UNIVERSITY OF CALIFORNIA, SAN DIEGO

Discovery and Investigation of the Thiazole/Oxazole Modified Microcin Natural Product Family

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Andrew L. Markley

Committee in charge:

Professor Jack E. Dixon, Chair
Professor Michael Burkart
Professor Edward Dennis
Professor Elizabeth A. Komives
Professor Victor Nizet

2011

The dissertation of Andrew Lute Markley is approved,

and it is acceptable in quality and form for publication on

microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2011

DEDICATION

I dedicate this work to my parents who enabled me to reach for my dreams and to my grandmother, Lottie for showing me how to live life the right way.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BSA	bovine serum albumin
CD	Circular Dichroism
CLS	clostridiolysin S
clos	clostridiolysin S associated gene
Cys	Cysteine
EDTA	Ethylenediaminetetraacetic Acid
erm	erythromycin
FMN	Flavin mononucleotide
FPLC	Fast Protein Liquid Chromatography
GAS	Group A *streptococcus*
Gly	Glycine
IPTG	Isopropyl β-D-1 thiogalactopyranoside
LC-MS	Liquid Chromatography Mass Spectrometry
list	listeriolysin S associated gene
NRPS	Nonribosomal peptide synthetase
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization with Time of Flight Detection
mcb	microcin B17 associated gene
MBP	maltose binding protein
ORF	Open reading frame
pag	*Pyrococcus furiosus* TOMM associated gene

PDB	Protein Data Bank
ppu	*Pseudomonas putida* TOMM associated gene
Q-TOF MS	Quadrapole Time-of- Flight Mass Spectrometer
sag	streptolysin S associated gene
Ser	Serine
SLS	streptolysin S
TCEP	tris-carboxyethylphosphine
TFA	Trifluoroacetic Acid
Thr	Threonine
TOMM	Thiazole/Oxazole Modified Microcin

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have many people to thank for helping me reach this stage in my career. First, I would like to thank my mentor, Jack Dixon for allowing me to join his lab. He has been a great example for me of how to do science the right way. I am particularly appreciative of Jack's support of me when I developed serious health problems in the second half of my graduate career. I would never have been able to complete my studies without his support and understanding and for that I am eternally grateful.

I would also like to thank my committee: Betsy Komives, Ed Dennis, Mike Burkart and Victor Nizet for giving their time to me and steering me in the right direction after listening to all of my off the wall ideas.

Everyone in the Dixon lab has played a role in my development throughout graduate school, though there are several people I would like to single out. Shaun Lee and Doug Mitchell were instrumental in my hands on education and development of my thesis project. We worked as a great team and they have had a hand in almost everything described in my dissertation. Seema Mattoo, Matt Gentry and Vinnie Tagliabracci taught me much of what I needed to know as a researcher and were always there for extra guidance. Finally, Joyce Limm and Phil Teh were great undergraduate researchers that provided important technical assistance.

The nature of my research project made it essential for me to reach out to other labs at UCSD. Pieter Dorrestein has been great to me, taking me under his wing for a number of years and allowing me to become an honorary member of his lab while I learned mass spectrometry. David Gonzalez and Michael Meehan in particular

provided a lot of help to me along the way. Majid Ghassemian in Betsy Komives' lab also provided me with great mass spectrometry guidance and access to his facility. I am also very appreciative of Susan Taylor, who opened up her lab to me during several projects. It was in her lab where I did surface plasmon resonance, florescence polarization and crystallography experiments. I am particularly appreciative of Cecilia Cheng for taking the time to train me on these procedures.

Funding for my research was provided by the Molecular Biophysics Training Program and the National Cancer Institute Training Grant, who I thank for allowing me to undertake this research and exposing me to a vast knowledge base.

Finally I would like to thank my family: John, Diane and Jessamyn for giving me support throughout all of my schooling. I also thank my girlfriend, Arthi, for keeping my going during the darkest times of graduate school and making sure I left lab from time to time. Finally I would like to thank my grandma, Lottie Sheehan, who unfortunately did not live to see me get my degree. She was like my second mother and prepared me well for school and for life.

Chapter II, in full, is a reprint that the dissertation author significantly contributed to as both a researcher and an author. The material appears in *Proceedings of the National Academies of Sciences of the USA (*S.W. Lee and D.A. Mitchell, **A.L. Markley**, M.E. Hensler, D. Gonzalez, A. Wohlrab, P.C. Dorrestein, V. Nizet, J.E. Dixon, (2008) Discovery of a Widely Distributed Toxin Biosynthetic Gene Cluster, **105**, 5879–5884.)

Chapter III, in full, is a reprint that the dissertation author significantly contributed to as both a researcher and an author. The material appears in the *Journal of Biological Chemistry (*D.A. Mitchell, S.W. Lee, M.A. Pence, **A.L**. **Markley,** J.D. Limm, V. Nizet, J.E. Dixon (2009), Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S, J. Biol. Chem. **284**(19),13004–13012.)

Chapter IV, in part, is a reprint that the dissertation author significantly contributed to as both a researcher and an author. The material appears in the *Journal of Biological Chemistry (*D.J. Gonzalez, S.W. Lee, M.E. Hensler, **A.L. Markley**, S. Dashesh, D.A. Mitchell, N. Bandiera, V. Nizet, J.E. Dixon, P.C. Dorrestein. (2010) Clostridiolysin S: A post-translationally modified biotoxin form clostridium botulinum, J. Biol. Chem. **285(**36), 28220–28228.)

Chapter VI, in full, is a reprint that the dissertation author was the principal researcher and author of. The material will appear in *Analytical Biochemisty* (**A.L. Markley**, E.R. Jensen, S.W. Lee. An E. coli-based bioengineering strategy to study Streptolysin S biosynthesis. Analytical Biochemistry, *in review.)*

VITA

2005 Bachelors of Science, Chemistry, Carnegie Mellon University

2009 Master of Science, Chemistry, University of California, San Diego

2011 Doctor of Philosophy, Chemistry, University of California, San Diego

PUBLICATIONS

A.L. Markley, E.R. Jensen, S.W. Lee. *An E. coli-based bioengineering strategy to study Streptolysin S biosynthesis.* Analytical Biochemistry, *submitted.*

R. Scholz, KJ Molohon, G.Nachtigall J. Vater, **A.L. Markley**, R.D. Süssmuth D.A. Mitchell, R. Borriss. *Plantazolicin, a novel microcin B17/streptolysin S-like natural product from Bacillus amyloliquefaciens FZB42.* J. Bacteriol **193**(1), 215–224 (2011).

D.J. Gonzalez, S.W. Lee, M.E. Hensler, **A.L. Markley**, S. Dashesh, D.A. Mitchell, N. Bandiera, V. Nizet, J.E. Dixon, P.C. Dorrestein. *Clostridiolysin S: A post-translationally modified biotoxin form clostridium botulinum,* J. Biol. Chem. **285(**36), 28220–28228 (2010).

D.A. Mitchell, S.W. Lee, M.A. Pence, **A.L**. **Markley,** J.D. Limm, V. Nizet, J.E. Dixon, *Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S*, J. Biol. Chem. **284**(19),13004–13012. (2009)

S.W. Lee and D.A. Mitchell, **A.L. Markley**, M.E. Hensler, D. Gonzalez, A. Wohlrab, P.C. Dorrestein, V. Nizet, J.E. Dixon, *Discovery of a Widely Distributed Toxin Biosynthetic Gene Cluster*, Proc. Natl. Acad. Sci. USA, **105**, 5879–5884 (2008).

D. Banerjee, **A.L. Markley**, T. Yano, A. Ghosh, P.B. Berget, E.G. Minkley Jr., S.K. Khetan, T.J. Collins. *Green Oxidation Catalysis for Rapid Deactivation of Bacterial Spores.* Angewandte Chemie International Edition.2006 Jun 12; 45(24):3974-7.

FIELDS OF STUDY

Major Field: Biochemistry

Studies in Biochemistry and Natural Products

Professor Jack Dixon

HONORS AND AWARDS

2011	Chancellor's Advisory Committee Award on Sustainability, Outstanding Graduate Student
2010-2011	National Cancer Institute Cancer Training Grant, trainee
2007-2009	Molecular Biophysics Training Grant, trainee
2005	Urey Fellowship

ABSTRACT OF THE DISSERTATION

Discovery and Investigation of the Thiazole/Oxazole Modified Microcin Natural Product Family

by

Andrew L. Markley

Doctor of Philosophy in Chemistry

University of California, San Diego, 2011

Professor Jack E. Dixon, Chair

Thiazole/Oxazole Modified Microcins (TOMMs) represent a large, widely distributed family of ribosomally encoded natural products. This family of enzymes selectively modify cysteine, serine and threonine amino acid residues to five-membered heterocyclic rings. A diverse set of bacterial organisms have utilized the TOMM biosynthetic pathway to create a multitude of natural products with a wide range of activities.

Chapter II describes how the TOMM family was created using bioinformatic searches to expand the known set of natural product biosynthetic gene clusters from four to over 250, each encoding a different natural product. Included in this new family is streptolysin S (SLS) a virulent toxin produced by human pathogenic bacteria *Streptococcus pyogenes*. We showed through *in vitro* analysis that SLS is a heterocycle containing TOMM natural product.

In Chapter III we used biochemical studies to identified the precursor peptide residues that were essential for SLS hemolytic activity. From this information we were able to generate artificial precursor peptides that had hemolytic activity upon TOMM posttranslational modification. It was shown that SLS biosynthesis relies on the same substrate binding motif as microcin B17

In Chapter IV a SLS-like TOMM cluster in *Clostridium botulinum* was characterized. Mass spectrometry was used to identify sites of heterocyclization in SLS as well as in its close relative clostridiolysin S.

Chapter V described the generation of an *E. coli* based bioengineering strategy for reconstituting SLS activity *ex vivo*, the first example of an active SLS analog being produced in another organism outside of the firmicutes phylum. This represents a method for generating high yields of TOMM natural products in a high throughput manner

In Chapter VI we validated the expression of the unannotated TOMM precursor peptides identified in Chapter II. RT-PCR, gene knockouts and *in vitro* activity assays were used to demonstrate expression.

Chapter I

Introduction

Natural products as drugs

Throughout human history, natural products have been at the origin of most therapeutic drugs (19). Advances in chemical purification techniques have allowed the traditional medicinal practices of tinctures and salves to give way to the purified pharmaceutical drugs that we have today (13). Starting with the isolation of morphine from poppy seeds in the early 1800's, scientists have relied predominantly on the "find and grind" method of natural product drug discovery. In this method, organic material with a desired biological activity is collected, mechanically homogenized, and fractionated by solubility, size and charge among other parameters. Biological activity assays guide the fractionation process to follow the natural product from crude extract to purified product. Typically plants have lead the way as the source for most of the early drugs owing to their widespread use in traditional medicine and the sheer mass of material that can be harvested. Recent development of undersea exploration techniques has added to the natural product discovery pipeline. With the advent of microbiology in the past century, microorganisms have replaced plants and animals as the most promising source of medically relevant natural products. This is due to the vast genetic diversity present among single celled organisms, however only a small minority of the microbial species have been successfully cultured on a large scale suitable for natural product discovery (11).

The need for new drugs, particularly those with antibiotic activity, has never been higher. Unprecedented use of antibiotics medically and in the modern meat production industry has caused antibiotic resistance to evolve in many strains of

pathogenic bacteria. One estimate puts the cost of treating patients affected by antibiotic resistant bacteria at $7 billion dollars annually (33). The discovery of new natural products with antimicrobial activity is very important as these molecules are commonly used as lead compounds from which medically useful analogues are developed. This study reports on the discovery of a new, widely distributed, class of natural product biosynthetic pathway and investigates whether “find and grind” natural product discovery can be complemented by a genome mining based drug discovery approach.

Secondary metabolism

The most commonly utilized biosynthetic pathway for producing molecules is ribosomal polypeptide biosynthesis, common to all living organisms (28). In this system, genetic information encoded by nucleotide chains directs the biosynthesis of an amino acid polymer in a highly controlled fashion. However, only 20 amino acids can be drawn upon to build the polypeptide chain, and the ribosome can only add those amino acids with one allowable linkage. In order to expand the chemical possibilities provided by ribosomal peptide synthesis, organisms have developed alternative biosynthetic pathways for creating useful secondary metabolites. The two most common classes of secondary metabolites are nonribosomal peptides and polyketides. Nonribosomal peptide synthesis (NRPS) is controlled by an operon of genes encoding a series of biosynthetic modules (43). The type and order of these modules dictate the molecular structure of the resultant natural product. Natural products formed in this way can be branched or cyclic and can contain a mixture of

natural and unnatural amino acids, vastly expanding the possible chemical space compared with ribosomally encoded peptides (16). Examples include bacitracin (12), gramicidin (42) and actinomycin (15). Polyketide synthesis (PKS) enables the formation of fatty acid chains using enzymes and proteins encoded by gene clusters similar to NRPS operons (e.g. erythromycin (40), tetracycline (24) and pikromycin (22)). Like in NRPS systems, control over the products of the biosynthesis pathway resides in the order of genes present in the operon. Because of their architectural similarity, NRPS and PKS modules are commonly combined together to form NRPS-PKS hybrid natural products (e.g. cyclosporine (1), polymyxin (17), and bleomycin (5)). NRPS and PKS biosynthetic systems allow for the production of diverse secondary metabolites at a cost of devoting megabase stretches of the genome to each molecule produced.

Ribosomally encoded bacteriocins

While the majority of biological cytotoxic compounds are produced through NRPS or PKS pathways, ribosomally encoded toxins, called bacteriocins, are widely produced and serve important functions for many organisms (10). Their small genetic footprint lowers the metabolic cost of production and increases the likelihood that the complete biosynthetic gene cluster will be transferred during horizontal gene transfer events (18).

Bacteriocins are produced with varying peptide length and posttranslational modification frequency, however they are structurally distinct from other peptides due to their higher backbone rigidity and increased resistance to proteolysis (23). The three

most common architectures that impart these traits on bacteriocins are disulfide bonds, thioether bonds, and aromatic heterocycles, which are commonly found in defensins, lantibiotics, and thiazole-oxazole modified microcins, respectively. Defensins constitute a large class of small, cysteine-rich peptides produced by eukaryotes (35). Defensins are typically 30-40 amino acids in length with three to four disulfide bond linkages. These molecules are believed to bind and insert themselves into cell membranes, causing ion leakage and cell death. Human neutrophils produce defensins locally in response to bacterial infection. (26). Lantibiotics are small peptides, typically 15-35 amino acids in length, produced by gram-positive bacteria containing alanine moieties linked by a thioether bond (23). The lantibiotic precursor peptide gene is bordered by a gene cluster that encodes the tailoring enzymes, exporters, and immunity proteins required for biosynthesis. Nisin, produced by *Lactococcus lactis*, is the most commonly studied lantibiotic and is widely used as a food preservative in the dairy industry (26). Finally, thiazole-oxazole modified microcins (TOMMs) are 30-70 amino acid molecules with cysteines, serines or threonines that have been heterocyclized into five-membered rings (25). These heterocycles add rigidity to the peptide backbone and makes an amide bond inaccessible to proteolysis. The prototypical and best studied TOMM natural product is microcin B17 (14).

