genes required for thiopeptide maturation, genes for host immunity, and other functions. The precursor peptide for thiopeptides and other RiPPs consists of an N-terminal leader that acts as a required docking site for the biosynthetic machinery, as well as a C-terminal core peptide that is the direct substrate for post-translational modifications (28, 29). The leader peptide is ultimately cleaved from the modified core during post-translational processing, releasing the active compound (see example in Figure 3-1B). In 2009, multiple independent researchers reported this general scheme for thiopeptide synthesis (66, 72, 73, 128, 129). In the years since, several additional biosynthetic principles have been elucidated (77, 98, 130). The process appears to involve three core steps: (i)

cyclodehydration/oxidation of Cys residues to form thiazole heterocycles (Ser and Thr residues may also be modified to oxazoline or oxazole); (ii) dehydration of Ser and Thr residues yielding dehydroalanines and dehydrobutyrines, respectively; and (iii) cycloaddition of two of the dehydrated residues to yield the final macrocyclic thiopeptide scaffold (see Figure 3-1B).

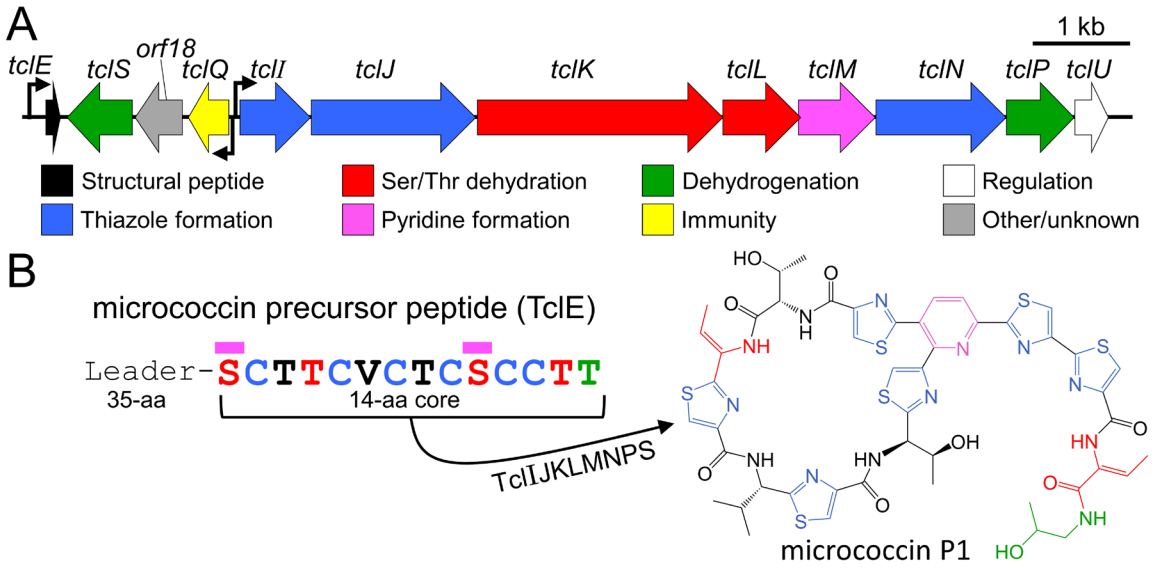


Figure 3-1. The *M. caseolyticus* str. 115 *tcl* gene cluster is responsible for production of the antibiotic MP1. (A) The *tcl* gene cluster from *M. caseolyticus* (GenBank accession number: KM613043) with predicted functions. Bent arrows represent putative promoters. (B) Sequence of the micrococcin precursor peptide TclE. The 14-amino acid core peptide gives rise to the bioactive thiopeptide MP1. Magenta bars indicate residues joined to form the macrocycle

Specific physical interactions between thiopeptide biosynthetic enzymes and cognate leader peptides are generally thought to license each processing enzyme to act relatively promiscuously on multiple substrate moieties within the core (42, 131, 132). This means that a single biosynthetic active site can work processively at multiple locations on the core peptide. For a typical thiopeptide gene cluster in nature, one often encounters numerous additional genes that obscure the simplicity of central thiopeptide chemistry. The recent *in vitro* reconstitution of the thiomuracin biosynthetic pathway illustrates this simplicity (77). In that study, five different

protein enzymes, along with an additional RRE-containing protein (RiPP precursor peptide Recognition Element), were found to be sufficient for the synthesis of a bioactive thiopeptide scaffold. This and other studies pursuing minimal thiopeptide biosynthetic pathways (126) inspire efforts to create easily manipulated *in vivo* platforms for the investigation of thiopeptide biology and for the derivation of novel antimicrobial compounds.

In this paper we address the biosynthesis of micrococcin, the founding member of the thiopeptide antibiotic family (93, 99, 133). The genetic basis of its synthesis was first discovered in the producer strain *Bacillus cereus* ATCC 14579 (66). The cluster of 24 *tcl* genes in this strain controls the production of a mixture of structurally similar compounds including several thiocillins (hence the *tcl* gene designation), micrococcin P1 (MP1), and micrococcin P2 (MP2). MP1 and MP2 share a common structure and intracellular target but differ only by a pair of hydrogen atoms, as will be described later. A number of incisive studies have delineated key players from the 24-gene *tcl* cluster and provided insights into how the Tcl processing machinery tolerates alterations in the core peptide sequence (67-69, 118). More recently, we have identified a second micrococcin gene cluster in a strain of *Macrococcus caseolyticus* str. 115 (76). This gene cluster is considerably smaller (12 genes, Figure 3-1) and was found to produce only a single thiopeptide product, MP1. We report here our effort to minimize and modularize the *M. caseolyticus* micrococcin gene cluster for expression in the laboratory model organism *Bacillus subtilis*, enabling rapid genetic analysis and pathway engineering.

3.3 Materials and Methods

Basic bacterial strains and culture conditions. Bacterial strains and plasmids used in this study are summarized in Table 3-1. The micrococcin-producing strain 115, previously identified as *S. epidermidis* (102, 103), was investigated by multi-locus sequence typing (MLST) and reassigned as *M. caseolyticus*. MLST primers and results are summarized in Tables 3-S1 and 3-S2. Plasmid DNA from *M. caseolyticus* 115 was used as template for *tcl* gene amplification. The thiopeptide expression chassis *B. subtilis* 168 was used for promoter analysis, and a micrococcin-resistant derivative (BS 168R) with a single-codon deletion in the *rplK* gene (ΔP22) was used for micrococcin expression experiments. The indicator strain used in all spot-on-lawn bioassays was *Staphylococcus aureus* ATCC 6538. All plasmids in this study were selected and maintained in *E. coli* DH5α. All strains listed above were maintained in liquid or solidified lysogeny broth (LB) unless described otherwise. Antibiotics used were ampicillin (Ap, 100 μg/ml), chloramphenicol (Cm, 5 μg/ml) and spectinomycin (Sp, 80 μg/ml). All bacterial cultures were grown at 37°C.

***B. subtilis* transformation.** Transformation of *B. subtilis* was accomplished using the previously described starvation-induced competence protocol (76). Following selection, colonies were verified for correct integration of transgenes using a modified colony PCR protocol, as follows: 1 ml of a saturated overnight culture was pelleted and resuspended in 200 μl of *B. subtilis* lysis buffer (50 mM Tris with 0.5 mg/ml lysozyme). Cells were then incubated at 30°C for 20 min and vortexed briefly before boiling for 3 min to complete cell lysis. Lysates were cleared by microcentrifugation (16,100 x *g*, 5 min), and 1 μl of supernatant was used as template for PCR with primers flanking the integration junctions.

Strain construction for *tcl* promoter analysis. To generate the reporter strains used in promoter analysis experiments, a two-step integration strategy was used. First, integration of the *thrC*::P_{xylA}-*tclU* insert was accomplished using a derivative of the *thrC* integrative plasmid pBS4S (134) in which the plasmid was modified to include the xylose-responsive P_{xylA} promoter and its cognate repressor (*xylR*) upstream of the putative regulatory gene *tclU* (full sequences provided in Figure 3-S1). The correctly assembled construct was transformed into *B. subtilis* 168 and the resulting strain was used as the recipient strain for subsequent transformations. To generate the promoter-*lacZ* fusion plasmids pLacZ_E, pLacZ_Q and pLacZ_I , the *amyE* integrative vector pDG1661 (135) was modified to include promoters P_{tclE}, P_{tclQ} and P_{tclI}. These promoters were amplified from *M. caseolyticus* plasmid DNA and ligated into pDG1661 within the polylinker upstream of the *spoVG-lacZ* reporter gene (full sequences provided in Figure 3-2). The promoter inserts consisted of approximately 500 bp of sequence immediately upstream of the start codons of the corresponding genes. Once constructed, each plasmid was transformed into the *B. subtilis thrC*::P_{xylA}-*tclU* strain described above to yield the strains used in β-galactosidase assays.

β-galactosidase assays. *B. subtilis* strains were grown overnight in LB-Sp, then diluted 1:100 into fresh antibiotic-free broth and cultured until late-exponential phase ($OD_{600} \approx 1.0$). Cultures were then placed on ice for 10 min to arrest growth before quantifying OD_{600} values. 150 µl of culture was added to 750 µl of buffer Z (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 20 mM KCl, 1 mM $MgSO_4$, 50 mM β-Mercaptoethanol, pH 7.0) before adding 10 µl of toluene to permeabilize cells. Suspensions were vortexed for 15 s, warmed to 30°C for 5 min, combined with 150 µl of o-nitrophenyl-β-galactoside solution (ONPG; 4 mg/ml in buffer Z) and incubated at 30°C for 30-40 min. Reactions were stopped by adding 400 µl of stop buffer (1M Na_2CO_3). Insoluble debris

was sedimented by microcentrifugation (16,100 x *g*, 5 min) before measuring OD_{420} values. A sample processed in identical manner but omitting cells was used as a blank. Miller units were calculated according to (136).

Construction of pLEGO and derivatives. To construct pLEGO, numerous alterations were made to the *amyE* integrative vector pDG1662 (135). In brief, the high-copy ColE1 replication origin was replaced with the lower-copy p15A origin from pACYC184 (137) followed by removal of the β-lactamase (*bla*) gene and destruction of many backbone restriction sites. The small polylinker within the integrating region of the plasmid was expanded to include the following restriction sites; BamHI, AvrII, SphI, KpnI, PstI, SacI, SalI, AatII, XhoI, BglII and EcoRI. Finally, the *xylR*-P_{xylA} promoter cassette from pHCMC04 (138) was introduced into the polylinker between BamHI and AvrII sites. We disrupted XhoI, EcoRI, and SalI sites within *xylR* in the process. This construct, termed pLEGO_proto, served as the base upon which pLEGO was constructed. To complete the assembly of pLEGO, the eight selected *tcl* processing genes (*tclIJKLMNPS*) were amplified from *M. caseolyticus* str. 115 and introduced between appropriate restriction sites in the pLEGO_proto polylinker. In several cases, mutagenic PCR was employed to disrupt restriction sites within *tcl* genes without altering the amino acid sequence of gene products. The complete sequence of pLEGO is provided in Figure 3-S3A. Epitope-tagged derivatives of pLEGO-encoded enzymes were made by re-amplifying *tcl* genes from pLEGO with primers designed to append the 8 amino-acid (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) *Strep*-tag (sequences given in Figure 3-S3B). For pLEGO derivatives lacking *tclP* or *tclS*, the genes were removed by digesting pLEGO at the appropriate flanking restriction sites followed by treatment with Klenow polymerase and religation.

Table 3-1. Strains and plasmids used in this study

Strain or plasmid[a]	Description[b]
Strains	
MC 115	*M. caseolyticus* 115. MP1 producer, source of all *tcl* genes used in this study
BS 168	*B. subtilis* 168. Heterologous host for promoter activity assays
BS 168R	Micrococcin-resistant (*rplK*ΔP22) derivative of BS 168 used for heterologous expression
SA 6538	*S. aureus* ATCC 6538. Indicator strain for spot-on-lawn bioassays
Plasmids	
pDG1661	*spoVG-lacZ* reporter for amyE integration; Cm^r in *B. subtilis*, Ap^r or Sp^r in E. coli
pLacZ_E	pDG1661 with P*tclE-spoVG-lacZ*
pLacZ_Q	pDG1661 with P*tclQ-spoVG-lacZ*
pLacZ_I	pDG1661 with Pt*clI-spoVG-lacZ*
pBS4S	*B. subtilis thrC* integration vector; Sp^r in *B. subtilis*, Ap^r in *E. coli*
pPxyl_U	pBS4S derivative with P*xylA-tclU*
pTclE	pBS4S derivative with Pstrong-*tclE*; Spr
pTclE_His	pTclE with His6-*tclE* (N-terminal tag)
pTclE_GST	pTclE with GST-*tclE* (N-terminal tag)
pDG1662	a*myE* integration vector; Cm^r in *B. subtilis*, Ap^r or Sp^r in *E. coli*
pLEGO_proto	pDG1662 derivative with p15A ori, P*xylA* promoter and new MCS; Cm^r
pLEGO	pLEGO_proto with P*xylA-tclIJKLMNPS*
pLEGO_strepI	pLEGO with Strep-*tclI* (N-terminal tag)
pLEGO_strepJ	pLEGO with Strep-*tclJ* (N-terminal tag)
pLEGO_strepK	pLEGO with Strep-*tclK* (N-terminal tag)
pLEGO_Kstrep	pLEGO with *tclK*-Strep (C-terminal tag)
pLEGO_strepL	pLEGO with Strep-*tclL* (N-terminal tag)
pLEGO_Lstrep	pLEGO with *tclL*-Strep (C-terminal tag)
pLEGO_strepM	pLEGO with Strep-*tclM* (N-terminal tag)
pLEGO_Nstrep	pLEGO with *tclN*-Strep (C-terminal tag)
pLEGO_strepP	pLEGO with Strep-*tclP* (N-terminal tag)
pLEGO_Pstrep	pLEGO with *tclP*-Strep (C-terminal tag)
pLEGO_strepS	pLEGO with Strep-*tclS* (N-terminal tag)
pLEGO_Sstrep	pLEGO with *tclS*-Strep (C-terminal tag)
pLEGO_ΔS	pLEGO with *tclS* removed
pLEGO_ΔP	pLEGO with *tclP* removed

[a] strains and plasmids not from this study are cited in Materials and Methods.
[b] Ap^r, ampicillin resistance; Cm^r, chloramphenicol resistance; Sp^r, spectinomycin resistance.

Construction of pTclE and derivatives. The plasmid pTclE was engineered as a derivative of the *thrC* integrative vector pBS4S (134). The strong constitutive *B. subtilis* promoter P_{strong} (iGEM biological parts registry BBa_K780003) was introduced between NotI and XbaI sites of pBS4S using hybridizing oligonucleotides, and the *tclE* gene was cloned downstream between XbaI and PstI sites to generate the final construct, pTclE. The complete sequence of pTclE is provided in Figure 3-S4A. Plasmid derivatives for appending hexahistidine (His_6) or glutathione S-transferase (GST) tags to the N-terminus of TclE were generated in a similar manner, by incorporating the tag in the primer (His_6) or using overlap extension PCR (GST). The GST tag was amplified from pGEX-4T-1 (GE Life Sciences, Marlborough, MA, USA). Detailed sequences are provided in Figure 3-S4B.

Micrococcin susceptibility bioassays. Methanolic extracts from *B. subtilis* strains expressing *tcl* genes were obtained as follows. After overnight growth in antibiotic-supplemented LB, cells were diluted 1:60 into 6 ml of LB supplemented with xylose (1% w/v). After an additional 8 h of growth, 4.8 ml of each culture was pelleted using iterative microcentrifugation (16,100 x *g*, 2 min for each run) and pellets were stored at -20°C prior to extraction. Extractions were accomplished by suspending thawed cell pellets in 1 ml of HPLC-grade methanol and vortexing vigorously for 3 min. Suspensions were left to incubate at room temperature for 5 min, followed by vortexing for an additional 3 min. Insoluble material was sedimented by microcentrifugation (16,100 x *g*, 5 min) and supernatants were transferred to fresh microcentrifuge tubes. These extracts were concentrated at room temperature to 80-µl volumes using a SpeedVac DNA 120 (Savant Instruments Inc., Holbrook, NY, USA) and stored at 4°C until assayed. The bioactivity of each extract was evaluated using a spot-on-lawn bioassay as described previously (76). The spot-on-lawn indicator strain *S. aureus* ATCC 6538 was grown overnight and diluted 1:20 into

fresh LB before spreading 200 µl onto LB-agar and allowing to dry. Once dry, 5-µl spots of concentrated extracts were placed upon bacterial lawns, allowed to dry and then incubated at 37°C. After 8 h, plates were assessed for zones of inhibition and imaged.

Extraction and electrospray ionization mass spectrometric (ESI-MS) analysis of MP1 and MP2. For each MP1- or MP2-producing strain, a single colony was grown overnight in LB-Sp and used to inoculate LB-Sp supplemented with 1% xylose (1 L, 37°C, 200 rpm, 16 h). The cells were harvested by centrifugation (7,878 x *g*, 15 min), resuspended in 25 ml of methanol, vortexed for 30 s, and gently agitated on a bench-top shaker for 20 min at room temperature. Sodium sulfate (10 g) was added to the suspended cells before centrifugation (16,420 x *g*, 5 min). The methanol fraction was evaporated to < 1 ml and 30 ml of ethyl acetate was added. The mixture was transferred to a separatory funnel, washed with 30 ml of water, and subsequently washed with 30 ml of saturated sodium chloride. The combined aqueous washes were back extracted with ethyl acetate. The ethyl acetate layers were combined, dried with sodium sulfate, filtered, and the volatiles were removed under reduced pressure. The resulting oily residue was re-dissolved in 50 µl of methanol, and 5 µl was transferred to a 96-well sample plate and evaporated under a stream of air. The residue was taken up in 20 µl of DMSO, injected (10 µl) into a Waters XBridge C18 column (3.5 µm, 4.6 x 50 mm), and separated using a Waters Alliance 2795 HPLC system (Waters, Milford, MA, USA). The following linear gradient was used: 5% to 100% B over 8 min (A = water/0.1% formic acid, B = acetonitrile/0.1% formic acid, 1 ml/min). The eluent was analyzed by UV absorbance (Waters 2996 diode array UV detector) and ESI+ mass analysis (Waters ZQ4000 quadrupole mass spectrometer). MP1 and MP2 products eluted between 7.3 and 7.6 min.

3.4 Results

Identification of essential elements of an engineered micrococcin expression system. To aid in the initial design of a minimized *B. subtilis* micrococcin expression system, we combined bioinformatic analyses (76), structural modeling and a detailed literature survey to identify essential elements within the *M. caseolyticus tcl* gene cluster (shown in Figure 3-1A). Homology arguments (66, 77, 96) suggest that TclI/J/N are involved in converting each of the core peptide Cys residues to thiazoles (Figure 3-1B, blue) with TclI as the Ocin-ThiF-like RRE containing protein, TclJ performing ATP-dependent heterocyclization (to thiazoline) and TclN oxidizing the thiazoline groups to thiazoles. TclK/L are likely responsible for the four Ser/Thr dehydration reactions (Figure 3-1B, red), with TclM catalyzing peptide macrocyclization via the two dehydrated serines (magenta). TclP and TclS have homology with short-chain dehydrogenases, and we hypothesized that these enzymes are involved in decarboxylation of the C-terminal Thr residue (Figure 3-1B, green) as has also been proposed for the homologous proteins in the *B. cereus tcl* cluster (66). TclQ is a ribosomal L11 homologue that we have shown to be involved in immunity (76), TclU is a putative transcriptional regulator (discussed below), and the function of the protein potentially encoded by *orf18* is unclear.

Identification of *tcl* promoter elements expressed in *B. subtilis*. The *tcl* gene cluster is organized into three apparent transcriptional units controlled by promoters upstream of *tclE*, *tclQ*, and *tclI* (see Figure 3-1A). The *tclE* unit is monocistronic, encoding only the TclE precursor peptide; the *tclQ* unit is a three-gene operon encoding TclQ, Orf18, and TclS; the *tclI* transcriptional unit encodes eight proteins, most of which are predicted to participate in post-translational processing of the TclE core peptide. The final gene in the *tclI* operon (*tclU*) encodes a putative MerR-type transcriptional regulator. Sequence alignments and structural modeling of

TclU (not shown) indicate the presence of a conserved helix-turn-helix DNA binding domain; however, the effector domain typically found in MerR regulators appears to be truncated or absent in TclU, making it difficult to predict its regulatory role with respect to *tcl* gene expression.

In transferring a functional *tcl* gene cluster into *B. subtilis*, transcriptional compatibility was a critical consideration. We sought to test whether the three *tcl* promoters described above are active in *B. subtilis* and whether TclU influences their activity. As an initial step, we employed a *lacZ* reporter assay to monitor activity of the three predicted *tcl* promoters in *B. subtilis*. 500-bp regions immediately upstream of the *tclE, tclQ* and *tclI* start codons were fused to the *spoVG-lacZ* reporter and these fusions were integrated into the *B. subtilis* chromosome. In each of these reporter strains, *tclU* (controlled by the xylose-inducible promoter P_{xylA}) was integrated at a second chromosomal locus (Figure 3-2A). Promoter strength was quantified on the basis of β-galactosidase activity in the presence or absence of TclU induction (Figure 3-2B). From these results we conclude that P_{tclE} and P_{tclQ} exhibit modest to strong activity, whereas P_{tclI} is inactive. TclU appears to play only a minor role, if any, in the regulation of these promoters in *B. subtilis*, justifying its exclusion from the heterologous expression system.

Genetic refactoring of the *tcl* gene cluster supports production of MP1 in *B. subtilis.* We next turned to the construction of a functional *tcl* gene cluster in the laboratory model organism *B. subtilis* str. 168. We sought to accomplish three primary objectives: (i) minimization of the ensemble of *tcl* genes required for micrococcin production, (ii) modularization of these genes to permit rapid manipulation of the pathway, and (iii) transcriptional optimization of the precursor peptide-encoding gene *tclE*, and the *tcl* biosynthetic operon.

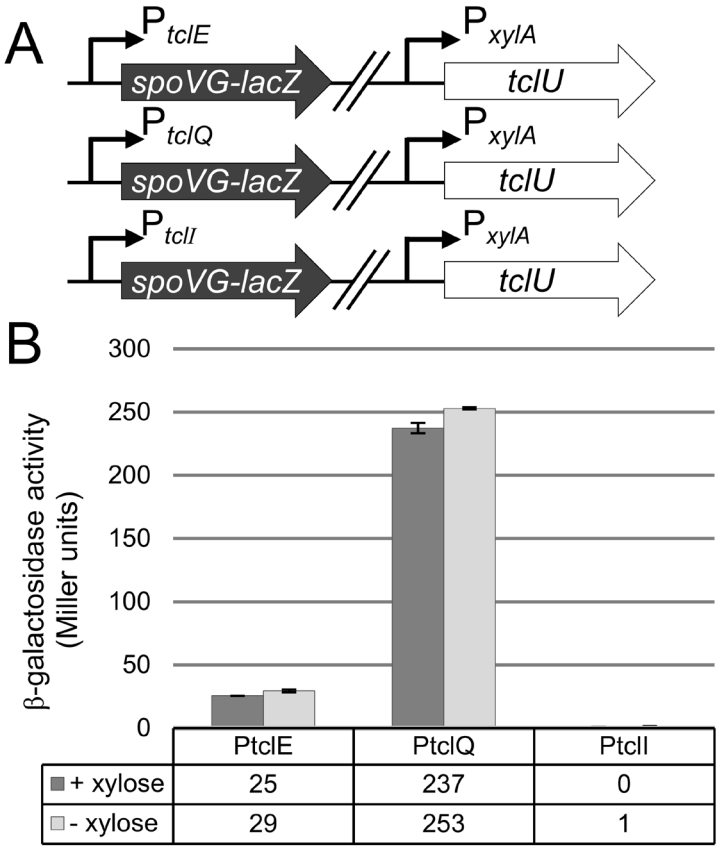


Figure 3-2. Analysis of *M. caseolyticus tcl* promoter activity in *B. subtilis* str. 168. (A) The promoter regions upstream of *tclE*, *tclQ*, and *tclI* were fused to a *lacZ* reporter and integrated into the *B. subtilis* chromosome. P*xylA*-*tclU* was similarly integrated at a separate locus for inducible expression of *tclU*. (B) *β*-galactosidase activity was measured to quantify *tcl* promoter strength in the presence (+ xylose) or absence (- xylose) of *tclU* expression. Error bars represent standard deviation from the mean, using three biological replicates per condition.

We decided on a two-plasmid system allowing independent manipulation of different components of the pathway. In a preliminary expression system, the naturally co-transcriptionally arranged genes *tclIJKLMNP* were introduced downstream of P_{xylA} in a vector designed for transgene integration at the *B. subtilis amyE* locus. The *tclU* gene was omitted in this construct. The remaining four genes (*tclE*, *tclS*, *orf18*, and *tclQ*) were introduced into a second vector designed for integration at the *thrC* locus. These genes were maintained in their original configuration (see Figure 3-1A) under the control of their native promoters (P_{tclE} and P_{tclQ}). Integration of these two preliminary plasmids into *B. subtilis*, followed by growth in the

presence of xylose, led to production of a methanol-extractable product with bioactivity against wild-type *B. subtilis*, based on a qualitative spot-on-lawn assay (not shown).

With the success of the proof-of-concept experiment described above, we further reduced and modularized our two-plasmid heterologous expression system. This system is summarized in Figure 3-3. For the *amyE* integration plasmid, the putative biosynthetic genes were all arranged as a single operon under the control of the xylose-inducible P_{xylA} promoter. To accomplish this, *tclS* was moved from the *thrC* integration plasmid described above, completing the eight-gene *tclIJKLMNPS* biosynthetic operon. This operon was punctuated by unique intergenic restriction sites, multiple overlapping regions between *tcl* open reading frames were resolved, and artificial ribosome binding sites were installed. As such, each biosynthetic gene constitutes a translationally independent unit that can be altered or removed while minimally affecting expression of neighboring genes. This final plasmid (pLEGO) is represented in Figure 3-3A and its full sequence is provided in Figure S3. The *thrC* integration plasmid was dedicated to *tclE* expression. This required removal of *tclS* (moved to pLEGO), *orf18* (a gene of unknown function), and the immunity gene *tclQ*. We alleviated the requirement for *tclQ* by selecting a spontaneous micrococcin-resistant *B. subtilis* mutant, which was used for all subsequent micrococcin expression experiments. This strain (BS 168R) was found to possess a single-codon deletion in *rplK*, which encodes the known ribosomal protein target of micrococcin (80). Given the relatively weak expression level of the native *tclE* promoter in *B. subtilis* (see Figure 3-2B), we replaced it with a strong constitutive promoter (P_{strong}). The resulting P_{strong}-*tclE* expression plasmid (termed pTclE) is represented in Figure 3-3B. Its complete sequence is given in Figure 3-S4.

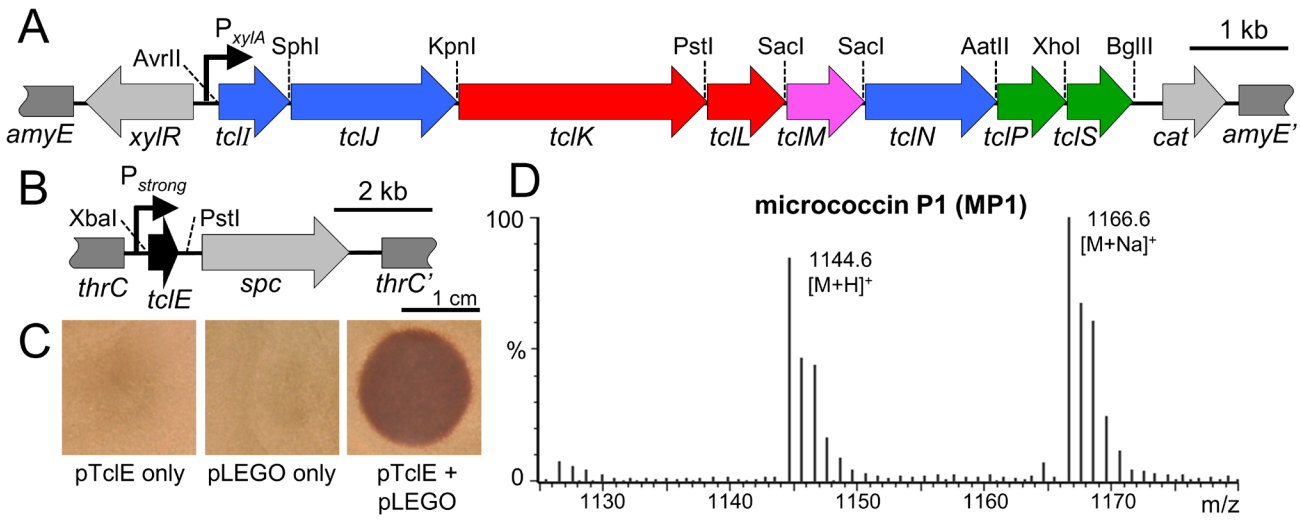


Figure 3-3. Minimization and modularization of the *tcl* biosynthetic pathway in *B. subtilis*. Gene map of the tcl region of pLEGO showing the xylose-inducible promoter (bent arrow), relevant processing genes (colored arrows), unique restriction sites, and homology regions (dark grey) for recombination into the *B. subtilis* chromosome. Light grey arrows indicate the xylose-responsive regulatory gene (*xylR*) and the chloramphenicol resistance determinant (*cat*). (B) Gene map of the TclE-encoding portion of pTclE showing the constitutive promoter (bent arrow), *tclE* (black arrow), homology regions for integration into the *B. subtilis* chromosome (dark grey) and the spectinomycin resistance determinant (*spc*, light grey). (C) Spot-on-lawn bioassays with methanolic extracts from strains transformed with pTclE only, pLEGO only, or both. (D) ESI-MS analysis of the bioactive extract from the pTclE + pLEGO strain. MP1: observed 1144.6 $[M+H]^+$, calculated 1144.2 $[M+H]^+$; observed 1166.6 $[M+Na]^+$, calculated 1166.2 $[M+Na]^+$

The pLEGO and pTclE plasmids were transformed into BS 168R and transformants were evaluated for micrococcin production (Figure 3-3C and D). Methanolic extracts from strains transformed with pTclE-only, pLEGO-only, or pTclE + pLEGO were tested for bioactivity against the *S. aureus* reference strain ATCC 6538. Only the strain with both transgenic components supported the production of a bioactive compound (see Figure 3-3C). Mass analysis of this methanol-extractable product was consistent with its identity as MP1 (expected *m/z* 1144.2 $[M+H]^+$; observed *m/z* 1144.6 $[M+H]^+$), in agreement with earlier work characterizing the product of this pathway (76). In addition, methanolic extracts from strains missing either TclI or TclP lacked bioactivity and were deficient in compounds detectable by MS within the expected mass range for MP1 and MP2 (not shown). Taken together, these results support the

identity of the antibacterial compound as MP1 and demonstrate the success of this highly engineered system in reconstituting the *M. caseolyticus* micrococcin pathway in *B. subtilis*.

Testing the utility of the heterologous, modularized genetic system. The successful reconstitution of this pathway in *B. subtilis* opens the door for detailed mechanistic studies of this pathway and its constituents. Such studies will inevitably require purification of substrate, enzymes or both, and would be benefitted by a better understanding of how Tcl proteins in this pathway tolerate N- or C-terminal affinity tags. To address this, we tagged each pathway component at one or both termini. As illustrated in Figure 3-1B, the TclE precursor peptide consists of a 35 amino-acid N-terminal leader (which presumably recruits the biosynthetic machinery) and the C-terminal 14-amino acid core peptide. Because this core peptide undergoes C-terminal decarboxylation as part of its maturation process, we did not attempt to fuse tags to this end of TclE. We were gratified to observe that fusions of both glutathione S-transferase (GST) and hexahistidine (His_6) to the N-terminus of TclE result in production of bioactive compound (Figure 3-4A), suggesting that such tags do not interfere with post-translational processing and would be suitable for purifying TclE intermediates at various stages of maturation. We also tested each of the biosynthetic Tcl proteins for their ability to tolerate the 8 amino-acid *Strep*-tag at their N- or C- termini. Based on the ability to produce biologically active micrococcin, all eight proteins are compatible with this tag in at least one orientation (Figure 3-4B). Of 12 *Strep*-tagged variants tested, only one (TclP-*Strep*) failed to support micrococcin biosynthesis. Taken together, these results demonstrate that every protein component in this pathway can be modified in a manner that would allow for affinity purification and detection with commercial reagents. These results also demonstrate the relative ease with which

alterations can be made to individual components in this expression system without perturbing the rest of the pathway.