Microcin B17

Microcin B17 (MccB17) is a DNA gyrase poison produced by *Escherichia coli* harboring the pMccB17 plasmid. The MccB17 operon is made up of 7 genes (*mcbABCDEF)* that are all necessary for MccB17 production (2, 29, 30, 39). The

precursor peptide, McbA, is a 69 amino acid peptide with a 26 residue amino terminus that is required for substrate binding and a 43 residue carboxyl terminus which contains 8 heterocycles in the active natural product (38). A trimeric complex of McbBCD binds McbA and selectively installs heterocycles before an unknown protease cleaves off the amino terminus (20). Each protein in the trimeric complex is responsible for a different step in the posttranslational processing. Heterocycles are preferentially installed at cysteine, serine and threonine residues that have a preceding glycine. The preceding glycine lacks steric hindrance, enabling higher rotational freedom around the amide bond into the correct orientation for heterocycle formation (32). In this reaction, the -SH or -OH group of cysteine and serine respectively attacks the carbonyl carbon from the preceding glycine residue forming an oxazoline or thiazoline moiety (20). Then a flavin containing dehydrogenase aromatizes these five-membered rings to oxazoles and thiazoles, respectively. The function of McbD is not yet fully understood, however it likely plays a role in passing the thiazolines and oxazolines from McbB to McbC as the complex moves down the prepropeptide from the N-terminus to the C-terminus (3). Once the heterocycles are installed and the signal peptide has been removed, McbE and McbF then export the active MccB17 peptide outside of the cell where it is taken up by nearby rival enterobacteria. MccB17 causes double stranded breaks in the chromosomal DNA of bacteria that do not harbor the *mcbG* immunity gene (39).

When we began our bioinformatic studies to assemble the TOMM family, only three other known microcin B17 like natural products were known. Intriguingly, even

this small sample size represented great diversity in host species and molecular structure. One TOMM family member, goadsporin is a linear peptide natural product with six heterocycle modifications produced by *Streptomyces sp.*, an actinobacterium (27). The remaining two natural products, patellamide (31) and trichamide (34) were found in cyanobacteria. These natural products contained multiple heterocycles and were also macrocyclized into a ring structure. This diversity led us to believe that many more family members were yet to be discovered.

Streptolysin S

One TOMM biosynthetic gene cluster was found in Group A *Streptococcus* (GAS), one of the most prevalent human pathogens responsible for over 500,000 deaths a year (7, 9). The hallmark of GAS is the ability to clear a lytic zone around colonies plated on blood agar media. The toxin attributed to this phenotype, streptolysin S (SLS) allows the bacteria to develop a niche in tissue and fend off autoimmune responses; however, to date its chemical structure remains elusive (21).

Streptolysin S (SLS) is a ribosomally encoded, cytolytic virulence factor secreted by the leading human pathogen, Group A *Streptococcus*. SLS has been shown to be necessary and sufficient for the light colored zone of erythrocyte clearing by GAS grown on blood agar plates. This activity, known as the beta-hemolytic phenotype, contributes to the ability of GAS to evade the human immune system (8, 21). GAS infection in humans most commonly manifests itself as pharyngitis or impetigo; however, in some cases, GAS can become invasive resulting in serious diseases such as rheumatic fever, necrotizing fasciitis, and toxic shock syndrome (4,

6). Currently GAS is treated with β-lactam antibiotics, which are still effective at controlling the infection in most cases. In the developing world, where access to antibiotics is limited, rates of invasive GAS infection are markedly higher, and GAS related death is common. Alternative GAS therapies are becoming increasingly important, even in western nations, due to a combination of β-lactam allergies and emerging resistance of GAS to common antibiotics (36). SLS is an attractive target for GAS-specific therapeutics because the factor itself contributes to virulence. Furthermore, because SLS expression has no effect on the viability of GAS cells, the development of resistant strains is unlikely (41). The activity of SLS has been well studied since its discovery in the 1930's; however, its structure and biosynthetic mechanism had not been determined (37). As reported in the second chapter of my thesis, my colleagues and I showed that SLS is a ribosomally encoded peptide with heterocycle posttranslational modifications (PTMs) at cysteine and serine residues. Structural determination of SLS would be an important step toward the rational design of a chemical vaccine against GAS.

Acknowledgements

Carolyn Worby and Vincent Tagliabracci critically edited this manuscript, for which I am grateful.

References

1. **Agathos, S., C. Madhosingh, J. Marshall, and J. Lee.** 1987. The Fungal Production of Cyclosporinea. Annals of the New York Academy of Sciences **506:**657-662.

2. **Baquero, F., and F. Moreno.** 1984. The microcins. FEMS Microbiology Letters**:**117-124.

3. **Belshaw, P. J., R. S. Roy, N. L. Kelleher, and C. T. Walsh.** 1998. Kinetics and regioselectivity of peptide-to-heterocycle conversions by microcin B17 synthetase. Chemistry & biology **5:**373-84.

4. **Bisno, A. L., M. A. Gerber, J. M. Gwaltney, Jr., E. L. Kaplan, and R. H. Schwartz.** 2002. Practice guidelines for the diagnosis and management of group A streptococcal pharyngitis. Infectious Diseases Society of America. Clin Infect Dis **35:**113-25.

5. **Calcutt, M. J., and F. J. Schmidt.** 1994. Bleomycin biosynthesis: structure of the resistance genes of the producer organism. Ann N Y Acad Sci **721:**133-7.

6. **Carapetis, J. R.** 2007. Rheumatic heart disease in developing countries. N Engl J Med **357:**439-41.

7. **Carapetis, J. R., A. C. Steer, E. K. Mulholland, and M. Weber.** 2005. The global burden of group A streptococcal diseases. Lancet Infect Dis **5:**685-94.

8. **Datta, V., S. Myskowski, L. Kwinn, D. Chiem, N. Varki, R. Kansal, M. Kotb, and V. Nizet.** 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Molecular Microbiology **56:**681-695.

9. **Datta, V., S. M. Myskowski, L. A. Kwinn, D. N. Chiem, N. Varki, R. G. Kansal, M. Kotb, and V. Nizet.** 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Mol Microbiol **56:**681-95.

10. **Farkas-Himsley, H.** 1980. Bacteriocins—are they broad-spectrum antibiotics? Journal of Antimicrobial Chemotherapy **6:**424.

11. **Foissner, W., H. Berger, K. Xu, and S. Zechmeister-Boltenstern.** 2005. A huge, undescribed soil ciliate (Protozoa: Ciliophora) diversity in natural forest stands of Central Europe. Biodiversity and Conservation **14:**617-701.

12. **Froyshov, O.** 1974. Bacitracin biosynthesis by three complementary fractions from Bacillus licheniformis. FEBS Letters.

13. **Hamilton, G. R., and T. F. Baskett.** 2000. In the arms of Morpheus the development of morphine for postoperative pain relief. Can J Anaesth **47:**367-74.

14. **Heddle, J. G., S. J. Blance, D. B. Zamble, F. Hollfelder, D. A. Miller, L. M. Wentzell, C. T. Walsh, and A. Maxwell.** 2001. The antibiotic microcin B17 is a DNA gyrase poison: characterisation of the mode of inhibition. Journal of Molecular Biology **307:**1223-34.

15. **Keller, U.** 1987. Actinomycin synthetases. Multifunctional enzymes responsible for the synthesis of the peptide chains of actinomycin. Journal of Biological Chemistry.

16. **Kleinkauf, H., and H. Dohren.** 1990. Nonribosomal biosynthesis of peptide antibiotics. Eur J Biochem **192:**1-15.

17. **Komura, S., and K. Kurahashi.** 1980. Biosynthesis of polymyxin E by a cell-free enzyme system. II. Synthesis of enzyme-bound octanoyldiaminobutyric acid. J Biochem **88:**285-8.

18. **Leikoski, N., D. P. Fewer, and K. Sivonen.** 2009. Widespread occurrence and lateral transfer of the cyanobactin biosynthesis gene cluster in cyanobacteria. Applied and Environmental Microbiology **75:**853-7.

19. **Li, J. W.-H., and J. C. Vederas.** 2009. Drug discovery and natural products: end of an era or an endless frontier? Science **325:**161-5.

20. **Li, Y. M., J. C. Milne, L. L. Madison, R. Kolter, and C. T. Walsh.** 1996. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. Science **274:**1188-93.

21. **Lin, A., J. Loughman, B. Zinselmeyer, M. Miller, and M. Caparon.** 2009. Streptolysin S Inhibits Neutrophil Recruitment During the Early Stages of Streptococcus pyogenes Infection. Infection and Immunity.

22. **Maezawa, I., T. Hori, A. Kinumaki, and M. Suzuki.** 1973. Biological conversion of narbonolide to picromycin. J Antibiot (Tokyo) **26:**771-5.

23. **McAuliffe, O., R. P. Ross, and C. Hill.** 2001. Lantibiotics: structure, biosynthesis and mode of action. FEMS Microbiology Reviews **25:**285-308.

24. **McCormick, J. R., and E. R. Jensen.** 1965. Biosynthesis of the Tetracyclines. 8. Characterization of 4-Hydroxy-6-Methylpretetramid. J Am Chem Soc **87:**1794-5.

25. **Melby, J. O., N. J. Nard, and D. A. Mitchell.** 2011. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. Current opinion in chemical biology.

26. **Nolan, E., and C. Walsh.** 2008. How Nature Morphs Peptide Scaffolds into Antibiotics. ChemBioChem.

27. **Onaka, H., M. Nakaho, K. Hayashi, Y. Igarashi, and T. Furumai.** 2005. Cloning and characterization of the goadsporin biosynthetic gene cluster from Streptomyces sp. TP-A0584. Microbiology (Reading, England) **151:**3923-33.

28. **Ramakrishnan, V., and S. W. White.** 1998. Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. Trends Biochem Sci **23:**208-212.

29. **Roy, R. S., A. M. Gehring, J. C. Milne, P. J. Belshaw, and C. T. Walsh.** 1999. Thiazole and oxazole peptides: biosynthesis and molecular machinery. Natural Product Reports **16:**249-63.

30. **San Millán, J. L., R. Kolter, and F. Moreno.** 1985. Plasmid genes required for microcin B17 production. Journal of Bacteriology **163:**1016-20.

31. **Schmidt, E., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. Haygood, and J. Ravel.** 2005. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proceedings of the National Academy of Sciences of the United States of America **102:**7315-20.

32. **Sinha Roy, R., P. J. Belshaw, and C. T. Walsh.** 1998. Mutational analysis of posttranslational heterocycle biosynthesis in the gyrase inhibitor microcin B17: distance dependence from propeptide and tolerance for substitution in a GSCG cyclizable sequence. Biochemistry **37:**4125-36.

33. **Sipahi, O. R.** 2008. Economics of antibiotic resistance. Expert review of Anti-infective Therapy **6:**523-539.

34. **Sudek, S., M. Haygood, D. Youssef, and E. Schmidt.** 2006. Structure of Trichamide, a Cyclic Peptide from the Bloom-Forming Cyanobacterium Trichodesmium erythraeum, Predicted from the Genome Sequence. Applied and Environmental Microbiology **72:**4382.

35. **Taylor, K., P. E. Barran, and J. R. Dorin.** 2008. Structure-activity relationships in beta-defensin peptides. Biopolymers **90:**1-7.

36. **Thigpen, M. C., C. L. Richards, Jr., R. Lynfield, N. L. Barrett, L. H. Harrison, K. E. Arnold, A. Reingold, N. M. Bennett, A. S. Craig, K. Gershman, P. R. Cieslak, P. Lewis, C. M. Greene, B. Beall, and C. A. Van**

Beneden. 2007. Invasive group A streptococcal infection in older adults in long-term care facilities and the community, United States, 1998-2003. Emerg Infect Dis **13:**1852-9.

37. **Todd, E.** 1938. The differentiation of two distinct serological varieties of streptolysin, streptolysin O and streptolysin S. J. Pathol. Bacteriol **47:**423–445.

38. **Videnov, G., D. Kaiser, M. B. and, and G. Jung.** 1996. Synthesis of the DNA Gyrase Inhibitor Microcin B17, a 43-Peptide Antibiotic with Eight Aromatic Heterocycles in its Backbone. Agnew. Chem. Int. Ed. Engl. **35:**1506-1508.

39. **Vizán, J. L., C. Hernández-Chico, I. del Castillo, and F. Moreno.** 1991. The peptide antibiotic microcin B17 induces double-strand cleavage of DNA mediated by E. coli DNA gyrase. EMBO J **10:**467-76.

40. **Weber, J. M., J. O. Leung, G. T. Maine, R. H. Potenz, T. J. Paulus, and J. P. DeWitt.** 1990. Organization of a cluster of erythromycin genes in Saccharopolyspora erythraea. J Bacteriol **172:**2372-83.

41. **Wessels, M. R.** 2005. Streptolysin S. The Journal of infectious diseases **192:**13-5.

42. **Yamada, M., H. Itoh, and K. Kurahashi.** 1967. Cell-Free Synthesis of Gramicidin S*. Biochemistry.

43. **Zuber, P.** 1991. Non-ribosomal peptide synthesis. Curr Opin Cell Biol **3:**1046-50.

Chapter II

Discovery of a Widely Distributed Toxin Biosynthetic Gene Cluster

Introduction

Streptolysin S (SLS), from the human pathogen *Streptococcus pyogenes*, is a ribosomally synthesized, secreted toxin responsible for the classical β-hemolytic phenotype of bacterial colonies grown on blood agar media (4, 17). *S. pyogenes* is associated with a wide spectrum of diseases ranging from simple pharyngitis to life-threatening necrotizing fasciitis. The expression of SLS promotes virulence in animal models of invasive infection (4, 9). The molecular structure of SLS has remained elusive for more than a century; however, SLS is known to be an oxygen-stable, non-immunogenic, cytolysin (16). The gene locus responsible for SLS biosynthesis (streptolysin S-associated genes, *sag*) was recently identified through transposon and targeted mutagenesis coupled with heterologous expression (2, 12, 17, 27). The *sag* locus consists of nine genes (A-I) and allelic exchange mutagenesis has shown *sagA-G* to be essential for the cytolytic phenotype of *S. pyogenes*. The first gene, *sagA*, encodes a 53 amino acid protoxin (Fig. 2.1A). *SagA* is followed by a set of three genes (*sagBCD*) exhibiting low sequence identity to three genes found in the *E.coli mcb* gene cluster that produces the DNA gyrase inhibitor, microcin B17 (MccB17, Fig. 2.5) (12, 20). *SagBCD* encode a cyclodehydratase (SagC; 13% identical to McbB), dehydrogenase (SagB; 22% identical to McbC), and a "docking" protein (SagD; 18% identical to McbD) (Figs. 2.6-2.9). The final genes in each cluster are dedicated ATP-binding cassette (ABC) transporters.