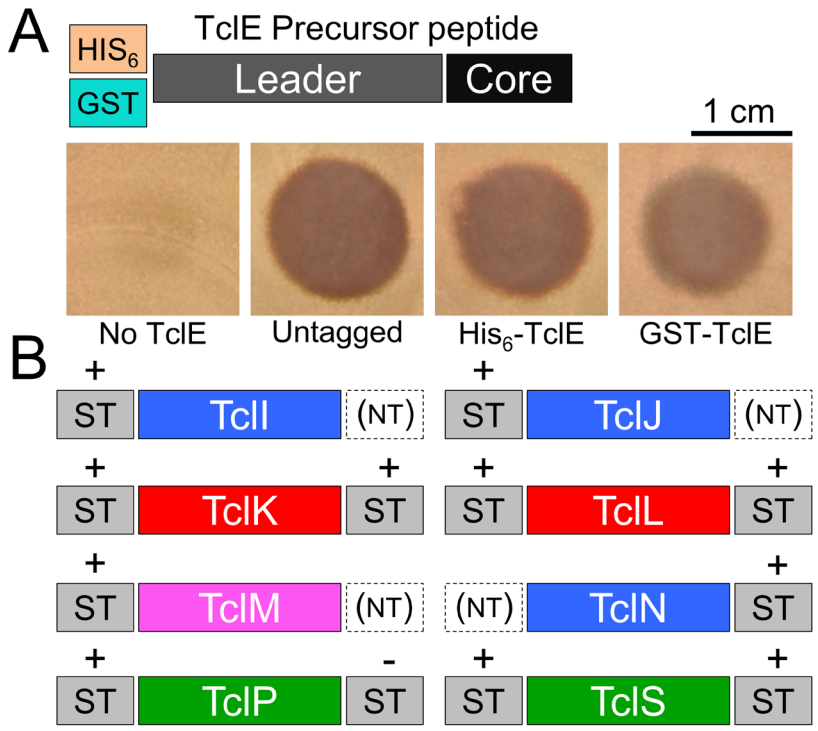


Figure 3-4. Each component of the micrococcin biosynthetic pathway is amenable to affinity tagging. (A) Graphical representation of the affinity tags fused to the N-terminus of the precursor peptide TclE. Spot-on-lawn assays were used to test the bioactivity resulting from tagged TclE. (B) Representations of the various locations in which Strep-tags (ST) were fused to Tcl processing proteins. The effect of these tags on bioactivity is also indicated; (+) indicates the fusion-protein was functional, (-) indicates the fusion abolished bioactivity, and (NT) indicates the tag was not tested.

Accumulation of MP1 versus MP2 is determined by the activity of the putative C-terminal dehydrogenase TclS. The modularity of this micrococcin production system also allows one to rapidly perform specific *tcl* gene removal experiments. Of the eight processing proteins in this system, six have predictable roles in thiazole installation (TclI/J/N), Ser/Thr dehydration (TclK/L) and macrocyclization (TclM). We preserved the two putative short-chain dehydrogenases (TclP and TclS) in our expression system with the assumption that they are somehow involved in decarboxylation of the core peptide C-terminal Thr residue; however, defined roles for these enzymes have remained unclear. To address this, either *tclP* or *tclS* was removed from pLEGO by simple restriction digestion and religation. The resulting plasmids were integrated into a *B. subtilis* strain background harboring the *GST-tclE* cassette. Methanol extracts from these strains were then tested for bioactivity by spot-on-lawn assay (Figure 3-5A) and the presence of thiopeptide compounds by ESI-MS. Interestingly, extract from the *tclP* deletion strain showed no bioactivity and no detectable thiopeptide product; however, the *tclS* deletion mutant produced a bioactive compound with a mass consistent with MP2 rather than MP1 (Figure 3-5B). MP2 has a methylketone group at the C-terminus, whereas MP1 is reduced to the alcohol at the same position. We conclude therefore that TclS is responsible for the conversion of MP2 to MP1 (see Figure 3-5C). TclP appears to play an essential role in micrococcin production, acting at some point prior to TclS in the biosynthetic pathway. Its presumed function as a short-chain dehydrogenase points to a possible role for TclP in the initial oxidative decarboxylation of the C-terminal Thr, as illustrated in Figure 3-5C.

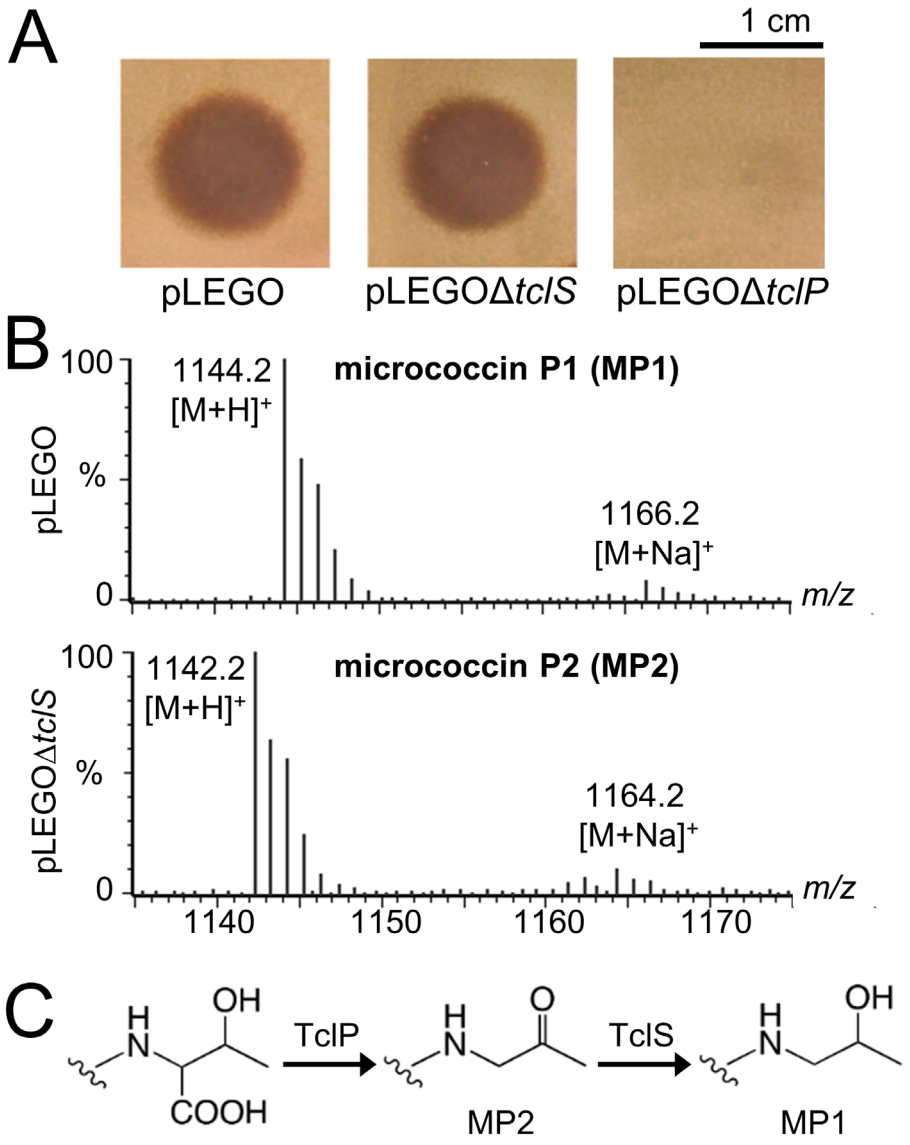


Figure 3-5. Analysis of the role of TclS in micrococcin C-terminal processing. (A) Spot-on-lawn assays with methanolic extracts from *B. subtilis* strains transformed with pLEGO and pLEGO-derivatives lacking *tclS* and *tclP*. In these experiments the pTclE plasmid had been modified to express the GST-TclE fusion. (B) ESI-MS analysis of purified extracts from pLEGO and pLEGOΔ*tclS* strains. MP1: observed 1144.2 $[M+H]^+$, calculated 1144.2 $[M+H]^+$. MP2: observed 1142.2 $[M+H]^+$, calculated 1142.2 $[M+H]^+$. (C) Proposed roles of TclP and TclS in C-terminal processing of MP1 and MP2.

3.5 Discussion

Heterologous thiopeptide expression serves several purposes: (i) it allows confirmation of the sufficiency of a given gene cluster to facilitate production of a specific thiopeptide product; (ii) with the gene cluster cloned on a plasmid, it allows the researcher to probe the biosynthetic logic of the pathway with greater facility; and (iii) it opens the door to the rational design and large-scale production of new thiopeptide analogs with improved pharmacological or bioactivity properties. Attempts to heterologously express diverse thiopeptide gene clusters have been met with mixed success (55, 56, 86, 126, 129). In most cases, *Streptomyces* species such as *S. lividans* and *S. coelicolor* have been employed. These expression hosts can be limiting due to slow growth rate and potential incompatibility with genes derived from distantly related organisms such as the low-GC firmicutes. It is remarkable that *B. subtilis* has rarely (if ever) been utilized for thiopeptide production. The *B. subtilis* strain used in this study is fast-growing and straightforward to genetically manipulate, with its natural competence, high recombinogenicity, and a wealth of genetic tools developed by a large community of researchers. *B. subtilis* is particularly attractive for the production of thiopeptides whose native producers (such as *Bacillus, Staphylococcus,* and *Macrococcus* spp.) share close evolutionary relatedness.

Several levels of potential incompatibility should be considered with respect to heterologous thiopeptide production. These include codon bias, transcriptional efficiency, self-intoxication by the bioactive product, and cofactor compatibility. Worthy of special mention is the $tRNA^{glu}$ glutamate donor that is employed by RiPP Ser/Thr dehydratase systems (TclK and TclL in the work presented here). Initially demonstrated for the two-domain NisB lantibiotic dehydratase (97), $tRNA^{glu}$ is presumably the source of the glutamyl group used to activate Ser/Thr residues for subsequent dehydration in many RiPP pathways. In the *in vitro* reconstituted

thiomuracin biosynthetic system (77), all thiopeptide biosynthetic components could be produced and purified from *E. coli*; however, the *E. coli* $tRNA^{glu}$ was not a suitable substrate for Ser/Thr dehydration, so a hybrid system utilizing the *E. coli* aminoacyl $tRNA^{glu}$ synthetase and *Thermobispora bispora* $tRNA^{glu}$ was required. $tRNA^{glu}$ is therefore a crucial consideration in transferring thiopeptide biosynthetic systems across bacterial taxa. In moving the micrococcin pathway from *M. caseolyticus* to *B. subtilis*, we were satisfied with the similar codon preferences of the two organisms and we were able to override transcriptional incompatibilities by using promoters with predictable behavior in the producer strain. We were encouraged by the close similarity in $tRNA^{glu}$ sequence (a 1-bp difference located in the acceptor stem, see Figure 3-S5), and successful micrococcin production in *B. subtilis* indicates there was not a problem with $tRNA^{glu}$ incompatibility.

Our functional genetic reconstruction of the *M. caseolyticus* micrococcin pathway in *B. subtilis* defines a minimally sufficient gene set for micrococcin biosynthesis. By preselecting a micrococcin-resistant variant of the producer strain we could omit immunity functions from the engineered gene cluster. We also excluded genes for micrococcin export; thus it is not surprising that micrococcin could be successfully extracted from *B. subtilis* cell pellets. It is not known whether our strains secrete residual micrococcin into the medium either passively or via endogenous efflux systems. Our data show that the minimal genetic requirement for conversion of TclE precursor peptide to micrococcin (MP2) is the seven-gene set *tclIJKLMNP*, with *tclS* responsible for reduction of MP2 to MP1. TclI, which we predict to be an Ocin-ThiF-like RRE containing protein, is required for micrococcin production (not shown), TclJ and TclN constitute a putative thiazole installation module, TclK and TclL constitute a Ser/Thr dehydration system,

TclM is predicted to catalyze the [4+2] cycloaddition required for macrocycle formation, with TclP likely involved in oxidative decarboxylation of the core peptide C-terminus.

The genetic refactoring described here enables unprecedented control over the properties of this system, as demonstrated by our affinity tagging of every protein component and our ability to easily remove the processing enzymes TclP and TclS. These experiments constitute the beginning of an ongoing series of tests that are allowing us to dissect the entire micrococcin biosynthetic pathway in great detail. These experiments capitalize on the ability to genetically block the pathway and rapidly purify the affinity-tagged precursor peptide TclE at various stages of maturation. This approach may be employed to assess the biosynthetic consequences of numerous *tcl* gene deletions and point mutations, all of which are easily incorporated using standard molecular cloning techniques. We hope to be able to leverage this system to study physical interactions between pathway components, to explore how well the biosynthetic proteins tolerate core peptide sequence variation, and to create a system for combinatorial biosynthesis of never-seen-by-nature bioactive compounds.

3.6 Acknowledgements

We would like to thank Paul Price and JC Price for useful discussions and insights. We acknowledge the *Bacillus* Genetic Stock Center for access to critical research tools. This work was funded by a grant from the BYU College of Life Sciences Vaccine Royalties Fund (to R.A.R.), a BYU Graduate Studies Fellowship Award (to P.R.B.), and private donations through the UCSF Foundation (to S.M.M.).

3.7 Supplemental Material

Supplemental Material for:

Reconstitution and Minimization of a Micrococcin Biosynthetic Pathway in *Bacillus subtilis*

Philip R. Bennallack[a], Kathryn D. Bewley[b], Mark A. Burlingame[b], Richard A. Robison[a], Susan M. Miller[b], and Joel S. Griffitts[a#]

[a]Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT, USA; [b]Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, CA, USA

#Correspondence should be addressed to Joel S. Griffitts, joelg@byu.edu

LIST OF SUPPLEMENTAL TABLES AND FIGURES

Table 3-S1: Primers used for multi-locus sequence analysis

Primer name	Primers Sequence (5' - 3')[a]	Product Size (bp)
groEL for	GAAATTGCTGGTGACGGTACGACAAC	601
groEL rev	GGTCACCAAATCCTGGTGCTT	
ef4 For	CAACGTTTAAAGCGTCAGGAACG	1267
ef4 Rev	CAGCACCCACATAATCATTCGG	
rplK for	GCACGTATATGAGTGGGAGGG	578
rplK rev	GGTTTTAGCGGAATTTCCTCCC	
ctaD for	GCCGTTACTGTTTGCATATATG	1664
ctaD rev	GCCATTATTTAACACCTCTTTCG	
dnaK for	GTCAATGGTTAAAGTAAACGAATG	1975
dnaK rev	TACTATATCAGTGTATCGCATGTT	
gyrR for	ATTCAATGTATCAGAAGGACAGA	2027
gyrR rev	CAATTTTGAATCATGGTCTAAAGC	

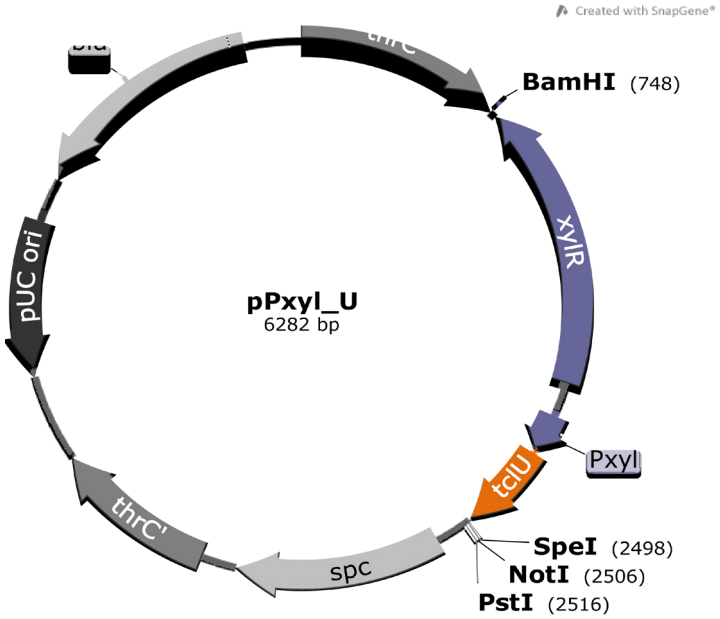
[a] Primers were designed using the *M. caseolyticus* JCSC5402 genome (Genbank ref: NC_011999.1)

Table 3-S2. BLASTp output from multi-locus sequence typing of strain 115

Target	Top Hits[a]	Coverage (%)	Identity (%)
groEL (heat shock protein 60)	1. *Macrococcus caseolyticus* JCSC5402	100	90
	2. *Staphylococcus caseolyticus* ATCC 13548	96	89
	3. *Macrococcus* sp. 1123N2A	95	89
ef4 (elongation factor 4)	1. *Macrococcus caseolyticus* JCSC5402	100	88
	2. *Bacillus thuringiensis* str. Bt185	90	77
	3. Bacillus cereus str. NJ-W	97	76
rplK (50S ribosomal protein L11)	1. *Macrococcus caseolyticus* JCSC5402	100	99
	2. *Staphylococcus haemolyticus* str. S167	96	88
	3. *Staphylococcus haemolyticus* str. Sh29/312/L2	96	88
ctaD (cytochrome C oxidase)	1. *Macrococcus caseolyticus* JCSC5402	99	89
	2. Bacillus sp. FJAT-22090	68	81
	3. Lysinibacillus sphaericus str. 2362	88	78
dnaK (mol. chaperone DnaK)	1. *Macrococcus caseolyticus* JCSC5402	100	95
	2. *Staphylococcus agnetis* str. 908	100	84
	3. *Staphylococcus hyicus* ATCC 11249	100	84
gyrB (DNA gyrase subunit B)	1. *Macrococcus caseolyticus* JCSC5402	100	93
	2. *Staphylococcus capitis* str. AYP1020	90	78
	3. *Staphylococcus hominis* str. DAR3383	94	77

[a] For simplicity, only top three BLASTp hits are shown

Figure 3-S1. Plasmid map and sequence of pPxyl_U used to generate *B. subtilis thrC*::*xylR*-P$_{xylA}$-*tclU-spc* mutants strains. Key features are annotated, as well as relevant unique restriction sites (bold). The complete nucleotide sequence of the plasmid is provided below. The sequence of *tclU* is distinguished by lowercase font and the sequence of P$_{xylA}$ is underlined.

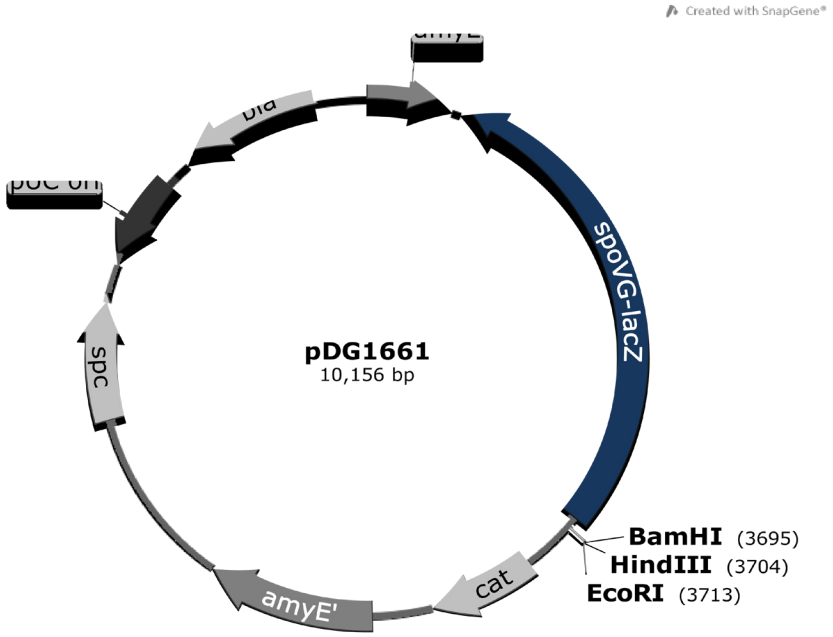


AATTCATGTAAAAGATGAGGTTGGTTCATTCTCGAAAATTACATCTGTGTTCTCAGAGCGGGGCGTGACGTTTGAAAAAATCCTTC
AGCTGCCAATTAAAGGCCATGATGAGTTAGCTGAAATCGTAATTGTCACACATCATACATCAGAAGCTGATTTCAGTGATATCCTG
CAAAACCTAAATGATTTGGAAGTCGTTCAAGAAGTCAAAAGCACATATCGTGTAGAAGGGAACGGTTGGAGCTAATGTGGAAAGGA
CTTATCCATCAATATAAAGAATTTTTACCTGTAACAGATCAAACACCGGCGCTAACTTTACATGAAGGAAACACACCTCTTATTCA
CCTGCCGAAGCTGTCTGAGCAGCTCGGAATTGAGCTTCATGTCAAAACGGAAGGCGTCAATCCTACGGGATCATTTAAAGATCGCG
GAATGGTTATGGCTGTGGCAAAGGCAAAAGAAGAAGGCAATGACACGATTATGTGCGCGTCAACAGGTAACACTTCCGCTGCTGCG
GCAGCATATGCAGCCCGTGCTAACATGAAATGCATTGTCATCATCCCGAACGGAAAAATTGCATTTGGAAAACTCGCTCAAGCTGT
CATGTACGGAGCCGAGATTATCGCAATTGACGGAAACTTTGACGATGCGCTTAAAATTGTCCGTTCCATCTGTGAGAAATCACCGA
TTGCCCTTGTCAACTCAGTCAACCCTTACCGCATTGAAGGCCAAAAAACTGCTGCCTTCGGATCCTCCTAACTTATAGGGGTAACA
CTTAAAAAAGAATCAATAACGATAGAAACCGCTCCTAAAGCAGGTGCATTTTTTCCTAACGAAGAAGGCAATAGTTCACATTTATT
GTCTAAATGAGAATGGACTCTAGAAGAAACTTCGTTTTTAATCGTATTTAAAACAATGGGATGAGATTCAATTATATGATTTCTCA
AGATAACAGCTTCTATATCAAATGTATTAAGGATATTGGTTAATCCAATTCCGATATAAAAGCCAAAGTTTTGAAGTGCATTTAAC
ATTTCTACATCATTTTTATTTGCGCGTTCCACAATCTCTTTTCGAGAAATATTCTTTTCTTCTTTAGAGAGCGAAGCCAGTAACGC
TTTTTCAGAAGCATATAATTCCCAACAGCCTCGATTTCCACAGCTGCATTTGGGTCCATTAAAATCTATCGTCATATGACCCATTT
CCCCAGAAAAACCCTGAACACCTTTATACAATTCGTTGTTAATAACAAGTCCAGTTCCAATTCCGATATTAATACTGATGTAAACG
ATGTTTTCATAGTTTTTTGTCATACCAAATACTTTTTCACCGTATGCTCCTGCATTAGCTTCATTTTCAACAAAAACCGGAACATT
AAACTCACTCTCAATTAAAAACTGCAAATCTTTGATATTCCAATTTAAGTTAGGCATGAAAATAATTTGCTGATGACGATCTACAA
GGCCTGGAACACAAATTCCTATTCCGACTAGACCATAAGGGGACTCAGGCATATGGGTTACAAAACCATGAATAAGTGCAAATAAA
ATCTCTTTTACTTCACTAGCGGAAGAACTAGACAAGTCAGAAGTCTTCTCGAGAATAATATTTCCTTCTAAGTCGGTTAGAATTCC
GTTAAGATAGTCGACTCCTATATCAATACCAATCGAGTAGCCTGCATTCTTATTAAAAACAAGCATTACAGGTCTTCTGCCGCCTC
TAGATTGCCCTGCCCCAATTTCAAAAATAAAATCTTTTTCAAGCAGTGTATTTACTTGAGAGGAGACAGTAGACTTGTTTAATCCT
GTAATCTCAGAGAGAGTTGCCCTGGAGACAGGGGAGTTCTTCAAAATTTCATCTAATATTAATTTTTGATTCATTTTTTTTACTAA
AGCTTGATCTGCAATTTGAATAATAACCACTCCTTTGTTTATCCACCGAACTAAGTTGGTGTTTTTTGAAGCTTGAATTAGATATT
TAAAAGTATCATATCTAATATTATAACTAAATTTTCTAAAAAAAACATTGAAATAAACATTTATTTTGTATATGATGAGATAAAGT
TAGTTTATTGGATAAACAAACTAACTCAATTAAGATAGTTGATGGATAAACTTGTTCACTTAAATCAAAGGAGGTGGAGCTCatgg
taaaaatatataccacaaaagaattaagcgacttccttggcatatctgccaaaattataagagattacgacaatttcggtctaatt
tctgtacaaagagacagtaacaattatagaatatttacttctgatgacgttagactcattgcctgtattagattattatctctata
tggaaaatcattccaagagataaaaacaataatttttaagtcaaatactgatgaaatccaagaatattctatcaattcaggaaaca
aaaaattaatagatataaataataaaatttcaattatagatgaaaaaattaaatctctccaatcaagtaacataatataaGGTACC
TCGACTAGTAGCGGCCGCTGCAGGATAAAAAATTTAGAAGCCAATGAAATCTATAAATAAACTAAATTAAGTTTATTTAATTAACA
ACTATGGATATAAAATAGGTACTAATCAAAATAGTGAGGAGGATATATTTGAATACATACGAACAAATTAATAAAGTGAAAAAAAT
ACTTCGGAAACATTTAAAAAATAACCTTATTGGTACTTACATGTTTGGATCAGGAGTTGAGAGTGGACTAAAACCAAATAGTGATC

TTGACTTTTTAGTCGTCGTATCTGAACCATTGACAGATCAAAGTAAAGAAATACTTATACAAAAAATTAGACCTATTTCAAAAAAA
ATAGGAGATAAAAGCAACTTACGATATATTGAATTAACAATTATTATTCAGCAAGAAATGGTACCGTGGAATCATCCTCCCAAACA
AGAATTTATTTATGGAGAATGGTTACAAGAGCTTTATGAACAAGGATACATTCCTCAGAAGGAATTAAATTCAGATTTAACCATAA
TGCTTTACCAAGCAAAACGAAAAAATAAAAGAATATACGGAAATTATGACTTAGAGGAATTACTACCTGATATTCCATTTTCTGAT
GTGAGAAGAGCCATTATGGATTCGTCAGAGGAATTAATAGATAATTATCAGGATGATGAAACCAACTCTATATTAACTTTATGCCG
TATGATTTTAACTATGGACACGGGTAAAATCATACCAAAAGATATTGCGGGAAATGCAGTGGCTGAATCTTCTCCATTAGAACATA
GGGAGAGAATTTTGTTAGCAGTTCGTAGTTATCTTGGAGAGAATATTGAATGGACTAATGAAAATGTAAATTTAACTATAAACTAT
TTAAATAACAGATTAAAAAAATTATAAAAAAAATTGAAAAAATGGTGGAAACACTTTTTTCAATTTTTTTGTTTTATTATTTAATAT
TTGGGAAATATTCATTCTAATTGGTAATCAGATTTTAGAAAACAATAAACCCTTGCATAGGGGGATCTCGAGGCCTTCCGAAAATG
CGCGGCTTTGAAGCTGAGGGTGCGGCGGCAATCGTGCGCAATGAAGTGATTGAAAATCCGGAAACAATAGCGACAGCCATTCGTAT
CGGAAACCCGGCAAGCTGGGACAAAGCTGTAAAGGCAGCCGAGGAATCCAATGGGAAAATTGACGAAGTCACTGATGATGAAATCC
TTCACGCATATCAGCTGATCGCCCGTGTAGAAGGCGTGTTTGCAGAACCAGGTTCTTGCGCGTCTATCGCAGGAGTGCTGAAACAG
GTGAAATCCGGAGAAATTCCGAAAGGCAGCAAGGTCGTAGCTGTGTTAACAGGAAACGGACTGAAAGATCCGAACACAGCGGTCGA
CATTTCAGAAATCAAGCCTGTCACATTGCCGACTGATGAAGACAGCATCCTTGAATATGTAAAAGGAGCGGCCCGTGTATGAACGA
AGCCGACATGCTGTTCTCTGTCACTGTTCCCGGAAGCACAGCTAACCTAGGCCCCGGCTTTGATTCAGTCGGAATGGCGCTCAGCA
GATATTTGAAGCTGACCGTCTTTGAAAGCGACAAATGGTCTTTTGAGGCTGAAACAGAAACAGTCGCCGGAATTAATTCGAGCTCC
TTTAACTTCCTCACCAGTAGTTGTATCGGTACCATAAGTAGAAGCAGCAACCCAAGTAGCTTTACCAGCATCCGGTTCAACCAGCA
TAGTAAGAATCTTACTGGACATCGGCAGTTCTTCGAACAGTGCGCCAACTACCAGCTCTTTCTACAGTTCATTCAGGGCACCGGAG
AACCTGCGTGCAATCCATCTTGTTCAATCATGCGAAACGATCCTCATCCTGTCTCTTGATCCATGGATTACGCGCGACCTTCAGCA
TCGCCGGCATGTCCCCCTGGCGGACGGGAAGTATCCAGCTCGAGGTCGGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC
CTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA
AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTC
TCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCG
ACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATT
TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGATAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC
GCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAA
ATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAG
CGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCC
AGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG
AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTT
TGCGCAACGTTGTTGCCATTGCTGCCGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGA
TCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGC
CGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTG
AGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCA
CATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGC
AAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTAT
CAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA
AGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAG
AATT

Figure 3-S2. Plasmid map and sequences of the constructs used for reporter assays. (A) The plasmid pDG1661, which was modified to produce the plasmids pLacZ_E, pLacZ_Q and pLacZ_I. These derivatives were used to generate the *B. subtilis* 168 *lacZ* reporter strains shown in Figure 3-2A. Shown on the plasmid map are key features including the *spoVG-lacZ* reporter gene and polylinker (bold). The complete nucleotide sequence of pDG1661 is provided below with *spoVG-lacZ* (3'–5') shown in lowercase font and the EcoRI, HindIII and BamHI sites in the polylinker are highlighted in yellow. (B) The *tcl* promoter sequences introduced into pDG1661. Restriction sites utilized are highlighted in yellow. Promoter sequences are in 5'-3' orientation.