Analogous to the *sag* locus, the first open-reading frame (orf) of the MccB17-producing gene cluster, *mcbA*, encodes a 69-amino acid precursor protein (5, 28).

Walsh and colleagues have shown that McbA is extensively posttranslationally modified by the McbBCD synthetase complex (12). These modifications convert four serines and four cysteines into oxazole and thiazole heterocycles, respectively (Fig. 2.1B and Fig. 2.5). Unmodified McbA exhibits no effect on DNA gyrase (14).

Results and Discussion

Heterocycle formation by SagBCD

Given similarities in organization of the *mcb* and *sag* gene clusters and the observed homology of SagBCD to McbBCD, we hypothesized that the SagBCD synthetase complex would serve to posttranslationally modify a toxin precursor in the same manner as McbBCD. The *sagBCD* genes were cloned individually and purified as fusions to maltose-binding protein (MBP). Because numerous attempts to observe SagA by mass spectrometry failed, we employed McbA, which is amenable to mass spectrometry, to detect and confirm heterocycle formation by SagBCD (11). After removal of the MBP tag, recombinant McbA has a calculated molecular mass of 6293 Da due to addition of Gly-Ser-His to the N-terminus. For each heterocycle formed, a loss of 18 Da (oxazoline/thiazoline) or 20 Da (oxazole/thiazole) is expected from the parent peptide (Fig. 2.1B). Treatment of McbA with recombinant SagBCD resulted in the formation of four new masses differing from the precursor peptide by multiples of 20 Da (Fig. 2.1C-D). These masses are within error for linear mode matrix assisted laser desorption/ionization (MALDI) mass spectrometry and correspond to heterocycle formation at four residues of McbA. The fourth heterocycle was not

visible when the synthetase concentration was reduced (data not shown) and only unmodified McbA was observed when SagBCD was omitted from the reaction. These results provide the first evidence that SagBCD functions in a manner analogous to McbBCD.

SagBCD converts SagA into a cytolysin

Since recombinant SagBCD was active *in vitro*, we next sought to confirm that SagA could be transformed into an active cytolysin in a SagBCD-dependent manner. SagA was produced as an MBP-fusion protein and then subjected to modification by the addition of SagBCD in an *in vitro* synthetase reaction. Following this reaction, samples were tested for lytic activity against sheep erythrocytes. In this assay, SLS extracted from *S. pyogenes* cultures caused rapid hemolysis (data not shown). The addition of SagA alone failed to induce lysis, but robust hemolysis was observed after treatment with SagBCD (Fig. 2.2A). All three synthetase proteins were required for the lytic phenotype, indicating a cooperative role in toxin maturation.

To determine whether SagBCD-treated SagA exhibited a broader cytolytic phenotype characteristic of the native SLS toxin, reaction mixtures were applied to HEK293a cells. Cells treated with samples containing both SagA and all of the synthetase components progressively lost their focal contacts and eventually detached from the tissue culture wells (Fig. 2.2B). Actin staining revealed massive cytoskeletal collapse consistent with severe membrane damage. HEK cells incubated with solutions missing any component of the synthetase complex, or the protoxin precursor, were indistinguishable from untreated cells (Fig. 2.2B). Lactate dehydrogenase (LDH) release into the media provided quantitative confirmation that cytotoxicity required

both the SagA substrate and the SagBCD complex (Fig. 2.2C). SagA contains seven cysteines that serve as potential sites for thiazole formation by SagBCD. Mutation of all seven cysteines to serine (SagA-panC/S) completely abolished cytolytic activity, indicating that SagA-panC/S is either not recognized as a substrate by SagBCD or that the membrane damaging phenotype requires cysteine modification. Taken together, these experiments describe the first *in vitro* reconstitution of SLS activity. Furthermore, we demonstrate that SagBCD functions similarly to McbBCD by installing heterocycles on a peptidic toxin precursor.

The *sagBCD* gene clusters are widely distributed among prokaryotes

The similarities between the SLS and MccB17 biosynthetic operons were intriguing, given that *(i)* SLS is produced by a Gram-positive organism whereas MccB17 is from a Gram-negative organism and *(ii)* Microcins have heretofore been classified as unique to Gram-negative bacteria (22). This suggested that other prokaryotes could use related machinery to introduce Ser/Thr/Cys-derived heterocycles into a wide variety of ribosomally produced peptides. We therefore initiated a comprehensive survey of the public genomic databases to identify related biosynthetic gene clusters (1). This search revealed that similar gene clusters are widespread among prokaryotes, most of which are not associated with a function (Fig. 2.3A). The genes encoding the SagB-like dehydrogenase (Fig. 2.6), SagC-like cyclodehydratase (Fig. 2.7), and SagD-like protein (Fig. 2.8) were present as adjacent orfs in a diverse group of prokaryotic organisms spanning six phyla. The SagC and SagD orthologs were also sometimes found fused as a single orf (Fig. 2.9).

Of particular interest is the fact that gene clusters that highly resemble *sagA-I* are present in major mammalian pathogens such as *Clostridium botulinum*, *Listeria monocytogenes*, and *Staphylococcus aureus* RF122 (Fig. 2.3A). Similar gene clusters are also found in distantly related prokaryotes, such as cyanobacteria (21, 26) and archaea (*e.g. Pyrococcus furiosus* DSM 3638). Other family members are found throughout proteobacteria. For instance, a gene cluster identical in arrangement to *E. coli mcbA-G* is present in the plant symbiont, *Pseudomonas putida* KT2440 (22). *P. putida* colonizes the nutrient-rich rhizosphere of plants and induces plant growth while secreting antibiotics to limit the growth of competing soil bacteria (6, 15). Furthermore, two other plant symbionts, *Bacillus amyloliquefaciens* FZB42 (Bam) and *Bradyrhizobium japonicum* USDA110, encode analogous synthetase proteins. *Bacillus thuringiensis* harbors a similar operon; although not a plant symbiont, *B. thuringiensis* is of agricultural interest owing to its ability to secrete toxins that are effective against a variety of pests, such as moths, butterflies, flies, mosquitoes, and beetles (3). The conservation of this synthetase gene cluster across prokaryotes- in particular, amongst both pathogenic and symbiotic bacteria- suggests that these gene products play an important role in the establishment of a survival niche for these organisms.

Presence of protoxins across prokaryotic lineages

McbA and SagA encode protoxins of 69 and 53 residues, respectively (Fig. 2.3B). These are extremely short orfs, which many gene-identification algorithms do not always recognize. For this reason, manual orf searches were performed in the intergenic regions for each biosynthetic cluster identified. Short orfs encoding proteins

of 50-70 residues that are rich in Cys/Ser/Thr were present in all organisms containing *sagBCD*-like genes as adjacent orfs (Fig. 2.3B). Although the functions of these potential protoxins are not known, based on similarity to McbA and SagA, we speculate that some will function as membrane damaging agents (*e.g. S. aureus* and *L. monocytogenes*) or DNA gyrase poisons (*P. putida*). Secondary metabolites produced by biosynthetic clusters such as these have historically been an abundant source of pharmaceuticals. The bacterial products of these gene clusters are potential candidates for novel antibiotic design as well as promising targets for vaccine development.

Most of the gene clusters identified contain orfs encoding additional modifying enzymes, suggesting that in addition to heterocycle formation, these protoxins will undergo further modification. For example, clusters containing acetyltransferases and lantibiotic dehydratases, such as the goadsporin-producing organism *Streptomyces* sp. TP-A0584 (18), will likely produce toxins with acetylated and dehydrated amino acids (Fig. 2.5). Furthermore, methylated toxins are expected from gene clusters with methyltransferases. There are also gene clusters containing radical S-adenosylmethionine (SAM)-like proteins, which catalyze the formation of many types of chemical linkages in complex natural products (25). Inclusion of these ancillary enzymes is expected to dramatically increase the chemical diversity of the toxins, perhaps for niche purposes. The cyanobacteria *Prochloron didemni* (21) and *Trichodesmium erythraeum* IMS101 (26) have been previously identified as sources of cyclic peptides with Ser/Thr/Cys-derived heterocycles (Fig. 2.5). The *sag*-like biosynthetic clusters of *T. erythraeum* and *P. didemni* harbor an additional gene encoding a putative serine protease (Fig. 2.3A). Although the exact function of this

protein is unclear, it is proposed to catalyze a head-to-tail macrocyclization reaction by using the N-terminus of a peptide, instead of water, to hydrolyze the acyl-enzyme intermediate. After this reaction, the volume occupied by the toxin is much smaller and may eliminate the need for an elaborate export system. The lack of ABC transporters in these gene clusters supports this notion (Fig. 2.3A). Based upon a similar arrangement of genes, we hypothesize that *Rhodobacter sphaeroides*, *Lyngbya* sp. PCC 8106, and *Microcystis aeruginosa* will also form macrocyclized peptides.

The heterocycle synthetase complex displays functional promiscuity

A gene cluster that is highly similar to the *sag* cluster was found in the pathogenic bacterium, *Clostridium botulinum* (Fig. 2.3A). The organization and amino acid sequence of proteins present in the *C. botulinum* cluster, *closA-I*, displays the highest similarity to *sagA-I*. The protoxin, which we have designated ClosA, shares many characteristics with SagA (39% identical) including the presence of numerous adjacent cysteines, other conserved heterocyclizable residues, and a putative protease site for removal of a leader peptide (Fig. 2.3B). To explore the possibility that *closA* encodes an SLS-like cytolysin, recombinant MBP-ClosA was prepared and subjected to *in vitro* synthetase assays with SagBCD. Hemolysis was only observed in samples containing both ClosA and the SagBCD synthetase complex (Fig. 2.4A). Microscopy revealed that cytotoxicity in HEK cells treated with ClosA + SagBCD was identical to cells treated with SagA + SagBCD (Fig. 2.2B and 4A). These data demonstrate SagBCD accepts substrates beyond its cognate peptide, SagA.

Given the ability of SagBCD to accept McbA and ClosA as alternative substrates, we assessed whether other synthetases were permissive in their substrate

selectivity. To this end, the heterocycle-forming synthetases from *E. coli* (McbBCD) and the thermophilic archaeon, *P. furiosus* (PagBCD, identity to corresponding Sag: 17-24%), were prepared as MBP fusions. SagA and ClosA were then tested for erythrocyte lytic activity after reaction with McbBCD and PagBCD. The PagBCD complex efficiently converted both SagA and ClosA into a hemolysin (Fig. 2.4B). In contrast, the McbBCD complex converted only ClosA into a potent lytic species. It is intriguing that McbBCD accepted and converted an alternate substrate into a cytolytic factor given that its biological function is to produce a DNA gyrase inhibitor (8). Therefore, we conclude that the mechanism of toxin action is governed by the protoxin amino acid sequence and that the synthetase proteins are functionally redundant. The extent of substrate tolerance and the kinetics and selectivity of ring formation are currently under investigation.

In sum, our bioinformatic survey has uncovered the presence of *sag*-like gene clusters in a myriad of prokaryotes, leading us to define a new class of bacteriocin. This class is characterized by a biosynthetic gene cluster that encodes a small protoxin and three adjacent synthetase proteins that endow an organism with the ability to cyclize Ser/Thr/Cys residues from a ribosomally synthesized protoxin scaffold. The finding that the heterocycle synthetase genes are exclusively found as adjacent orfs will facilitate the identification of orthologous biosynthetic clusters in emerging genomes. Using the protoxin and synthetase proteins from the human pathogen *S. pyogenes*, we demonstrate the first *in vitro* reconstitution of streptolysin S (SLS) activity. Furthermore, we show that the synthetase enzymes are functionally redundant and catalyze heterocycle formation on alternative substrates, despite their significantly

distinct evolutionary lineages. Our data suggest that all of the newly identified gene clusters are responsible for the production of bacteriocins containing at least one Ser/Thr/Cys-derived heterocycle. Many of the gene clusters encode ancillary modifying enzymes that will install further modifications onto the bacteriocin, thus increasing the chemical diversity of this family. Further insights into this bacteriocin family will lead to the identification of novel targets for antibiotic and vaccine development.

Materials and Methods

Bioinformatics.

The amino acid sequences of SagBCD from *S. pyogenes* (17) and McbBCD from *E. coli* (19) were initially used as a BLAST query against the non-redundant protein sequence database of the National Center for Biotechnology Information (1). Highly similar proteins were identifiable in the firmicutes phylum, specifically in *C. botulinum*, *S. aureus* RF122, and *L. monocytogenes*. The surrounding orfs for each identified gene were analyzed for the presence of all three synthetase enzymes using the Genomic Context feature of Entrez Gene and the SEED tool from The Fellowship for the Interpretation of Genomes. Gene clusters that contained all three synthetase enzymes were also found to contain other bacteriocin-type operon components, such as a probable protease, immunity protein, and ABC transporters. In cases where the automatic annotation algorithms do not identify the smaller toxin structural genes (*e.g. Listeria*, ListA), manual inspection of all predicted orfs in the local intergene region

was performed and these were subsequently scored using ClustalW alignment to known toxin structural genes from *S. pyogenes* and *E. coli*. *S. iniae* has been shown to contain a nearly identical gene cluster to *S. pyogenes* and was therefore not used for bioinformatics searching (7, 13). Furthermore, another *Streptococcus* species, *S. dysgalactiae* sp. *equisimilis*, contains a SagA and SagB homolog (9). Due to an incomplete genome sequence, we cannot confirm the presence of SagC-I at this time.

Wider subsequent searches of the non-redundant protein sequence database (nr, NCBI) were initiated using all proteins confirmed to have a *sag*-like gene locus (with *bcd* genes located directly adjacent to or within a few orfs of each another). A lower limit of 30% amino acid similarity was chosen as a threshold to assign a candidate protein as being homologous (BLOSUM30 matrix). This second search identified gene clusters from three other organisms that produce a heterocycle-containing bacteriocin with a known structure (Fig. 2.5), *Streptomyces* sp. TP-A0584 (goadsporin) (10), *P. didemni* (patellamide A/C) (21), and *T. erythraeum* IMS101 (trichamide) (26) and from additional organisms.

Mass Spectrometry

Sample Preparation. Recombinant MBP-McbA (110 μL at 40 μM) was proteolytically digested with 50 NIH units of thrombin (Sigma-Aldrich) for 4 h at room temperature in thrombin cleavage buffer: 50 mM tris (pH 7.8) and 10 mM $CaCl_2$. During the course of the reaction, cut McbA precipitates (pI ~ 7) as a white solid. Precipitated McbA was subsequently harvested by centrifugation and washed with deionized water (2-3 x 100 μL) to remove the majority of buffer, salts, detergent, cut MBP, and uncut MBP-McbA. This precipitate was then resolubilized in 60%

MeCN / 2% formic acid (40 μL). Samples of MBP-McbA that were treated with BCD synthetase complex (Sag and Pag) were first reacted, as described in "*In vitro* synthetase assay", before thrombinolysis.