A



ATGTTTGCAAAACGATTCAAAACCTCTTTACTGCCGTTATTCGCTGGATTTTTATTGCTGTTTCATTTGGTTCTGGCAGGACCGGC
GGCTGCGAGTGCTGAAACGGCGAACAAATCGAATGAGCTTACAGCACCGTCGATCAAAAGCGGAACCATTCTTCATGCATGGAATT
GGTCGTTCAATACGTTAAAACACAATATGAAGGATATTCATGATGCAGGATATACAGCCATTCAGACATCTCCGATTAACCAAGTA
AAGGAAGGGAATCAAGGAGATAAAAGCATGTCGAACTGGTACTGGCTGTATCAGCCGACATCGTATCAAATTGGCAACCGTTACTT
AGGTACTGAACAAGAATTTAAAGAAATGTGTGCAGCCGCTGAAGAATATGGCATAAAGGTCATTGTTGACGCGGTCATCAATCATA
CCACCAGTGATTATGCCGCGATTTCCAATGAGGTTAAGAGTATTCCAAACTGGACACATGGAAACACACAAATTAAAAACTGGTCT
GATCGATAGTACATAATGGATTTCCTTACGCGAAATACGGGCAGACATGGCCTGCCCGGTTATTATtatttttgacaccagaccaa
ctggtaatggtagcgaccggcgctcagctgaaattccgccgatactgacgggctccaggagtcgtcgccaccaatccccatatgga
aaccgtcgatattcagccatgtgccttcttccgcgtgcagcagatggcgatggctggtttccatcagttgctgttgactgtagcgg
ctgatgttgaactggaagtcgccgcgccactggtgtgggccataattcaattcgcgcgtcccgcagcgcagaccgttttcgctcgg
gaagacgtacggggtatacatgtctgacaatggcagatcccagcggtcaaaacaggcggcagtaaggcggtcgggatagttttctt
gcggccctaatccgagccagtttacccgctctgctacctgcgccagctggcagttcaggccaatccgcgccggatgcggtgtatcg
ctcgccacttcaacatcaacggtaatcgccatttgaccactaccatcaatccggtaggttttccggctgataaataaggttttccc
ctgatgctgccacgcgtgagcggtcgtaatcagcaccgcatcagcaagtgtatctgccgtgcactgcaacaacgctgcttcggcct
ggtaatggcccgccgccttccagcgttcgacccaggcgttagggtcaatgcgggtcgcttcacttacgccaatgtcgttatccagc
ggtgcacgggtgaactgatcgcgcagcggcgtcagcagttgttttttatcgccaatccacatctgtgaaagaaagcctgactggcg
gttaaattgccaacgcttattacccagctcgatgcaaaaatccatttcgctggtggtcagatgcgggatggcgtgggacgcggcgg
ggagcgtcacactgaggttttccgccagacgccactgctgccaggcgctgatgtgcccggcttctgaccatgcggtcgcgttcggt
tgcactacgcgtactgtgagccagagttgcccggcgctctccggctgcggtagttcaggcagttcaatcaactgtttaccttgtgg
agcgacatccagaggcacttcaccgcttgccagcggcttaccatccagcgccaccatccagtgcaggagctcgttatcgctatgac
ggaacaggtattcgctggtcacttcgatggtttgcccggataaacggaactggaaaaactgctgctggtgttttgcttccgtcagc
gctggatgcggcgtgcggtcggcaaagaccagaccgttcatacagaactggcgatcgttcggcgtatcgccaaaatcaccgccgta
agccgaccacgggttgccgttttcatcatatttaatcagcgactgatccacccagtcccagacgaagccgccctgtaaacggggat
actgacgaaacgcctgccagtatttagcgaaaccgccaagactgttacccatcgcgtgggcgtattcgcaaaggatcagcgggcgc
gtctctccaggtagcgaaagccatttttttgatggaccatttcggcacagccgggaagggctggtcttcatccacgcgcgcgtacat
cgggcaaataatatcggtggccgtggtgtcggctccgccgccttcatactgcaccgggcgggaaggatcgacagatttgatccagc
gatacagcgcgtcgtgattagcgccgtggcctgattcattccccagcgaccagatgatcacactcgggtgattacgatcgcgctgc

accattcgcgttacgcgttcgctcatcgccggtagccagcgcggatcatcggtcagacgattcattggcaccatgccgtgggtttc
aatattggcttcatccaccacatacaggccgtagcggtcgcacagcgtgtaccacagcggatggttcggataatgcgaacagcgca
cggcgttaaagttgttctgcttcatcagcaggatatcctgcaccatcgtctgctcatccatgacctgaccatgcagaggatgatgc
tcgtgacggttaacgcctcgaatcagcaacggcttgccgttcagcagcagcagaccattttcaatccgcacctcgcggaaaccgac
atcgcaggcttctgcttcaatcagcgtgccgtcggcggtgtgcagttcaaccaccgcacgatagagattcgggatttcggcgctcc
acagtttcgggttttcgacgttcagacgtagtgtgacgcgatcggcataaccaccacgctcatcgataatttcaccgccgaaaggc
gcggtgccgctggcgacctgcgtttcaccctgccataaagaaactgttacccgtaggtagtcacgcaactcgccgcacatctgaac
ttcagcctccagtacagcgcggctgaaatcatcattaaagcgagtggcaacatggaaatcgctgatttgtgtagtcggtttatgca
gcaacgagacgtcacggaaaatgccgctcatccgccacatatcctgatcttccagataactgccgtcactccaacgcagcaccatc
accgcgaggcggttttctccggcgcgtaaaaatgcgctcaggtcaaattcagacggcaaacgactgtcctggccgtaaccgaccca
gcgcccgttgcaccacagatgaaacgccgagttaacgccatcaaaaataattcgcgtctggccttcctgtagccagctttcatcaa
cattaaatgtgagcgagtaacaacccgtcggattctccgtgggaacaaacggcggattgaccgtaatgggataggttacgttggtg
tagatgggcgcatcgtaaccgtgcatctgccagtttgaggggacgacgacagtatcggcctcaggaagatcgcactccagccagct
ttccggcaccgcttctggtgccggaaaccaggcaaagcgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgc
gggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccggtcg
acccgtaatcttacgtcagtaacttccacAGTAGTTCACCACCTTTTCCCTATATAAAAGCATTAGTGTATCAACAAGCTGGGGAT
CCTAGAAGCTTATCGAATTCTCATGTTTGACAGCTTATCATCGGCAATAGTTACCCTTATTATCAAGATAAGAAAGAAAAGGATTT
TTCGCTACGCTCAAATCCTTTAAAAAAACACAAAAGACCACATTTTTTAATGTGGTCTTTATTCTTCAACTAAAGCACCCATTAGT
TCAACAAACGAAAATTGGATAAAGTGGGATATTTTTAAAATATATATTTATGTTACAGTAATATTGACTTTTAAAAAAGGATTGAT
TCTAATGAAGAAAGCAGACAAGTAAGCCTCCTAAATTCACTTTAGATAAAAATTTAGGAGGCATATCAAATGAACTTTAATAAAAT
TGATTTAGACAATTGGAAGAGAAAAGAGATATTTAATCATTATTTGAACCAACAAACGACTTTTAGTATAACCACAGAAATTGATA
TTAGTGTTTTATACCGAAACATAAAACAAGAAGGATATAAATTTTACCCTGCATTTATTTTCTTAGTGACAAGGGTGATAAACTCA
AATACAGCTTTTAGAACTGGTTACAATAGCGACGGAGAGTTAGGTTATTGGGATAAGTTAGAGCCACTTTATACAATTTTTGATGG
TGTATCTAAAACATTCTCTGGTATTTGGACTCCTGTAAAGAATGACTTCAAAGAGTTTTATGATTTATACCTTTCTGATGTAGAGA
AATATAATGGTTCGGGGAAATTGTTTCCCAAAACACCTATACCTGAAAATGCTTTTTCTCTTTCTATTATTCCATGGACTTCATTT
ACTGGGTTTAACTTAAATATCAATAATAATAGTAATTACCTTCTACCCATTATTACAGCAGGAAAATTCATTAATAAAGGTAATTC
AATATATTTACCGCTATCTTTACAGGTACATCATTCTGTTTGTGATGGTTATCATGCAGGATTGTTTATGAACTCTATTCAGGAAT
TGTCAGATAGGCCTAATGACTGGCTTTTATAATATGAGATAATGCCGACTGTACTTTTTACAGTCGGTTTTCTAATGTCACTAACC
TGCCCCGTTAGTTGAAGAAGGTTTTTATATTACAGCTCCAGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACA
GGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGG
CGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCA
ACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACATGGATGAGCGATGATGATATCC
GTTTAGGCTGGGCGGTGATAGCTTCTCGTTCAGGCAGTACGCCTCTTTTCTTTTCCAGACCTGAGGGAGGCGGAAATGGTGTGAGG
TTCCCGGGGAAAAGCCAAATAGGCGATCGCGGGAGTGCTTTATTTGAAGATCAGGCTATCACTGCGGTCAATAGATTTCACAATGT
GATGGCTGGACAGCCTGAGGAACTCTCGAACCCGAATGGAAACAACCAGATATTTATGAATCAGCGCGGCTCACATGGCGTTGTGC
TGGCAAATGCAGGTTCATCCTCTGTCTCTATCAATACGGCAACAAAATTGCCTGATGGCAGGTATGACAATAAAGCTGGAGCGGGT
TCATTTCAAGTGAACGATGGTAAACTGACAGGCACGATCAATGCCAGGTCTGTAGCTGTGCTTTATCCTGATGATATTGCAAAAGC
GCCTCATGTTTTCCTTGAGAATTACAAAACAGGTGTAACACATTCTTTCAATGATCAACTGACGATTACCTTGCGTGCAGATGCGA
ATACAACAAAAGCCGTTTATCAAATCAATAATGGACCAGACGACAGGCGTTTAAGGATGGAGATCAATTCACAATCGGAAAAGGAG
ATCCAATTTGGCAAAACATACACCATCATGTTAAAAGGAACGAACAGTGATGGTGTAACGAGGACCGAGAAATACAGTTTTGTTAA
AAGAGATCCAGCGTCGGCCAAAACCATCGGCTATCAAAATCCGAATCATTGGAGCCAGGTAAATGCTTATATCTATAAACATGATG
GGAGCCGAGTAATTGAATTGACCGGATCTTGGCCTGGAAAACCAATGACTAAAAATGCAGACGGAATTTACACGCTGACGCTGCCT
GCGGACACGGATACAACCAACGCAAAAGTGATTTTTAATAATGGCAGCGCCCAAGTGCCCGGTCAGAATCAGCCTGGCTTTGATTA
CGTGCTAAATGGTTTATATAATGACTCGGGCTTAAGCGGTTCTCTTCCCCATTGAGGGCAAGGCTAGACGGGACTTACCGAAAGAA
ACCATCAATGATGGTTTCTTTTTTGTTCATAAATCAGACAAAACTTTTCTCTTGCAAAAGTTTGTGAAGTGTTGCACAATATAAAT
GTGAAATACTTCACAAACAAAAAGACATCAAAGAGAAACATACCCTGCAAGGATGCTGATATTGTCTGCATTTGCGCCGGAGCAAA
CCAAAAACCTGGTGAGACACGCCTTGAATTAGTAGAAAAGAACTTGAAGATTTTCAAAGGCATCGTTAGTGAAGTCATGGCGAGCG
GATTTGACGGCATTTTCTTAGTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATG
CCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCT
AACTTCGATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCG
CCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTACTGAAGTGGATTTCTTTAAGAGCTC
CTTTAACTTCCTCACCAGTAGTTGTATCGGTACCATAAGTAGAAGCAGCAACCCAAGTAGCTTTACCAGCATCCGGTTCAACCAGC
ATAGTAAGAATCTTACTGGACATCGGCAGTTCTTCGAACAGTGCGCCAACTACCAGCTCTTTCTGCAGTTCATTCAGGGCACCGGA
GAACCTGCGTGCAATCCATCTTGTTCAATCATGCGAAACGATCCTCATCCTGTCTCTTGATCCATGGATTACGCGTTAACCCGGGC
CCGCGGATGCATATGATCAGATCTTAAGGCCTAGGTCTAGAGGATCGATCTGTATAATAAAGAATAATTATTAATCTGTAGACAAA
TTGTGAAAGGATGTACTTAAACGCTAACGGTCAGCTTTATTGAACAGTAATTTAAGTATATGTCCAATCTAGGGTAAGTAAATTGA
GTATCAATATAAACTTTATATGAACATAATCAACGAGGTGAAATCATGAGCAATTTGATTAACGGAAAAATACCAAATCAAGCGAT
TCAAACATTAAAAATCGTAAAAGATTTATTTGGAAGTTCAATAGTTGGAGTATATCTATTTGGTTCAGCAGTAAATGGTGGTTTAC
GCATTAACAGCGATGTAGATGTTCTAGTCGTCGTGAATCATAGTTTACCTCAATTAACTCGAAAAAAACTAACAGAAAGACTAATG
ACTATATCAGGAAAGATTGGAAATACGGATTCTGTTAGACCACTTGAAGTTACGGTTATAAATAGGAGTGAAGTTGTCCCTTGGCA
ATATCCTCCAAAAAGAGAATTTATATACGGTGAGTGGCTCAGGGGTGAATTTGAGAATGGACAAATTCAGGAACCAAGCTATGATC
CTGATTTGGCTATTGTTTTAGCACAAGCAAGAAAGAATAGTATTTCTCTATTTGGTCCTGATTCTTCAAGTATACTTGTCTCCGTA
CCTTTGACAGATATTCGAAGAGCAATTAAGGATTCTTTGCCAGAACTAATTGAGGGGATAAAAGGTGATGAGCGTAATGTAATTTT

AACCCTAGCTCGAATGTGGCAAACAGTGACTACTGGTGAAATTACCTCGAAAGATGTCGCTGCAGAATGGGCTATACCTCTTTTAC
CTAAAGAGCATGTAACTTTACTGGATATAGCTAGAAAAGGCTATCGGGGAGAGTGTGATGATAAGTGGGAAGGACTATATTCAAAG
GTGAAAGCACTCGTTAAGTATATGAAAAATTCTATAGAAACTTCTCTCAATTAGGCTAATTTTATTGCAATAACAGGTGCTTACTT
TTCTGGAGTTCTTTAGCAAATTTTTTTATTAGCTGAACTTAGTATTAGTGGCCATACTCCTCCAATCCAAAGCTATTTAGAAAGAT
TACTATATCCTCAAACAGGCGGTAACCGGCCTCTTCATCGGGAATGCGCGCGACCTTCAGCATCGCCGGCATGTCCCCCTGGCGGA
CGGGAAGTATCCAGCTCGAGGTCGGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAC
GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCT
CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTA
GGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCC
AGTTACCTTCGGAAAAAGAGTTGATAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGT
TAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCA
TAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAC
CCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGC
CTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTG
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCC
ATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTAT
GGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAG
AATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTC
ATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACC
CAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAA
GGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGA
TACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAAC
CATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTAACAAAATTCTCCAGTCTTC
ACATCGGTTTGAAAGGAGGAAGCGGAAGAATGAAGTAAGAGGGATTTTTGACTCCGAAGTAAGTCTTCAAAAAATCAAATAAGGAG
TGTCAAGA

B pLacZ_E Insert (5’-3’):

...AAGCTTGTCTTGATGATGGGTCGTCCAAGAATAGATTCTGTGATTTTATCCCGAATAATATTGATAAAAAAATTATTCAG
GATAAAATCAATAAGTCAAACAATAAAAAATAGAAACAAATGTTGCTTATCACTGGTATCTAGGATTCAGATTTCAAGATAAG
ATTCCTCATTTCACTAACTTCGGTAAAAATTATATTAGAAGATTCAGTGATAATACACTTTTCGAAGAAAAATTTGAACACAC
TCTATAGTAATTTGGATTATTAAAAATATCTAAATAAATAATAAAATACAAAATTATACTATTCATTACATAACGAAAGAAAA
ATATAAATGGAGAAGAATTCAATTCCGTTTGAACATACTTTCTTATTTAACGGAAAAAATGTTCTATTTTTGTAAGTATTATG
TAAATTGTTGAAACTTAGAGACATTTCTGATAAATTTAAAGAAGGCATAATAATATATGCCTTCAAGGAGGTGGAAAAAATGGG
GATCC...

pLacZ_Q Insert (5’-3’):

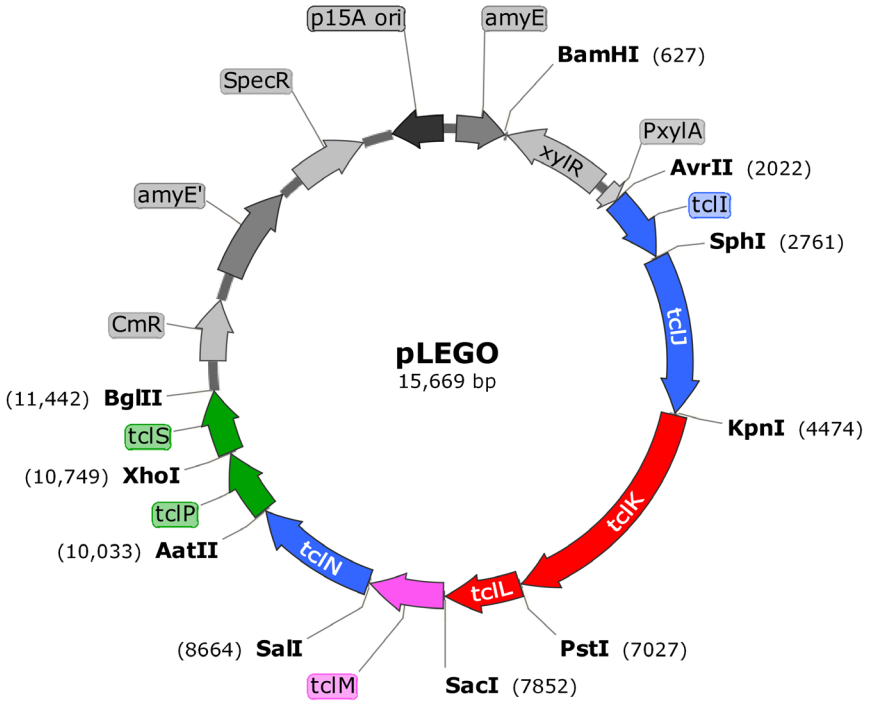
...AAGCTTAGTGTTTGCAAACTCATTTTCCATGAGATTATATGTCTTTATGTCTATTTTATTATTATTTAAGAATGCATCTTTAG
ATAATTGTATGTTTTTTGAATTCAAATATAATATTTTGATTTCAAGTTCACTTAAATTATCTAGTCTGTTTTCAGAAGCATAAATA
AAGTTTTTTTCTTTCAAAACATTTATAATATTTTTACAAAAAAGAGCAACTTTTTTATTTTTAATATTTTCTATAATTTGATTTTC
TGTATAGGAACCATTGAGATACTCAATTAAGTTTTCAAAGACATAGTAACTTTTATGAGTTTTTACTTCTAAATTTGTAAAATTAT
TTTTGAATAAAATACCTTCATTATGTGTTTTTATGTAAAATGTATCTTTTTTAAATAGTAAATACATATTAGCCCTCTTTCTTTTT
CTTGATTTTTTAGTTACACTAGTGTATAGTTCTTTAGGATATATTTAAATAGAAAATTATTGAAAGTGAGTGACGAATAATGGGAT
CC...

pLacZ_I Insert (5’-3’):

...GAATTCCTAGCAGTACCTTCAATCATTGACATTGCTGTTTCAATAGAACTTGCATTCAGATCAGGCATTTTTATTTCTGC
AATCTCTTGTAAATCTTTCTTTGAAATACTTCCTACAATTTCTTTGTTTGGAACACCAGATCCTTTATTGATATGTGCTTTCT
CTTTTATTAATTCAGCTGTGGGAGGAGTTTTTAGTATAAAGTTAAAGCTTCGATCTTCATATACCGTAATTTCTGCAGGAATA
ACCCCAGACTTTTCTTTAGTCTGAGCATTATATTGTGTTACAAACTCCATAATATTAACTCCTGCTGGGCCTAATGCAGTACC
TACCGGAGGTGCTGGAGAAGCTTTTCCAGCTTCTAATTGAATATTAATAATTTGCTTTATAGCTTTTGCCATTATTCGTCACT
CACTTTCAATAATTTTCTATTTAAATATATCCTAAAGAACTATACACTAGTGTAACTAAAAAATCAAGAAAAAGAAAGAGGGC
TAATATGGGATCC..

Figure 3-S3. Plasmid map and sequences of pLEGO and its tagged derivatives. (A) pLEGO was engineered for xylose-inducible expression of the Tcl processing machinery. The genes encoding processing enzymes are colored by predicted function (see Figure 3-1A) and are each flanked by unique restriction sites (bold). The nucleotide sequence of pLEGO is shown below. All *tcl* genes are distinguished by lowercase font and adjacent restriction sites are highlighted in yellow. (B) Sequence context (5'-3') of all successfully installed N- and C-terminal *Strep*-tags. In each case, the nucleotide sequence of the *Strep*-tag is underlined, *tcl* coding sequences are distinguished by lowercase font and relevant restriction sites are highlighted in yellow. For N-terminal tags, only the 5' end of the gene is shown. For C-terminal tags, the 3' end is shown.

A



GCTCCAGTCTTCACATCGGTTTGAAAGGAGGAAGCGGAAGAATGAAGTAAGAGGGATTTTTGACTCCGAAGTAAGTCTTCAAAAAA
TCAAATAAGGAGTGTCAAGAATGTTTGCAAAACGATTCAAAACCTCTTTACTGCCGTTATTCGCTGGATTTTTATTGCTGTTTCAT
TTGGTTCTGGCAGGACCGGCGGCTGCGAGTGCTGAAACGGCGAACAAATCGAATGAGCTTACAGCACCGTCGATCAAAAGCGGAAC
CATTCTTCATGCATGGAATTGGTCGTTCAATACGTTAAAACACAATATGAAGGATATTCATGATGCAGGATATACAGCCATTCAGA
CATCTCCGATTAACCAAGTAAAGGAAGGGAATCAAGGAGATAAAAGCATGTCGAACTGGTACTGGCTGTATCAGCCGACATCGTAT
CAAATTGGCAACCGTTACTTAGGTACTGAACAAGAATTTAAAGAAATGTGTGCAGCCGCTGAAGAATATGGCATAAAGGTCATTGT
TGACGCGGTCATCAATCATACCACCAGTGATTATGCCGCGATTTCCAATGAGGTTAAGAGTATTCCAAACTGGACACATGGAAACA
CACAAATTAAAAACTGGTCTGATCGGATCCTCCTAACTTATAGGGGTAACACTTAAAAAAGAATCAATAACGATAGAAACCGCTCC
TAAAGCAGGTGCATTTTTTCCTAACGAAGAAGGCAATAGTTCACATTTATTGTCTAAATGAGAATGGACTCTAGAAGAAACTTCGT
TTTTAATCGTATTTAAAACAATGGGATGAGATTCAATTATATGATTTCTCAAGATAACAGCTTCTATATCAAATGTATTAAGGATA
TTGGTTAATCCAATTCCGATATAAAAGCCAAAGTTTTGAAGTGCATTTAACATTTCTACATCATTTTTATTTGCGCGTTCCACAAT
CTCTTTTCGAGAAATATTCTTTTCTTCTTTAGAGAGCGAAGCCAGTAACGCTTTTTCAGAAGCATATAATTCCCAACAGCCTCGAT
TTCCACAGCTGCATTTGGGTCCATTAAAATCTATCGTCATATGACCCATTTCCCCAGAAAAACCCTGAACACCTTTATACAATTCG
TTGTTAATAACAAGTCCAGTTCCAATTCCGATATTAATACTGATGTAAACGATGTTTTCATAGTTTTTTGTCATACCAAATACTTT
TTCACCGTATGCTCCTGCATTAGCTTCATTTTCAACAAAAACCGGAACATTAAACTCACTCTCAATTAAAAACTGCAAATCTTTGA
TATTCCAATTTAAGTTAGGCATGAAAATAATTTGCTGATGACGATCTACAAGGCCTGGAACACAAATTCCTATTCCGACTAGACCA
TAAGGGGACTCAGGCATATGGGTTACAAAACCATGAATAAGTGCAAATAAAATCTCTTTTACTTCACTAGCGGAAGAACTAGACAA
GTCAGAAGTCTTTTCGAGAATAATATTTCCTTCTAAGTCGGTTAGGATTCCGTTAAGATAGTCCACTCCTATATCAATACCAATCG
AGTAGCCTGCATTCTTATTAAAAACAAGCATTACAGGTCTTCTGCCGCCTCTAGATTGCCCTGCCCCAATTTCAAAAATAAAATCT
TTTTCAAGCAGTGTATTTACTTGAGAGGAGACAGTAGACTTGTTTAATCCTGTAATCTCAGAGAGAGTTGCCCTGGAGACAGGGGA
GTTCTTCAAAATTTCATCTAATATTAATTTTTGATTCATTTTTTTTACTAAAGCTTGATCTGCAATTTGAATAATAACCACTCCTT

TGTTTATCCACCGAACTAAGTTGGTGTTTTTTGAAGCTTGAATTAGATATTTAAAAGTATCATATCTAATATTATAACTAAATTTT
CTAAAAAAAACATTGAAATAAACATTTATTTTGTATATGATGAGATAAAGTTAGTTTATTGGATAAACAAACTAACTCAATTAAGA
TAGTTGATGGATAAACTTGTTCACTTAAATCAAAGGAGGTGGACCTAGGatgtatttactatttaaaaaagatacattttacataa
aaacacataatgaaggtattttattcaaaaataattttacaaatttagaagtaaaaactcataaaagttactatgtctttgaaaac
ttaattgagtatctcaatggttcctatacagaaaatcaaattatagaaaatattaaaaataaaaaagttgctctttttttgtaaaaa
tattataaatgttttgaaagaaaaaaacttttatttatgcttctgaaaacagactagataatttaagtgaacttgaaatcaaaatat
tatatttgaattcaaaaaacatacaattatctaaagatgcattcttaaataataataaaatagacataaagacatataatctcatg
gaaaatgagtttgcaaacactatgttaaaagaatattatttctctcctagcgagacagcaaatataaaaatatcactcggaaaatt
tgagagcgacattggctactatataatacctgacggagaatatttaatagtggggaaaagtaaaaaacagcacacacacacagtgt
cactatccacttatgaaatcccactacacgcatggttaatatgtctcaacattttattaagcgaactattcttatattgtacaagt
atatattctgaagataaattcgatgaaaattattacaagtttaatgtccaaacttttgaaggtgatttttttaaaaaagagaataa
gtaaGCATGCAAGGAGGTGTGTCTCatggaaatttttgagagtcctgaatttaatattattcgctatttatctaattcttatattt
ttaaatccacttccaaagaaccagacttttttaaattcaataaaagagcgaatattattaaatacaaatgatattatttataaaaaa
aatatacccgaaagaaatatttcatttcttagtagaatagaaaattctatagggaatttttatatgcaatcaaattagcgaatacga
ttacattatatatatatataatcataaaactaataaaattaaaaaagtaaattactttatatctcctcactactataagaaattac
caaatgatactactaattgttttaatatcataataaatgaattgaaaaattctaaatcacataactttagatcaaggaatatagac
agtatttataatcaattagataaatatttttttggataaaaacttaggaatttgtaatatgctcctagataactatgatggacccctt
cccaattagtgtagcaatgctaccattagataacggaaaagaagaacctggggtaggtagaactagaaaaatacaacgtagtagag
cagtagctctcttagaagcctatgaaagatattcagggcttgaacctagaggtaagaaaacaataaatcatacagaagatttaaaa
tcaaaaaaagttaacatgaactctttaattctacataataatccatttttaatttctcatggaattaaaaattcaaatttcactta
taacgatttactagacaatgagatcagttgggtcaaatgcctaaatttaaacactttccaaacactactaattcctgaacaatatg
cttactacggaattaacatctctaaccataaagaaaaggctataaatatagcttatgagatatctaatggttgctccgttggtaat
aactatttggaagttgtatattatggcctaatggaagttatcgagagagactcttttctctgtagttggtactttaatacaccaaa
agataaaatatcactgaaaaatgcaagtacctccataaaaaacctaatccaacaattcacaagttattataatgactataaattag
aattattttatttatataatgaatttaatattcctgtagtacttgctaccgtcacattaaaagaatcttcaacgaaaaaaatgaac
ttcatgtgtgctgctgctgctgatataaatatagaagacgcaatagaaaaatctatacatgaaataggtggtattctttttggact
aaataaaaaatttattgacaggtatcatgaattggaagcgatacgcaagaataatttagatgttaaaacaatggaagatcatactt
tagtctatggccttccagaacatagaacatatatccaacaaaaaatgaattacgaaaatatatacgattatgataaagaattaact
ccaaaaatattttacaaagaagtacaaaaattgataaaaaaaataagtacaactaaagatatattattagtggaccaaacaccatt
aatatctaaaaaaatagatttaaaggtaggtaaaattattgtacctggattattacctatgactttcggaaaatataatatcagag
taagtgaaaatagatatcacgaattatgtcatttctattcaaaagatttaattattgatcttaacccccatccatttccttaaGGT
ACCAAGGAGGTGTGTCTCatgataaaaaaaatgccttatagtataacgcgtaaatgtaacagattgaactcttttatagatataaa
catgaactttgaagaatttaaaaatacagtagaaaaaaaaaaattaagtatggaaaatgctaaagaacgtacaaaacaatttttat
atacagcagttaatcaagaaaaaaatgaagctaatcgaaaatatataattgacttaaaaagaaacatagataaaatatcatacgaa
aagctagataaatacaaacaaattaatattaatcatccaaattatcaacaatatatagaagatttacatgatatacttgaatatcg
aaattttgatttttctaaagaatataaaataatattagataaacatttacaagataatagagaagccgtattacagtcctttgaat
gcgaatattttaaaaatggagtagctatgtccaatagaaacattttaaaaaactattataagtatgtagtaaataataataataag
aaaagtaggatgacacaacaatccataaaaaattactacaatagaatgaatttaaaaccttctccatttagtacatttactactat
agaaataaatatagaggattatattgtctcagataataaacattcttatgaagtttattcagagtataatgaaacttattataaaa
taatagagcttttctttattagaattagacgctttagataatatcaaaaaaataaaactcaattcttcaatcagtaaaactaaagat
aattatcattatattaatgttgagatgacacctaaatatatgtactacaaagaaaaagagatgatattgaaagaaaatataacaat
caacgaatgtatacatataataaaaaaagaatgctttaattattctaagtcaatgaaatctattgaaaaaaatttaggttcaaaag
gcgttaatacatttctatatttgttaaaaaataatttcttatatctgagctttggaatagatgattatgagaaagacaaaattaac
aaactgtctgatatgataaaaactacacaaactcaagatttaacaataataaaagtaatacaaaatctagataagctatcacatct
acttaaccaaataaataacaaacctttacatataaagaggcactggtagaaaatatttatttacacactatagatttgttgaatt
taattggtaaacaagacattattaatactattgaaagtggtagtattatttatcaaaacaatattgatacacaaataacatttcct
aaagaaaaattagaggttagttttgagactgttgatattattcatagtattttcaggctatttgatgacaacattatagaaaaaat
atttatgaaaaagttcttcgaaaaaaattataataacactgttaaatttaaagatttttctaaacatttcaattctgtatataatg
agaaaaattgggaaatgattaaaaaaactacagaattacaaaccattaaaaaattgcgcgaagaatttttcacatatattaaaaat
aaaaagataaattctaaagatattaacttagatactcaatggcttaaaaagtttataagtacatttccagactatttagattgctg
ggaaagtttttctatgtattttcaatataacaaagaagataactattatattttaaatgatattggaccaggaatgggcagacata
tattgcgatatactagaaatttagacaataaaaatttgaaattctttactgatacatataaaaagaatattagtgaaattgagaat
agcaaaaagaaaaaatatgccgatatcagctctacattatctttaaacattaatttaaagagttgtagtttaaatactcaaattaa
ttatcctactaataataataaaaataactatatatatccgatctttttgtcacacaaaaaaatggtgtattaatacttgttgatg
aaaaaggtacggaaatcgatgttacacccactggtttttttatttccagctttagctcctgaaacatataaaatgttatgtcgattc
tcattctctaatggtgttcaagtaaatttatgggatagatactcttcatttattgactgtaatgaagaacattttcctaatatatt
tataaacaacgttttgattgaaagagaagtatttaatgagaaggcttctgatttacaattattattagagaacaatacctacgaaa
tgatggacttaaacgaatacttgtccaataaaattggaataaaagaacctttggaatttttctctaagtttacatacgatatagat
gccttaataaatgaaaatataataccagaagactggataaaacttattaaaaactccaaattaaggaaacctcaatactacaatct
caataactatctagacttcaaaaatttctcaaatattataaaaagtgtcgatagcgaacaatacgtaaaaattcaaaaagtgaatc
ctcgtcttaatgattcaaatgaatatatatacgatttttataaagaaggcgtatgaCTGCAGAAGGAGGTGTGTCTCatgaattgg
acaagcctacatatctttttattttgataataacaaagaactattagtaaaatgtatttataacatgtatttaaataatttttttga
taaatttttctacattaattattgggacggaggtcctcatatccgtctaagaatagctaatattacaaagaatgaaaaagaaataa

tcataaagacaatacaaaaatttattcaagaaaatccatctctttcaaaaataagtgaaaaagattaccttattacatcaaataat
tttgctgacaaagaaaattctgaaatttagagctgcaaaaaaataacactatactagaaatagaatataagcctgagatagacaa
atatttaaataatgaaggagtgcatatttcagaagatatttttttttatttcttctctaaattctctaacttatttaaaaaattctc
ctaataaagaaatggtttatatgcactcactacagtttgcaaactatatttttaaaatatttcttaaatactaaacaaacagtgttg
ttcttaaaagaatatgaaagatattggcattctttttcaaatagtgatattaaatccactaaacttattagaatacctcataaaaa
attcgcagaatttgatgatccatttttcaaaaataaaagcaaatttcactcatagcgaacaaatggggttcatttttaattatattc
atttaacgaacaatagattaggaatcaaacctttttgaagaagctattttatcaagtatattaactaaattttgggggggatgaatat
gaaaacaaaagtaataaatgaGAGCTCAAGGAGGTGTGTCTCatgaaaacaaaagtaataaatgatttttttattattgagtaatca
ctctaatgataatagtttttattttttagaaaatgataaattagaaatatttaatattaataatagcgacttaaagaaaactaaaa
gttttatgggtgataaaagtgacgaatacttatcaatctatttaaataaactaaatgacttctatgagaatatgatgtatctacaa
gttaataattattcagtattccaaacagagttatttaaatttatgattaactattcagaatttaattatgaaagtttagaaagagg
aatgatttcttattgctctcattcagaaggtttcttaagcatccctaaaaatcaaaagttcaaaaaaatttttaaagaaggatatt
taaaaaacgaacacgttttagacttaattattaataaccgaaaagattcatttttttacacatatcatattgataccataatctct
gaattaaaaccatgtattagaaactctattaaaaaaaatgaaattcacttttgaatattgatcactcaaaaaataacgaccaatt
aactagtgattttcaccaacatatgctttccaacgaaaaatttttaaaatttatgagatgtgatatagactttctaacttctagat
ttcttacaatagctcaatacttttattaaaaaacatgggaataagtaatattaacagatatttcacttgctatttaacatataaa
tctctatccaatttcacatctaaaaatccaaatgatttaattaaatatttcaaagaagactagGTCGACAAGGAGGTGTGTCTCat
ggacatctctaagttcttgtataatttacactataatccaggagaagttgtgtcagcatcctatacaatagaagacactattcaaa
ggaatagcgagggtttttataaaggatatggcatagatttttaaaattacaacaaaaaagccctatcgtaaaagttattcttaaa
agttacggtgatattttcttcaatagggtagaaaataagaaaaaacctcttatcttttgtagaaagatgacgccgtcaggtggtgg
tttatatccaattaatatttttatttgtacaaattttaagaatcgcattgcattatttcaatttgactttaaaaggaacttattaa
tatttataaaatatattaatattgaaataaataatgaatgtactaaattgtacttagttccttgttatactaggaattattttaaa
tataaagaattctcctatagactatgtcctctagatactggttatttaatctcaactcttctatataattttagtgttgagaatat
tacattcaaacttttctataaaattaaacaaaaatagcgatattacagatgttttaaacgaaataggatgcgaagaaattccatatt
ctatcattgaactaaatgaaagtttaaatttagataatttaagtttagaacattatgatacggagtcttacttttttaatcctaat
aaagtaagaaatctattagaaatagacactcttatacatcaagaataccataaagatataaatatcaattttaacaacgaaaataa
attatttgaaaaattcgaaatacaaaaaagaatatctccgggtggcgaatttattcaaaattccaaagtagaacaagaatctataa
ataagttcatatctttaataatgcagtataaaaataaaagtaatttttctttctgagtacatactattaaatctgatagatgtacag
aataaacgcataattaacttatctgctagtgaatttttaagctataaaaataacgtttctattgaattcatagacaaacaacttac
cagaagaaattttaacttattagcagtcccttatattctttacgtaggggtgaatgaggaaaaaattaaagagaattatagcaaca
attattttaagatttccagaataatagcaggattctggagtggagtagtaagtatactatctgctcaatgcggtctatctacccac
ccaatgatgagttataatgcaagagaattagaagaatatatttttaagaatagatattctatattaaatcaaattgtaataggtgg
taatattacaacaaatagaatggatagtatgctaatccggagtgatttataaGACGTCAAGGAGGTGTGTCTCatgaatatactaa
ttgtaggagcttctagcgaaattgcacattatataattaactatcataaaattaaggatcaagtatttctacttgatcttccttct
cagatagataatctaaaaaaatgggatactaactttgatacactagacgtacaaaataataaagaaattgaaacttattttaaaga
atgtaatataacatttgataaattatactatttagttggaataaatactatgaaaaatggtttagatttaattcacaggaatggg
acaatataatggggactaatctcaaaagcttttatttctttgtgaaggaattcacaaagaaaaacgtaattaataatatacctgct
accatagttagtattgcatctcaacatggcgtcgtagctaatgcctatagaactccttattgtgtaagtaaagctggtttaataca
tttaaccagagttttagcactagaactatcattatatgacattagggttaattgtgtctctcctggctttattttaaattccaaaa
gtcatgaattcttaaataatcctaaagtaaaaaaagaatatttaagcaaaaccccctctccaaagatatatcacacctaacgaagta
gcaaattgctgtatttttttaaataattctacatcaataactggacaaaatttaataatcgatggcggttatacaatatggtaaCT
CGAGAAGGAGGTGTGTCTCatggataattgtataatatatggaggaacaagtcctattgctttacaatttattaaacataatatca
ataaaatgaatatttttgtttttgtccgaaataaagataggtttattgaacagttatcagattatacacctttaaaaaataatatt
gatatatttgaaggtaatatattatcagttgaagacattaataatttttttaattatctttcaacaaagataattaaaattgatat
atttatcaatttaatatctattatgtctgtaaatagatttgaggaagatttattagctactagaaaaattattgatacaaatgtgt
atggatcttcttatatattacaaaagagtatccccttctttggtgaaggatctagtgtgatatatattagttctcatacttccgac
atagcccctgttaatggagtattatacgctgccagcaagagctatatgaatactttgttgaaaggaatttacaatgaatctattta
ttaccaaaaaccgattaggatactaaatatactagctggtccaattgatactgcattctataataattctctttctgcaaatataa
agaaaatacaagtaatagagatagtagattctataatatctttcctaaataatagtattacattttgtgagataaaattacttcct
taaAGATCTCAGGAATTCTCATGTTTGACAGCTTATCATCGGCAATAGTTACCCTTATTATCAAGATAAGAAAGAAAAGGATTTTT
CGCTACGCTCAAATCCTTTAAAAAAACACAAAAGACCACATTTTTTAATGTGGTCTTTTATTCTTCAACTAAAGCACCCATTAGTT
CAACAAACGAAAATTGGATAAAGTGGGATATTTTTAAAATATATATTTATGTTACAGTAATATTGACTTTTAAAAAAGGATTGATT
CTAATGAAGAAAGCAGACAAGTAAGCCTCCTAAATTCACTTTAGATAAAAATTTAGGAGGCATATCAAATGAACTTTAATAAAATT
GATTTAGACAATTGGAAGAGAAAAGAGATATTTAATCATTATTTGAACCAACAAACGACTTTTAGTATAACCACAGAAATTGATAT
TAGTGTTTTATACCGAAACATAAAACAAGAAGGATATAAATTTTACCCTGCATTTATTTTCTTAGTGACAAGGGTGATAAACTCAA
ATACAGCTTTTAGAACTGGTTACAATAGCGACGGAGAGTTAGGTTATTGGGATAAGTTAGAGCCACTTTATACAATTTTTGATGGT
GTATCTAAAACATTCTCTGGTATTTGGACTCCTGTAAAGAATGACTTCAAAGAGTTTTATGATTTATACCTTTCTGATGTAGAGAA
ATATAATGGTTCGGGGAAATTGTTTCCCAAAACACCTATACCTGAAAATGCTTTTTCTCTTTCTATTATTCCATGGACTTCATTTA
CTGGGTTTAACTTAAATATCAATAATAATAGTAATTACCTTCTACCCATTATTACAGCAGGAAAATTCATTAATAAAGGTAATTCA
ATATATTTACCGCTATCTTTACAGGTACATCATTCTGTTTGTGATGGTTATCATGCAGGATTGTTTATGAACTCTATTCAGGAATT
GTCAGATAGGCCTAATGACTGGCTTTTATAATATGAGATAATGCCGACTGTACTTTTTACAGTCGGTTTTCTAATGTCACTAACCT
GCCCCGTTAGTTGAAGAAGGTTTTTATATTACAGCTCCAGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAG
GTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGC

GTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGGCGGCCGCTCGACATGGATGAGCGATGA
TGATATCCGTTTAGGCTGGGCGGTGATAGCTTCTCGTTCAGGCAGTACGCCTCTTTTCTTTTCCAGACCTGAGGGAGGCGGAAATG
GTGTGAGGTTCCCGGGGAAAAGCCAAATAGGCGATCGCGGGAGTGCTTTATTTGAAGATCAGGCTATCACTGCGGTCAATAGATTT
CACAATGTGATGGCTGGACAGCCTGAGGAACTCTCGAACCCGAATGGAAACAACCAGATATTTATGAATCAGCGCGGCTCACATGG
CGTTGTGCTGGCAAATGCAGGTTCATCCTCTGTCTCTATCAATACGGCAACAAAATTGCCTGATGGCAGGTATGACAATAAAGCTG
GAGCGGGTTCATTTCAAGTGAACGATGGTAAACTGACAGGCACGATCAATGCCAGGTCTGTAGCTGTGCTTTATCCTGATGATATT
GCAAAAGCGCCTCATGTTTTCCTTGAGAATTACAAAACAGGTGTAACACATTCTTTCAATGATCAACTGACGATTACCTTGCGTGC
AGATGCGAATACAACAAAAGCCGTTTATCAAATCAATAATGGACCAGAGACGGCGTTTAAGGATGGAGATCAATTCACAATCGGAA
AAGGAGATCCATTTGGCAAAACATACACCATCATGTTAAAAGGAACGAACAGTGATGGTGTAACGAGGACCGAGAAATACAGTTTT
GTTAAAAGAGATCCAGCGTCGGCCAAAACCATCGGCTATCAAAATCCGAATCATTGGAGCCAGGTAAATGCTTATATCTATAAACA
TGATGGGAGCCGAGTAATTGAATTGACCGGATCTTGGCCTGGAAAACCAATGACTAAAAATGCAGACGGAATTTACACGCTGACGC
TGCCTGCGGACACGGATACAACCAACGCAAAAGTGATTTTTAATAATGGCAGCGCCCAAGTGCCCGGTCAGAATCAGCCTGGCTTT
GATTACGTGCTAAATGGTTTATATAATGACTCGGGCTTAAGCGGTTCTCTTCCCCATTGAGGGCAAGGCTAGACGGGACTTACCGA
AAGCTCGATCGAGCCACCTACTGAAGTGGATTTCTTTAAGCTAGAGGATCGATCTGTATAATAAAGAATAATTATTAATCTGTAGA
CAAATTGTGAAAGGATGTACTTAAACGCTAACGGTCAGCTTTATTGAACAGTAATTTAAGTATATGTCCAATCTAGGGTAAGTAAA
TTGAGTATCAATATAAACTTTATATGAACATAATCAACGAGGTGAAATCATGAGCAATTTGATTAACGGAAAAATACCAAATCAAG
CGATTCAAACATTAAAAATCGTAAAAGATTTATTTGGAAGTTCAATAGTTGGAGTATATCTATTTGGTTCAGCAGTAAATGGTGGT
TTACGCATTAACAGCGATGTAGATGTTCTAGTCGTCGTGAATCATAGTTTACCTCAATTAACTCGAAAAAAACTAACAGAAAGACT
AATGACTATATCAGGAAAGATTGGAAATACGGATTCTGTTAGACCACTTGAAGTTACGGTTATAAATAGGAGTGAAGTTGTCCCTT
GGCAATATCCTCCAAAAAGAGAATTTATATACGGTGAGTGGCTCAGGGGTGAATTTGAGAATGGACAAATTCAGGAACCAAGCTAT
GATCCTGATTTGGCTATTGTTTTAGCACAAGCAAGAAAGAATAGTATTTCTCTATTTGGTCCTGATTCTTCAAGTATACTTGTCTC
CGTACCTTTGACAGATATTCGAAGAGCAATTAAGGATTCTTTGCCAGAACTAATTGAGGGGATAAAAGGTGATGAGCGTAATGTAA
TTTTAACCCTAGCTCGAATGTGGCAAACAGTGACTACTGGTGAAATTACCTCGAAAGATGTCGCTGCTGAATGGGCTATACCTCTT
TTACCTAAAGAGCATGTAACTTTACTGGATATAGCTAGAAAAGGCTATCGGGGAGAGTGTGATGATAAGTGGGAAGGACTATATTC
AAAGGTGAAAGCACTCGTTAAGTATATGAAAAATTCTATAGAAACTTCTCTCAATTAGGCTAATTTTATTGCAATAACAGGTGCTT
ACTTTTCTGGAGTTCACTAGTGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAG
TGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCCTCGCTCACTG
ACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATA
CTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCA
AATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGC
CTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGC
TCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGA
AAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAA
ACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAA
ACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATC
AGATAAAATATTTCTAG

B Sequence context (5'-3') of all successfully installed N-terminal *Strep*-tags.

Strep-tclI: …CCTAGGatgtcatggagccatccgcaatttgaaaaaggcgcaggcgcatatttacta…

Strep-tclJ: …GCATGCAAGGAGGTGTGTCTCatgtcatggagccatccgcaatttgaaaaaggcgcaggcgcagaaatttttt…

Strep-tclK: …GGTACCAAGGAGGTGTGTCTCatgtcatggagccatccgcaatttgaaaaaggcgcaggcgcaataaaaaaa…

Strep-tclL: …CTGCAGAAGGAGGTGTGTCTCatgtcatggagccatccgcaatttgaaaaaggcgcaggcgcaaattggaca…

Strep-tclM: …GAGCTCAAGGAGGTGTGTCTCatgtcatggagccatccgcaatttgaaaaaggcgcaggcgcaaaaacaaaa…

Strep-tclP: …GACGTCAAGGAGGTGTGTCTCatgtcatggagccatccgcaatttgaaaaaggcgcaggcgcaaatatacta…

Strep-tclS: …CTCGAGAAGGAGGTGTGTCTCatgtcatggagccatccgcaatttgaaaaaggcgcaggcgcagataattgt…

Sequence context (5'-3') of all successfully installed C-terminal *Strep*-tags:

tclK-Strep: …gaaggcgtaGGCGCAGGCGCATGGAGCCATCCGCAATTTGAAAAAtaaCTGCAG…

tclL-Strep: …agtaataaaGGCGCAGGCGCATGGAGCCATCCGCAATTTGAAAAAtaaGAGCTC…

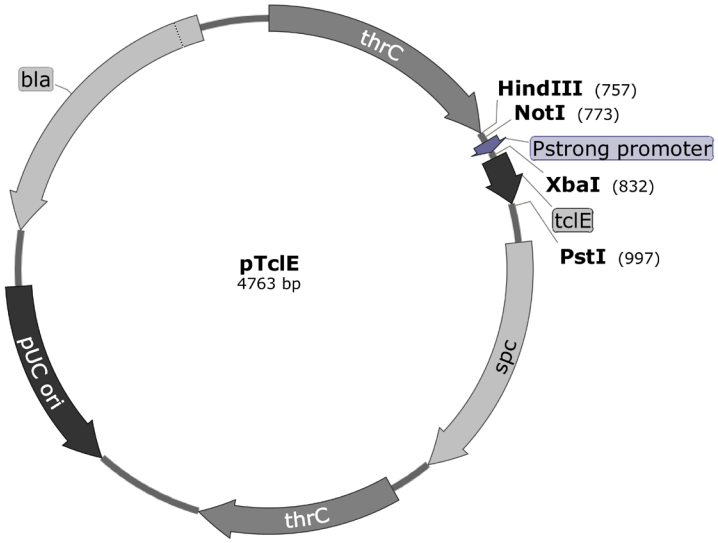
tclN-Strep: …agtgatttaGGCGCAGGCGCATGGAGCCATCCGCAATTTGAAAAAtaaGACGTC…

tclP-Strep: …acaatatggGGCGCAGGCGCATGGAGCCATCCGCAATTTGAAAAAtaaCTCGAG…

tclS-Strep: …ttacttcctGGCGCATGGAGCCATCCGCAATTTGAAAAAtaaAGATCT…

Figure 3-S4. Plasmid map and sequence of pTclE and its derivatives. (A) pTclE was used to introduce *tclE* into the *thrC* locus of *B. subtilis* 168R. The plasmid map shows key features including the P_{strong} promoter and the micrococcin precursor gene, *tclE* (black arrow). The complete plasmid sequence is shown below with the P_{strong} promoter underlined, *tclE* distinguished in lowercase font, and useful restriction sites highlighted yellow. (B) Relevant portions of the *His_6-tclE* and *GST-tclE* inserts in the pTclE derivatives pTclE_His and pTclE_GST. In each case the upstream restriction site is highlighted in yellow, affinity tags are underlined, and the first three codons of *tclE* are shown. The GST tag includes a TEV cleavage site (italicized) for tag removal.

A



AATTCATGTAAAAGATGAGGTTGGTTCATTCTCGAAAATTACATCTGTGTTCTCAGAGCGGGGCGTGACGTTTGAAAAAATCCTTC
AGCTGCCAATTAAAGGCCATGATGAGTTAGCTGAAATCGTAATTGTCACACATCATACATCAGAAGCTGATTTCAGTGATATCCTG
CAAAACCTAAATGATTTGGAAGTCGTTCAAGAAGTCAAAAGCACATATCGTGTAGAAGGGAACGGTTGGAGCTAATGTGGAAAGGA
CTTATCCATCAATATAAAGAATTTTTACCTGTAACAGATCAAACACCGGCGCTAACTTTACATGAAGGAAACACACCTCTTATTCA
CCTGCCGAAGCTGTCTGAGCAGCTCGGAATTGAGCTTCATGTCAAAACGGAAGGCGTCAATCCTACGGGATCATTTAAAGATCGCG
GAATGGTTATGGCTGTGGCAAAGGCAAAAGAAGAAGGCAATGACACGATTATGTGCGCGTCAACAGGTAACACTTCCGCTGCTGCG
GCAGCATATGCAGCCCGTGCTAACATGAAATGCATTGTCATCATCCCGAACGGAAAAATTGCATTTGGAAAACTCGCTCAAGCTGT
CATGTACGGAGCCGAGATTATCGCAATTGACGGAAACTTTGACGATGCGCTTAAAATTGTCCGTTCCATCTGTGAGAAATCACCGA
TTGCCCTTGTCAACTCAGTCAACCCTTACCGCATTGAAGGCCAAAAAACTGCTGCCTTCGGATCCTAGAAGCTTATCGAATTCGCG
GCCGCTTCTAGGCTTGACAAATTGCAGTAGGCATGACAAAATGGACTCAAGGAGGTGTCTAGAATGGGAtccgaattccaaacaaa
caatatcgaaggtttagatgtcactgatttagaatttatcagtgaagaagttactgaaaaagacgagaaagaaatcatgggtgctt
cttgtactacatgtgtttgtacatgcagttgttgtacaacttaaGGCTGCAGGATAAAAAATTTAGAAGCCAATGAAATCTATAAA
TAAACTAAATTAAGTTTATTTAATTAACAACTATGGATATAAAATAGGTACTAATCAAAATAGTGAGGAGGATATATTTGAATACA
TACGAACAAATTAATAAAGTGAAAAAAATACTTCGGAAACATTTAAAAAATAACCTTATTGGTACTTACATGTTTGGATCAGGAGT
TGAGAGTGGACTAAAACCAAATAGTGATCTTGACTTTTTAGTCGTCGTATCTGAACCATTGACAGATCAAAGTAAAGAAATACTTA
TACAAAAAATTAGACCTATTTCAAAAAAAATAGGAGATAAAAGCAACTTACGATATATTGAATTAACAATTATTATTCAGCAAGAA
ATGGTACCGTGGAATCATCCTCCCAAACAAGAATTTATTTATGGAGAATGGTTACAAGAGCTTTATGAACAAGGATACATTCCTCA
GAAGGAATTAAATTCAGATTTAACCATAATGCTTTACCAAGCAAAACGAAAAAATAAAAGAATATACGGAAATTATGACTTAGAGG
AATTACTACCTGATATTCCATTTTCTGATGTGAGAAGAGCCATTATGGATTCGTCAGAGGAATTAATAGATAATTATCAGGATGAT
GAAACCAACTCTATATTAACTTTATGCCGTATGATTTTAACTATGGACACGGGTAAAATCATACCAAAAGATATTGCGGGAAATGC
AGTGGCTGAATCTTCTCCATTAGAACATAGGGAGAGAATTTTGTTAGCAGTTCGTAGTTATCTTGGAGAGAATATTGAATGGACTA
ATGAAAATGTAAATTTAACTATAAACTATTTAAATAACAGATTAAAAAAATTATAAAAAAATTGAAAAAATGGTGGAAACACTTTT
TTCAATTTTTTTGTTTTATTATTTAATATTTGGGAAATATTCATTCTAATTGGTAATCAGATTTTAGAAAACAATAAACCCTTGCA
TAGGGGGATCTCGAGGCCTTCCGAAAATGCGCGGCTTTGAAGCTGAGGGTGCGGCGGCAATCGTGCGCAATGAAGTGATTGAAAAT
CCGGAAACAATAGCGACAGCCATTCGTATCGGAAACCCGGCAAGCTGGGACAAAGCTGTAAAGGCAGCCGAGGAATCCAATGGGAA
AATTGACGAAGTCACTGATGATGAAATCCTTCACGCATATCAGCTGATCGCCCGTGTAGAAGGCGTGTTTGCAGAACCAGGTTCTT
GCGCGTCTATCGCAGGAGTGCTGAAACAGGTGAAATCCGGAGAAATTCCGAAAGGCAGCAAGGTCGTAGCTGTGTTAACAGGAAAC
GGACTGAAAGATCCGAACACAGCGGTCGACATTTCAGAAATCAAGCCTGTCACATTGCCGACTGATGAAGACAGCATCCTTGAATA
TGTAAAAGGAGCGGCCCGTGTATGAACGAAGCCGACATGCTGTTCTCTGTCACTGTTCCCGGAAGCACAGCTAACCTAGGCCCCGG

CTTTGATTCAGTCGGAATGGCGCTCAGCAGATATTTGAAGCTGACCGTCTTTGAAAGCGACAAATGGTCTTTTGAGGCTGAAACAG
AAACAGTCGCCGGAATTAATTCGAGCTCCTTTAACTTCCTCACCAGTAGTTGTATCGGTACCATAAGTAGAAGCAGCAACCCAAGT
AGCTTTACCAGCATCCGGTTCAACCAGCATAGTAAGAATCTTACTGGACATCGGCAGTTCTTCGAACAGTGCGCCAACTACCAGCT
CTTTCTACAGTTCATTCAGGGCACCGGAGAACCTGCGTGCAATCCATCTTGTTCAATCATGCGAAACGATCCTCATCCTGTCTCTT
GATCCATGGATTACGCGCGACCTTCAGCATCGCCGGCATGTCCCCCTGGCGGACGGGAAGTATCCAGCTCGAGGTCGGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGG
CTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACT
TATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCT
AACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGATAGCTCTTG
ATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAG
ATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGG
ATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCA
ATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTA
CGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATA
AACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGC
TAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCCGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGT
CCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC
ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTC
TCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCT
TCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAA
ATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAG
GCGTATCACGAGGCCCTTTCGTCTTCAAGAATT

B pTclE_His insert, His_6-**TclE:**

…TCTAGA ATG GGC TCA CAT CAT CAT CAT CAT CAT GGC GGC GGA **tcc gaa ttc**…

pTclE_GST insert, GST-*TEV*-**TclE**:

…TCTAGAATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTT
GGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAAC
AAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACA
CAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCG
TGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATA
GTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAA
GATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGAC
GCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAA
AAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTT
GCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA GGA GGA *GAA AAC CTG TAT*
TTT CAA GGC GGA **tca gaa ttc…**

Figure 3-S5. Alignment of $tRNA^{glu}$ DNA sequences between relevant bacterial strains. Deviations from the *B. subtilis* trnSL_glu2 sequence are highlighted in yellow. All nucleotide sequences were obtained from GenBank. Sequences shown are representative of one of multiple identical copies of each $tRNA^{glu}$ gene from each species. Abbreviations are: *Bs, B. subtilis* str. 168; *Mc*, *M. caseolyticus* JCSC5402; *Ec*, *E. coli* MG1655.

```
>Bs|trnSL_glu2      GGCCCGTTGG TCAAGCGGTT AAG-ACACCG CCCTTTCACG...
>Mc|MCCL_tRNA01     GGCCCCTTGG TCAAGCGGTT AAG-ACACCG CCCTTTCACG...
>Ec|gltT            GTCCCCTTCG TCTAGAGGCC CAGGACACCG  CCTTTCACG...

>Bs|trnSL_glu2   ...GCGGTAACAC GGGTTCGAAT CCCGTACGGG TCA(cca)
>Mc|MCCL_tRNA01  ...GCGGTAACAC GGGTTCGAAT CCCGTAGGGG TCA(cca)
>Ec|gltT         ...GCGGTAACAG GGGTTCGAAT CCCCTAGGGG ACG(cca)
```

Chapter 4. Capture of micrococcin biosynthetic intermediates reveals C-terminal processing as an obligatory step for *in vivo* maturation

The following chapter is taken from a manuscript, which was accepted for publication in the Proceedings of the National Academy of Sciences and published electronically as an early edition article on Oct 17, 2016. This work has been greatly benefitted by collaboration with Susan Miller and her postdoc Kathryn Bewley at the University of California San Francisco. Kathryn and I contributed equally to the experimental work and are listed as co-first authors on the manuscript.

4.1 Summary

Thiopeptides, including micrococcins, are a growing family of bioactive natural products that are ribosomally synthesized and heavily modified. Here we use a refactored, modular *in vivo* system containing the micrococcin P1 (MP1) biosynthetic genes from *M. caseolyticus* str 115 in a genetically tractable *B. subtilis* strain to parse the processing steps of this pathway. By fusing the micrococcin precursor peptide to an affinity tag and coupling it with catalytically defective enzymes, biosynthetic intermediates were easily captured for analysis. We found that two major phases of molecular maturation are separated by a key C-terminal processing step. Phase-I conversion of six Cys residues to thiazoles (TclIJN) is followed by C-terminal oxidative decarboxylation (TclP). This TclP-mediated oxidative decarboxylation is a required step for the peptide to progress to Phase II. In Phase II, Ser/Thr dehydration (TclKL) and peptide macrocycle formation (TclM) occurs. A C-terminal reductase, TclS, can optionally act on the substrate peptide, yielding MP1, and is shown to act late in the pathway. This comprehensive characterization of the MP1 pathway prepares the way for future engineering efforts.

4.2 Introduction

Thiazolyl peptides (thiopeptides) constitute a family of ribosomally synthesized and post-translationally modified peptides (RiPPs), with precursor peptides comprised of an N-terminal leader peptide and a C-terminal Cys/Ser/Thr-rich core (17, 139). The core peptides are extensively modified with formation of azol(in)es, dehydroalanine(Dha)/dehydrobutyrine(Dhb) and a pyridine/piperidine, en route to macrocyclization and cleavage of the leader to yield a bioactive product. Although the mature products possess potent activity towards several targets, most notably gram-positive bacteria, most lack sufficient solubility to be useful pharmaceuticals (70). Their ribosomal synthesis, however, makes them attractive targets for genetic engineering of both the core peptide and the biosynthetic enzymes to create diverse thiopeptide chemical libraries. Previous efforts to explore pathway promiscuity and define enzyme function *in vivo* have coupled expression of core peptide variants and/or gene deletions with organic extraction of pathway products (67, 68, 71, 118). These studies succeed in identifying extractable products and intermediates that have been prematurely cleaved from the leader peptide, but overlook peptide intermediates that are chemically very different from product molecules. *In vitro* studies of thiopeptide biosynthetic components have provided information about functional sets of enzymatic processing (77, 98), but complete reconstitution remains a challenging endeavor. As an alternative approach, we recently reported an *in vivo* heterologous expression system that allows both efficient genetic changes in the core and biosynthetic enzymes, and isolation and characterization of biochemical intermediates in micrococcin biosynthesis (57).

Micrococcin (Figure 4-1A), an archetypal thiopeptide, contains six thiazoles, two Thr-derived Dhb residues, and a central pyridine ring formed from two Ser-derived Dha residues. It is additionally oxidatively decarboxylated at the C-terminal Thr residue, initially to a ketone to form micrococcin P2 (MP2) which may be re-reduced to an alcohol, yielding micrococcin P1 (MP1). Micrococcin/thiocillin (*tcl*) biosynthetic gene clusters have been identified in *Macrococcus caseolyticus* strain 115 (*Mc*) and *Bacillus cereus* ATCC 14579 (*Bc*) (66, 76). The *Bc* cluster is comparatively large (24 chromosomal genes) and controls the synthesis of a mixture of eight micrococcin/thiocillin variants, whereas the simpler *Mc* gene cluster (12 plasmid-encoded genes) dictates the synthesis of a single micrococcin (MP1) product (66, 76).

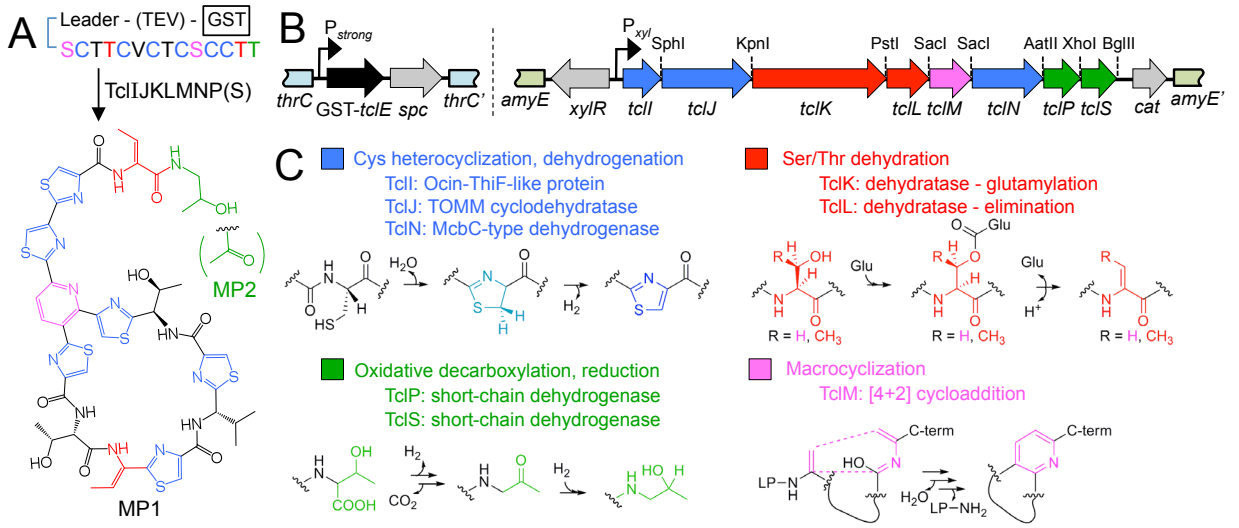


Figure 4-1. Micrococcin structure and post-translational processing.(A) Diagram depicting conversion of the TclE precursor peptide to mature MP1 or MP2. (B) Engineered gene clusters for micrococcin production in *B. subtilis*. Left, the constitutive Pstrong promoter drives expression of GST-TclE from a cassette integrated into the *thrC* locus; right, the xylose-inducible P_{xyl} promoter drives expression of tcl biosynthesis genes from a cassette integrated into the *amyE* locus. A TEV-protease cleavage site is incorporated between the GST-affinity tag and the TclE precursor peptide. (C) Summary of the seven types of chemical transformations required in micrococcin maturation and the enzymes predicted to carry them out.

M*c* requires only 9 of its 12 *tcl* genes to synthesize MP1 (Figure 4-S1) (57), and these genes can be easily manipulated as synthetic gene clusters (Figure 4-1B) in our heterologous *Bacillus subtilis* producer strain expressing the precursor peptide TclE and the 8 biosynthetic proteins TclIJKLMNPS. The functions of these proteins are straightforward to predict by their homology with proteins from other RiPP pathways (Figure 4-1C). TclI is an Ocin-ThiF-like protein with a proposed RiPP precursor peptide recognition element (28), and together with TclJ, a Thiazole/Oxazole-Modified Microcin (TOMM) family cyclodehydratase, converts core peptide cysteines to thiazolines. TclN, an FMN-binding McbC-type dehydrogenase, should subsequently oxidize thiazolines to thiazoles. TclK and TclL show homology to the N- and C-terminal domains, respectively, of NisB, a lantibiotic Ser/Thr dehydratase (139), implicated in dehydrating core peptide Ser/Thr residues by glutamylation (TclK) followed by elimination (TclL). The TclM homolog from the *Bc* pathway has been shown to catalyze a [4+2] cycloaddition reaction that forms the central pyridine ring, simultaneously closing the peptide macrocycle and releasing the leader peptide (98, 118). TclP and TclS, both short-chain dehydrogenases, are proposed to catalyze the C-terminal oxidative decarboxylation (TclP) and subsequent re-reduction (TclS) (57).

Although a foundational understanding of micrococcin biosynthesis exists, we still have an incomplete understanding of the requirements and ordering of the Tcl protein activities. In implementing this heterologous expression system, we made two key observations (57): *i*) TclP, predicted to be involved only in C-terminal tailoring, is absolutely required for production of an extractable bioactive product, and *ii*) an N-terminal glutathione S-transferase (GST) fusion to the TclE precursor peptide does not interfere with micrococcin biosynthesis, allowing non-extractable intermediates to be isolated and characterized. To investigate the role of TclP-

mediated oxidative decarboxylation in the overall processing of micrococcin, we employed this system to capture pathway intermediates from strains genetically blocked for specific Tcl enzyme activities. This unique capability to capture GST-tagged intermediates has allowed us to discern in great detail the ordering of micrococcin processing steps (see proposed model in Figure 4-2), with the striking observation that the two major phases of core peptide processing (Phase I: TclIJN-mediated Cys cyclization and Phase II: TclKL-mediated Ser/Thr dehydration and TclM-mediated macrocyclization) are punctuated by a TclP-dependent decarboxylation event. This oxidative decarboxylation constitutes a biochemical checkpoint that can occur only after Cys cyclization, and is required for substrate peptide progression to Phase II reactions.

4.3 Materials and Methods

Generation of *tcl* expression plasmids and bioassays. All *tcl* mutant strains were constructed by making alterations in pLEGO (Figure 4-S2) and transforming the resulting plasmids into a *B. subtilis* P_{strong}-*GST-tclE* recipient. Appropriate mutations were made in *tcl* genes using overlap-extension PCR. Amplified mutant alleles were reintroduced into pLEGO using restriction digestion to remove the native gene and ligating the altered gene in its place. Detailed descriptions of pLEGO-derived plasmids primers used in their construction are summarized in Tables 4-S1 and 4-S2. For pLEGO derivatives containing *tcl* gene deletions, DNA segments were removed by restriction digestion followed by treatment with Klenow polymerase and religation. All pLEGO alterations were confirmed by PCR and Sanger sequencing. Successful integration into *B. subtilis* was confirmed by PCR with primers spanning chromosomal integration junctions. Bioassays of methanolic extracts from *B. subtilis* micrococcin expression strains were described previously (57, 76).

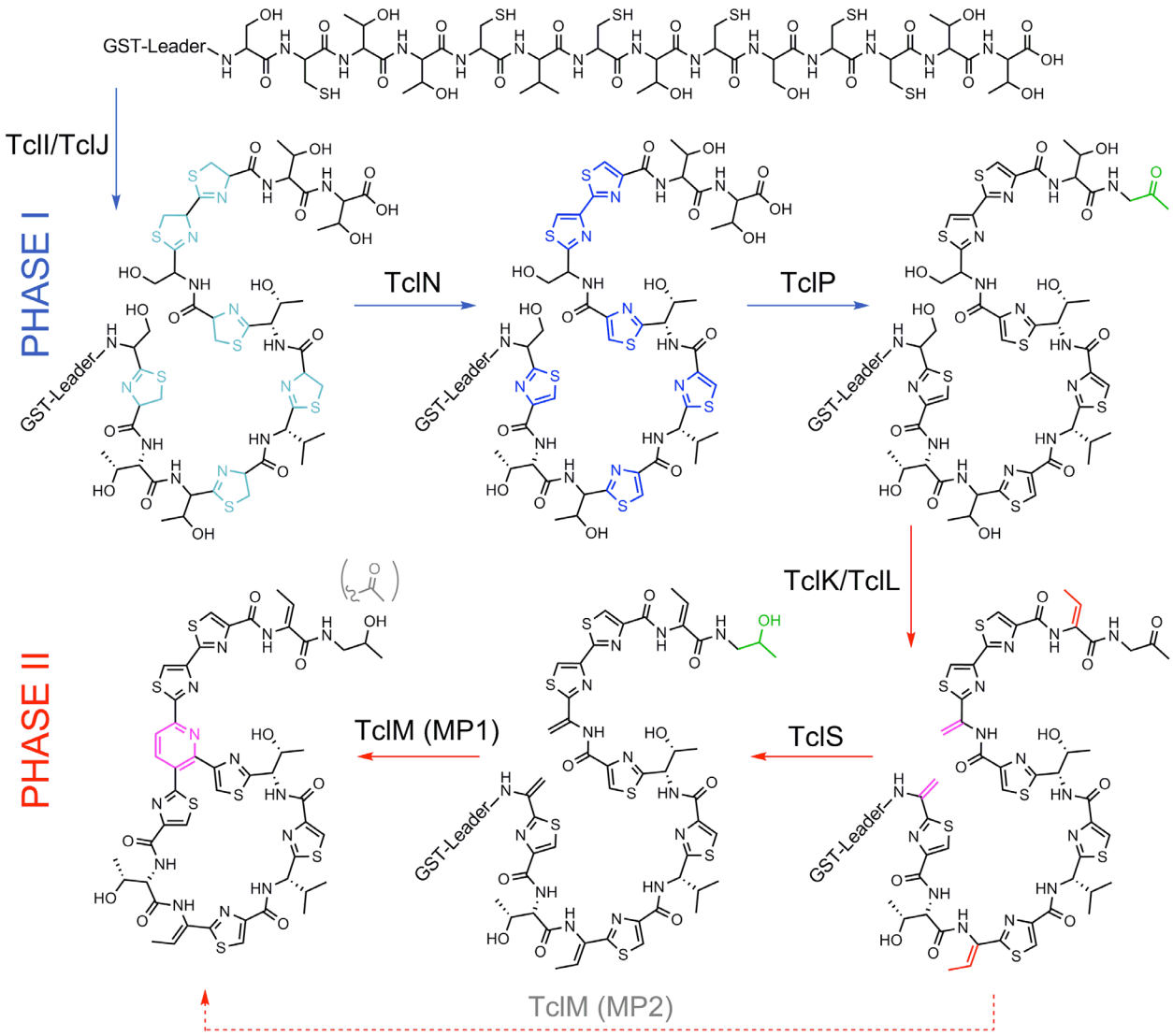


Figure 4-2. A proposed model for micrococcin biosynthesis. The two major processing phases are delineated by TclP-mediated oxidative decarboxylation. In the absence of TclS macrocyclization occurs (dotted arrow) to yield MP2.

Purification of processed peptide intermediates. *B. subtilis tcl* mutant or deletion strains were grown as previously described for either 8 or 16 h (57). More details are included in the Supporting Information, but briefly, cell pellets were resuspended, sonicated, centrifuged and the clarified lystate was incubated with glutathione resin. After washing, the GST-peptide was eluted off the resin beads with 10 mM glutathione, buffer exchanged, concentrated and stored at –80 °C. For ΔTclM strains only, the TEV protease cleavage was carried out on the resin.

Sample preparation for mass spectrometry. The concentration of peptide was estimated using the value supplied by GE-Healthcare for GST (1 Abs_{280} ≈ 0.5 mg/mL GST) prior to cleavage. Each 10-30 µL aliquot of GST-peptide was cleaved with 2 µg TEV protease (pRK793, (140)) and 0.5 mM DTT at room temperature for 1 h. In samples from 16 h cultures, TEV and GST were removed by precipitation with acetonitrile (ACN) (50% final concentration) and centrifugation (2 min, 9503 × g). Supernatant with soluble peptide was diluted with water to final 20% ACN concentration. To minimize loss of lower yields of isolated peptides from 8 h cultures, TEV and GST were not precipitated prior to MS analysis. These samples were acidified with 0.1% TFA, desalted and concentrated with a C18 Zip-Tip into 75% ACN: 25% H_2O (0.1% TFA).

MALDI-MS. A 0.5 µL saturated α-cyano-4-hydroxycinnamic acid (CHCA) matrix in 50:50 ACN:H_2O (0.1% TFA) was mixed with 0.5 µL of Zip-Tipped peptide sample on the MALDI plate. Calibration of the MALDI-MS instrument (AXIMA Performance, Shimadzu Biotech) was performed with a mixture of insulin (Sigma Aldrich), Calibration Mixture 2 (AnaSpec) and CHCA matrix. The spectra were collected in linear mode with the pulsed extraction set to *m/z* 5200.

High resolution electrospray ionization liquid chromatography-mass spectrometry. A Thermo Scientific EASY-Spray source and column (PepMap, C18, 3 µm, 100A, 75 µm x 15 µm) was used with a Waters Nano Acquity HPLC system and a Thermo Scientific LTQ-Orbitrap XL mass spectrometer for LCMS analysis. Zip-tipped samples were diluted as appropriate in water/0.1% formic acid and injected onto the column, which was maintained at 45 ºC. All modified peptides eluted between 38 – 39.5 % B (A = water/0.1% formic acid, B = ACN/0.1% formic acid) during a 30 min gradient of 30-45% B (total run time of method = 55 min).

4.4 Results

Prediction and creation of catalytically inactivated Tcl enzyme variants. The ability of the *Mc* biosynthetic enzymes to process the affinity tagged GST-TclE precursor peptide provides an avenue for isolation and analysis of all biosynthetic intermediates along the pathway prior to the macrocyclization step. To capitalize on this opportunity, we generated separate *B. subtilis* strains, each of which couples expression of GST-TclE with either a point mutation(s) or gene deletion designed to knock out the activity of one of the seven enzymes in the pathway (Figures 4-1, 4-S2). Mutational predictions were achieved by aligning Tcl sequences with proteins of known structure using HHpred (141) and using the top-scoring alignments to generate structural models using Modeller (142) (Figures 4-S3 through S8). Loss of micrococcin formation in GST-TclE strains harboring individual mutations was then verified by evaluating the bioactivity of methanol extracts against a *Staphylococcus aureus* indicator strain (57).

TclJ is homologous with the Mg^{2+}-ATP-dependent TOMM cyclodehydratases, which have been shown to lose activity upon substitution of Mg^{2+}-binding glutamic acid residues with alanine (143, 144). Our structural model, based on LynD from the cyanobactin RiPP pathway (144) (Figure 4-S3), predicted alteration of TclJ E319 would have the same effect, and a TclJ E319A mutation abolished formation of extractable bioactive compounds (Figure 4-3). TclK and TclL correspond to the well characterized N- and C-terminal domains of the Ser/Thr dehydratase NisB from the nisin RiPP pathway (139, 145), which allowed us to design mutations in TclK (R181A) and TclL (R49A) that disrupt enzymatic activity and abolish bioactivity in extracts (Figures 4-3 & 4-S4 through S5). Structural alignment of TclN, a putative FMN-dependent nitroreductase, with another nitroreductase (Ava_2154 from *Anabaena variabilis*) revealed two conserved residues that likely coordinate the phosphate of FMN (Figure 4-S6). Expression of a

double-substitution variant of TclN (R269A/S271A) abolished formation of bioactive product (Figure 4-3). The structure of TclP was modeled with (*R*)-hydroxypropyl-coenzyme M dehydrogenase (146) (Figure 4-S7), and of the conserved active site residues (147), N97, S126 and Y139, only the N97L mutation gave complete loss of activity in extracts and was used in this study (Figure 4-3). Since the cycloaddition enzyme TclM does not have homologs of known structure and its catalytic mechanism is not well understood, we evaluated sequence alignments with related enzymes, including TbtD involved in thiomuracin biosynthesis. Based on these alignments (Figure 4-S8) and *in vitro* mutational studies with TbtD (77), we predicted that TclM Y222A and R235A substitutions might perturb activity. Extracts from both of these mutants retained bioactivity (Figure 4-S8), thus we used a deletion mutant (ΔTclM) for these studies (Figure 4-3). The above mutations were all introduced in the presence of TclS. Since we previously showed that macrocycle formation can occur in the absence of TclS (57), where appropriate, the above mutations were also incorporated into the pathway lacking the *tclS* gene.

TclP mediates the transition between two phases of core peptide processing. GST-TclE peptides isolated from the different genetically trapped *B. subtilis* producer strains were cleaved with TEV (tobacco etch virus) protease and analyzed for changes in mass that, coupled with knowledge of activities of orthologs from other RiPP systems, allow us to deduce the type and extent of post-translational modification that occur upstream of the genetically-defective step. Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) was initially employed for analysis of pathway intermediates, and high-resolution electrospray ionization liquid chromatography-mass spectrometry (ESI-LCMS) was used to further resolve the identity of processed peptides. This approach gave rise to a well-resolved model for the sequential processing events leading to micrococcin product formation (Figure 4-2).

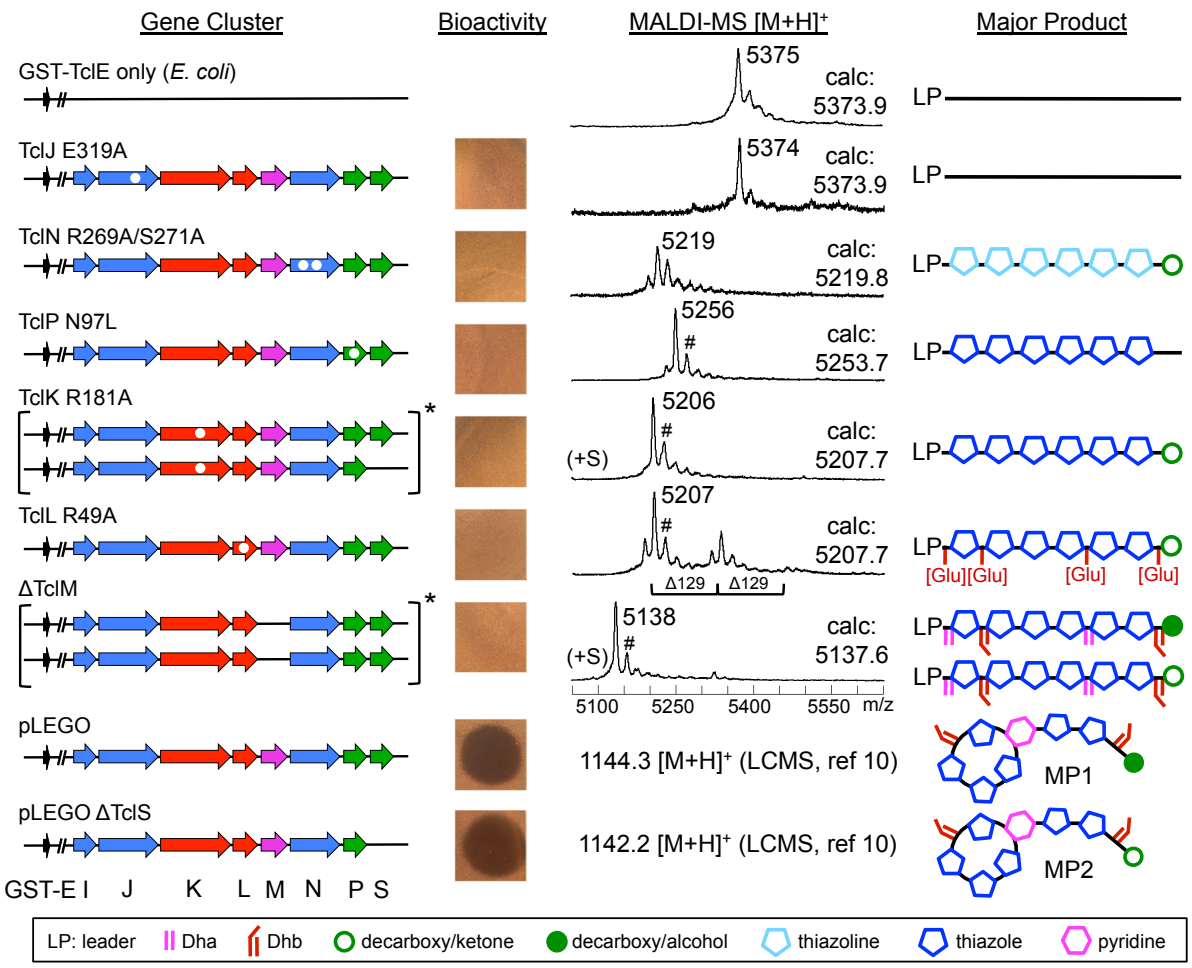


Figure 4-3. Dissection of micrococcin biosynthesis by genetic disruption, capture of biosynthetic intermediates, and mass analysis. Amino acid substitutions are indicated with white dots within Tcl proteins (arrows). Bioassays of methanol extracts from each strain are shown along with MALDI-MS data for modified peptides obtained after TEV-cleavage of GST-tagged intermediates. Observed m/z values are given next to major peaks, and calculated m/z values (calc) are indicated. Products consistent with mass values of the major peaks are diagrammed on the right, with chemical group symbols shown in the accompanying key. Asterisk (*) denotes mass comparisons that required high-resolution ESI-LCMS (see Figure 4-4). Non-proton adduct ions were observed (#) and found to be +17 amu via ESI-LCMS. Previous studies showed only MP1 is produced from the full set of genes, while only MP2 is formed in the ΔTclS strain (57).

From our modelling, TclJ is expected to convert the six core-peptide Cys residues to thiazolines, and by analogy to results from the recent *in vitro* thiomuracin study (77), may initiate processing. Indeed, the TclE peptide remains unmodified in the TclJ-defective strain (Figure 4-3), confirming its role in catalyzing the first step in the pathway. TclN is expected to

carry out the FMN-dependent oxidation of thiazolines to thiazoles. Isolated peptide from the TclN-defective strain gave an envelope of ion peaks discussed in more detail below, however, the major ion peak has a mass consistent with a peptide containing 6 thiazolines and, surprisingly, a decarboxylated C-terminus (Figure 4-3). This result confirms both the roles of TclJ and TclN and the timing of thiazole installation as the first step in the pathway. However, the observation of decarboxylated peptide suggests TclP also acts very early in the pathway. Strikingly, the TclP-inactivated strain accumulates core peptide modified only with six thiazoles (Figure 4-3), indicating that not only is TclP responsible for the decarboxylation step, it is required for robust downstream Ser/Thr dehydration and subsequent macrocyclization. TclP-mediated decarboxylation therefore appears to act as a functional link between two distinct phases of the overall pathway: TclIJN-dependent thiazole installation with TclP-dependent decarboxylation (Phase I), and TclKLM-dependent dehydration/macrocyclization (Phase II).

Characterization of Phase II processing by TclK, TclL, and TclM. The TclK enzyme is expected to initiate Ser/Thr dehydration steps via side chain glutamylation, so peptide from the TclK-defective strain identifies all processing steps that precede Ser/Thr dehydration. This strain accumulated peptide with a mass consistent with six thiazoles and complete removal of the C-terminal carboxyl group (Figure 4-3), confirming the placement of the TclP-mediated decarboxylation prior to Ser/Thr dehydration. TclL is expected to complete the Ser/Thr dehydration by elimination of glutamyl ester (Glu) adducts introduced by TclK. Thus, the TclL-defective strain should accumulate decarboxylated and thiazole-modified peptide intermediates containing one or more glutamylations. MALDI-MS data from this strain show a series of peaks (Δ129 amu) consistent with decarboxylated peptides harboring six thiazoles and up to three glutamylations (Figures 4-3 & 4-S5). A similar pattern of multiple glutamylations has previously

been observed for non-thiopeptide RiPP pathways (145, 148); however, in the *in vitro* reconstituted thiomuracin system, only a monoglutamylated intermediate accumulates when the TclL ortholog (TbtC) is omitted (77). This discrepancy may be due to the different reaction conditions (*in vitro* vs. *in vivo*), or to an intrinsic difference in how this two-step dehydration process is coordinated in the two biosynthetic systems. To shed light on this, we repeated our analysis using a strain with TclL deleted rather than catalytically-inactivated. Isolated peptide from the TclL-deletion strain also exhibited multiple glutamylations (Figure 4-S5), suggesting there may be some inherent difference in the way glutamylation/elimination steps are coordinated between the two systems.

TclM from the *Bc* pathway has been shown to complete the thiocillin/micrococcin maturation process by catalyzing formation of the central pyridine ring from the Dha residues generated by the TclKL dehydratase module (98). Thus, a TclM-defective strain should accumulate a linear peptide with complete side chain processing. In initial experiments with this strain grown for 16 h, we detected only TclE leader peptide lacking the C-terminal core (Figure 4-S8). This result is consistent with results from the *Bc tclM* deletion strain where after their typical 3-day growth, the major extractable compound was the fully processed linear core resulting from proteolytic cleavage between the leader and modified core (118). Taken together, we postulated this results from a slow, unidentified proteolytic process and that a shortened culture time would allow capture of the non-proteolyzed intermediate. Indeed, by decreasing culture time from 16 h to 8 h, we successfully isolated the expected fully-processed linear core still attached to the leader (Figure 4-3).

Timing of TclS-mediated C-terminal reduction. Thus far, we have mapped the timing of six of the seven enzymatic reactions in the pathway. The remaining step involves the TclS-catalyzed

re-reduction of the TclP-generated C-terminal ketone to an alcohol, a change in mass of only 2 amu that requires high resolution ESI-LCMS to distinguish in peptides of this size. Although the TclS-catalyzed reaction is not required for macrocycle formation, when TclS is present, only the alcohol product is observed suggesting the enzyme activity is robust. Here we address the placement of TclS activity with respect to the other Tcl enzymatic steps. We first observed that peptide from the ΔTclM strain had a C-terminal alcohol, and confirmed the reduction was TclS-mediated using a ΔTclMΔTclS strain which yielded a peptide with a C-terminal ketone, as expected (Figure 4-4A). Next we tested whether TclS acted prior to Phase II by analyzing peptides from TclK-defective strains with and without TclS. In this case, the results were less straightforward. Both TclK-defective strains (±TclS) grown for 16 h had peptide peaks with identical isotope patterns that were consistent with a mixture of C-terminal ketone and alcohol (Figure 4-4B). To test for possible slow-acting, TclS-independent reduction, we shortened the culture time to 8 h, and indeed found a shift toward the oxidized form for both strains (Figure 4-4B). This TclS-independent, slow-acting redox phenomenon was also observed in TclL- and TclN-defective strains (Figures 4-5 & 4-S5) and may be attributable to weak reductase activity by TclP or to some unknown cellular component. Taken together, these observations show that TclS activity is biochemically distinct and temporally uncoupled from that of TclP, that it requires a Ser/Thr dehydrated substrate, and that it does not require prior macrocyclization (Figure 4-2).

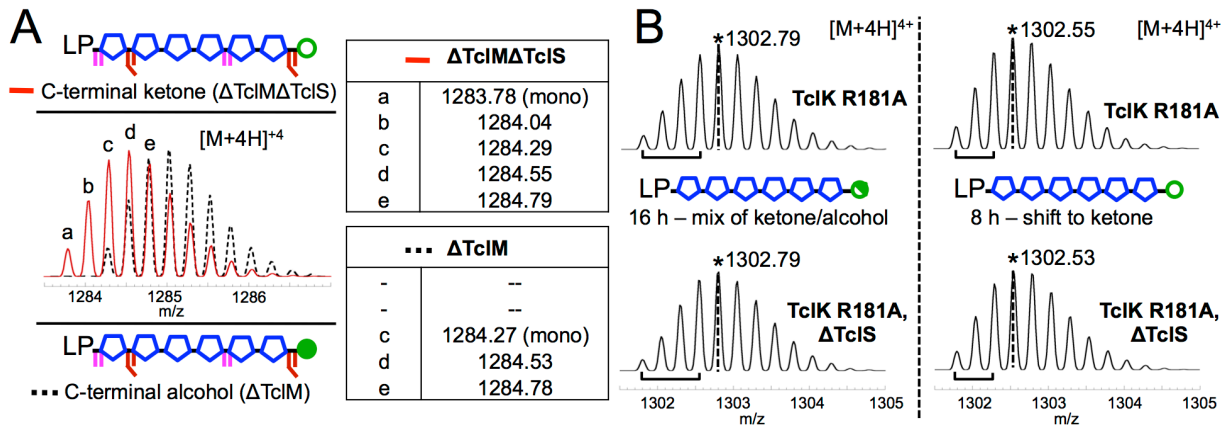


Figure 4-4. ESI-LCMS analysis of C-terminal redox state in peptides from TclM- and TclK-defective strains. (A) Data for peptides captured from ΔTclM (black dashed) and ΔTclMΔTclS (red) strains. Relative intensities of isotopic peaks indicate the presence of a single product in each case. (B) In the TclK-defective strain, maximum abundance isotope shifts to higher mass with longer culture time (16 h vs. 8 h) in both the presence and absence of TclS. Asterisk (*) denotes m/z for the most abundant isotope for $[M+4H]^{4+}$ ions.

TclN-dependent thiazoline oxidation facilitates efficient progression to downstream modifications. In addition to the timing of TclS, analysis of the ESI-LCMS data for peptide from TclN-defective strains at 8 h and 16 h allowed us to probe more deeply the transition from Phase I to Phase II. Disruption at this step gives rise to a complex mix of products that vary in the number of modifications and in C-terminal decarboxylation status, both of which were found to be dependent on culture time (Figure 4-5). After 8 h, ESI-LCMS data show two series of peaks corresponding to decarboxylated intermediates (more intense, green peaks) and non-decarboxylated intermediates (less intense, red peaks). Within each series, peaks are separated by 18 amu (deconvoluted), indicative of products with different numbers of dehydrations up to 6 — suggesting all are due to cyclodehydration of cysteines to thiazolines. Several differences emerge with 16 h of culture time: (*i*) the proportion of decarboxylated peptides (i.e., ratio of green:red) increases, (*ii*) decarboxylated products shift to higher degrees of dehydration, up to 8 (Figure 4-5, peaks a and b), and (*iii*) a trace of thiazole formation is detected (Figure 4-5, 16 h, inset). The two additional degrees of dehydration may be explained by Ser/Thr

conversion to either Dha/Dhb, or to oxazolines. Of these two types of modifications, only Dha/Dhb are resistant to mild acid hydrolysis. Incomplete hydrolysis of 16-h peptides treated with 10% formic acid supports the accumulation of ~2 Dha/Dhb residues (Figure 4-S6). Additionally, we observe ~Δ307 satellite peaks that likely arise from glutathione addition to Dha/Dhb residues during the GST purification process. Taken together, these results indicate that blocking the pathway at the thiazoline stage still allows TclP-mediated decarboxylation, albeit less efficiently, but without thiazole formation, progress to Ser/Thr dehydration is very limited. Therefore, robust Phase II Ser/Thr processing requires both decarboxylation and complete thiazole installation.

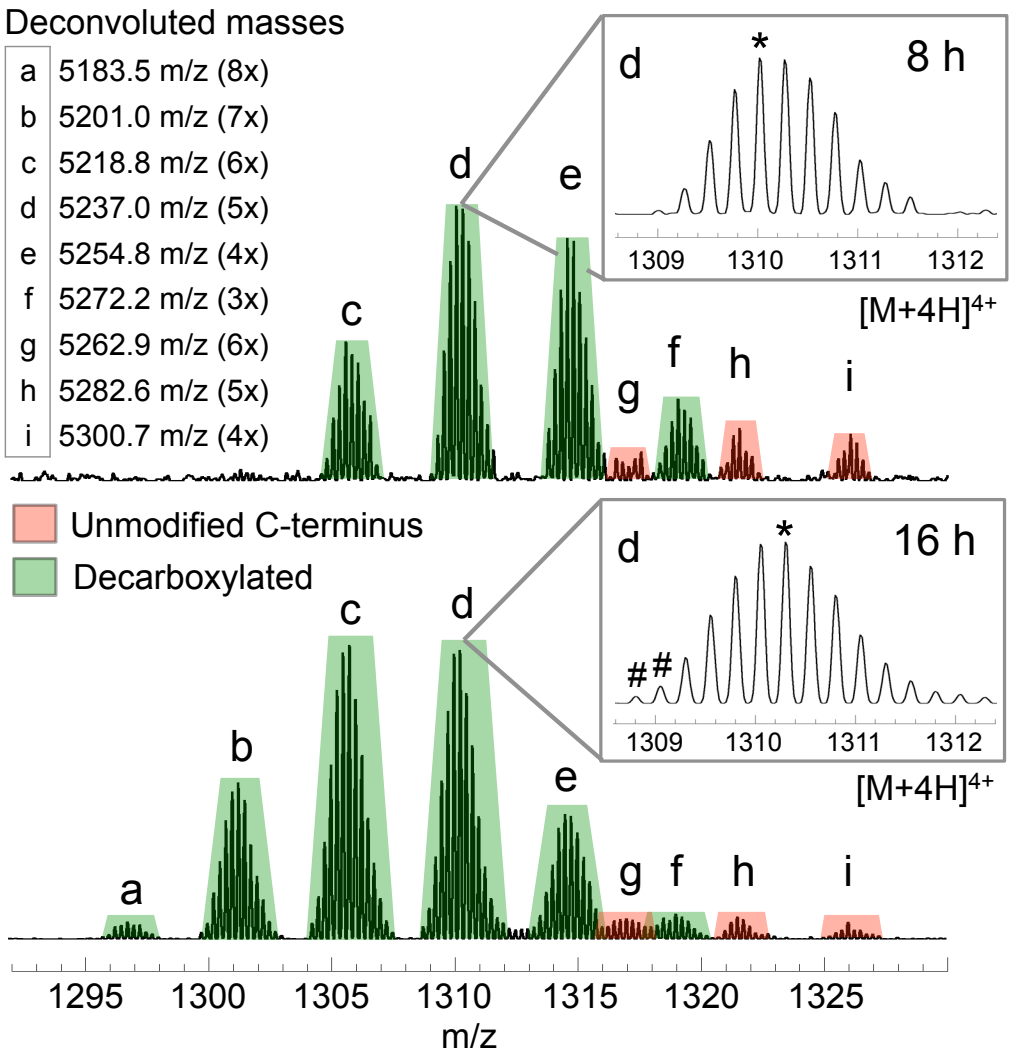


Figure 4-5. ESI-LCMS data for modified peptides isolated from the TclN-defective strain after shorter (8 h, top) or longer (16 h, bottom) culture times. Decarboxylated peptide series (green peaks) and non-decarboxylated series (red peaks) display within-series mass differences (deconvoluted) of 18 amu. Inset table reports deconvoluted masses of peaks a-i, and degree of dehydration (3x-8x) is indicated for each peak. Subtle isotopic details (d, insets) provide additional information about peptide processing (see text), including a minor amount of thiazole formation (#) and a shift in the most abundant isotope (*).

4.5 Discussion

This genetically refactored system is the first thiopeptide analytical system that allows for *in vivo* processing of a tagged precursor peptide and isolation of intermediates at each step in the pathway. Using this approach we have determined the sequential functions of all *Mc* Tcl enzymes in the micrococcin biosynthetic pathway, and resolved the specific placement of TclP and TclS C-terminal oxidoreductase activities. Our GST-TclE pull-down data support the following model (Figure 4-2): First, TclI/TclJ cyclizes Cys residues to thiazolines. Although the data suggest all 6 thiazolines can be installed without oxidation, the installation appears to be slow and most likely occurs in a distributive fashion in conjunction with TclN-dependent oxidation. Once TclN converts thiazolines to thiazoles (or potentially a subset of the six), TclP readily oxidizes the C-terminus and decarboxylation occurs, licensing the core peptide for Phase II modifications. Next, two Ser and two Thr residues are dehydrated to Dha and Dhb, respectively, by the coordinated activities of TclK and TclL. TclS reduces the C-terminal ketone to an alcohol, an optional step that occurs only *after* Dha/Dhb formation. We show TclS can act on the fully processed linear peptide, but the data are inconclusive as to whether it can also act after macrocyclization. Finally, TclM performs the macrocyclization step, cleaving the leader from the final MP1 (alcohol) or MP2 (ketone) product.

Perhaps the most striking principle emerging from these experiments is that the TclP-catalyzed oxidative decarboxylation step is a prerequisite for Phase II core peptide processing. At present it is unknown whether TclP catalyzes just the oxidation of the C-terminal Thr, or if it additionally catalyzes the decarboxylation step. Whether TclP is completely absent (ΔTclP) or catalytically inactivated (TclP N97L), the accumulating core peptide intermediate is the same (Figure 4-S7), strongly suggesting that TclP catalytic activity (and not just its presence to

stabilize a multiprotein complex) is crucial for progression to Phase II processing. The presence of only trace amounts of Ser/Thr dehydration from TclP-defective strains indicates that TclK (the initiator of Phase II reactions) requires substrate with a decarboxylated C-terminus.

Our finding that C-terminal decarboxylation is an obligatory step is consistent with results obtained with the thiopeptide TP-1161, where disruption of the single dehydrogenase, TpaJ, also abolished formation of detectable products in extracts (71), though in this case an accumulated intermediate was not characterized. Previous data from the *Bc* Tcl system showed core peptide variants T14A and ΔT14 greatly impacted the ability to form extractable product (68), further corroborating our results. The other major C-terminal modifications found in thiopeptides are amides and carboxylic acids mostly arising from cleavage of C-terminal residues (52). These modifications, or lack thereof, appear to be more permissive to processing (52, 149), suggesting differences in the substrate recognition requirements for Ser/Thr dehydratases from different thiopeptide pathways.

Analysis of the peptides from the TclN-defective strain gives us further understanding of the substrate requirements for TclP and TclK. The Phase I TclN-catalyzed oxidation of thiazolines to thiazoles results in substantial rigidification of core peptide structure (68), which influences the suitability of the core peptide as a substrate for both TclP and TclK. Comparison of the accumulated peptides from TclN-defective versus TclK-defective strains showed incomplete decarboxylation in the less rigid thiazoline-rich peptide intermediates but complete decarboxylation in the more rigid thiazole-containing intermediate suggesting that core peptide rigidification by TclN results in a better substrate for TclP-dependent decarboxylation. A rigidified substrate is also preferred by TclK, as evidenced by very slow and incomplete accumulation of Dha/Dhb groups in the decarboxylated peptides from the TclN-defective strain.

From the ESI-LCMS data, we also found accumulation of peptide species in the TclN-defective strain with fewer than six thiazolines, whereas blocking the pathway *after* TclN results in robust installation of all six thiazoles. Although partial hydrolysis of thiazolines during workup cannot be ruled out, it seems unlikely given the stark difference in peptide species observed at 8 h vs. 16 h using the same workup. These observations support a model in which TclIJ and TclN act in a coordinated fashion, distributively processing each Cys residue prior to C-terminal decarboxylation. This reflects similar conclusions from studies with TOMM (42, 150) and microcin B17 (43) heterocyclization enzymes. However, further study is needed to unambiguously identify the mixture of peptides arising from the TclN-defective strain and contextualize them within the proposed cooperativity of TclIJN.

The comprehensive study reported here demonstrates the capability of this new analytical system for investigation of the principles of thiopeptide biosynthesis. This work paves the way for further inquiry into the mechanisms of Tcl enzymes with respect to substrate specificity, and provides an opportunity to further engineer the Tcl core peptide and biosynthetic genes to generate diverse thiopeptide derivatives.

4.6 Acknowledgments

We would like to thank T. Acker for the sample of purified TEV protease. This work was supported by a grant from the BYU College of Life Sciences Vaccine Royalties Fund (to R.A.R.), a BYU Graduate Studies Fellowship Award (to P.R.B.), private donations through the UCSF Foundation (to S.M.M.), NIH P41 grant P41GM103481 (LTQ-Orbitrap XL LCMS, UCSF Mass Spec Facility), and UCSF Research Resource Program Shared Equipment Award funded by the Chancellor (AXIMA MALDI-TOF).

Supplemental Material for:

Capture of micrococcin biosynthetic intermediates reveals C-terminal processing as an obligatory step for *in vivo* maturation

Kathryn D. Bewley[a,1], Philip R. Bennallack[b,1], Mark A. Burlingame[a], Richard A. Robison[b], Joel S. Griffitts[b], and Susan M. Miller[a,2]

[a]Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, CA 94158; [b]Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT 84602; [1]K.D.B and P.R.B contributed equally to this work.

[2]Correspondence should be addressed to Susan Miller, Susan.MillerPhd@ucsf.edu

Supplemental Materials and Methods

Construction, expression and purification of *E. coli*-produced GST-TclE

To optimize conditions for expression and analysis of the precursor peptide (TclE), a GST-TclE translational fusion was created by overlap-extension PCR and introduced into the high copy *E. coli* expression vector (pJG542) under control of an IPTG-inducible promoter. The glutathione S-transferase (GST) encoding gene was sourced from pGEX-4T-1 (GE Life Sciences, Marlborough, MA, USA) and amplified using oPB281 and oPB288 (Table 4-S2). The *tclE* gene was amplified from pBac115 (76) using oPB282 and oPB289. Once assembled and the sequence verified, expression of GST-TclE was achieved by transforming the construct into Rosetta 2(DE3) pLysS cells (EMD Biosciences), inoculating a single colony into a 25 mL overnight culture (LB media, 37 °C, carbenicillin 100 mg/L and chloramphenicol 34 mg/L) followed by inoculation into a 1-L fermentation flask (LB media, 37 °C, carbenicillin 100 mg/L and chloramphenicol 34 mg/L). Cells were grown until the OD_{600} = 0.6-0.65 upon which 0.25 mM IPTG was added to induce expression for 3 hours at 37 °C. The cells were then pelleted, and GST-TclE was purified as described for GST-TclE peptides isolated from *B. subtilis* mutant strains.

Purification of processed peptide intermediates.