The MALDI target (stainless steel, Applied Biosystems) was prepared by allowing a saturated solution of sinapic acid (Sigma-Aldrich) in 50% MeCN / 0.1% TFA to fully dry on the target (1.5 μL per sample, in duplicate). The precipitated McbA sample was then diluted (1:1) in the saturated matrix solution before spotting on top of the dried spot (1 μL in duplicate).

Instrument Settings. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectra were acquired on a Voyager DE-STR instrument (Applied Biosystems) in linear positive mode. Specific settings were typically as follows: 2000-2100 laser power, 93% grid, 0.12% guide wire, 400 nsec delay, mass window 3000-18000 Da.

***In vitro* Synthetase Assay.**

Synthetase reactions using MBP-tagged SagA and SagBCD, and other fusion proteins were performed in a manner described earlier (23, 24). Reaction mixtures consisted of MBP-SagA (10 μM) and MBP-SagBCD (2 μM each) in a total volume of 100 μL synthetase buffer (50 mM Tris HCl, pH 7.5, 125 mM NaCl, 20 mM $MgCl_2$, 2 mM ATP, 10 mM DTT). Reactions were allowed to proceed at 37 °C for 16 h unless otherwise stated. These reactions were used for hemolytic assays and LDH release assays to quantify membrane damage. Omission of either ATP or DTT resulted in loss of hemolytic activity (data not shown).

Hemolytic Assay.

Fresh defibrinated whole sheep blood was obtained from Hemostat laboratories (San Diego, CA). Whole blood was washed 3x in PBS and diluted to a final concentration of 1:25 v/v in PBS. Prepared whole blood (100 μL) was then placed into individual wells of a flat bottom 96 well microtiter plate (Costar, Corning, Lowell, MA). Reactions from *in vitro* synthetase assays were then added directly to the wells and the mixture was reacted for 16 h at 37 °C unless otherwise noted. Following incubation, plates were centrifuged at 500 x g for 10 min and an aliquot of supernatant (100 μL) was placed in a separate microtiter plate for measuring hemoglobin absorbance at 450 nm on a Victor3 microplate reader (Perkin Elmer, Waltham, MA).

Phylogenetic Tree Assembly

Phylogenetic trees were constructed using the ClustalW server located at http://align.genome.jp. Slow/accurate pairwise alignments were used. All other parameters were the default options. The trees were visualized by selecting ‘‘unrooted N-J tree.’’

Cloning

The genes encoding SagA (locus tag, Spy_0738), SagB (Spy_0739), SagC (Spy_0740), and SagD (Spy_0741) were previously amplified from Streptococcus pyogenes genomic DNA and cloned into the group A streptococcal expression vector, pDCerm (1). For recombinant expression of these proteins in *E. coli*, the SagA-D inserts were amplified by PCR using the following primers (IDT), which contain 5’ BamHI and 3’ NotI restriction endonuclease sites (left to right, 5’ to 3’): SagA_fwd,

GAG GGATCC ATG TTA AAA TTT ACT TCA AAT ATT TTA G; SagA_rev, ACA GCG GCC GCT TAT TTA CCT GGC GTA TAA CTT CC; SagB_fwd, CGG ATC CAT GTC ATT TTT TAC AAA GG; SagB_rev, AAA AGC GGC CGC CTA TTG AGA CTC CTT AGT TCC; SagC_fwd, AAA AGG ATC CAT GAA ATA TCA ACT TAA TAG TAA TG; SagC_rev, AAA AGC GGC CGC TCG ATT ACT CGT CAA GGA G; SagD_fwd, AAA AGG ATC CAT GTT ATA CTA TTA TCC TTC TTT TAA TC; SagD_rev, AAA AGC GGC CGC GAA TTC TAA GGC ATT GG.

The amplified inserts were purified by using a 1% agarose gel, excised, and gel-extracted (Invitrogen). Purified inserts and a modified pET28 vector (EMD Chemicals) containing an N-terminal maltose-binding protein (MBP) fusion with thrombin and TEV protease sites were double digested (BamHI/NotI; NEB) as per the manufacturer's instructions. Digested vector was treated with calf intestine alkaline phosphatase (CIP; Promega) for 30 min at 37°C before purifying digested inserts and vector on a 1% agarose gel. Digested plasmid and insert were gel-extracted and ligated by using T4 DNA ligase (NEB) before transformation into chemically competent DH5α. Plasmid DNA was isolated by miniprep kit (Invitrogen) and sequenced at Eton Bioscience using the T7 forward, T7 reverse, and MBP forward (5'-ATG AAG CCC TGA AAG ACG-3') primers to verify the presence of the gene and to ensure no mutations were introduced during the cloning. The gene encoding ClosA, locus tag CLI_0566 (2), from *Clostridium botulinum* F str. Langeland was chemically synthesized (Genscript). The construct was provided as a GST fusion in the pGS21a vector. ClosA was then subcloned into the pET28-MBP vector and sequenced, as described above, using the following primers: ClosA_fwd, AAG GAT CCA TGC

TGA AAT TTA ACG AAC ATG TGC TGA CC; ClosA_rev, AAG CGG CCG CTT AGT TGC CGC CCT GAC CCG C. Genes encoding the synthetases from Pyrococcus furiosus (PagB locus tag is PF0001, PagC is PF0003, and PagD is PF0002) amplified from genomic DNA (American Type Culture Collection) and inserted into an MBP-modified pET15 vector (3, 4).

This vector is similar to pET28-MBP but lacks the hexahistidine tags and TEV protease site. The primers used were as follows: PagB_fwd, AAG GAT CCA TGG TTC ATT ATT TTG TGA AGG TAA C; PagB_rev, AAG CGG CCG CTC ACA CTA ATT TAC CCA TTC CTC C; PagC_fwd, AAA AGC TTG GAT CCA TGG CAA ATT CAA AAA GGA AGA ATG CAC G; PagC_rev, AAG CGG CCG CTC ATT TAA ACC ACC TCC CCT CAC AGA CAT TAC; PagDfwd, AAG GAT CCA TGA GGA TAA ATT GTC ATG TTA GCA ATA TTG AAG; PagD_rev, AAG CGG CCG CTT ATG GAT ATG GAT GAG GAA TAC ACC TTA GC.

Expression and Purification of Recombinant Proteins

Constructs containing confirmed sequence were transformed into BL21(DE3)RIL cells under selection with 50 μg/ml kanamycin (or 100 μg/ml ampicillin) and 35 μg/ml chloramphenicol. A starter culture (10 ml) was grown overnight and used to inoculate 2 × 1 liter of LB medium for each protein. These cells were grown at 37°C to an OD600 of 0.7. Cultures expressing the protoxins were induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3–4 h at 37°C. Cultures expressing the BCD synthetase enzymes were lowered to 22°C before

inducing expression with 0.4 mM IPTG for 16 h. The cells were harvested by centrifugation, and the pellets were stored at -80°C until purification.

Cell pellets were resuspended in MBP lysis buffer [50 mM Tris, pH 7.5/0.5 M NaCl/2.5% (vol/vol) glycerol/0.1% (vol/vol) Triton X-100] along with 5 mg/ml lysozyme and EDTA-free Complete Protease Inhibitor Mixture Tablets (Roche) for 45 min at 4°C. The resuspended cells were lysed by three rounds of sonication at 4°C. Centrifugation for 40 min at 40,000 × g yielded supernatant that was immediately gravity-loaded onto a column packed with 8 ml of amylose resin (NEB) and prequilibrated with MBP lysis buffer. The loaded column was washed with 10–15 column volumes of ice-cold wash buffer A: 50 mM Tris (pH 7.5), 0.4 M NaCl, 0.5 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP), 2.5% glycerol, and 0.1% Triton X-100. A second wash step included 1 column volume of ice-cold wash buffer B (lacks Triton X-100) before elution with 5 column volumes of 50 mM Tris (pH 7.5), 0.15 M NaCl, 0.5 mM TCEP, 10 mM maltose, and 2.5% glycerol. Elution fractions containing a band of the correct molecular weight, as determined by Coomassie-stained SDS/PAGE, were pooled, buffer-exchanged, and concentrated by using 50-kDa molecular mass cutoff concentrators (Millipore). The final storage buffer was 50 mM 4-(2-hydroxyethyl)-1-piperazineethane sulfonic acid (Hepes, pH 7.5), 50 mM NaCl, 0.25 mM TCEP, and 2.5% glycerol. All proteins containing the N-terminal MBP fusion tag overexpressed to sufficient levels to yield ~95% pure protein after single column purification. The fractions containing homogeneous proteins were then aliquoted and stored at -80°C until needed. Concentrations were assessed by using the Bradford method (BSA standard) and denatured absorbance at 280 nm (6M guanidine

HCl). Typical yields: protoxin, 50 mg/liter; dehydrogenase 30 mg/liter; cyclodehydratase, 10 mg/liter; docking protein, 15 mg/liter.

Cell Culture and Treatments

Human embryonic kidney cells (HEK293) were maintained in a 5% CO2, water-saturated atmosphere at 37°C in DMEM with glutamine (Gibco) supplemented with 5% FBS, penicillin (100 units/ml), and streptomycin (100 μg/ml). Cytotoxicity Assay. Lactate dehydrogenase (LDH) levels were determined by using a colorimetric assay (Roche). HEK293a cells were grown to 70–80% conf luence in 24-well microtiter plates (Falcon, BD Biosciences) as described above. Cells were washed three times in assay medium consisting of phenol-red free DMEM (Mediatech), and treated with in vitro synthetase reactions for 10 h unless otherwise stated. Cells were treated with Triton X-100 (Sigma–Aldrich) (1%, vol/vol, in medium) for 15 minutes as a positive control. Culture supernatants were prepared as per the manufacturer's instructions, and the absorbance at 490 nm was recorded (n = 3). Background was subtracted from control reactions; all readings were averaged and plotted as the mean ± standard deviation.

Cell Microscopy

HEK293a cells were grown to 50% confluence on BIOCOAT fibronectin-coated coverslips (BD Biosciences) and treated with synthetase reactions for 4–6 h. The coverslips were washed three times with PBS, fixed in paraformaldehyde (4% in PBS, wt/vol) for 20 min at 22°C, and permeabilized with isotonic PBS containing normal goat serum (3% vol/vol; Gibco, Invitrogen) and Triton X-100 (0.2% vol/vol)

for 1 h. The samples were subsequently treated with rhodamine phalloidin (1:1,000 in PBS; Molecular Probes) for 10 min to visualize actin filaments and anti-vinculin antibody (1:3,000) for cytoplasmic staining. Samples were washed extensively with PBS before mounting in Gelvatol. Images were acquired by using a ×40 objective on a Zeiss Axiovert System. The images were subsequently exported to Adobe Photoshop 7.0 for image preparation.

Acknowledgements

We thank Joyce Limm for technical assistance and Christopher Walsh (Harvard Medical School) for the *mcbA-D* clones. We also wish to thank Elizabeth Komives and Michael VanNieuwenhze (Indiana University) for their helpful suggestions. Seema Mattoo critically reviewed this manuscript, for which we are grateful. This research was supported by grants from the NIH, the Walther Cancer Institute, the Ellison Foundation, and the Howard Hughes Medical Institute (JED). DAM is supported by a Hartwell Foundation postdoctoral fellowship. SWL is supported by the UCSD Heme and Blood Proteins postdoctoral training grant. ALM is supported by the UCSD Molecular Biophysics Training Program.

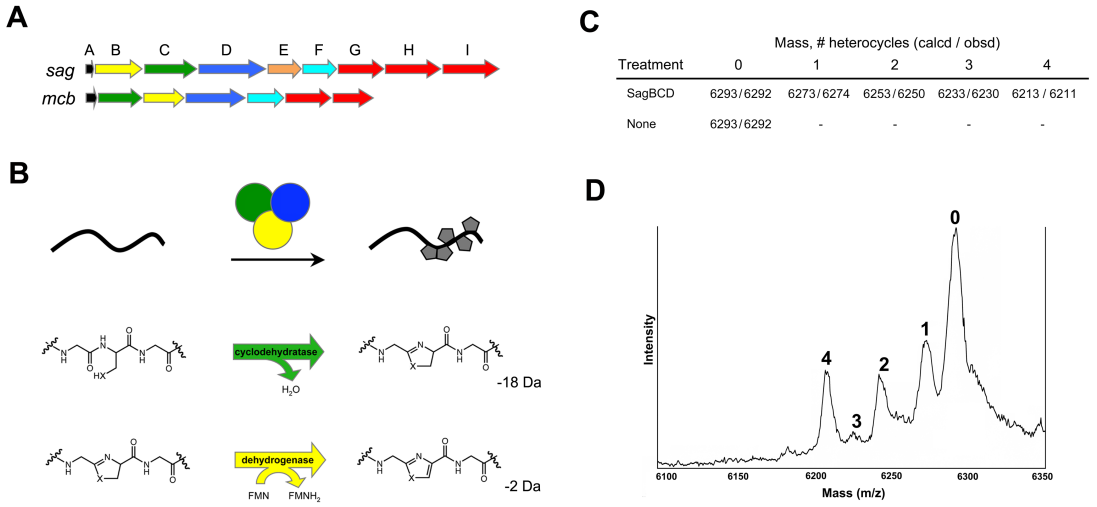


Figure 2.1 Conservation of toxin biosynthesis operons in *S. pyogenes* and *E. coli*.
(A) Genetic organization of the streptolysin S-associated gene cluster (*sagA-I*) from *S. pyogenes* and the *E. coli* microcin B17 gene cluster (*mcbA-G*). (B) Through the action of a trimeric synthetase complex, oxazole and thiazole heterocycles are incorporated into a peptidic protoxin scaffold (black) and are active *in vitro*. Chemical transformations carried out by SagC/McbC (green, cyclodehydratase) and SagB/McbC (yellow, dehydrogenase) orthologs are shown. Molecular weight change for each reaction is shown in Daltons. SagD/McbD (blue) serves as an enzymatic scaffold and facilitates substrate binding. (C) Heterocycle formation on *E. coli* McbA by *S. pyogenes* SagBCD. Calculated (calcd) and observed (obsd) molecular masses for MBP-McbA after reaction with SagBCD and thrombinolysis. Formation of a single thiazole/oxazole leads to the loss of 20 Da from the mass of the peptide. Linear mode MALDI-TOF mass spectrum of McbA treated with SagBCD. The number of heterocycles on the protoxin peptide is indicated above their respective mass (singly charged species).