For each *B. subtilis tcl* mutant or deletion strain, a single colony was grown overnight in LB medium (25 ml, 37 °C, 200 rpm) with spectinomycin (80 mg/L). This was used to inoculate LB medium (1 L, 37 °C, 200 rpm, 16 h or 8 h) with spectinomycin (80 mg/L) and xylose (1%) to induce production of the biosynthetic enzymes. Cells were harvested by centrifugation for 15 min at 7,878 × g. Cell pellets were frozen for later use. Thawed pellets were resuspended in 50

mM sodium phosphate buffer with 100 mM NaCl, pH 7.5 (NaPiCl). DNase and a protease inhibitor tablet (Roche) were added, and the cells were sonicated for 2 min (duty cycle, 50%; output, 6) on ice using a Branson Sonifier 450, followed by centrifugation for 20 min at 32,539 × g. The clarified supernatant was incubated with glutathione resin (1 mL of resin per liter of culture) for 1 h with gentle shaking at room temperature. The slurry was collected by centrifugation for 5 min at 700 × g. The resin was washed with 3 × 10 mL of NaPiCl, centrifuging after each wash. To elute the GST-peptide, the resin was incubated with 4 × 1 mL of 50 mM Tris HCl, pH 8 with 10 mM reduced glutathione for 10 min at room temperature before centrifuging. The combined 4 mL of elution supernatant was syringe filtered to remove any remaining resin beads and was then concentrated using a 10 kDa MW cutoff Amicon spin filter (2,600 × g). The concentrated GST-peptide was buffer exchanged into 10 mM NaPi, pH 7.5 buffer using the same Amicon spin filter, concentrated to <100 µL, aliquoted and frozen at –80 °C. For ΔTclM strains only, tobacco etch virus (TEV) protease cleavage of the peptide was carried out on the resin. To do this, 20 µg of TEV protease in 1 mL of NaPiCl containing 0.5 mM dithiothreitol (DTT) was added and incubated at room temperature for 1 h. The resin was centrifuged and supernatant was collected. The resin was washed with 2 × 1 mL of NaPiCl without DTT or protease. The 3 mL were combined, filtered and concentrated to <100 µL with a 3 kDa MW cutoff Amicon spin filter, aliquoted and frozen at –80 °C.

Alignments and Data Analysis

The structural alignments generated by HHpred are displayed in this Appendix using JalView (151). The structural models from Modeller were overlaid with their template structures using Chimera (152) and its MatchMaker feature, but ultimately illustrated in this Appendix using Pymol (PyMOL Molecular Graphics System, Version 1.7.4.1 Schrödinger, LLC.). The

alignment of TclM and its homologues was generated with Clustal Omega (153) and displayed using JalView. Deconvolution of ESI-MS data was achieved with MagTran (The Magic Transformer), a software package based on the Zscore algorithm (154).

Figure 4-S1. Comparison of Tcl proteins from *M. caseolyticus* 115 and *B. cereus* ATCC 14579. (*A*) Alignment of TclE precursor peptides from *B. cereus* and *M. caseolyticus*. The core peptide is shown in red, and the leader peptide is shown in black. (*B*) The predicted functions of micrococcin biosynthetic proteins from *B. cereus* (*Bc*) and *M. caseolyticus* (*Mc*).

A

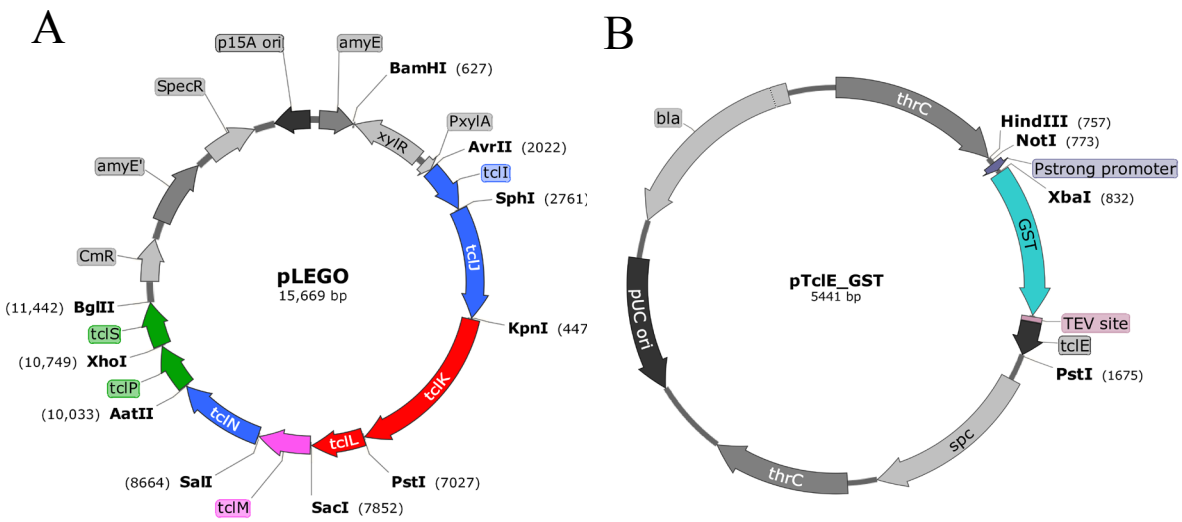
B. cereus ATCC 14579:
MGSEIKKALNTLEIEDFDAIEMVDVDAMPENEALEIMGASCTTCVCTCSCCTT
M. caseolyticus str. 115:
MGSEFQTNNIEGLDVTDLEFISEEVTEKDEKEIMGASCTTCVCTCSCCTT

B

***Mc* Gene Product (size, aa, GenBank ID)**	***Bc* Gene Product (size, aa, GenBank ID)**	**Predicted function**	**Identity/similarity/ coverage %[a]**
TclI (242) AIU53944.1	TclI (503) AAP11955.1	Ocin-ThiF-like protein	36/45/23
TclJ (563) AIU53945.1	TclJ (660) WP_001031656.1[b]	TOMM cyclodehydratase	35/50/87
TclK (843) AIU53946.1	TclK (871) AAP11953.1	dehydratase	28/47/92
TclL (267) AIU53947.1	TclL (323) AAP11952.1	dehydratase	28/45/95
TclM (264) AIU53948.1	TclM (325) AAP11951.1	aza-Diels-Alderase	32/52/79
TclN (447) AIU53949.1	TclN (522) AAP11950.1	McbC-type dehydrogenase	25/43/95
TclP (232) AIU53950.1	TclP (256) AAP11948.1	Short-chain dehydrogenase	38/57/99
TclS (223) AIU53941.1	TclS (237) AAP11945.1	Short-chain dehydrogenase	23/44/81

[a] identity, similarity and coverage % numbers are based on a blastp comparison of *B. cereus* ATCC 14579 and *M. caseolyticus* str. 115 *tcl* gene products.
[b] There is some discrepancy in the length of *Bc*TclJ. It was originally deposited as BC5085 (547 aa), but it was later determined this was an N-terminal truncation and *Bc*TclJ is actually longer (660 aa). The 547 aa version has the GenBank ID AAP11954.1, while the 660 aa version has GenBank ID WP_001031656.1.

Figure 4-S2. Integration plasmids and GST-TclE precursor peptide gene/protein sequences. (*A*) Plasmid map of pLEGO harboring the biosynthetic genes (*tcl*IJKLMNPS) for expression with a xylose-inducible promoter after integration into the *amyE* locus of *B. subtilis* (57). (*B*) Plasmid map of pTclE_GST, for *thrC* integration and constitutive expression of the precursor peptide, TclE. A TEV-cleavage site was engineered between the N-terminal GST tag and TclE. (*C*) DNA sequence of the TEV-TclE portion of GST-TclE (spacer regions = lower case, TEV recognition site = underlined yellow highlight, TclE = grey highlight) and the resulting peptide after TEV cleavage (core peptide = underlined). The sequence of the untagged precursor peptide (TclE) is shown for comparison.



C---Glutathione-S-transferase---

ggaggaGAAAACCTGTATTTTCAAGGCggaTCAGAATTCCAAACAAACAATA

TCGAAGGTTTAGATGTCACTGATTTAGAATTTATCAGTGAAGAAGTTACTGAAAAAGACGAGAAAGAAATCATGGGT

GCTTCTTGTACTACATGTGTTTGTACATGCAGTTGTTGTACAACTTAA

GST-TclE Post TEV cleavage:

GGSEFQTNNIEGLDVTDLEFISEEVTEKDEKEIMGASCTTCVCTCSCCTT

Untagged TclE:

MSEFQTNNIEGLDVTDLEFISEEVTEKDEKEIMGASCTTCVCTCSCCTT

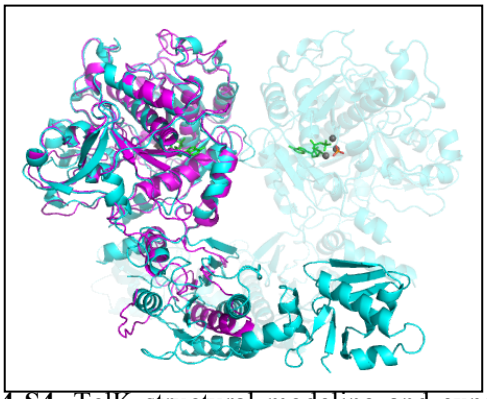
Figure 4-S3. TclJ alignment and structural model. (*A*) TclJ is predicted to have a structure similar to ~450 residues of the C-terminus of LynD (PDB: 4V1T). The HHpred alignment of TclJ and LynD (probability >99%) identifies two conserved sequences involved in Mg^{2+} binding in the active site of LynD (LynD

numbering): 423 EAXER 427, 548 EXXER 552; as well as a conserved C-terminus (572 PXPF 575) that also lies in the active site (144). The red asterisk (*) shows the predicted Mg-binding residue in TclJ (E319) mutated in this study. (*B*) Overlay of the Modeller structure of TclJ (magenta) derived from the HHpred alignment with LynD (cyan). Residues 107-562 of TclJ comprise the predicted globular domain with the catalytic active site. The disparate size and large gaps in sequences between the N-termini of the two proteins gives less confidence in the predicted structure of the ~100 N-terminal residues of TclJ. (*C*) Close up of the ATP binding site in LynD (cyan) overlaid with TclJ (magenta): ADP (green), phosphate (orange), Mg^{2+} (grey spheres), and E548/E319 (cyan/magenta) (144). E548/E319 is from one of the two conserved EXXER motifs, which have been shown to coordinate Mg^{2+} in the homologous BalhD protein (143).

A

```
TclJ   4 FESPEFNIIRYLSNSYI--FK-STSKEPDFLNSIKERILLNTN---------------------43
4V1T 174 LESQQTWLLVKPVGSVLWLGPVFVPGKTGCWDCLAHRLRGNREVEASVLRQKQAQQQRNGQSGSV 238

TclJ  44 -----DIIYKKNIPERNISFLSRIENSIGNFIC----NQI----------SEYDYIIYIYNHKT 88
4V1T 239 IGCLPTAR------ATLPSTLQTGLQFAATEIAKWIVKYHVNATAPGTVFFPTLDGKIITLNHSI 297

TclJ  89 NKIKKVNYFISPHYYKK-LPND---TTNCFNIIINELKN--SKSHNFRSRNIDSIYNQLDKYFLD 147
4V1T 298 LDLKSHILIKRSQCPTCGDPKILQHRGFEPLKLESRPKQFTSDG-GHRGTTPEQTVQKY-QHLIS 360

TclJ 148 KNLGICNMLLDNYD--G-PFPISVAMLPLDN---------GKEEPGVGRTRKIQRSRAVALLEAY 200
4V1T 361 PVTCVVTELVRITDPANPLVHTYRAGHSFGSATSLRGLRNTLKHKSSCKGKTDSQSKASGLCEAV 425

TclJ 201 ERYSGLEPRGKKTINHTE--DLKSKKVNMNSLILHNNPFLISHGIKN---------SNFTY--ND 252
4V1T 426 ERYSGIFQGD-EPRKRATLAELGDLAIHPEQCLCFSDG-----QYANRETLNEQATVAHDWIPQR 484

TclJ 253 -LLDNEISWVKCLNLNTFQTLLIPEQYAYYGINISNHKEKAINIAYEISNGCSVGNNYLEVVYYG 316
4V1T 485 FDASQAIEWTPVWSLTEQTHKYLPTALCYYHYPLP----PEHRFARGDSNGNAAGNTLEEAILQG 545

TclJ 317 LMEVIERDSFLCSWYFNTPKDKISLKNASTSIKNLIQQFTSYYNDYKLELFYLYNEFNIPVVLAT 381
4V1T 546 FMELVERDGVALWWYNRLRRPAVDLGSFNEPYFVQLQQFYR-ENDRDLWVLDLTADLGIPAFAGV 609

TclJ 382 VTLKESSTKKMNFMCAAAADINIEDAIEKSIHEIGGILFGLNKKFIDRYHELEAIRKNNLDVKTM 446
4V1T 610 SNRKTG--SSERLILGFGAHLDPTIAILRAVTEVNQIGLELDKVPDE---NLKSDATDWLITEKL 669

TclJ 447 EDHTLVYGLPEHRTYIQQKMNYENIYDYDKEL--TPKIFYKEVQKLIKKI-STTKDILLVDQTPL 508
4V1T 670 ADHPYLLPDT----------TQPLKTAQDYPKRWSDDIYTDVMTCVNIAQQAGLETLVIDQTRP 723

TclJ 509 ISKKIDLKVGKIIVPGLLPMTFGKYNIRVSENRY-----HELCHFY-S-KDLIIDLNPHPF 562
4V1T 724 DI---GLNVVKVTVPGMRHFWSR------FGEGRLYDVPVKLGWLDEPLTEAQMNPTPMPF 775
```

B



C

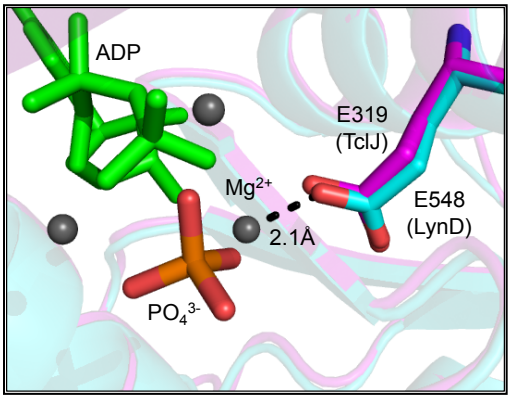


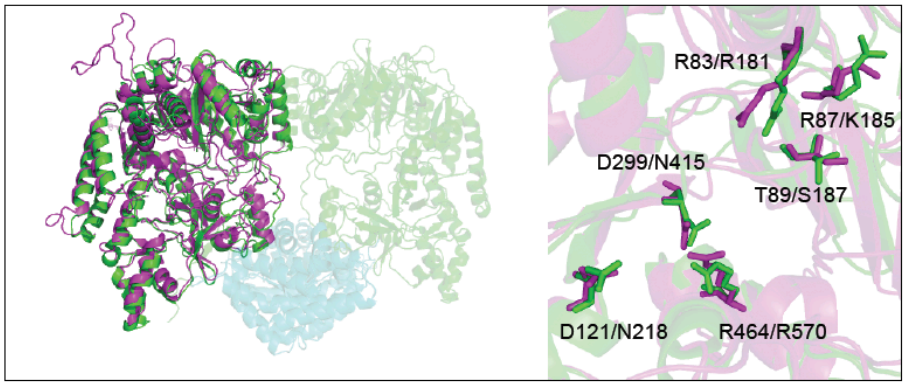
Figure 4-S4. TclK structural modeling and expanded mass spectrometry data. (*A*) HHpred alignment of TclK with N-terminal domain of NisB (probability >99%). Based on inactivation of NisB by an R83A mutation (145), we mutated conserved R181 to alanine for this study (*). Other residues identified in NisB as important for glutamylation of the NisA peptide are labeled

(*) (145). (*B, left*) Overlay of TclK model (magenta, from HHpred alignment and Modeller) with NisB dimer (PDB: 4WD9, C-terminus = cyan, N-terminus = green). Most of the TclK model aligns well, but several loop regions differ. (*B, right*) Closeup of TclK model (magenta) with active site of NisB (green) showing aligned residues that are important for glutamylation activity. (*C*) We observed a degradation product in most samples prepared for mass spectrometry (peak 1 in MALDI and peaks a & d in ESI). The *m/z* of the MALDI-MS peak is consistent with a peptide with the same modifications as the main peptide (peak 2 in MALDI & b in ESI) but missing the first 8 residues of the TEV-cleaved leader (GGSEFQTN, see Figure 4-S2). Peaks a & d in the ESI-MS spectrum are the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions of the degradation product. Since this product appears in the GST-purified peptide sample, and results from cleavage of the peptide bond between two asparagines (N8 and N9, see Figure 4-S2B), which are susceptible to acid-catalyzed cleavage (155), we propose this results from TFA-catalyzed cleavage during workup. (*D*) Expanded windows of ESI mass spectrometry data for samples from 16 h cultures: TclK R181A (*top, left*), TclK R181A ΔTclS (*bottom, left*); and from 8 h cultures: TclK R181A (*top, right*) and TclK R181 ΔTclS (*bottom, right*). The insets show isotopic peaks for peak b in each panel (same as in Figure 4B) labeled 1-10 with *m/z* values listed in the tables. The average mass of a 6-thiazole, decarboxylated peptide is 5206.7 and its $[M+4H]^{4+}$ ion is peak b. Peak c is an unknown contaminant observed in almost every LCMS run from the various mutant strains that elutes over large time window.

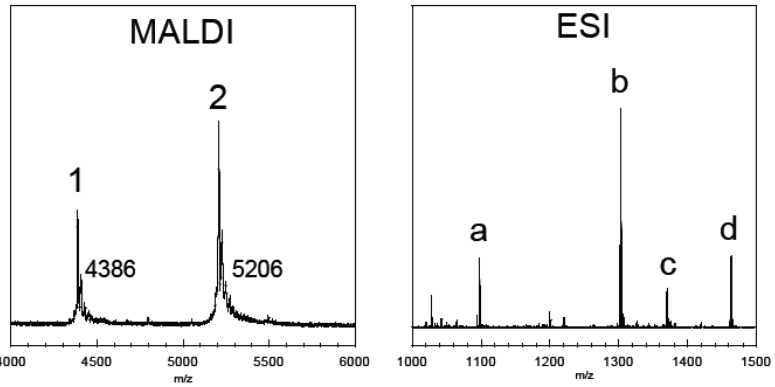
A

```
TclK  129 REAVLQSFECEYFKNGVAMSNRNILKNYYKYVVN-NNNKKSRMTQQSIKNYYNRMNLKPSPFSTF 192
4wd9   30 TQVIETVSKNKVFLEQLLLANPKLYDVMQKYNAGLLKKKRVKKLFESIYKYYKRSYLRSTPFGLF  94

TclK  193 TTIEINIEDYIVSDNK---HS-YEVYSEYNETYYKIIELSLLELDALDNIKKIKLNSSISKTKDN 253
4wd9   95 SETSIGVFSK---SSQYKLMGKTTKGIRLDTQWLIRLVHKMEVD--FSKKLSFTRNNANYKFGDR 154

TclK  254 YHYINVEMTPKYMYYKEKEMILKENITINECIHII----KKECFNYSK-SMKSIEKN-------- 305
4wd9  155 VFQVYTINSS-----------------ELEEVNIKYTNVYQIISEFCENDYQKYEDICETVTLCY 202

TclK  306 --LGSKGVNTFLY-LLKNNFLYLSFGI----DDYEKDKINKLSDMIKTTQTQDLTIIKVIQNLDK 363
4wd9  203 GDEYRELSEQYLGSLIVNHYLISNLQKDLLSDFSWNTFLTKVEAIDEDK--------KYIIPLKK 259

TclK  364 LSHLLNQINNKPFTYKEALVENIYLHTIDLLNLIGKQDIINTIESGSIIYQNN-IDTQITFPKEK 427
4wd9  260 VQKFIQEYSEIEIGEGIEKLKEIYQEMSQI----------LENDNYIQIDLISDSEINFDVKQ 312

TclK  428 LEVSFETVDIIHSIFRLFDDNIIEKIFMKKFFEKNYNNT--VKFKDFSKHFNSVYNEKNWEMIKK 490
4wd9  313 KQQLEHLAEFLGNTTKS--VRRTYLDDYKDKFIEKYGVDQEVQITELFDSTFGIGAPYNYNHPRN 375

TclK  491 T-----TELQTIK-KLREEFFTYIKNKKINSKDINLDTQWLKKFISTFP-DYLDCWESFSMYFQY 548
4wd9  376 DFYESEPSTLYYSEEEREKYLSMYVEAVKNHNVINL--DDLESHYQKMDLEKKSELQGLELFLNL 438

TclK  549 NK--EDNYYILNDIGP--GMGRHILRYTRNLDNKNL--KFFTDTYKKNISEIENSKKKYADISS 607
4wd9  439 AKEYEKDIFILGDIVGNNNLGGASGRFSALSPELTSYHRTIVDSVER----ENENKEITSCEIVF 499

TclK  608 -T---LSLNINLKSCSLNTQINYPTNNNKNNYN---ISDLFVTQKNGVLILVDEKGTEIDVTPTG 665
4wd9  500 LPENIRHANVMHTSIMRRKVLPFFTSTSHNEVLLTNIYIGIDEK-EKFYARDISTQEVLKFYITS 563

TclK  666 FLFPALAPETYKMLCRFSFSNGVQVNLWDRYSSFIDCNEEHFPNIFINNVLIEREVFNEKASDLQ 730
4wd9  564 MYNKTLFSNELRFLYEISLDDKFGNLPWELIY----RDFDYIPRLVFDEIVISPAKWKIWGRDVN 624

TclK  731 LLLENNTYEMMDLNEYLSNKIGIKEPLEFFSKFTYDIDALINENIIPEDWIKLIKNSKLRKPQYY 795
4wd9  625 SKM--------TIRELIQS---KEIPKEFYIVNGD-----------------------NKVYL 653

TclK  796 NLNNYLDFKNFSNII----KSVDSEQYVKIQKVNPRLND---------SNEYIYDFYKE      841
4wd9  654 SQKNPLDMEILESAIKKSSKR---KDFIELQEYFEDE-NIINKGEKGRVADVVVPFIRT      708
```

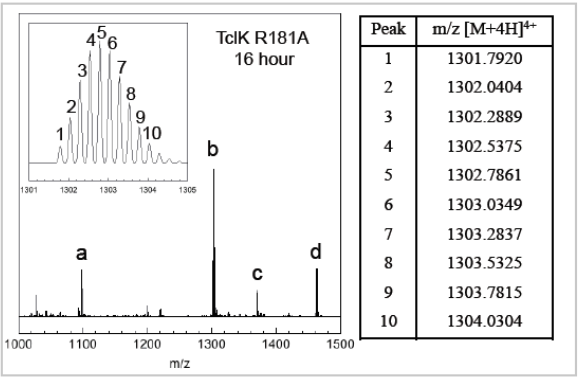
B



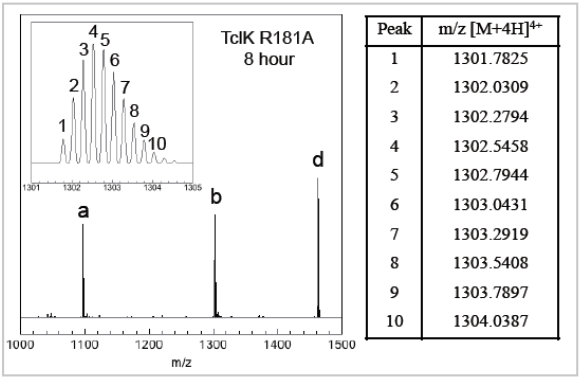
C



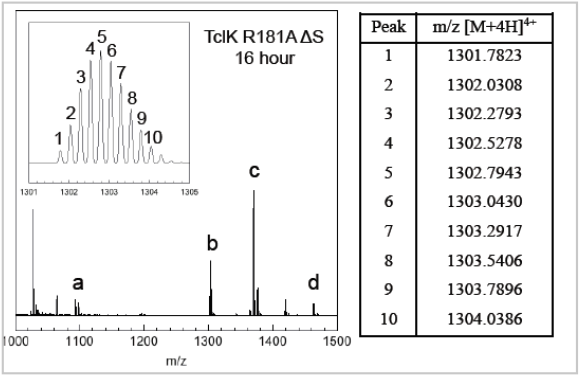
D



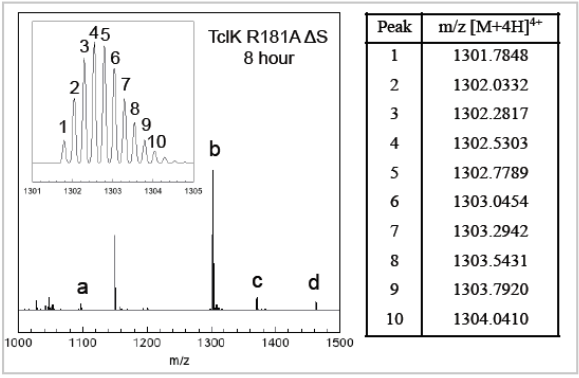
Peak	m/z [M+4H]$^{4+}$
1	1301.7920
2	1302.0404
3	1302.2889
4	1302.5375
5	1302.7861
6	1303.0349
7	1303.2837
8	1303.5325
9	1303.7815
10	1304.0304



Peak	m/z [M+4H]$^{4+}$
1	1301.7825
2	1302.0309
3	1302.2794
4	1302.5458
5	1302.7944
6	1303.0431
7	1303.2919
8	1303.5408
9	1303.7897
10	1304.0387



Peak	m/z [M+4H]$^{4+}$
1	1301.7823
2	1302.0308
3	1302.2793
4	1302.5278
5	1302.7943
6	1303.0430
7	1303.2917
8	1303.5406
9	1303.7896
10	1304.0386



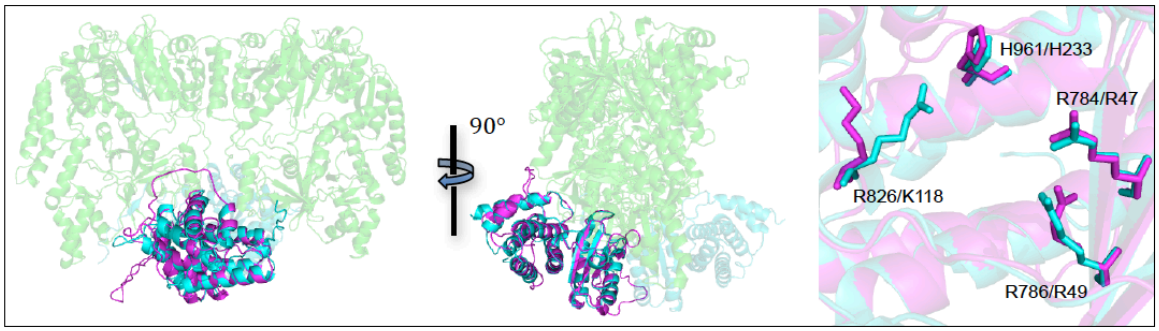
Peak	m/z [M+4H]$^{4+}$
1	1301.7848
2	1302.0332
3	1302.2817
4	1302.5303
5	1302.7789
6	1303.0454
7	1303.2942
8	1303.5431
9	1303.7920
10	1304.0410

Figure 4-S5. TclL structural modeling and expanded mass spectrometry data. (*A*) HHpred alignment of TclL with the C-terminal ~250 amino acids of NisB (PDB: 4WD9; probability >99%). Key residues identified in this domain of NisB as important for elimination of glutamic acid from glutamylated Ser and Thr residues to form dehydroalanines and dehydrobutyrines, respectively, are marked (*) (139, 145). A TclL R49A (*) mutation was used in this study to abolish enzyme activity resulting in the buildup of glutamylated peptide. (*B*) (*left & middle)* Overlay of the TclL structural model (Modeller) from the HHpred alignment (magenta) with the NisB dimer (green = N-term, cyan = C-term). (*right*) Close up showing alignment of key (*) residues (TclL – magenta, NisB – cyan). (*C*) MALDI-MS data for modified peptide from 16-h culture of TclL R49A mutant. (*inset*) Expanded view of lower intensity ion peaks: 2 Glu (*m/z* 5464, calc. *m/z* 5465.9), 3 Glu (*m/z* 5595, calc. *m/z* 5595.0), and 4 Glu (calc. *m/z* 5724.1, barely above the noise). (*D*) Comparison of MALDI-MS data for processed peptides from TclL R49A and ΔTclL strains. TEV-cleaved peptide from ΔTclL strain shows a simplified set of peaks, missing the lower (–18) *m/z* peak (*) that most likely arises from weak elimination activity of TclL R49A toward a single glutamylated Ser or Thr residue, which was also observed with the corresponding mutation in NisB (139, 145). Data indicate TclL is not required to form a complex with TclK (or TclIJNP) since all other modifications are the same in the two strains. (*E*) ESI-MS data for peptides (0 Glu, 1Glu and 2 Glu) from 16-h cultures of TclL R49A, (*insets*) isotopic details of the major ion peak (*) in each. Obs. values refer to the average mass obtained from deconvoluting the envelope of isotopic peaks. Calc. values refer to the average mass of the 0, 1 and 2 Glu peptides containing a C-terminal ketone (*F*) ESI-MS data for peptides from 8-h cultures of TclL R49A, (*insets*) isotopic details of the major ion peak (*) in each. Comparison of insets in (*E*) and (*F*) shows the shift from lower *m/z* at 8 h to higher *m/z* at 16 h for the most abundant isotope (*insets*, *) that arise from the slow, perhaps nonspecific, reduction of the C-terminal ketone to the alcohol.

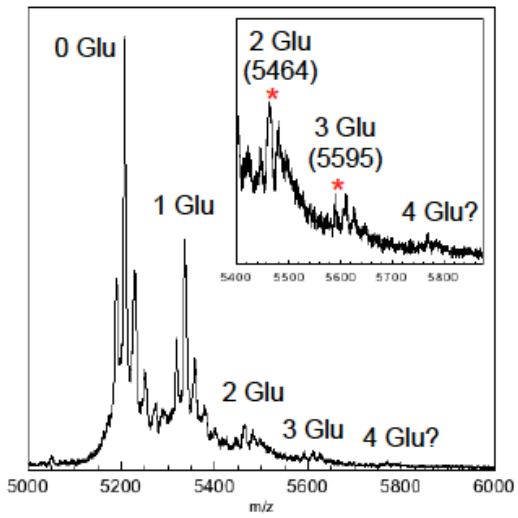
A

```
TclL    1 MNWTSLHIFYFDN-NKELLVKCIYNMY--LNNFFDKFFYINYWDGGPH 45
4WD9  735 NEWLYLKLYISINRQNEFLLSYLPDIQKIVANLGGNLFFLRYTDPKPH 782

TclL   46 IRLRIANITKNEKEIIIKTIQKFIQENPSLSKISEKDYLITSNNFADK 93
4WD9  783 IRLRIKCS---DLFLAYGSILEILKRSRKN------------------- 809

TclL   94 ENSEILELQKNNTILEIEYKPEIDKYLNNEGVHISEDIFFISSLNSLT 141
4WD9  810 --------RIMSTFDISIYDQEVERYGGFDTLELSEAIFCADSKIIPN 849

TclL  142 YLKNSPN---KEMVYMHSLQFANYILKY-F-LNTKQTVLFLKEYERYW 184
4WD9  850 LLTLIKDTNNDWKVDDVSILVNYLYLKCFFQNDNKKILNFLNLVSTKK 897

TclL  185 HSFSNSDIKSTKLI-RIPHK----KFA---------EFDDPFSKIKAN 218
4WD9  898 VK--------ENVNEKIEHYLKLLKVNNLGDQIFYDKNFKELKHAIKN 937

TclL  219 -----FTHS-E-Q--MGFIFNYIHLTNNRL-GIKPFEEAILSSILTKF 256
4WD9  938 LFLKMIAQDFELQKVYSIIDSIIHVHNNRLICIERDKEKLIYYTLQRL 985

TclL  257 WGDEYENKSNK                                      267
4WD9  986 FVSEEYMK---                                      993
```

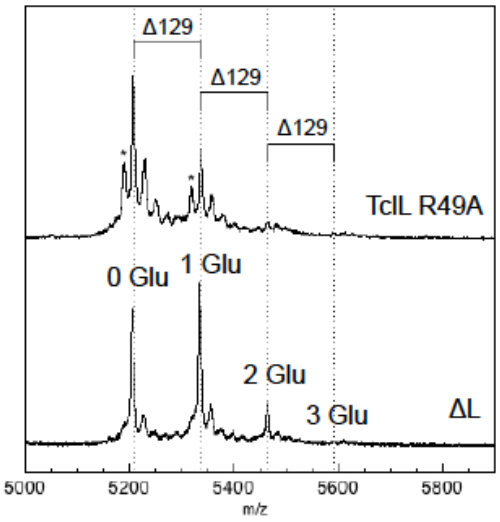
B



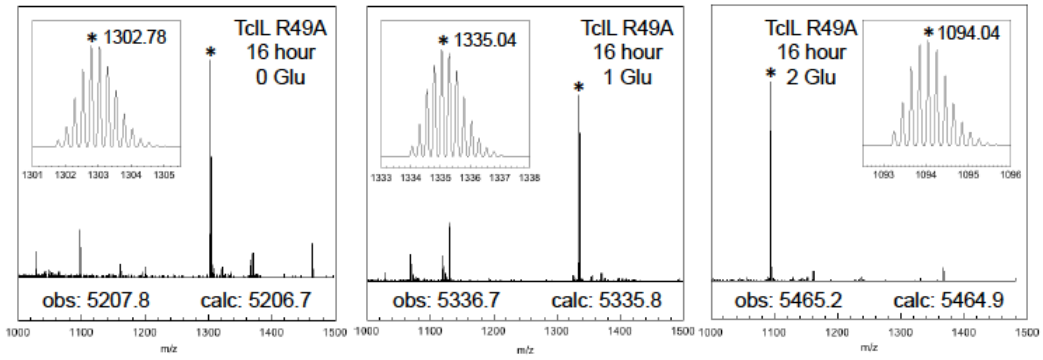
C



D



E



F

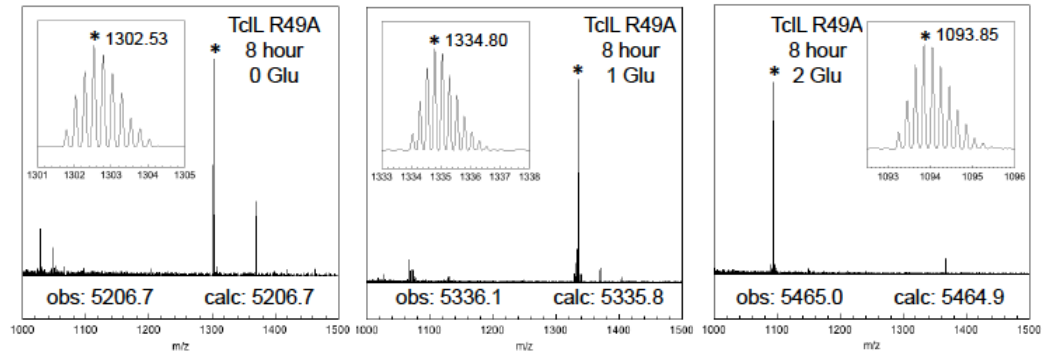
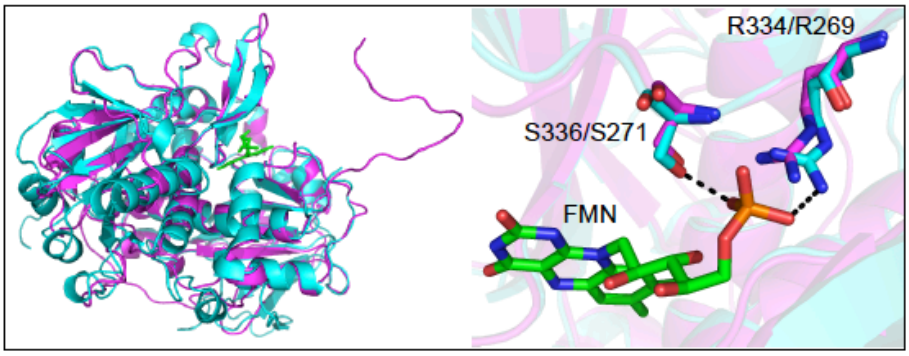


Figure 4-S6. TclN structural modeling and MALDI-MS data after treatment of peptide with formic acid or glutathione. (*A*) HHpred alignment of TclN with a putative nitroreductase from *Anabaena variabilis* ATCC 29413 (ava_2154, PDB: 3EO7, probability >99%). (*B, left*) Overlay of TclN model (Modeller, magenta) with structure of the single-FMN-containing putative nitroreductase (PDB: 3EO7, cyan). FMN is shown in green. (*B, right*) Close-up showing FMN binding site and key phosphate-binding residues in 3EO7 structure. An R269A/S271A (* in *A*) double mutant of the corresponding residues in TclN was made to disrupt FMN binding in order to inactivate the enzyme. (*C*) Comparison of MALDI-MS data for TclN R269A/S271A peptide from 16-h culture before (*top*) and after (*bottom*) incubation with 10% formic acid (>24 h) to rehydrate/ring-open the thiazolines. Incomplete conversion of peptide to a single peak with 0 dehydrations suggests small amounts of thiazole and Dha/Dhb are introduced over the longer culture time. (*D*) MALDI-MS data for TclN R269A/S271A peptide also show peaks consistent with glutathione (GSH ~Δ307) additions, suggesting small amounts of Dha/Dhb modifications are present in the sample (lower *m/z* set is from NN peptide cleavage) (156).

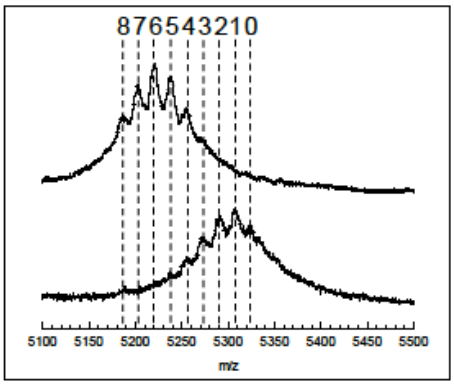
A

```
TclN    3 ISKFLYNLHYNPGEVVSASYTIEDTIQRNSEGFYKGYGIDFL-KLQQK 49
3EO7   10 AQHYHERTKYDPETIASKSQRLDWAKQPVPFKEYKIGSAIDLKPYLQ- 56

TclN   50 SPIVKVI----------LKSYGDIFFN--RVENKKK--PLIFCRKMTPS 84
3EO7   57 --ETPEVFVNDTNGQWWQRLSRLLFRSYGLTARMPSMGNTVYLRAAPS 102

TclN   85 GGGLYPINIFICTNF-KN-RIALFQFDFKRNLL----------I--116
3EO7  103 AGGLYPAEVYVVSRGTPLLSPGLYNYQCRTHSLIHYWESDVWQSLQEA 150

TclN  117 --FIKYINIEINNECTK-LYLVPCYTRNYFKYKEFSYRLCPLDTGYLI 161
3EO7  151 CFWHPALES-----TQLAIIVTAVFYRSAWRYEDRAYRRICLDTGHLL 193

TclN  162 STLLYNFSVENITFKLSIKLNKNSDITDVLNEIGCEEIPYSIIELNE-208
3EO7  194 GNIELSAAITDYRPHLIGGFID-EAVNDLLYIDPLQEGAIAVLPLADL 240

TclN      ------------------------------------------------
3EO7  241 LDIQQNISPGCTALPSATETNYPQVPDGELLKYFHHHTQISASITGKL 288

TclN  209 ----------SLNLDNLSLEHYDTESYFFNPNKVRNLLEIDTLIHQEY 246
3EO7  289 NLPTVIQEKSLEDKYNFPFCLKISTVS-APIYWGENLSDLEITMHK--333
                                    *  *
TclN  247 HKDININFNNENKLFEKFEIQKRISPGGEFIQNSKVEQESINKFISLI 294
3EO7  334 ----------------------RRST--RAYNGEELTFDELKALLDFT 357

TclN  295 MQYKNK-----SNFLSE--------YILLNLIDVQNKRIINLS--ASE 327
3EO7  358 YQPQNYIDQSLDNSPDYFDLNLIETFIAVCGVQGLEAGCYYYAPKAQE 405

TclN  328 FLSYKNNVSIEFIDKQLTRRNFNLLAVPYILYVGVNEEKIKENYSNNY 375
3EO7  406 LRQIRFKNFRRREL-HFLCLGQELGRDAAAVIFHTSDLKSAIAQYGDRV 452

TclN  376 FKISRIIACFWSGVVSILSAQCGLSTHPMMSYNARELEEYIF-KNRYS 422
3EO7  453 YRYLHMDACHLGQRLNLAAIQLNLGVSGIGGFFDDQVNEVLGIPNDEA 500

TclN  423 ILNQIVIG                                        430
3EO7  501 VIYITTLG                                        508
```

B



C



D

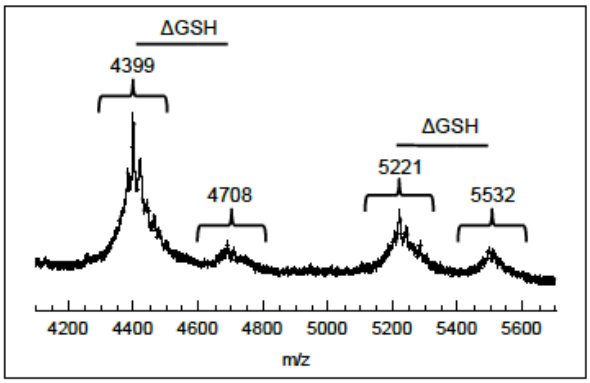
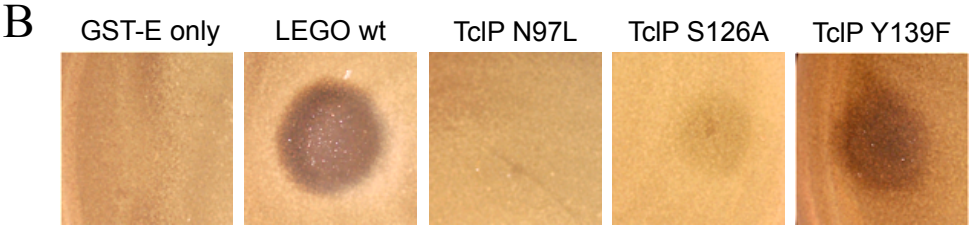


Figure 4-S7. TclP structural modeling, mutant bioassays, and expanded mass spectrometry data. (*A*) TclP is annotated as a short-chain dehydrogenase. It was aligned with (*R*)-hydroxypropyl-coenzyme M dehydrogenase (PDB: 2CFC) using HHpred (probability >99%). Three conserved active site residues shown to abolish activity in other short-chain dehydrogenases (147) were chosen for mutation: Asn97(*), S126(*) and Y139(*). (*B*) The methanol extract from TclP N97L showed no detectable bioactivity, however, extracts from TclP S126A and TclP Y139F retained activity. (*C*) (*left*) Overlay of TclP model (magenta) (Modeller) with tetrameric (*R*)-hydroxypropyl-coenzyme M dehydrogenase (2CFC, cyan), NAD^+ is shown as green sticks. (*middle*) Close-up showing four conserved residues important for catalysis in other short-chain dehydrogenases (146, 147, 157). (*right*) Alternative orientation of active site, showing H-bonds from side chain of Asn114 (Asn97 in TclP) to backbone of Ile90 in 2CFC. (*D*) ESI mass spectrum of TEV-cleaved purified GST-peptide from TclP N97L mutant; major peak has *m/z* expected for peptide with an intact threonine at the C-terminus (i.e., no decarboxylation). Since no C-terminal processing has occurred, the isotopic pattern (*inset*) is consistent with a single product rather than a mix of alcohol and ketone as found for intermediates that have undergone decarboxylation. (*E*) MALDI-MS data for PP from ΔTclP (*top*) and TclP N97L (*bottom*) both show no oxidative decarboxylation activity as the major product is consistent with a peptide containing 6 thiazoles. Peaks consistent with minor amounts of 1-2 Ser/Thr dehydration are present (*) along with peaks for subsequent glutathione additions (#).

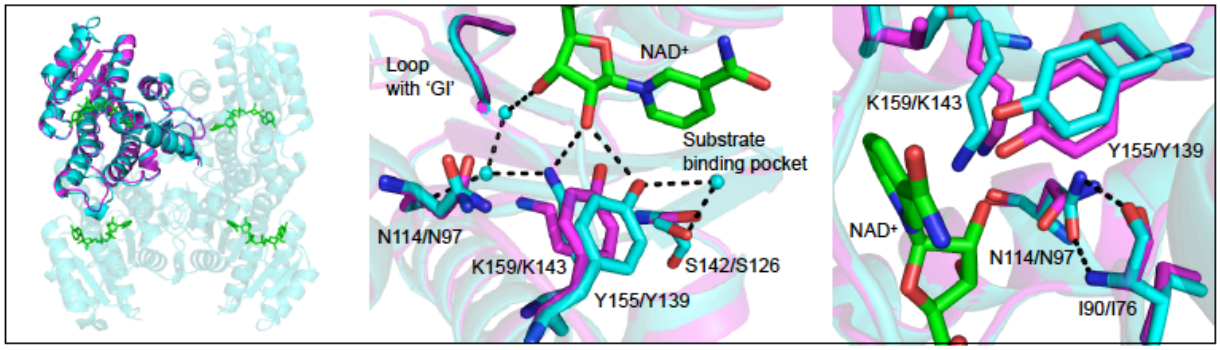
A

```
TclP    1 MNILIVGASSEIAHYIINYHK-IKDQVFLLDLPSQID-----N-LK-  39
2CFC    3 RVAIVTGASSGNGLAIATRFLARGDRVAALDLSAETLEETARTHWHA  49

TclP   40 KWD----TNFDTLDVQNNKEIETYFKECNITFDKLYYLVGINTM---  79
2CFC   50 YADKVLRVRADVADEGDVNAAIAATMEQFGAIDVLVNNAGITGNSEA  96
                                  *
TclP   80 KNGLDFNSQEWDNIMGTNLKSFYFFVKEFTKKNVINNI--PATIVSI 124
2CFC   97 GVLHTTPVEQFDKVMAVNVRGIFLGCRAVLPH---MLLQGAGVIVNI 140
           *                 *
TclP  125 ASQHGVVANAYRTPYCVSKAGLIHLTRVLALELSLYDIRVNCVSPGF 171
2CFC  141 ASVASLVAFPGRSAYTTSKGAVLQLTKSVAVDYAGSGIRCNAVCPGM 187

TclP  172 ILNSKSHEFLNNPKVKKEYLSKTPLQRYITPNEVANCCIFL--NNST 216
2CFC  188 IETPMTQWRLDQPELRDQVLARIPQKEIGTAAQVADAVMFLAGEDAT 234

TclP  217 SITCQNLIIDGGYTIW                               232
2CFC  235 YVNGAALVMDGAYTAI                               250
```

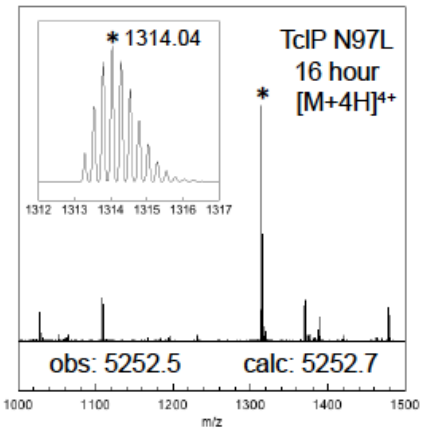
B



C



D



E

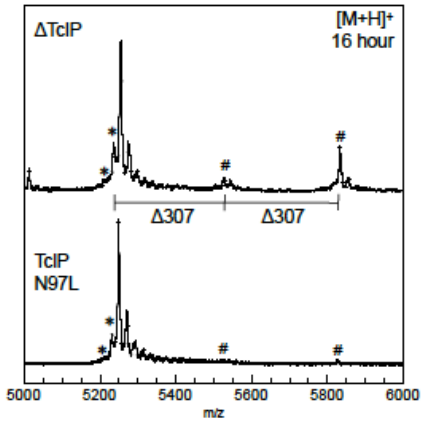
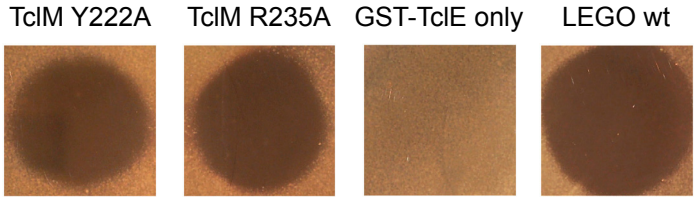


Figure 4-S8. Alignments, mutant bioactivity assays and mass spectrometry data for TclM. Currently no homologous crystal structures are available for structure modeling of TclM. (*A*) Clustal Omega sequence alignment of several TclM homologs from different thiopeptide clusters including McTclM: GenBank AIU53948.1, BcTclM: GenBank AAP11951.1, TbtD: GenBank ADG87279.1, LazC: GenBank BAO57436.1, BerD: GenBank AGN11669.1, NocO: GenBank ADR01090.1, NosO: GenBank ACR48344.1, GetF: GenBank AEM00619.1. Blue highlighting indicates sequence conservation (light = low, dark = high). Red (*) indicate amino acids that are the most important for catalytic activity in TbtD (77). (*B*) Bioassays of methanol extracts indicate mutations of two conserved residues, Y222A and R235A, in TclM did not abolish activity. (*C*) Expanded window of ESI mass spectrometry data for TEV-cleaved peptides isolated from ΔTclM and ΔTclMΔTclS deletion mutants (same data as Figure 4-4). (*D*) (*Left*) MALDI-MS data for TEV-cleaved peptide from a 16-h culture of ΔTclM mutant strain. The long culture resulted in cleavage of the (presumably processed) core peptide leaving just the leader peptide (`GGSEFQTNNIEGLDVTDLEFISEEVTEKDEKEIMGA`, expected *m/z* 3976 after TEV cleavage) attached to the GST-tag. (*Right & table*) MALDI-MS data for TEV-cleaved peptide obtained from an 8-h culture of the ΔTclM mutant strain that was eluted from the affinity resin with 10 mM reduced glutathione prior to TEV-cleavage. Peaks d-g are consistent with full length modified precursor with 0-3 additions of glutathione (Δ307 *m/z*) to the Dha/Dhb residues. Peaks a-c (*) are consistent with 1-3 GSH addition products of the NN cleavage product described in Figure 4-S4.

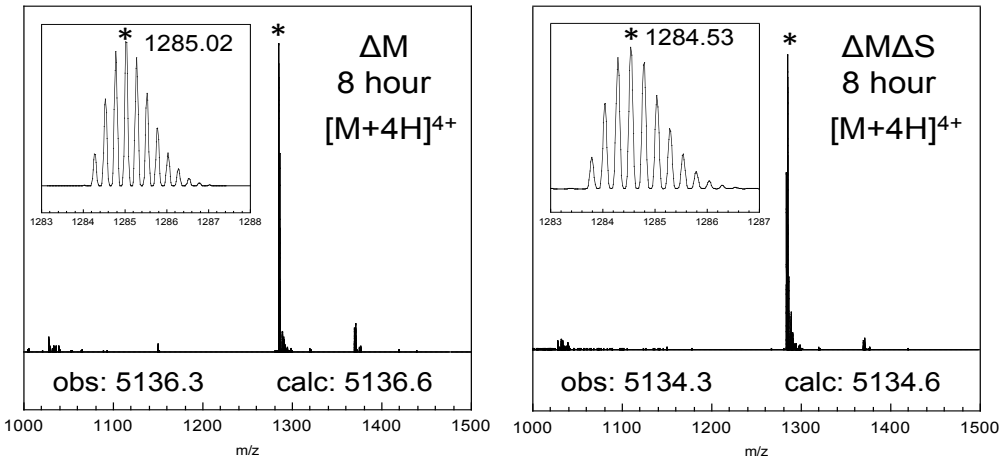
A

```
McTclM   92 EFNYESLERGMISYCSHSEGFLSIPK-NQKFKKIFKEGYLKNEHVLDLII-----NNRKDS-  146
BcTclM  135 HYNYDSIKKGYLSYASHVNGFFTRWKDPNKIRDIFHKNYLNNKEYLESKVSEIIDNNNRSS-  195
TbtD    175 ALSTGGLPVARTSLRSHAEAYLARRSDGVRLRELWRDHYARNREAFTERLIAVASSAESAEN  236
LazC    191 TAHPDGLVGAHYSVLSHLEDFLVHEDPDGSLRAAFERRWEQSGRAVTALVGRIADGGA----  248
BerD    190 YEFRELLSLRLLSYRSHFEAIYLRTKDPQSFDAACARFYEQVGA----GVREFITACGDPDD  247
NocO    135 -RSGDGPPLSFLSFRSHAEAFLATTRDPNAARHAFDTRYTDHRTTVEAAVRAILLDGDPG--  193
NosO    179 EELREGVPLGFLSYRSHAEAFLASSRDPKAAQAMMDAKYTRAAATLERLVDGVLTQCEERGP  240
GetF     77 EFGRGGLVSAALSFRSHAEAYLNLEAA-PDERAAWDAAARASAPALRRRLLAVATGPD----  133

McTclM  147 ---FFYTYHIDTIISELKPCIRNSIKKNEIHFLNID-------------HSKNN----DQLT  188
BcTclM  196 ------LSELSDIITEMKKEMTTDIEKGNLHVFNIE-------------LLQKPGERDFLEK  238
TbtD    237 GAHLPHVREWVRRLRPIRERARALLESGELTLEY-A-SPA---------EGARDLP-SLAEV  286
LazC    249 ---RDWERDWAHWSATAWSLAERRLTAGADLGGRHAEYRERAEALGDPATAERWNAELRTRY  307
BerD    248 DPADEMVRLWTKSITSESSHLAENFSDGSVVNAGH--TLE---------DLVRKRG-APVEP  297
NocO    194 -----DAAPWSAAARAAKPRFTAGFASADLVAHT--GYTR---------DHLRERT-D-FAD  237
NosO    241 V-VSLPARQWYEAMRAAKPAVTELFRAGTDLALDTEEQPP---------DTGPDGK-G-LSE  290
GetF    134 --RPAYARDWLGLITPLVRAAEQAQRRGELALPT---------------LAEGFSS-DLTER  177

                                  *                   *
McTclM  189 SDFHQHMLSNEKFLKFMRCDIDFLTSRFLTIAQYFLLKNMGISNINRYFTCYLTYKSLSNFT  250
BcTclM  239 SQFHKTILNNPDFSNFMNKDINFLGSRLITVFTYLLIRNLGIQNKDRYLLCYYIYKIIEEKY  300
TbtD    287 SAFHRELESRPEWAR-LRDSPAFGAYRLVINCTYLHLTRLGLTPHQRFLVCHLAADAAADVY  347
LazC    308 SEFHRMLQRADPDGR-MWHRPDYLINRAGTNGLYRLLAICDVRPMERYLAAHLLVRSVPELT  368
BerD    298 TRFH--TPPSPELDRLMHRDADFLAFRLQTSLLYSCLYTLGFSLAERYVFCYVVARANEDVC  357
NocO    238 NPFHSRAGASEQLQAYLGGDPSFLATRLLTSLLYVTLHSSGVSLMQRYFLCHAIAKACESIY  299
NosO    291 SAFHRIVEGSDGLRDFLDRDPSFLATRLLTSLLYLSLSSVGIALAERYFLCYAVSRACESIF  352
GetF    178 SAFHRGLAGSTSWED-VRTSDWFVLYRFAINLLYLQLSRLGVKPVGRYRLCHLVATALDQRE  238

McTclM  251 SKNPNDLIKYFKED-------------------------  264
BcTclM  301 NIDTLELIRDFGKGRD---NNVEDL----QRY-------  325
TbtD    348 GIAAHEEVAT---R-------------------------  358
LazC    369 GHRWQTLLGAAEQPGGPEQSGAAGATGGAGRTKLEGAA  406
BerD    358 GKSMKELQDELDGLAR---SMASGSTKTAE---------  384
NocO    300 HVDSMSLLAD---LAV---G-------------------  313
NosO    353 DTDALTVLSG---LAR---TSLAS---------------  370
GetF    239 APTTTAEEGD---S-------------------------  249
```

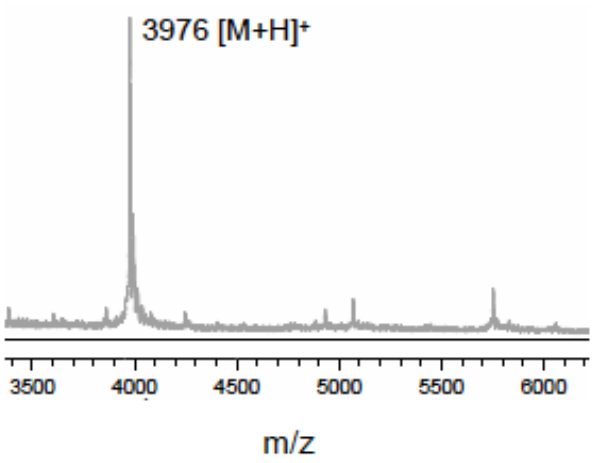
B

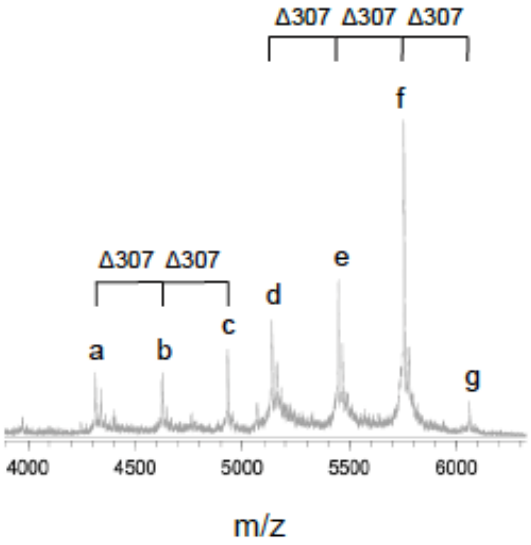


C



D





peak	$[M+H]^+$	# of GSH adducts
a	4319*	1
b	4626*	2
c	4932*	3
d	5138	0
e	5447	1
f	5754	2
g	6062	3

Table 4-S1. Plasmids used to generate *tcl* mutant strains.

Plasmid	Genotype[a]	Description[b]
L7152D7	TclJ E319A	TclJ E319A was amplified by OE-PCR using oPB465 as external forward primer, oPB471 as internal reverse primer, oPB470 as internal forward primer and oPB474 as external reverse primer
L7153C5	TclK R181A	TclK R181A was amplified by OE-PCR using oPB475 as external forward primer, oPB478 as internal reverse primer, oPB477 as internal forward primer and oPB476 as external reverse primer
L7154C1	TclL R49A	TclL R49A was amplified by OE-PCR using oPB447 as external forward primer, oPB485 as internal reverse primer, oPB484 as internal forward primer and oPB483 as external reverse primer
L769A1	TclM Y222A	TclM Y222A was amplified by OE-PCR using oPB500 as external forward primer, oPB534 as internal reverse primer, oPB533 as internal forward primer and oPB011 as external reverse primer
L769B1	TclM R235A	TclM R235A was amplified by OE-PCR using oPB500 as external forward primer, oPB536 as internal reverse primer, oPB535 as internal forward primer and oPB011 as external reverse primer
L769C1	TclP N97L	TclP N97L was amplified by OE-PCR using oPB507 as external forward primer, oPB538 as internal reverse primer, oPB537 as internal forward primer and oPB506 as external reverse primer
L769D1	TclP S126A	TclP S126A was amplified by OE-PCR using oPB507 as external forward primer, oPB540 as internal reverse primer, oPB539 as internal forward primer and oPB506 as external reverse primer
L769E1	TclP Y139F	TclP Y139F was amplified by OE-PCR using oPB507 as external forward primer, oPB542 as internal reverse primer, oPB541 as internal forward primer and oPB506 as external reverse primer
L783A1	TclN R269A/S271A	TclN R269A/S271A was amplified by OE-PCR using oPB448 as external forward primer, oPB549 as internal reverse primer, oPB550 as internal forward primer and oPB503 as external reverse primer
L762A1	ΔTclM	pLEGOΔTclM was constructed by treating pLEGO with SacI/SalI/Klenow and religating plasmid
L762D1	ΔTclS	pLEGOΔTclS was constructed by treating pLEGO with XhoI/BglII/Klenow and religating plasmid
L783G2	ΔTclL	pLEGOΔTclL was constructed by treating pLEGO with PstI/SacI/Klenow and religating plasmid
L794C2	ΔTclM,ΔTclS	pLEGOΔTclMS was contructed by treating pLEGOΔTclM (L762A1) with XhoI/BglII/Klenow and religated

[a] Genotypes represent variations in pLEGO, which integrate into the *amyE* locus of *B. subtilis*; [b] Full primer sequences are provided in Table 4-S2

Table 4-S2. Primers used in this study.

Primer	Sequence (5'-3')
oPB281	CCGCCTTGAAAATACAGGTTTTCTCCTCCTTTTGGAGGATGGTCGC
oPB282	CCTGTATTTTCAAGGCGGATCAGAATTCCAAACAAACAAT
oPB288	gcgTCTAGAAAGGaGGtgGAGGTCATGTCCCCTATACTAGGTTA
oPB289	cgcGGATCCTTAAGTTGTACAACAACTGCA
oPB011	GCTGACACAACTTCTCCTGG
oPB447	GTGAATCCTCGTCTTAATGATTCA
oPB448	CTCTATCCAATTTCACATCTAAAAATCC
oPB465	CAAGTTTAATGTCCAAACTTTTGAAGGTG
oPB470	GTATATTATGGCCTAATGGCAGTTATCGAGAGAGACTCTT
oPB471	AAGAGTCTCTCTCGATAACTGCCATTAGGCCATAATATAC
oPB474	CTGTTACATTTACGCGTTATACTATAAGGC
oPB475	GTACCTGGATTATTACCTATGACTTTCGG
oPB476	AAGATATGTAGGCTTGTCCAATTCA
oPB477	CCATAAAAAATTACTACAATGCAATGAATTTAAAACCTTCTC
oPB478	GAGAAGGTTTTAAATTCATTGCATTGTAGTAATTTTTTATGG
oPB483	GATAAGTATTCGTCACTTTTATCACCC
oPB484	GGTCCTCATATCCGTCTAGCAATAGCTAATATTACAAAGA
oPB485	TCTTTGTAATATTAGCTATTGCTAGACGGATATGAGGACC
oPB500	CGAACAATAGATTAGGAATCAAACC
oPB503	CTATCTGAGAAGGAAGATCAAGTAG
oPB506	GTAAAGCAATAGGACTTGTTCCTCC
oPB507	GGTCTATCTACCCACCCAATGATG
oPB533	TTTCTTACAATAGCTCAAGCCTTTTTATTAAAAAACATGG
oPB534	CCATGTTTTTTAATAAAAAGGCTTGAGCTATTGTAAGAAA
oPB535	CATGGGAATAAGTAATATTAACGCATATTTCACTTGCTAT
oPB536	ATAGCAAGTGAAATATGCGTTAATATTACTTATTCCCATG
oPB537	GACAATATAATGGGGACTCTTCTCAAAAGCTTTTATTTC
oPB538	GAAATAAAAGCTTTTGAGAAGAGTCCCCATTATATTGTC
oPB539	CCATAGTTAGTATTGCAGCTCAACATGGCGTC
oPB540	GACGCCATGTTGAGCTGCAATACTAACTATGG
oPB541	CTAATGCCTATAGAACTCCTTTTTGTGTAAGTAAAGCTGG
oPB542	CCAGCTTTACTTACACAAAAAGGAGTTCTATAGGCATTAG
oPB549	AAATTCGCCACCCGGAGCTATTGCTTTTTGTATTTCGAAT
oPB550	ATTCGAAATACAAAAAGCAATAGCTCCGGGTGGCGAATTT

Chapter 5. Future work and conclusions

The work described herein constitutes the most detailed analysis of the micrococcin biosynthetic pathway to date and establishes the most tractable genetic system for manipulation of any thiopeptide pathway. The versatility of these tools is highlighted by the speed and ease with which we have been able to make changes in pathway components and analyze the resulting products. We have also described a previously unknown mechanism of thiopeptide immunity in the form of a ribosomal target replacement mechanism. This mechanism is clearly also exploited by the *B. cereus* thiocillin producing strain which harbors two copies of the immunity gene (*tclQ* and *tclT*), although it is unknown whether other thiopeptide producers use this or similar strategies for self protection. While this enhanced understanding of thiopeptide production and immunity will inform future efforts to understand and engineer similar RiPP pathways, there is still much that is unknown or underdeveloped. Below we describe a few examples of current and future projects designed for this purpose, highlighting some of the most intriguing questions regarding thiopeptide biology that remain to be answered.

5.1 Investigating the regulation of TclQ-mediated resistance

A fascinating yet incompletely addressed question pertains to how TclQ-mediated resistance is regulated in the native MP1 producing *M. caseolyticus* strain. This strain constitutively expresses and secretes bioactive micrococcin yet is immune to its effects. We have demonstrated that the observed immunity can be explained by the co-expression of an L11 protein variant (TclQ) which functions at the ribosome but blocks the action of micrococcin (76). However, loss of the TclQ-encoding replicon results in cells that are defective in immunity implying that the native L11 protein still binds micrococcin (see Figure 2-2A). It is puzzling that an antibiotic-producing

bacteria would retain both sensitive and resistant copies of its protein target within the same cell, especially since spontaneous resistance is so easily achieved by relatively minor mutations in *rplK*, none of which appear to confer a noticeable fitness defect. One possible explanation is that macrocyclization (and by extension, bioactivity) coincides with export from cells and that TclQ largely serves as a backup immunity function in the event that antibiotic concentrations overwhelm the cell's ability to export the active compound. Another plausible explanation is that TclQ-expression is regulated at a transcriptional level, and is only expressed when micrococcin production is occurring. This hypothesis would suggest that micrococcin production in nature is not always constitutive, and that it benefits cells to only express TclQ when needed. The observed presence (but unexplained role) of TclU — a predicted MerR-type transcriptional regulator — in both the *B. cereus* and *M. caseolyticus* clusters gives support to such an idea, however, attempts to define its function in *B. subtilis* were not fruitful. Efforts to study the role of TclU in regulating TclQ and/or L11 expression in the native strain background are more likely to be informative and should be pursued.

One final explanation is that resistance occurs at a population level but that at a single cell level, cells display heterogeneity in terms of resistance. Perhaps the native L11 protein binds to ribosomes with higher affinity and confers greater fitness than TclQ, but is selected against in the presence of antibiotic. It could be that such competing fitness effects results in heterogeneous cell populations that collectively display the observed resistance phenotype. Distinguishing between these hypotheses would be greatly aided by transcriptomic and proteomic approaches where the relative expression and stoichiometry of TclQ and L11 can be compared. *In vitro* ribosome binding assays could also be employed to test whether differences in binding affinity contribute to this phenomenon. It is worth noting that each of these hypotheses relies upon the

assumption that the native L11 protein confers a fitness advantage over TclQ. While this appears to hold true in *B. subtilis* (Fig. 2-6), competition experiments conducted in *M. caseolyticus* could be used to validate this assumption.