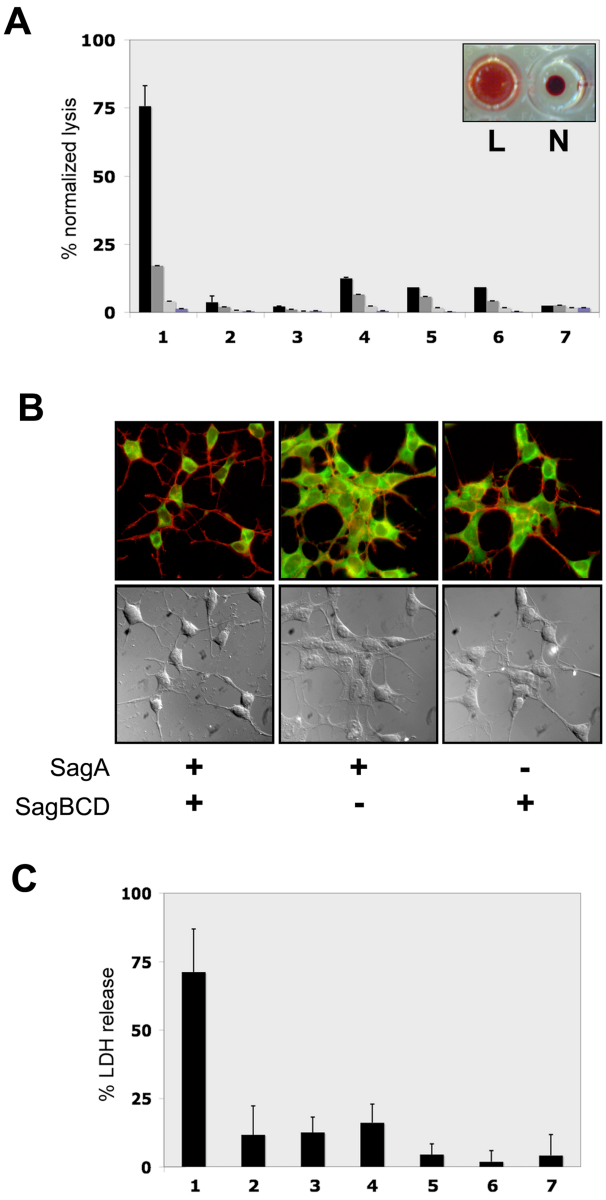


Figure 2.2 Cytolytic activity of *in vitro* synthetase reactions.
(A) Hemolytic assays of SagA + SagBCD synthetase reactions in microtiter wells containing defibrinated sheep blood. Bars indicate lysis normalized to a positive control (triton X-100). Levels indicated are 1:1, 3:4, 1:2 and 1:4 ratios of synthetase reaction to blood (left to right) of 16 h reactions (n = 3). Samples: 1. SagA + SagBCD, 2. SagA alone, 3. SagBCD alone, 4. SagA + SagBC, 5. SagA + SagCD, 6. SagA + SagBD, 7. vehicle. Inset demonstrates typical appearance of lytic (L) and nonlytic (N) reactions. (B) Fluorescence microscopy and DIC images of HEK293a cells treated as indicated. Actin filaments (red) and cytoplasm (green) are merged in the upper panels, and DIC images are in the lower panels. **C**, LDH release assay of HEK293a cells treated with SagA + SagBCD. Lysis is measured as A_{490} normalized to positive control (triton X-100). Samples: 1. SagA + SagBCD, 2. SagA alone, 3. SagA + SagBC, 4. SagA + SagCD, 5. SagA + SagBD, 6. SagBCD alone, 7. SagA-panC/S + SagBCD.

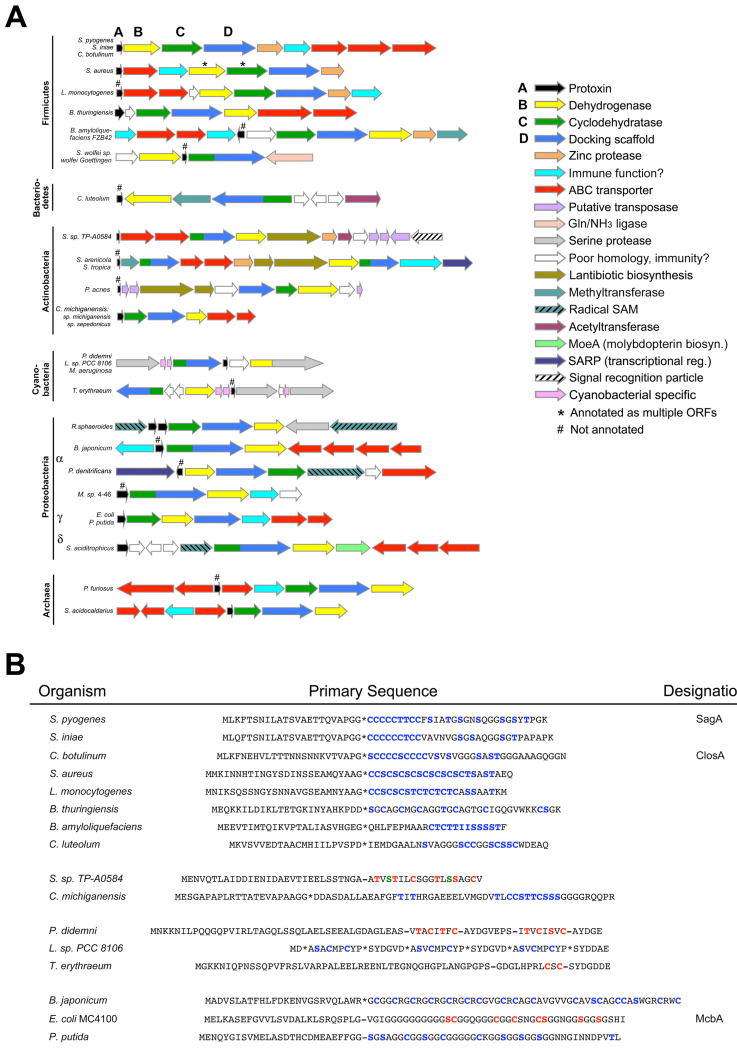


Figure 2.3 The biosynthetic operon for producing thiazole/oxazole-containing toxins is widely distributed.

(A) Gene clusters from organisms containing SLS- and MccB17-like bacteriocins. Members are sorted by prokaryotic phylum. Relative gene length and directionality are shown (scale for actinobacteria and cyanobacteria is reduced by 50%). Each gene cluster contains a protoxin ("A", black), dehydrogenase ('B", yellow), cyclodehydratase ("C", green), and a docking scaffold ("D", blue). These genes produce both single domain and fusion proteins (*e.g. S. wolfei* and *C. luteolum* CD). Numerous ancillary enzymes are included and increase the chemical diversity of the toxin family. (B) Select protoxin amino acid sequences. Predicted leader peptide cleavage sites are denoted with an asterisk. Hyphens indicate a known cleavage site. Potential sites of heterocyclization are indicated in blue and known sites of heterocycle formation are indicated in red. Green text signifies conversion to dehydroalanine. Toxins similar to SagA (top) are predicted cytolysins.

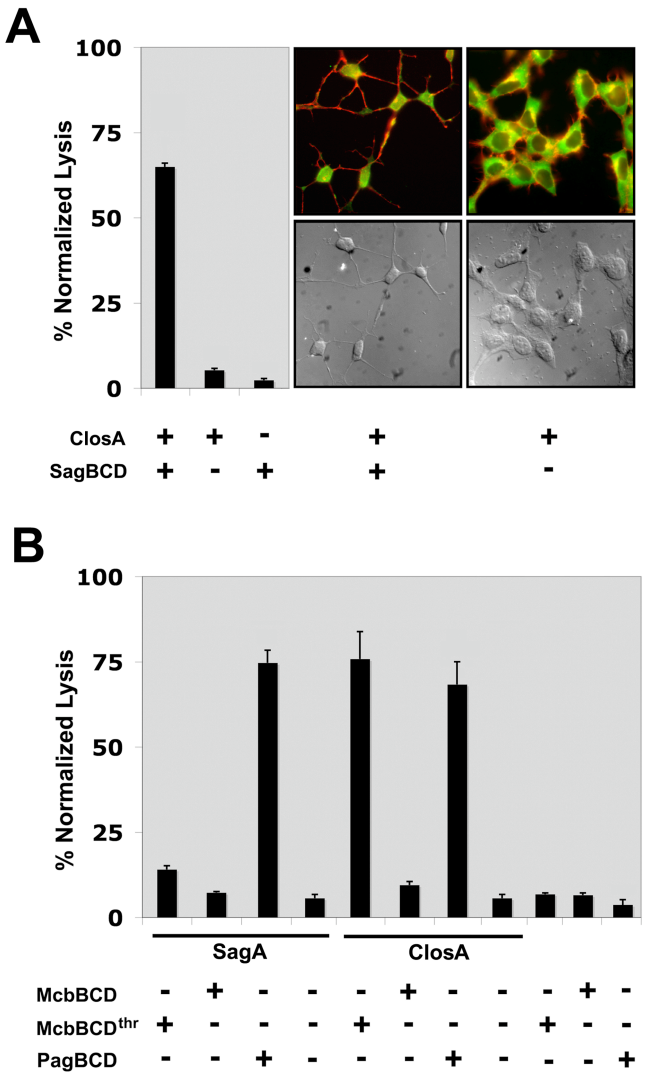


Figure 2.4 Disparate heterocycle synthetases accept substrates from related and distant prokaryotes.
(A) The *S. pyogenes* synthetase SagBCD accepts the *C. botulinum* protoxin and produces a cytolytic toxin. Left, Hemolytic assay; Right, Fluorescence microscopy and DIC images of HEK293a cells treated as indicated. (B) Hemolytic assay of SagA and ClostA treated with the synthetase complex from *E. coli* (McbBCD) and the hyperthermophile, *P. furiosus* (PagBCD). Recombinant MBP-McbD must be first treated with thrombin (thr) before synthetase activity is observed.

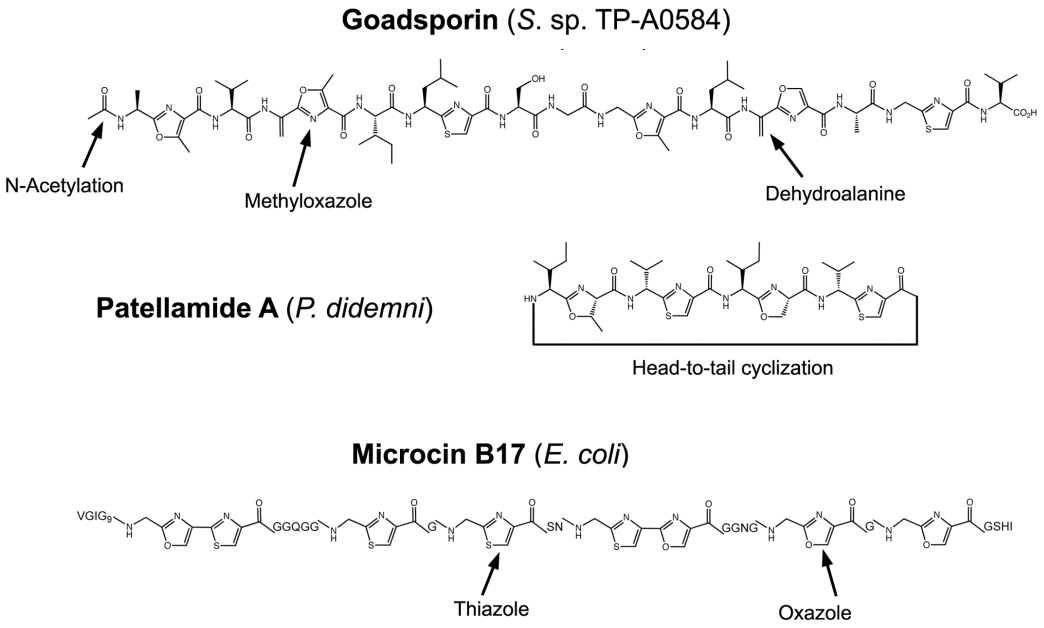


Figure 2.5 Molecular structures of compounds synthesized by a sag-like gene cluster. Posttranslational modifications are indicated. Dehydroalanines and thiazoles are derived from cysteine, oxazole from serine, and methyloxazole from threonine. Patellamide A contains two reduced heterocycles: a methyloxazoline and oxazoline. These are derived from threonine and serine, respectively, and are not labeled for clarity.

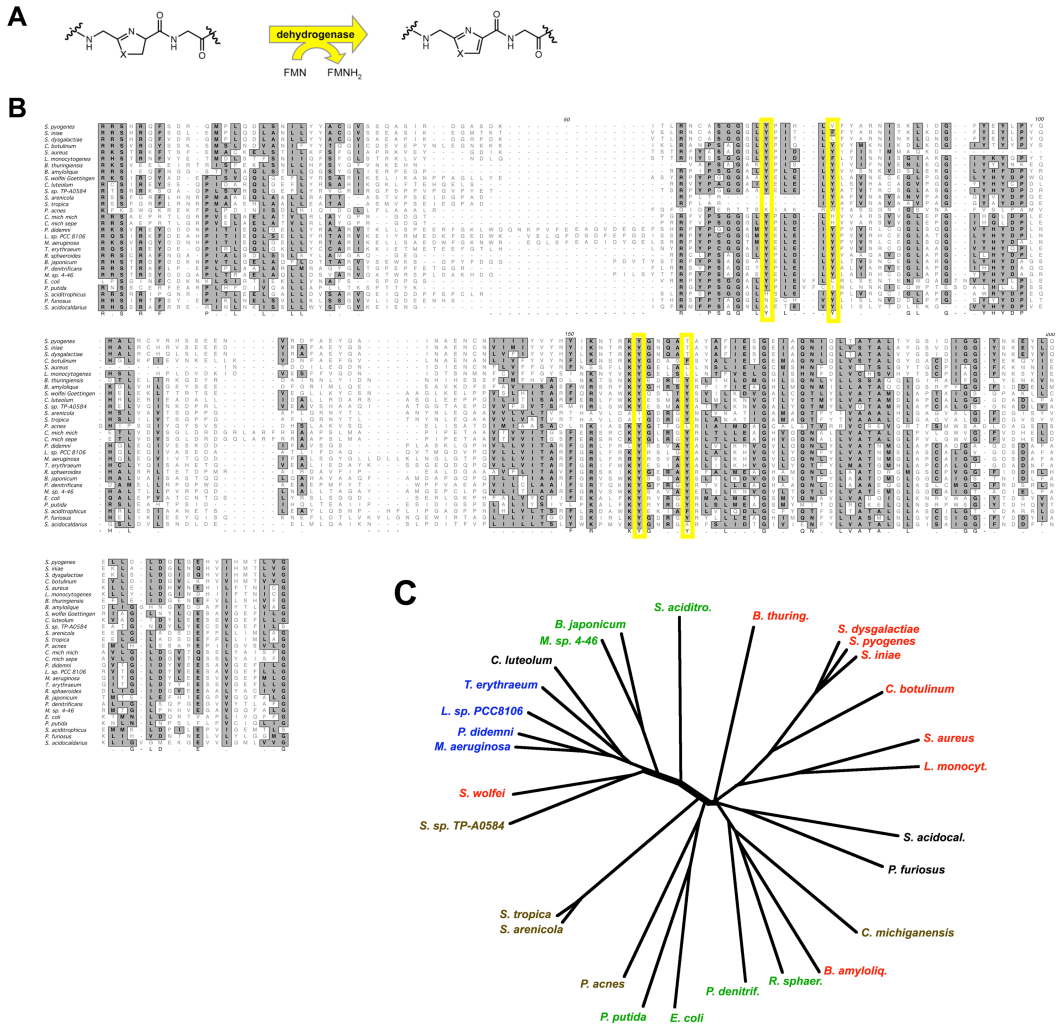


Figure 2.6. Multiple sequence alignment and phylogenetic analysis of the dehydrogenase. (A) The chemical transformation carried out by SagB orthologs. Oxazoline (X = O) and thiazoline (X = S) heterocycles are oxidized by two electrons to oxazole and thiazole, respectively. During this oxidative aromatization reaction, oxidized flavin mononucleotide (FMN) is reduced by two electrons to $FMNH_2$. Molecular oxygen restores the flavin to the fully oxidized state *in vitro*; *in vivo*, the oxidant remains unknown. (B) ClustalW alignment of ~30 SagB orthologs (includes *S. dysgalactiae*). Because of divergence at the N and C termini of the proteins, only 200 of the more highly conserved residues are shown in this alignment for each protein. This region contains several positions that contain nearly invariant tyrosines (yellow boxes). FMN binding sites are diverse, but usually the cofactor is π-stacked between Y or W. There are also invariant arginine residues, which may play a role in coordinating the phosphate group of FMN. Note that the *S. aureus* protein is annotated as multiple ORFs (SAB1378c, SAB1377c, and SAB1376c). **C,** Phylogenetic analysis of ClustalW aligned SagB homologs reveals evolutionary clustering. The labels are sorted by phylum and are colored as follows: red, firmicutes; green, proteobacteria; blue, cyanobacteria; brown, actinobacteria; black, other.