5.2 Exploring biophysical interactions between Tcl processing proteins

Another poorly understood aspect of RiPP biosynthesis relates to how enzymes physically interact to coordinate so many posttranslational modifications on so few core peptide residues. The versatility of the *B. subtilis* micrococcin platform allows for such analysis to be conducted quite easily using co-immunoprecipitation (co-IP) studies. We have previously demonstrated that each Tcl protein can tolerate affinity tags on at least one terminus (Fig. 3-4). Exploiting this feature, one could purify individual proteins by affinity chromatography (e.g. Strep-TclJ) and test whether other Tcl pathway members co-purify. Identification of interacting proteins could be accomplished initially by LC-MS then validated using western blots with complementary epitope tags for detection (e.g. FLAG, HA, His_6). Anticipated interactions may include TclIJN, which together coordinate the complete conversion of all core cysteines to thiazoles, and TclKL (which function as two domains of the same protein in the homologous lanthipeptide systems). Other potential interactions might exist between TclI- and TclK- with the TclE leader. TclI and TclK are both predicted to possess putative RiPP recognition elements (RREs) but direct binding to the TclE leader peptide has not been demonstrated. Alternative strategies for probing protein-protein interactions include split-protein reporter assays (e.g. bacterial adenylate cyclase two-hybrid, split-GFP etc.) and fluorescence anisotropy.

5.3 Pushing the limits of precursor peptide permissiveness

It is known that many RiPP enzymes tolerate significant substrate variation making them attractive targets for peptide engineering studies. In the case of thiopeptides, the ability of modifying enzymes to recognize precursors with altered core peptide sequences has been demonstrated in numerous studies (30, 67-69), however in most cases, changes have been minimal, testing only the effects of relatively minor substitutions or insertions one at a time. With the exception of the recent *in vitro* thiomuracin work, the scope of these studies has likely been limited by the difficulty of genetically manipulating producing strains or by a limited ability to purify untagged intermediates. The availability of our versatile *B. subtilis* platform overcomes both of these obstacles due to the ease with which genetic changes can be incorporated and by the ability to purify and both leaderless macrocyclic peptides and linear intermediates. We are in the process of designing experiments to test the effects of radical changes to the TclE precursor peptide, both in the leader and in the core peptide sequence. Truncations and alanine scans of the leader peptide will likely provide important information about key residues for binding and define the minimally sufficient leader for different stages of processing. Similarly, efforts to scramble the TclE core peptide sequence and to create fusions of other RiPP cores with the TclE leader will likely provide greater insight into the fine details of Tcl substrate recognition and processing.

5.4 Exploring unknown chemical space via massively parallel peptide depletion sequencing

From an engineering perspective, the permissiveness of RiPP core peptides to sequence variation is one of the most attractive features of these pathways. This property, combined with their ribosomal synthesis, allows for the generation of tremendous chemical complexity via simple

genetic changes. Several studies have demonstrated the potential for small-scale RiPP diversification (8, 55), however the availability of a well-studied, genetically tractable system presents the opportunity to generate random RiPP libraries on an entirely different level.

Current dogma strongly suggests that Phase-I reactions (Cys-thiazole conversion and C-terminal decarboxylation) are least sensitive to core peptide sequence changes and thus the best to target for initial peptide diversification strategies. One can imagine a system in which the TclE precursor peptide is expressed with multiple residues randomized by incorporating degenerate codons at these positions. When expressed in the presence of TclIJNP, these enzymes can act to cyclize and decarboxylate these random peptides and create a library of highly modified and diverse peptides, which can be screened for novel biological properties. We are currently building such a platform in our lab using a bipartite *E. coli* expression system, where the Phase-I enzymes (TclIJNP) are encoded on a single, tightly inducible plasmid, with the precursor peptide (TclE) encoded on a second inducible, compatible vector. The inducible control of these plasmids confers another major advantage: an ability to selectivity repress expression of toxic compounds, making this system a powerful tool to not only generate large libraries of highly modified cyclic peptides but also a way to rapidly evaluate their bioactivity. Consider the following example of how such an experiment could be accomplished: (1) a library of peptide-encoding *E. coli* is generated using the codon randomization strategy outlined above. (2) The library is split between two culture flasks: one is cultured in the presence of inducer causing toxic peptide-producing cells to die, the other culture is left uninduced and serves as a reference culture. (3) After ample time for self-inhibition is allowed, the pooled plasmid DNA from each flask is subjected to PCR to amplify the precursor encoding gene, followed by deep sequencing to identify all prospective peptides in that population. (4) The sequence profiles of the uninduced

(reference) and induced cultures are compared and toxic peptides are identified on the basis of their absence (or depletion) of their encoding sequences in the induced culture.

Such a strategy has the potential to uncover a large number of cytotoxic compounds that can be validated by recloning these sequences and testing their effects in monoculture. Validated leads could then be further investigated using MS and NMR for structural identification followed by transcriptomics/metabolomics and resistance studies to explore mechanisms of action. We have already shown that the Phase-I Tcl enzymes function correctly in *E. coli* (unpublished data) and proof-of-concept experiments with random linear peptides have shown that this is a viable strategy for rapidly generating peptide diversity and simultaneously screening for bioactivity.

5.5 Altering the substrate specificities of Tcl processing enzymes

Recent studies using the *in vitro* thiomuracin system have shown that the Phase I reactions depend largely on the leader peptide for substrate recognition and are quite tolerance to changes in the core peptide sequence (30). In contrast, Phase II reactions (Ser/Thr dehydration and macrocyclization) are reportedly less leader-dependent and are more dependent on core peptide structure. The increased selectivity of Phase II enzymes is further evidenced by observations that in all thiopeptide systems, Cys modification occurs almost ubiquitously across the core whereas Ser/Thr dehydration only occurs on specific residues. As additional details of biosynthetic processing are uncovered, opportunities to engineer enzymes with relaxed substrate specificities will inevitably emerge. Since many thiopeptide chemistries are uncommon in nature, being able to direct these enzymes to any substrate of interest would provide new tools for biotechnology and allow us to expand our peptide diversification and screening platforms for new levels of chemical diversity. Engineering of these enzymes may not be straightforward, however structural

modeling combined with mutant studies will likely eventually lead to variants with new and more desirable specificities and affinities.

5.6 Conclusions

Combined with the work presented in these research chapters, such future studies are likely to yield greater insight into the biosynthesis of these complex natural products and unlock greater chemical diversity than has been previously observed in a laboratory environment. Equipped with new tools and guided by nature, it is highly likely that RiPP systems like these will serve as a rich source of new compounds for future generations and will help to reinvigorate a vastly depleted antibiotic supply.

REFERENCES

1. CDC (2013) *Antibiotic resistance threats in the united states, 2013* (Centres for Disease Control and Prevention, US Department of Health and Human Services).

2. Magiorakos AP*, et al.* (2012) Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect* 18(3):268-281.

3. Falagas ME & Karageorgopoulos DE (2008) Pandrug resistance (PDR), extensive drug resistance (XDR), and multidrug resistance (MDR) among Gram-negative bacilli: Need for international harmonization in terminology. *Clin Infect Dis* 46(7):1121-1122

4. Silver LL (2011) Challenges of antibacterial discovery. *Clin Microbiol Rev* 24(1):71-109.

5. Chioro A*, et al.* (2015) Antimicrobial resistance: A priority for global health action. *Bull World Health Organ* 93(7):439-439.

6. United Nations (2016) High-level meeting on antimicrobial resistance. in *United Nations General Assembly* (UN Headquarters, New York).

7. Fleming AG (1929) Responsibilities and opportunities of the private practitioner in preventive medicine. *Can Med Assoc J* 20(1):11-13.

8. Ruffner DE, Schmidt EW, & Heemstra JR (2015) Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. *ACS Synth Biol* 4(4):482-492.

9. Schatz A, Bugie E, & Waksman SA (1944) Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Exp Biol Med* 55(1):66-69.

10. Lewis K (2013) Platforms for antibiotic discovery. *Nat Rev Drug Discov* 12(5):371-387.

11. Walsh CT & Wencewicz TA (2013) Prospects for new antibiotics: A molecule-centered perspective. *J Antibiot (Tokyo)* 67(1):7-22.

12. Cragg GM & Newman DJ (2013) Natural products: A continuing source of novel drug leads. *Biochim Biophys Acta* 1830(6):3670-3695.

13. Nichols D*, et al.* (2010) Use of iChip for high-throughput in situ cultivation of "uncultivable" microbial species. *Appl Environ Microbiol* 76(8):2445-2450.

14. Gavrish E*, et al.* (2014) Lassomycin, a ribosomally synthesized cyclic peptide, kills *Mycobacterium tuberculosis* by targeting the ATP-dependent protease ClpC1P1P2. *Chem Biol* 21(4):509-518.

15. Ling LL*, et al.* (2015) Erratum: A new antibiotic kills pathogens without detectable resistance. *Nature* 520(7547):388.

16. Ling LL, *et al.* (2015) A new antibiotic kills pathogens without detectable resistance. *Nature* 517(7535):455-459.

17. Arnison PG, *et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(1):108-160.

18. Roberts MF (1998) Production of alkaloids in plant cell culture. *Alkaloids*, (Springer Science + Business Media), pp 159-197.

19. Schlager S & Drager B (2016) Exploiting plant alkaloids. *Curr Opin Biotechnol* 37:155-164.

20. Walsh CT, O'Brien RV, & Khosla C (2013) Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem Int Ed Engl* 52(28):7098-7124.

21. Chen H & Du L (2016) Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl Microbiol Biotechnol* 100(2):541-557.

22. Staunton J & Weissman KJ (2001) Polyketide biosynthesis: A millenium review. *Nat Prod Rep* 18(4):380-416.

23. Schwarzer D, Finking R, & Marahiel MA (2003) Nonribosomal peptides: From genes to products. *Nat Prod Rep* 20(3):275-287.

24. Staunton J & Wilkinson B (2001) Combinatorial biosynthesis of polyketides and nonribosomal peptides. *Curr Opin Chem Biol* 5(2):159-164.

25. Strieker M, Tanovic A, & Marahiel MA (2010) Nonribosomal peptide synthetases: Structures and dynamics. *Curr Opin Struct Biol* 20(2):234-240.

26. Skinnider MA, *et al.* (2016) Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proc Natl Acad Sci U S A*.

27. Metelev MV & Ghilarov DA (2014) Structure, function, and biosynthesis of thiazole/oxazole-modified microcins. *Mol Biol* 48(1):29-45.

28. Burkhart BJ, Hudson GA, Dunbar KL, & Mitchell DA (2015) A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat Chem Biol* 11(8):564-570.

29. Dunbar KL, Tietz JI, Cox CL, Burkhart BJ, & Mitchell DA (2015) Identification of an auxiliary leader peptide-binding protein required for azoline formation in ribosomal natural products. *J Am Chem Soc* 137(24):7672-7677.

30. Zhang Z, *et al.* (2016) Biosynthetic timing and substrate specificity for the thiopeptide thiomuracin. *J Am Chem Soc* In press, DOI: 10.1021/jacs.6b08987

31. Yang X & van der Donk WA (2013) Ribosomally synthesized and post-translationally modified peptide natural products: New insights into the role of leader and core peptides during biosynthesis. *Chemistry* 19(24):7662-7677.

32. Crone WJK, Leeper FJ, & Truman AW (2012) Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: Expanding the biosynthetic diversity of ribosomal peptides. *Chem Sci* 3(12):3516.

33. Gomez-Escribano JP, Song L, Bibb MJ, & Challis GL (2012) Posttranslational β-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chem Sci* 3(12):3522.

34. Hou Y*, et al.* (2012) Structure and biosynthesis of the antibiotic bottromycin D. *Org Lett* 14(19):5050-5053.

35. Huo L, Rachid S, Stadler M, Wenzel Silke C, & Müller R (2012) Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem Biol* 19(10):1278-1287.

36. Lee J, McIntosh J, Hathaway BJ, & Schmidt EW (2009) Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J Am Chem Soc* 131(6):2122-2124.

37. Chatterjee C, Paul M, Xie L, & van der Donk WA (2005) Biosynthesis and mode of action of lantibiotics. *Chem Rev* 105(2):633-684.

38. Knerr PJ & van der Donk WA (2012) Discovery, biosynthesis, and engineering of lantipeptides. *Annu Rev Biochem* 81:479-505.

39. Willey JM & van der Donk WA (2007) Lantibiotics: Peptides of diverse structure and function. *Annu Rev Microbiol* 61(1):477-501.

40. Schnell N*, et al.* (1988) Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature* 333(6170):276-278.

41. Cotter P, Hill C, & Ross R (2005) Bacterial lantibiotics: Strategies to improve therapeutic potential. *Curr Protein Pept Sci* 6(1):61-75.

42. Melby JO, Dunbar KL, Trinh NQ, & Mitchell DA (2012) Selectivity, directionality, and promiscuity in peptide processing from a *Bacillus* sp. Al hakam cyclodehydratase. *J Am Chem Soc* 134(11):5309-5316.

43. Milne JC*, et al.* (1999) Cofactor requirements and reconstitution of microcin B17 synthetase: A multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17. *Biochemistry* 38(15):4768-4781.

44. Li YM, Milne JC, Madison LL, Kolter R, & Walsh CT (1996) From peptide precursors to oxazole and thiazole-containing peptide antibiotics: Microcin B17 synthase. *Science* 274(5290):1188-1193.

45. Heddle JG, *et al.* (2001) The antibiotic microcin B17 is a DNA gyrase poison: Characterisation of the mode of inhibition. *J Mol Biol* 307(5):1223-1234.

46. Yorgey P, *et al.* (1994) Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. *Proc Natl Acad Sci U S A* 91(10):4519-4523.

47. Scholz R, *et al.* (2011) Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. *J Bacteriol* 193(1):215-224.

48. Hansen JN & Sandine WE (1994) Nisin as a model food preservative. *Crit Rev Food Sci Nutr* 34(1):69-93.

49. van Kraaij C, *et al.* (1999) Lantibiotics: Biosynthesis, mode of action and applications. *Nat Prod Rep* 16(5):575-587.

50. Cotter PD, Hill C, & Ross RP (2005) Food microbiology: Bacteriocins: Developing innate immunity for food. *Nat Rev Microbiol* 3(10):777-788.

51. Parks WM, Bottrill AR, Pierrat OA, Durrant MC, & Maxwell A (2007) The action of the bacterial toxin, microcin B17, on DNA gyrase. *Biochimie* 89(4):500-507.

52. Just-Baringo X, Albericio F, & Álvarez M (2014) Thiopeptide antibiotics: Retrospective and recent advances. *Mar Drugs* 12(1):317-351.

53. Donia Mohamed S, *et al.* (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* 158(6):1402-1414.

54. Velasquez JE & van der Donk WA (2011) Genome mining for ribosomally synthesized natural products. *Curr Opin Chem Biol* 15(1):11-21.

55. Young TS & Walsh CT (2011) Identification of the thiazolyl peptide GE37468 gene cluster from *Streptomyces* ATCC 55365 and heterologous expression in *Streptomyces lividans*. *Proc Natl Acad Sci U S A* 108(32):13053-13058.

56. Flinspach K, Kapitzke C, Tocchetti A, Sosio M, & Apel AK (2014) Heterologous expression of the thiopeptide antibiotic GE2270 from *Planobispora rosea* ATCC 53733 in *Streptomyces coelicolor* requires deletion of ribosomal genes from the expression construct. *PLoS ONE* 9(3):e90499.

57. Bennallack PR, *et al.* (2016) Reconstitution and minimization of a micrococcin biosynthetic pathway in *Bacillus subtilis*. *J Bacteriol* 198(18):2431-2438.

58. Kluskens LD, *et al.* (2005) Post-translational modification of therapeutic peptides by NisB, the dehydratase of the lantibiotic nisin. *Biochemistry* 44(38):12827-12834.

59. Moll GN, Kuipers A, & Rink R (2010) Microbial engineering of dehydro-amino acids and lanthionines in non-lantibiotic peptides. *Antonie van Leeuwenhoek* 97(4):319-333.

60. Rink R, *et al.* (2007) Production of dehydroamino acid-containing peptides by *Lactococcus lactis*. *Appl Environ Microbiol* 73(6):1792-1796.

61. Tianero MDB, Donia MS, Young TS, Schultz PG, & Schmidt EW (2012) Ribosomal route to small-molecule diversity. *J Am Chem Soc* 134(1):418-425.

62. Donia MS, Ravel J, & Schmidt EW (2008) A global assembly line for cyanobactins. *Nat Chem Biol* 4(6):341-343.

63. Donia MS*, et al.* (2006) Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol* 2(12):729-735.

64. Knappe TA*, et al.* (2011) Introducing lasso peptides as molecular scaffolds for drug design: Engineering of an integrin antagonist. *Angew Chem Intl Ed Eng* 50(37):8714-8717.

65. Pan SJ & Link AJ (2011) Sequence diversity in the lasso peptide framework: Discovery of functional microcin J25 variants with multiple amino acid substitutions. *J Am Chem Soc* 133(13):5016-5023.

66. Wieland Brown LC, Acker MG, Clardy J, Walsh CT, & Fischbach MA (2009) Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. *Proc Natl Acad Sci U S A* 106(8):2549-2553.

67. Acker MG, Bowers AA, & Walsh CT (2009) Generation of thiocillin variants by prepeptide gene replacement and in vivo processing by *Bacillus cereus*. *J Am Chem Soc* 131(48):17563-17565.

68. Bowers AA, Acker MG, Koglin A, & Walsh CT (2010) Manipulation of thiocillin variants by prepeptide gene replacement: Structure, conformation, and activity of heterocycle substitution mutants. *J Am Chem Soc* 132(21):7519-7527.

69. Bowers AA, Acker MG, Young TS, & Walsh CT (2012) Generation of thiocillin ring size variants by prepeptide gene replacement and in vivo processing by *Bacillus cereus*. *J Am Chem Soc* 134(25):10313-10316.

70. Just-Baringo X, Albericio F, & Álvarez M (2014) Thiopeptide engineering: A multidisciplinary effort towards future drugs. *Angew Chem Intl Ed Eng* 53(26):6602-6616.

71. Engelhardt K, Degnes KF, & Zotchev SB (2010) Isolation and characterization of the gene cluster for biosynthesis of the thiopeptide antibiotic TP-1161. *Appl Environ Microbiol* 76(21):7093-7101.

72. Liao R*, et al.* (2009) Thiopeptide biosynthesis featuring ribosomally synthesized precursor peptides and conserved posttranslational modifications. *Chem Biol* 16(2):141-147.

73. Kelly WL, Pan L, & Li C (2009) Thiostrepton biosynthesis: Prototype for a new family of bacteriocins. *J Am Chem Soc* 131(12):4327-4334.

74. Bagley MC, Dale JW, Merritt EA, & Xiong X (2005) Thiopeptide antibiotics. *Chem Rev* 105(2):685-714.

75. Puar MS, *et al.* (1998) Sch 40832: A novel thiostrepton from *Micromonospora carbonacea*. *J Antibiot (Tokyo)* 51(2):221-224.

76. Bennallack PR, Burt SR, Heder MJ, Robison RA, & Griffitts JS (2014) Characterization of a novel plasmid-borne thiopeptide gene cluster in *Staphylococcus epidermidis* strain 115. *J Bacteriol* 196(24):4344-4350.

77. Hudson GA, Zhang Z, Tietz JI, Mitchell DA, & van der Donk WA (2015) In vitro biosynthesis of the core scaffold of the thiopeptide thiomuracin. *J Am Chem Soc* 137(51):16012-16015.

78. Dixon PD, Beven JE, & Cundliffe E (1975) Properties of the ribosomes of antibiotic producers: Effects of thiostrepton and micrococcin on the organisms which produce them. *Antimicrob Agents Chemother* 7(6):850-855.

79. Bausch SL, Poliakova E, & Draper DE (2005) Interactions of the N-terminal domain of ribosomal protein L11 with thiostrepton and rRNA. *J Biol Chem* 280(33):29956-29963.

80. Porse BT, Cundliffe E, & Garrett RA (1999) The antibiotic micrococcin acts on protein L11 at the ribosomal GTPase centre. *J Mol Biol* 287(1):33-45.

81. Porse BT, Leviev I, Mankin AS, & Garrett RA (1998) The antibiotic thiostrepton inhibits a functional transition within protein L11 at the ribosomal GTPase centre. *J Mol Biol* 276(2):391-404.

82. Rosendahl G & Douthwaite S (1994) The antibiotics micrococcin and thiostrepton interact directly with 23S rRNA nucleotides 1067A and 1095A. *Nucl Acids Res* 22(3):357-363.

83. Selva E, *et al.* (1997) Targeted screening for elongation factor Tu binding antibiotics. *J Antibiot (Tokyo)* 50(1):22-26.

84. Heffron SE & Jurnak F (2000) Structure of an EF-Tu complex with a thiazolyl peptide antibiotic determined at 2.35 Å resolution: Atomic basis for GE2270A inhibition of EF-Tu *Biochemistry* 39(1):37-45.

85. Parmeggiani A & Nissen P (2006) Elongation factor Tu-targeted antibiotics: Four different structures, two mechanisms of action. *FEBS Lett* 580(19):4576-4581.

86. Malcolmson SJ, Young TS, Ruby JG, Skewes-Cox P, & Walsh CT (2013) The posttranslational modification cascade to the thiopeptide berninamycin generates linear forms and altered macrocyclic scaffolds. *Proc Natl Acad Sci U S A* 110(21):8483-8488.

87. Mizuhara N, *et al.* (2011) Antifungal thiopeptide cyclothiazomycin B1 exhibits growth inhibition accompanying morphological changes via binding to fungal cell wall chitin. *Bioorg Med Chem* 19(18):5300-5310.

88. Rogers MJ, Cundliffe E, & McCutchan TF (1998) The antibiotic micrococcin is a potent inhibitor of growth and protein synthesis in the malaria parasite. *Antimicrob Agents Chemother* 42(3):715-716.

89. Clough B, Strath M, Preiser P, Denny P, & Wilson IRJM (1997) Thiostrepton binds to malarial plastid rRNA. *FEBS Lett* 406(1-2):123-125.

90. Lee M*, et al.* (2016) Micrococcin P1, a naturally occurring macrocyclic peptide inhibiting hepatitis C virus entry in a pan-genotypic manner. *Antiviral Res* 132:287-295.

91. Hegde NS, Sanders DA, Rodriguez R, & Balasubramanian S (2011) The transcription factor FOXM1 is a cellular target of the natural product thiostrepton. *Nat Chem* 3(9):725-731.

92. Bhat UG, Halasi M, & Gartel AL (2009) Thiazole antibiotics target FOXM1 and induce apoptosis in human cancer cells. *PLoS ONE* 4(5):e5592.

93. Ciufolini MA & Lefranc D (2010) Micrococcin P1: Structure, biology and synthesis. *Nat Prod Rep* 27(3):330.

94. Li C & Kelly WL (2010) Recent advances in thiopeptide antibiotic biosynthesis. *Nat Prod Rep* 27(2):153-164.

95. Spedding G & Cundliffe E (1984) Identification of the altered ribosomal component responsible for resistance to micrococcin in mutants of *Bacillus megaterium*. *Eur J Biochem* 140(3):453-459.

96. Bennallack PR*, et al.* (2016) Capture of micrococcin biosynthetic intermediates reveals C-terminal processing as an obligatory step for in vivo maturation. *Proc Natl Acad Sci U S A* 113(44):12450-12455.

97. Ortega MA*, et al.* (2014) Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* 517(7535):509-512.

98. Wever WJ*, et al.* (2015) Chemoenzymatic synthesis of thiazolyl peptide natural products featuring an enzyme-catalyzed formal [4 + 2] cycloaddition. *J Am Chem Soc* 137(10):3494-3497.

99. Su TL (1948) Micrococcin. An antibacterial substance formed by a strain of *Micrococcus*. *Br J Exp Pathol* 29(5):473-481.

100. Shoji JI*, et al.* (1976) Studies on antibiotics from the genus *Bacillus*. Viii. Isolation of three new antibiotics, thiocillins I, II and III, related to micrococcin P. *J Antibiot (Tokyo)* 29(4):366-374.

101. Brookes P, Fuller AT, & Walker J (1957) 137. Chemistry of micrococcin P. Part i. *J Chem Soc (Resumed)*:689.

102. Meyers CM & Jensen MM (1987) Staphylococcosis of turkeys. 3. Bacterial interference as a possible means of control. *Avian Dis* 31(1):74.

103. Wilkinson DM & Jensen MM (1987) Staphylococcosis of turkeys. 4. Characterization of a bacteriocin produced by an interfering S*taphylococcus*. *Avian Dis* 31(1):80.

104. Nicoll TR & Jensen MM (1987) Staphylococcosis of turkeys. 5. Large-scale control programs using bacterial interference. *Avian Dis* 31(1):85.

105. Harms JM*, et al.* (2008) Translational regulation via L11: Molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. *Mol Cell* 30(1):26-38.

106. Baumann S, Schoof S, Harkal SD, & Arndt H-D (2008) Mapping the binding site of thiopeptide antibiotics by proximity-induced covalent capture. *J Am Chem Soc* 130(17):5664-5666.

107. Walter JD, Hunter M, Cobb M, Traeger G, & Spiegel PC (2011) Thiostrepton inhibits stable 70S ribosome binding and ribosome-dependent GTPase activation of elongation factor G and elongation factor 4. *Nuc Acids Res* 40(1):360-370.

108. Cundliffe E & Thompson J (1981) Concerning the mode of action of micrococcin upon bacterial protein synthesis. *Eur J Biochem* 118(1):47-52.

109. Dahl EL & Rosenthal PJ (2007) Multiple antibiotics exert delayed effects against the *Plasmodium falciparum* apicoplast. *Antimicrob Agents Chemother* 51(10):3485-3490.

110. Clough B, Rangachari K, Strath M, Preiser PR, & Iain Wilson RJM (1999) Antibiotic inhibitors of organellar protein synthesis in *Plasmodium falciparum*. *Protist* 150(2):189-195.

111. Kwok JMM*, et al.* (2008) Thiostrepton selectively targets breast cancer cells through inhibition of Forkhead box M1 expression. *Mol Cancer Ther* 7(7):2022-2032.

112. Lukashin A (1998) Genemark.Hmm: New solutions for gene finding. *Nucleic Acids Res* 26(4):1107-1115.

113. Niedermeyer THJ & Strohalm M (2012) mMass as a software tool for the annotation of cyclic peptide tandem mass spectra. *PLoS ONE* 7(9):e44913.

114. Walsh CT, Malcolmson SJ, & Young TS (2012) Three ring posttranslational circuses: Insertion of oxazoles, thiazoles, and pyridines into protein-derived frameworks. *ACS Chem Biol* 7(3):429-442.

115. Agrawal RK, Linde J, Sengupta J, Nierhaus KH, & Frank J (2001) Localization of L11 protein on the ribosome and elucidation of its involvement in EF-G-dependent translocation. *J Mol Biol* 311(4):777-787.

116. Carnio MC*, et al.* (2000) The macrocyclic peptide antibiotic micrococcin P1 is secreted by the food-borne bacterium *Staphylococcus equorum* ws 2733 and inhibits *Listeria monocytogenes* on soft cheese. *Appl Environ Microbiol* 66(6):2378-2384.

117. Baba T*, et al.* (2008) Complete genome sequence of *Macrococcus caseolyticus* strain JSCS5402, reflecting the ancestral genome of the human-pathogenic *Staphylococci*. *J Bacteriol* 191(4):1180-1190.

118. Bowers AA, Walsh CT, & Acker MG (2010) Genetic interception and structural characterization of thiopeptide cyclization precursors from *Bacillus cereus*. *J Am Chem Soc* 132(35):12182-12184.

119. Trzasko A, Leeds JA, Praestgaard J, LaMarche MJ, & McKenney D (2012) Efficacy of LFF571 in a hamster model of *Clostridium difficile* infection. *Antimicrob Agents Chemother* 56(8):4459-4462.

120. Leeds JA, Sachdeva M, Mullin S, Dzink-Fox J, & LaMarche MJ (2012) Mechanism of action of and mechanism of reduced susceptibility to the novel anti-*Clostridium difficile* compound LFF571. *Antimicrob Agents Chemother* 56(8):4463-4465.

121. Naidu BN*, et al.* (2004) Nocathiacin I analogues: Synthesis, in vitro and in vivo biological activity of novel semi-synthetic thiazolyl peptide antibiotics. *Bioorg Med Chem Lett* 14(22):5573-5577.

122. Vojcic L, Despotovic D, Martinez R, Maurer KH, & Schwaneberg U (2012) An efficient transformation method for *Bacillus subtilis* DB104. *Appl Microbiol Biotechnol* 94(2):487-493.

123. Zhang XZ & Zhang Y (2011) Simple, fast and high-efficiency transformation system for directed evolution of cellulase in *Bacillus subtilis*. *Microb Biotechnol* 4(1):98-105.

124. Tegos GP & Hamblin MR (2013) Disruptive innovations: New anti-infectives in the age of resistance. *Curr Opin Pharmacol* 13(5):673-677.

125. Wright GD (2015) Solving the antibiotic crisis. *ACS Infect Dis* 1(2):80-84.

126. Hayashi S*, et al.* (2014) Genome mining reveals a minimum gene set for the biosynthesis of 32-membered macrocyclic thiopeptides lactazoles. *Chem Biol* 21(5):679-688.

127. Li J*, et al.* (2012) Thiofinder: A web-based tool for the identification of thiopeptide gene clusters in DNA sequences. *PLoS ONE* 7(9):e45878.

128. Morris RP*, et al.* (2009) Ribosomally synthesized thiopeptide antibiotics targeting elongation factor Tu. *J Am Chem Soc* 131(16):5946-5955.

129. Engelhardt K*, et al.* (2010) Production of a new thiopeptide antibiotic, TP-1161, by a marine *Nocardiopsis* species. *Appl Environ Microbiol* 76(15):4969-4976.

130. Dunbar KL & Mitchell DA (2013) Revealing nature's synthetic potential through the study of ribosomal natural product biosynthesis. *ACS Chem Biol* 8(3):473-487.

131. Thibodeaux CJ, Ha T, & van der Donk WA (2014) A price to pay for relaxed substrate specificity: A comparative kinetic analysis of the class II lanthipeptide synthetases ProcM and HalM2. *J Am Chem Soc* 136(50):17513-17529.

132. Koehnke J*, et al.* (2013) The cyanobactin heterocyclase enzyme: A processive adenylase that operates with a defined order of reaction. *Angew Chem Intl Ed Eng* 52(52):13991-13996.

133. Walker J, Olesker A, Valente L, Rabanal R, & Lukacs G (1977) Total structure of the polythiazole-containing antibiotic micrococcin P. A 13C nuclear magnetic resonance study. *J Chem Soc Chem Commun* (20):706.

134. Radeck J*, et al.* (2013) The *Bacillus* biobrick box: Generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. *J Biol Eng* 7(1):29.

135. Guérout-Fleury A-M, Frandsen N, & Stragier P (1996) Plasmids for ectopic integration in *Bacillus subtilis*. *Gene* 180(1-2):57-61.

136. Smale ST (2010) β-Galactosidase assay. *Cold Spring Harbor Protocols* 2010(5):pdb.prot5423-pdb.prot5423.

137. Chang AC & Cohen SN (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. *J Bacteriol* 134(3):1141-1156.

138. Nguyen HD*, et al.* (2005) Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability. *Plasmid* 54(3):241-248.

139. Ortega Manuel A & van der Donk Wilfred A (2016) New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem Biol* 23(1):31-44.

140. Tropea JE, Cherry S, & Waugh DS (2009) Expression and purification of soluble His_6-tagged TEV protease. in *Methods in Molecular Biology* (Springer Science + Business Media), pp 297-307.

141. Soding J, Biegert A, & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server):W244-W248.

142. Šali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779-815.

143. Dunbar KL*, et al.* (2014) Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat Chem Biol* 10(10):823-829.

144. Koehnke J*, et al.* (2015) Structural analysis of leader peptide binding enables leader-free cyanobactin processing. *Nat Chem Biol* 11(8):558-563.

145. Garg N, Salazar-Ocampo LMA, & van der Donk WA (2013) In vitro activity of the nisin dehydratase NisB. *Proc Natl Acad Sci U S A* 110(18):7258-7263.

146. Krishnakumar AM, Nocek BP, Clark DD, Ensign SA, & Peters JW (2006) Structural basis for stereoselectivity in the (R)- and (S)-hydroxypropylthioethanesulfonate dehydrogenases. *Biochemistry* 45(29):8831-8840.

147. Filling C (2002) Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J Biol Chem* 277(28):25677-25684.

148. Ozaki T, *et al.* (2016) Back cover: Insights into the biosynthesis of dehydroalanines in goadsporin. *ChemBioChem* 17(3):268-268.

149. Tocchetti A, *et al.* (2013) Capturing linear intermediates and C-terminal variants during maturation of the thiopeptide GE2270. *Chem Biol* 20(8):1067-1077.

150. Melby JO, Li X, & Mitchell DA (2014) Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry* 53(2):413-422.

151. Waterhouse AM, Procter JB, Martin DM, Clamp M, & Barton GJ (2009) Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.

152. Pettersen EF, *et al.* (2004) UCSF chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.

153. Sievers F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.

154. Zhang Z & Marshall AG (1998) A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J Am Soc Mass Spectrom* 9(3):225-233.

155. Rawlings ND, Barrett AJ, & Bateman A (2011) Asparagine peptide lysases: A seventh catalytic type of proteolytic enzymes. *J Biol Chem* 286(44):38321-38328.

156. Rose NL, *et al.* (2003) Involvement of dehydroalanine and dehydrobutyrine in the addition of glutathione to nisin. *J Agric Food Chem* 51(10):3174-3178.

157. Sliwa DA, Krishnakumar AM, Peters JW, & Ensign SA (2010) Molecular basis for enantioselectivity in the (R)- and (S)-hydroxypropylthioethanesulfonate dehydrogenases, a unique pair of stereoselective short-chain dehydrogenases/reductases involved in aliphatic epoxide carboxylation. *Biochemistry* 49(16):3487-3498.