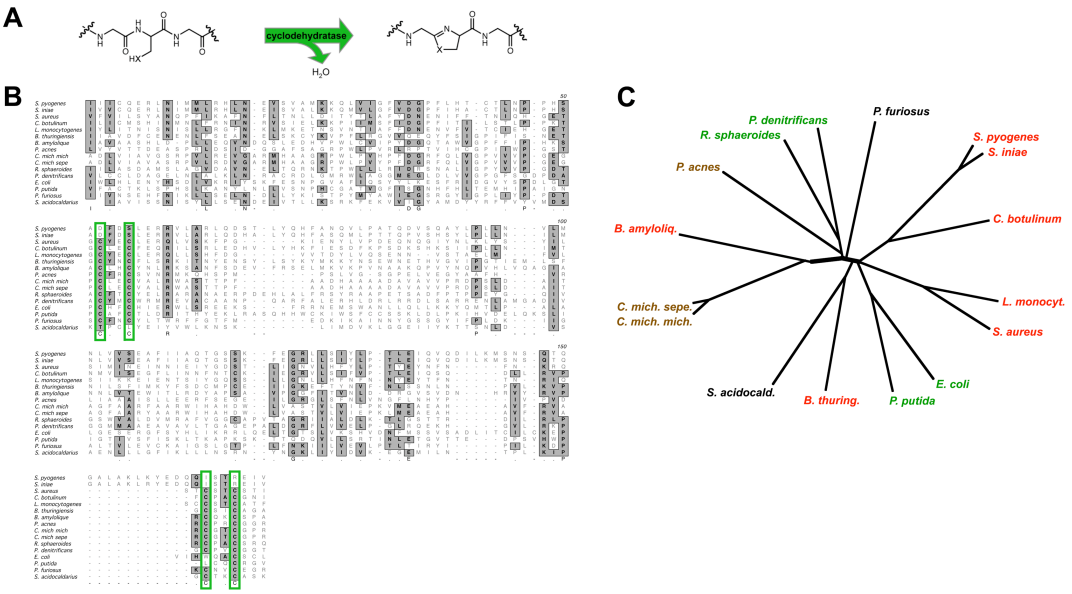


Figure 2.7 Multiple sequence alignment and phylogenetic analysis of the cyclodehydratase.

(A) The chemical transformation carried out by SagC orthologs. The cyclodehydratase catalyzes the conversion of serine to oxazoline (X = O), cysteine to thiazoline (X = S), and threonine to methyloxazoline with loss of water from the parent peptide. (B) ClustalW alignment of 16 SagC orthologs (the others are encoded as a fusion protein to SagD). Because of divergence at the N and C termini of the proteins, only ~160 of the more highly conserved residues are shown in this alignment for each protein. This region contains several positions that contain two nearly invariant CXXC motifs (green boxes) that likely serve to tetrahedrally coordinate a structural Zn^{2+}. The *Streptococcal* orthologs are the exception—they contain two nearby DXXE sequences (D, aspartic acid; E, glutamic acid), a known metal-binding motif. One of the DXXE sequences can be seen on the given alignment. **C,** Phylogenetic analysis of ClustalW aligned SagC proteins reveals evolutionary clustering. The labels are sorted by phylum and are colored as follows: red, firmicutes; green, proteobacteria; black, other. The cyanobacteria, actinobacteria, and bacteriodetes phyla have a tendency to fuse SagCD and are shown in a separate figure.

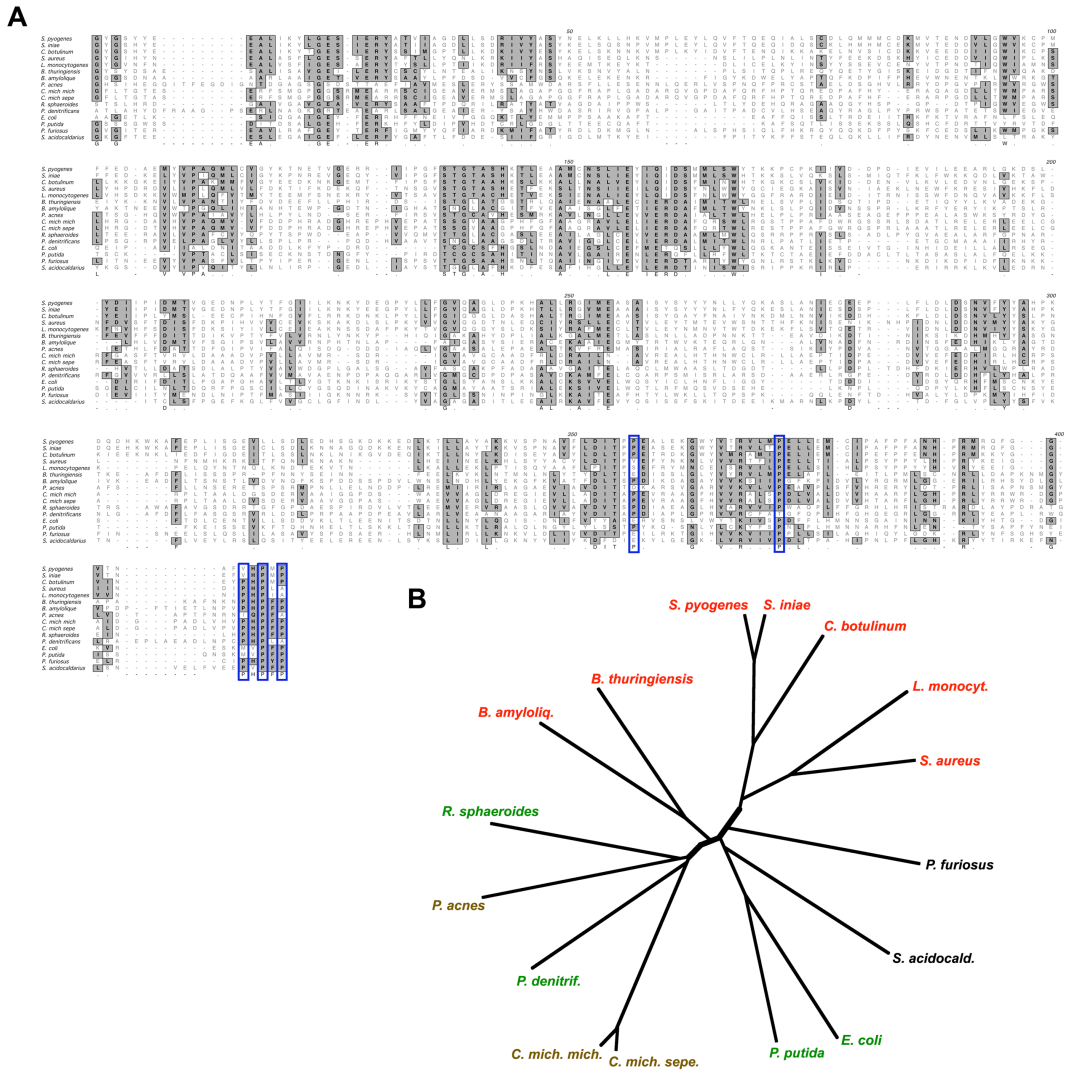


Figure 2.8 Multiple sequence alignment and phylogenetic analysis for the docking protein.

(A) ClustalW alignment of 16 SagD orthologs (the others are encoded as a fusion protein to SagC). Because of divergence at the N-terminus, the ~400 C-terminal residues are shown in this alignment. This region contains several invariant positions and a proline (P)-rich C-terminus. (B) Phylogenetic analysis of ClustalW aligned SagD proteins reveals evolutionary clustering and a conserved proline-rich C-terminus (blue boxes). The labels are sorted by phylum and are colored as follows: red, firmicutes; green, proteobacteria; black, other. The cyanobacteria, actinobacteria, and bacteriodetes phyla have a tendency to fuse SagCD and are shown in a separate figure.

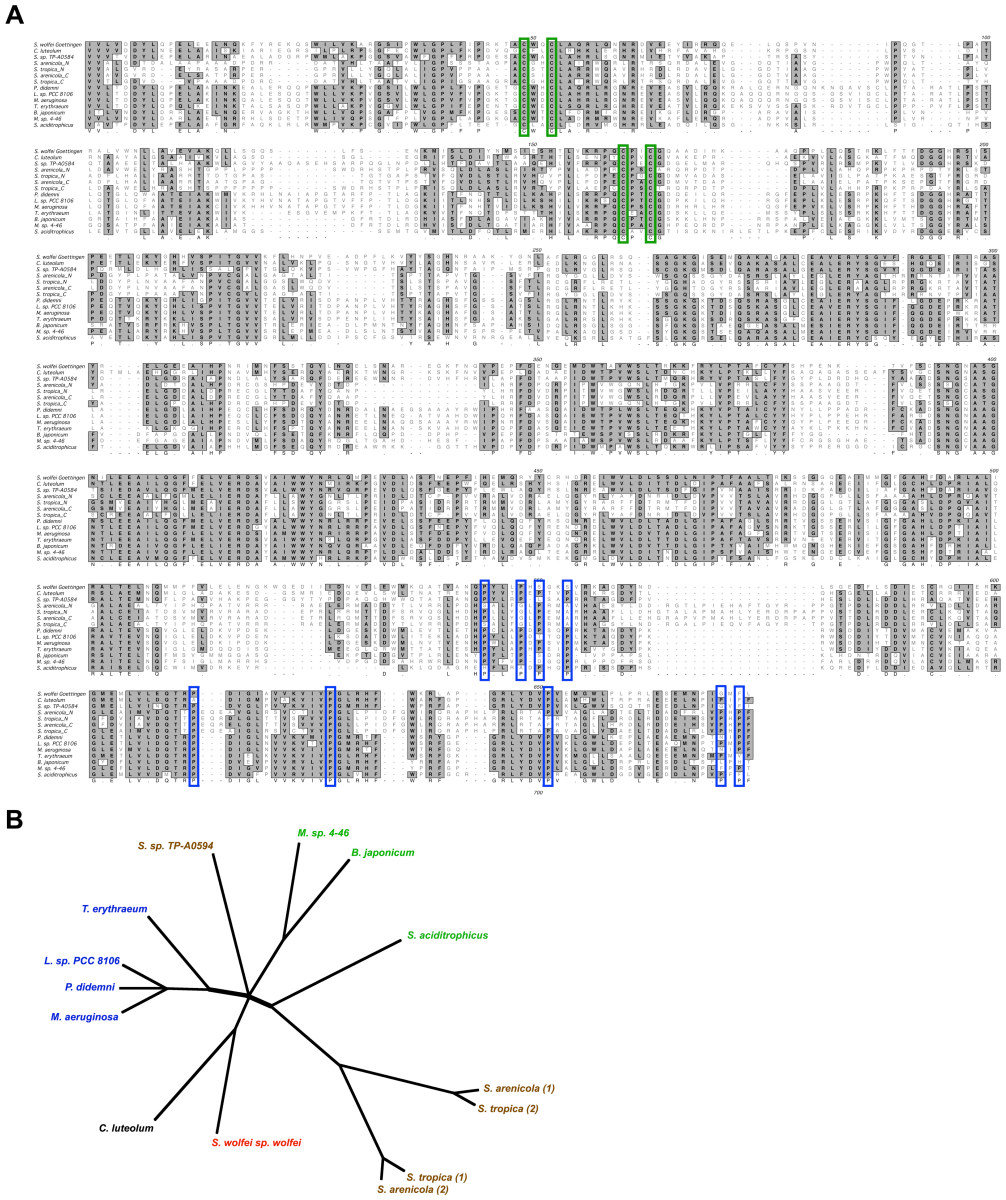


Figure 2.9 Multiple sequence alignment and phylogenetic analysis for the fused cyclodehydratase/docking protein.
(A) ClustalW alignment of 14 fused SagCD orthologs (the *Salinospora* clusters contain two copies of this protein). Because of divergence at the N-terminus, the ~700 C-terminal residues are shown in this alignment. The N-terminal part of this region contains invariant CXXC motifs that comprise a zinc-tetrathiolate (cyclodehydratase domain, green boxes). The C-terminal part of this alignment reveals a proline-rich region (docking domain, blue boxes). (B) Phylogenetic analysis of ClustalW aligned SagD proteins reveals evolutionary clustering. For *S. arenicola* and *S. tropica*, 1 designates the SagCD ortholog at the beginning of the operon and 2 represents the ortholog at the end of the operon. The labels are sorted by phylum and are colored as follows: brown, actinobacteria; green, proteobacteria; blue, cyanobacteria; red, firmicutes; black, bacteriodetes.

References

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-402.

2. **Betschel, S. D., S. M. Borgia, N. L. Barg, D. E. Low, and J. C. De Azavedo.** 1998. Reduced virulence of group A streptococcal Tn916 mutants that do not produce streptolysin S. Infect Immun **66:**1671-9.

3. **Broderick, N. A., K. F. Raffa, and J. Handelsman.** 2006. Midgut bacteria required for Bacillus thuringiensis insecticidal activity. Proc Natl Acad Sci U S A **103:**15196-9.

4. **Datta, V., S. M. Myskowski, L. A. Kwinn, D. N. Chiem, N. Varki, R. G. Kansal, M. Kotb, and V. Nizet.** 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Mol Microbiol **56:**681-95.

5. **Davagnino, J., M. Herrero, D. Furlong, F. Moreno, and R. Kolter.** 1986. The DNA replication inhibitor microcin B17 is a forty-three-amino-acid protein containing sixty percent glycine. Proteins **1:**230-8.

6. **Dos Santos, V. A., S. Heim, E. R. Moore, M. Stratz, and K. N. Timmis.** 2004. Insights into the genomic basis of niche specificity of Pseudomonas putida KT2440. Environ Microbiol **6:**1264-86.

7. **Fuller, J. D., A. C. Camus, C. L. Duncan, V. Nizet, D. J. Bast, R. L. Thune, D. E. Low, and J. C. De Azavedo.** 2002. Identification of a streptolysin S-associated gene cluster and its role in the pathogenesis of Streptococcus iniae disease. Infect Immun **70:**5730-9.

8. **Heddle, J. G., S. J. Blance, D. B. Zamble, F. Hollfelder, D. A. Miller, L. M. Wentzell, C. T. Walsh, and A. Maxwell.** 2001. The antibiotic microcin B17 is a DNA gyrase poison: characterisation of the mode of inhibition. J Mol Biol **307:**1223-34.

9. **Humar, D., V. Datta, D. J. Bast, B. Beall, J. C. De Azavedo, and V. Nizet.** 2002. Streptolysin S and necrotising infections produced by group G streptococcus. Lancet **359:**124-9.

10. **Igarashi, Y., Y. Kan, K. Fujii, T. Fujita, K. Harada, H. Naoki, H. Tabata, H. Onaka, and T. Furumai.** 2001. Goadsporin, a chemical substance which

promotes secondary metabolism and Morphogenesis in streptomycetes. II. Structure determination. J Antibiot (Tokyo) **54:**1045-53.

11. **Kelleher, N. L., P. J. Belshaw, and C. T. Walsh.** 1998. Regioselectivity and chemoselectivity analysis of oxazole and thiazole ring formation by the peptide-heterocyclizing microcin B17 synthetase using high-resolution MS/MS. Journal of the American Chemical Society **120:**9716-9717.

12. **Li, Y. M., J. C. Milne, L. L. Madison, R. Kolter, and C. T. Walsh.** 1996. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. Science **274:**1188-93.

13. **Locke, J. B., K. M. Colvin, N. Varki, M. R. Vicknair, V. Nizet, and J. T. Buchanan.** 2007. Streptococcus iniae beta-hemolysin streptolysin S is a virulence factor in fish infection. Dis Aquat Organ **76:**17-26.

14. **Milne, J. C., R. S. Roy, A. C. Eliot, N. L. Kelleher, A. Wokhlu, B. Nickels, and C. T. Walsh.** 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:**4768-81.

15. **Nelson, K. E., C. Weinel, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. A. Martins dos Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, L. Brinkac, M. Beanan, R. T. DeBoy, S. Daugherty, J. Kolonay, R. Madupu, W. Nelson, O. White, J. Peterson, H. Khouri, I. Hance, P. Chris Lee, E. Holtzapple, D. Scanlan, K. Tran, A. Moazzez, T. Utterback, M. Rizzo, K. Lee, D. Kosack, D. Moestl, H. Wedler, J. Lauber, D. Stjepandic, J. Hoheisel, M. Straetz, S. Heim, C. Kiewitz, J. A. Eisen, K. N. Timmis, A. Dusterhoft, B. Tummler, and C. M. Fraser.** 2002. Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. Environ Microbiol **4:**799-808.

16. **Nizet, V.** 2002. Streptococcal beta-hemolysins: genetics and role in disease pathogenesis. Trends Microbiol **10:**575-80.

17. **Nizet, V., B. Beall, D. J. Bast, V. Datta, L. Kilburn, D. E. Low, and J. C. De Azavedo.** 2000. Genetic locus for streptolysin S production by group A streptococcus. Infect Immun **68:**4245-54.

18. **Onaka, H., M. Nakaho, K. Hayashi, Y. Igarashi, and T. Furumai.** 2005. Cloning and characterization of the goadsporin biosynthetic gene cluster from Streptomyces sp. TP-A0584. Microbiology **151:**3923-33.

19. **Roy, R. S., A. M. Gehring, J. C. Milne, P. J. Belshaw, and C. T. Walsh.** 1999. Thiazole and oxazole peptides: biosynthesis and molecular machinery. Nat Prod Rep **16:**249-63.

20. **San Millan, J. L., R. Kolter, and F. Moreno.** 1985. Plasmid genes required for microcin B17 production. J Bacteriol **163:**1016-20.

21. **Schmidt, E. W., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood, and J. Ravel.** 2005. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proc Natl Acad Sci U S A **102:**7315-20.

22. **Severinov, K., E. Semenova, A. Kazakov, T. Kazakov, and M. S. Gelfand.** 2007. Low-molecular-weight post-translationally modified microcins. Mol Microbiol **65:**1380-94.

23. **Sinha Roy, R., P. J. Belshaw, and C. T. Walsh.** 1998. Mutational analysis of posttranslational heterocycle biosynthesis in the gyrase inhibitor microcin B17: distance dependence from propeptide and tolerance for substitution in a GSCG cyclizable sequence. Biochemistry **37:**4125-36.

24. **Sinha Roy, R., N. L. Kelleher, J. C. Milne, and C. T. Walsh.** 1999. In vivo processing and antibiotic activity of microcin B17 analogs with varying ring content and altered bisheterocyclic sites. Chem Biol **6:**305-18.

25. **Sofia, H. J., G. Chen, B. G. Hetzler, J. F. Reyes-Spindola, and N. E. Miller.** 2001. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. Nucleic Acids Res **29:**1097-106.

26. **Sudek, S., M. G. Haygood, D. T. Youssef, and E. W. Schmidt.** 2006. Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium Trichodesmium erythraeum, predicted from the genome sequence. Appl Environ Microbiol **72:**4382-7.

27. **Yorgey, P., J. Davagnino, and R. Kolter.** 1993. The maturation pathway of microcin B17, a peptide inhibitor of DNA gyrase. Mol Microbiol **9:**897-905.

28. **Yorgey, P., J. Lee, J. Kordel, E. Vivas, P. Warner, D. Jebaratnam, and R. Kolter.** 1994. Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. Proc Natl Acad Sci U S A **91:**4519-23.

Chapter III

Structural and Functional Dissection of the Heterocyclic Peptide Cytotoxin Streptolysin S

Introduction

Streptolysin S (SLS) is secreted by the human pathogen *Streptococcus pyogenes,* the causative agent of diseases ranging from pharyngitis to necrotizing fasciitis (27). SLS is a potent cytolysin that is ribosomally synthesized, extensively post-translationally modified, and exported to exert its effects on the target cell (7, 20). The expression of SLS promotes virulence in animal models of invasive infection and accounts for the hallmark zone of b-hemolysis surrounding colonies of these bacteria grown on blood agar (7, 28). An intriguing feature of SLS is its non-immunogenic nature (32). This characteristic is likely due to its small size and its capacity to lyse cells involved in both innate and adaptive immunity (13, 29). The b-hemolytic phenotype of *S. pyogenes* has been studied since the early 1900s but the molecular structure of SLS has remained elusive (25). In the last decade, transposon mutagenesis studies identified the gene encoding the SLS toxin precursor (*sagA*, for SLS-associated gene) and eight additional genes in an operon required for toxin maturation and export (5). Targeted mutagenesis of the *sag* operon yields non-hemolytic *S. pyogenes* mutants with markedly diminished virulence in mice (7). More recently, it was demonstrated that the protein products of *sagA-D* are sufficient for the *in vitro* reconstitution of cytolytic activity (20). The first gene product, SagA, serves as a structural template that after a series of tailoring reactions, matures into the active SLS metabolite (Fig. 3.1*A*). A trimeric complex of SagBCD catalyzes these tailoring reactions, which results in the conversion of cysteine, serine, and threonine residues to thiazole, oxazole, and methyloxazole heterocycles, respectively (20).

A DNA gyrase inhibitor, microcin B17, is produced by an orthologous biosynthetic cluster (*mcb*) found in a subset of *E. coli* strains (22, 38, 39). Microcin B17 contains four thiazole and four oxazole heterocycles, which are indispensable for biological activity. By analogy to microcin B17 and the lantibiotics, the heterocycles of SLS are formed on the C-terminus of SagA while the N-terminus serves as a leader peptide (8, 24, 35). The installation of thiazole and methyl- (oxazole) heterocycles restricts backbone conformational flexibility and provides microcin B17 and SLS with rigidified structures. The SLS heterocycles are formed via two distinct steps: SagC, a cyclodehydratase, generates thiazoline and (methyl)-oxazoline heterocycles while SagB, a dehydrogenase, removes two-electrons to afford the aromatic thiazole and (methyl)-oxazole (20, 26, 34). SagD is proposed to play a role in trimer formation and regulation (Fig. 3.1*A*). The final genes in the genetic cluster encode a predicted leader peptidase/immunity protein (SagE), a membrane-associated protein of unknown function (SagF), and three ABC transporters (SagGHI).

It is now appreciated that many other prokaryotes harbor similar genetic clusters for the synthesis of thiazole and (methyl)-oxazole heterocycles (6, 9, 20). Additional important mammalian pathogens such as *Listeria monocytogenes*, *Staphylococcus aureus*, and *Clostridium botulinum*, contain *sag*-like gene clusters that produce SLS-like cytolysins. These toxins are expected to promote pathogen survival and host cell injury during infection, but this has only been conclusively shown for *S. pyogenes* and *L. monocytogenes* (6, 7). Like *E. coli*, many other prokaryotes harbor a *sag*-like genetic cluster but are not known to produce cytolysins. Some examples are the goadsporin-producing organism, *Streptomyces* sp. TP-A0584 and cyanobactin

producers such as *Prochloron didemni* (15, 31, 36). The molecular targets of these secondary metabolites remain to be elucidated, but it is known that goadsporin exhibits antibiotic activity and the cyanobactin, patellamide D, reverses multiple-drug resistance in a human leukemia cell line (37). As genetic loci containing *sagBCD*-like genes have been widely disseminated in prokaryotes (20), nature appears to have found a preferred route to synthesizing such secondary metabolites.

In this work, we build upon our initial report on the *in vitro* reconstitution of SLS biosynthesis to uncover the requisite features of substrate selectivity and cytolytic activity. The impetus for defining substrate tolerance arose from earlier results showing that SagBCD accepts alternate substrates *in vitro* (20), as evidenced by two key experiments. First, SagBCD converted a non-cognate substrate, ClosA (*C. botulinum*), into a cytolytic entity. Second, mass spectrometry revealed heterocycle formation on the McbA (*E. coli*) peptide after SagBCD treatment (20). Here, we dissect the N-terminal leader peptide and C-terminal protoxin of SagA to define the residues necessary for conversion into SLS.

Experimental Procedures

In Vitro Synthetase Reactions

Protein preparation and synthetase reactions employing maltose-binding protein (MBP)-tagged substrate and SagBCD were performed as described earlier (20). Membrane damage was quantified by the erythrocyte lysis assay. In every case, omission of the substrate or the SagBCD synthetase resulted in no detectable hemolytic activity (data not shown).

Cytolytic Activity Assay

The *in vitro* and genetic reconstitution assays described below were performed at least three times for each substrate tested. Due to lot-to-lot variation in commercial blood sources and prep-to-prep variation in the specific activity of the synthetase complex, we have elected to report activity semi-quantitatively. *In vitro* hemolytic activity equal to wild-type SagA treated with SagBCD is reported as three plus signs (+++). Cytolytic activity that is approximately 30-70% of wild-type SagA is given as (++). Detectable activity that is less than 30% of SagA is thus a single plus sign (+) and non-detectable activity is a single minus sign (-). The cytolytic activity of mutant substrates tested via genetic reconstitution was scored in an analogous manner. All assays were internally normalized and baseline adjusted using two positive controls (triton X-100 and wild-type SagA treated with SagBCD) and two negative controls (substrate and SagBCD alone).

Transformation and Verification of *S. pyogenes* M1 sagA Mutants

The *sagA* allelic exchange mutant of *S. pyogenes* M1 was made electrocompetent using a previously published glycine/sucrose method (7). Maxiprepped pDCerm constructs (3 μg and 12 μg) were incubated with electrocompetent *S. pyogenes* M1 *DsagA* and electroporated using an Eppendorf 2510 electroporator set to 1.5 kV. These cells (50 μL and 150 μL) were then plated on Todd-Hewitt agar plates supplemented with 2 μg/mL erythromycin. Typically, 5-15 colonies would appear ~40 h post transformation. The insert size was evaluated by screening transformants by colony PCR. Clones harboring an appropriately sized insert were initially screened for cytolytic activity by streaking bacteria on blood agar

plates (Hardy Diagnostics). All clones were verified to be *S. pyogenes* (Group A *Streptococcus*) by using the BBL Streptocard Enzyme Latex Test (BD Diagnostics).

Cytolytic Assay of Genetically Reconstituted Mutants

Due to the possibility that non-physiological concentrations could lead to artifactual activity *in vitro*, the cytolytic activity of the peptide substrates were also tested using genetic reconstitution. This method requires that endogenous SagBCD accept the substrate. The method described below does not involve lysing the bacteria. Therefore, mutant substrates must also be proteolytically processed and accepted by the SagGHI export apparatus (ABC transporters). Due to toxicity in *E. coli* and transformation difficulties, intrinsically lytic SagA mutants were not tested by genetic reconstitution. Extracts containing BSA-stabilized SLS were prepared in the following manner. Overnight cultures (10 mL) of *S. pyogenes* M1 *DsagA* containing pDCerm-*sagA* plasmids were grown to OD_{600} ~0.6 in Todd Hewitt broth containing 2 µg/mL erythromycin. Cultures were treated with BSA (10 mg/mL) for 1 h at 37 °C and then centrifuged (6,000 x g, 10 min) before passing the supernatant through a 0.2 µm acrodisc syringe filter (Pall Corporation). These samples were centrifuged again (6,000 x g, 10 min) and the supernatants were assayed for hemolytic activity by addition to defribinated sheep blood (in V-bottom microtiter plates at 1:25 and 1:50 dilutions). The blood was treated for 2-4 h before assessing hemolytic activity as previously reported (20).

Assessment of SLS Mutants in a Murine Skin Infection Model

Experiments were performed using models reported previously (5, 14, 28). *S. pyogenes* M1 Δ*sagA* complemented with *sagA*-WT, *sagA*-S39A, and *sagA*-7C/S were

grown to log phase, harvested by centrifugation, washed, resuspended in PBS, and mixed 1:1 with Cytodex beads (1 mg/mL, Sigma). An inoculum of (1 x 10^7 colony-forming units in 100 µL) of *S. pyogenes* Δ*sagA* complemented with either *sagA*-S39A or the *sagA*-7C/S mutant was then injected subcutaneously into the right flank of 4-week-old male hairless crl:SKH1(hrhr)Br mice (n = 8 per group). *S. pyogenes* Δ*sagA* complemented with *sagA-wt* was injected into the left flank of the same animal for identical comparison. Animals were monitored daily for development of necrotic ulcers. At 4 d post-infection, all animals were sacrificed. Biopsies were performed on injection sites for histopathologic assessment (hematoxylin/eosin staining) after measuring the size of tissue ulcers.

Generation of ^{35}S-Met-SagBCD

Radioactive synthetase was prepared using maxiprepped pET28-MBP-SagBCD and *in vitro* transcription/translation (TNT) under T7 promoter control. Rabbit reticulocyte extract (RR, Promega) gave superior yield and purity to wheat-germ (WG) and S30 extracts. A typical 50 µL reaction was set up as follows: 1-3 µg of plasmid DNA was added to a master mix containing 2 µL of TNT buffer, 25 µL RR extract, 1 µL RNAsin, 1 µL minus Met amino acid mix, 1 µL of T7 RNA polymerase, 10 µCi of ^{35}S-Met (NEN), and DNAse/RNAse-free water. A small aliquot (0.5 µL) of the radiolabeled product was separated by SDS-PAGE, dried, and visualized by autoradiography using a Kodak BioMax low energy isotope intensifying screen (20 h, -80 °C).

Peptide Array

Immobilized peptides were synthesized on cellulose membranes using a MultiPep Autospot synthesis robot following the manufacturer's instructions (Intavis AG). Irradiation with 254 nm (UV) light gave an indication of the relative amount of peptide per spot. ^{35}S-Met-SagB, -C, and -D were allowed to bind to the array for 15 h at 4 °C in TBS/tween 20 (0.1% v/v) supplemented with 2.5% non-fat milk and 2.5% BSA (both w/w). The array was then extensively washed with TBS/tween (4 x 10 min, 23 °C) before exposing film with a Kodak BioMax low energy isotope intensifying screen (20 h, -80 °C). Under these conditions, SagB and SagD did not bind tightly to the array, as indicated by a weak radioactive signal, even after longer exposure times (4 d). Using SagBCD together led to a substantial amount of "radioactive precipitation" on areas of the cellulose membrane that did not contain peptide. Therefore, later experiments were carried out using SagC alone.

***E. coli* Plasmid Construction**

The genes for the ClosA, StaphA, ListA, SagA-StaphA, SagA-ListA, SagA-inverse, and SagA-scramble substrates were chemically synthesized at Genscript. These genes were provided on a pUC cloning vector and subcloned into the previously described pET28-MBP vector using 5'-BamHI and 3'-NotI restriction sites (20). Point mutations of SagA were introduced using the Stratagene Quikchange site-directed mutagenesis method as per the manufacturer's instructions. Positive clones were verified by DNA sequencing (Eton Biosciences) and transformed into chemically-competent BL21(DE3)RIPL *E. coli* for recombinant expression.

***S. pyogenes* Plasmid Construction**

DNA encoding the protoxin substrates were originally prepared in the pET28-MBP vector as described above. The substrate inserts were PCR amplified using PFU polymerase and primers containing 5'-XbaI and 3'-BamHI restriction sites. The amplified inserts were verified by separating the DNA using a 1% agarose gel and staining with ethidium bromide. After a PCR-clean up step (Qiagen), the *S. pyogenes* compatible vector pDCerm (erm, erythromycin resistance) and inserts were digested with XbaI (16 h, 37 °C) and then with BamHI (1 h, 37 °C) in Buffer 3 (NEB) supplemented with BSA. Ligation reactions were carried out using T4 ligase (NEB) for 16 h at 16 °C. The ligation products were then transformed into chemically-competent DH5a *E. coli*. A two-hour outgrowth was required to obtain transformed bacteria, which became visible on LB/erm agar plates approximately 36 h post transformation. Single transformants were grown to saturation (~36 h) in 10 mL of LB supplemented with 500 μg/mL erythromycin and miniprepped. Positive clones were verified by DNA sequencing (primer: 5'-CGG CAT AAA TCG CTC AGA ACA AGC TC-3'), grown to saturation in 1 L of LB with 500 μg/mL erythromycin, and maxiprepped (Invitrogen). Typical DNA yield was ~1 mg, as determined by absorbance measurement on a Nanodrop spectrophotometer.

Surface Plasmon Resonance

Surface plasmon resonance analysis was performed at 23 °C on a Biacore T100 (GE Healthcare) using SensorChip NTA at a flow rate of 25 μL/min. The running buffer was: 10 mM MOPS (pH 7.4), 150 mM NaCl, 50 μM EDTA, 0.005% (v/v) triton X-100, sterile-filtered, and degassed. Hexahistidine-tagged wild-type SagA

leader peptide (SagA-WT), F4A/I8A (SagA-FIA), and T17A/Q18A/V19A (SagA-TQV) mutants were synthesized at Genscript. Peptide sequences were: MLK**F**TSN**I**LA TSVAETTQVAPGGHHHHHH (3.16 kDa, 86% pure by HPLC), MLK**A**TSN**A**LA TSVAETTQVAPGGHHHHHH (3.04 kDa, 94% pure by HPLC) and MLKFTSNILA TSVAET**AAA**APGGHHHHHH (3.04 kDa, 88% pure by HPLC). These peptides were dissolved in running buffer at 200 nM. At this concentration, a 30 s injection gave a response of ~110 units on flow cells preloaded with Ni^{2+} (50 s injection of 500 μM $NiCl_2$ dissolved in the running buffer). A non Ni^{2+}-loaded flow cell was used as a negative control and served as a reference subtraction. Non His-tagged SagB, SagC, and SagD (from the pET15 vector) were then injected (individually and in combination) onto the flow cells (0.125 – 2.5 μM) loaded with each leader peptide. Association and dissociation curves were recorded in duplicate for 120 s and 240 s, respectively, with a 0.1 s data acquisition rate. At the end of each run, the Ni-NTA surface was stripped with a 50 s injection of regeneration buffer (350 mM EDTA in running buffer). Biacore evaluation software (BIAevaluation, version 3.2) was used to fit the association and dissociation curves of all injectants using a 1:1 binding model. Due to very slow off-rates, it was important to normalize the data to account for dissociation between Ni-NTA and the His tag. This was accomplished by adding a straight line with a slope equal to the dissociation rate of Ni-NTA:His_6 to the entire sensorgram (0.025-0.060 RU/s). The slope was determined using linear regression from the recorded response data, over the course of 100 s, before the analyte was injected. This adjustment was performed on all samples. Curve fitting analysis was performed after this baseline correction.

Results

The SagA Leader Peptide Provides Substrate Recognition

The precursor peptide from *Clostridium botulinum*, ClosA, shares significant amino acid sequence similarity with SagA (62%) and was converted to a cytolysin by recombinant SagBCD (20). The microcin B17 precursor, McbA, is only moderately similar (32%) but was also accepted as a substrate by SagBCD. Therefore, we hypothesized that SagBCD would accept numerous non-cognate substrates (Fig. 3.1*B*). To demonstrate that the permissive behavior of SagBCD on ClosA was not limited to *in vitro* biochemical studies with purified proteins, genetic complementation studies were performed using a *sagA* deletion mutant produced in an M1 serotype strain of *S. pyogenes* (D*sagA*). By transforming this strain with a plasmid encoding the desired peptide substrate, the *in vivo* selectivity of endogenous SagBCD was probed. b-hemolysis on blood agar was restored when either the wild-type *sagA* or *closA* gene was provided to the *S. pyogenes DsagA* mutant bacteria. To increase sensitivity, we adopted a liquid phase hemolysis assay. Here, the SLS is extracted from the *S. pyogenes* lipoteichoic acid layer using bovine serum albumin (BSA) as a carrier/stabilizer. This sample is mixed with sheep's blood and the amount of hemoglobin released from lysed erythrocytes is quantified by absorbance. Using this assay, we confirmed ClosA was converted to a cytolysin by endogenous SagBCD, but calculated the level of activity to be reduced by approximately 1/3 relative to that of SagA (Fig. 3.1*B*).

Interestingly, the related peptide substrates from *Staphylococcus aureus* (StaphA) and *Listeria monocytogenes* (ListA) did not exhibit detectable hemolytic activity with purified recombinant proteins or in genetically complemented *S. pyogenes* D*sagA* (Fig. 3.1*B*). Cotter and colleagues have recently shown that the *sag*-like genetic cluster from *Listeria monocytogenes* synthesizes an SLS-like cytolytic factor (6). Therefore, the lack of activity in our assays must originate at the level of substrate recognition or enzymatic tolerance. Previous work from the Walsh laboratory has shown that the N-terminal leader peptide of McbA is required for substrate recognition (24). The most important residues implicated in this process are McbA-F8 and -L12, which are proposed to lie on the same face of an alpha-helix (35). Alignment with other substrates shows that a similar motif (FxxxB, where x is any amino acid and B is a branched chain amino acid) is found in the leader peptides of SagA and ClosA but is lacking in StaphA and ListA (Fig. 3.1*B*). This finding prompted the construction of a double alanine mutant of the FxxxB motif (SagA-FIA) and chimeric substrates comprised of the SagA leader peptide fused to the StaphA and ListA C-terminus (Fig. 3.1*C*). If the SagA leader peptide contains adequate substrate recognition information, then activity should be restored upon treating the chimera with purified SagBCD. Furthermore, if the FxxxB motif is required for substrate recognition, then cytolytic activity should be reduced for similarly treated SagA-FIA. A full restoration of *in vitro* cytolytic activity was observed for both the SagA-ListA and SagA-StaphA chimera. In contrast, genetic complementation revealed that only SagA-StaphA retained activity, indicating that either the endogenous rules of cytolytic conversion are more restrictive or that the SagA-ListA substrate is not efficiently

exported. Enzymatic promiscuity could theoretically be amplified *in vitro* by the presence of large amounts of purified SagBCD. This is supported by the observation that SagA-FIA has detectable activity using purified enzymes but not under genetic reconstitution conditions (Fig. 3.1*C*).

Unnatural SagA analogs were prepared to shed light on the positional requirements of heterocycle formation. We first designed an artificial substrate, SagX, in an attempt to assess the minimum features for the formation of a cytolysin. Four criteria were used to design this potential substrate: *i.* the SagA leader peptide to provide recognition, *ii.* a stretch of contiguous heterocyclizable residues adjacent to the leader peptide cleavage site, *iii.* 30% glycine evenly distributed through the C-terminal half and, *iv.* a serine flanked by two glycine residues. Akin to SagA-ListA, this artificial substrate gave activity equal to SagA in the *in vitro* assay but no activity upon complementation of the D*sagA* mutant of *S. pyogenes*. The region of SagA between the cysteine-rich region and the last heterocyclizable residue ($_{33}$FSIA…GSYT$_{50}$) was further probed by designing two additional unnatural substrates. SagA-inverse simply inverted residues 33-50 of SagA, while in SagA-scramble this region was reordered in a randomized fashion. Despite having the highest protoxin similarity to wild-type SagA, the SagA-inverse substrate did not yield lytic activity in either assay. SagA-scramble was active *in vitro* but did not complement hemolytic activity in *S. pyogenes DsagA*. These data show that the leader peptide, especially the FxxxB motif, is important for substrate acceptance and that heterocyclizable positions must be available at precise positions in order to allow creation of an SLS-like cytolytic factor. Additionally, these data indicate that even

though SagBCD accepts artificial substrates, there is a limit to the enzymatic promiscuity. These limits are more pronounced during genetic complementation experiments most likely because substrate recognition and enzymatic tolerance are bypassed to an extent when using purified SagBCD. An alternative explanation is that the unnatural substrates are secreted from *S. pyogenes* with differing efficiencies.

Analysis of SagA Leader Peptide Binding Requirements

Because the above activity assays measure both substrate binding and heterocycle formation in a simultaneous and indirect fashion, peptide arrays were synthesized to directly evaluate binding. Based on a perceived importance of the FxxxB motif in SagBCD substrate acceptance, the first array consisted of a panel of leader peptides that contain the FxxxB motif. Next to each wild-type (WT) sequence, the FxxxB double alanine mutation was synthesized for SagA, ClosA, McbA, and a substrate from *Pyrococcus furiosus*, designated PagA (Fig. 3.2*A*). Equal peptide loading was confirmed by irradiating the array with UV (254 nm) light. To assess binding, ^{35}S-Met labeled SagBCD was prepared by *in vitro* transcription/translation (Fig. 3.5) and allowed to interact with the peptide array as described in the methods. Initially, the SagBCD complex was tested for binding in approximately a 1:1:1 ratio, as judged by the number of methionines and the intensity of exposure. Unfortunately, this led to a substantial amount of precipitation that could not be removed from the cellulose membrane and rendered the binding information unreliable. Therefore, SagB, SagC, and SagD were tested individually for binding. SagB and SagD did not bind tightly to the leader peptide array and were washed off before exposing the film (data not shown). SagC, however, bound with high affinity to the array. For the WT

sequences, SagC binding was highest to SagA, followed by ClosA and PagA (Fig. 3.2*A*). Binding to McbA was the weakest, as measured by autoradiographic intensity. Importantly, binding to each FxxxB mutant was reduced in comparison to the corresponding WT leader peptide.

The FxxxB Leader Motif is Necessary but not Sufficient for SagC Recognition

Although the FxxxB motif is an important determinant for directing SagC substrate binding, it does not contribute all of the interaction energy. This is illustrated by the observation that SagC bound to the SagA-WT leader peptide more efficiently than the other WT peptides (Fig. 3.2*A*). Furthermore, simple incorporation of FxxxB into the StaphA and ListA leader peptide did not provide sufficient affinity to detect SagC binding by this method (data not shown). These observations led to the hypothesis that additional residues of SagA contribute to SagC binding.

To elucidate these binding determinants, a second array was synthesized that consisted of an alanine scan of the SagA leader peptide (Fig. 3.2*B*). Each residue of SagA that is not naturally found as alanine was individually mutated to alanine and tested for SagC binding as above. As expected, mutation to alanine did not disrupt SagC affinity at every location. However, binding levels were significantly reduced when residues comprising the FxxxB motif were mutated (F4 and I8). It was also found that residues adjacent to the FxxxB sequence (L2, K3, and L9) and another binding site (T17, Q18, V19) were also critical for SagC binding (Fig. 3.2*B*).

An additional array was used to ascertain the contribution of each interaction site found in the leader peptide, and to also assess if the C-terminal region of SagA plays a role in binding. A panel of fourteen 10-mer peptides were included in the array

which scan the full length of SagA, except for the oxidatively prone, and synthetically challenging, cysteine-rich region (Fig. 3.2*C*). After binding SagC to this array, it was found that only the first peptide was capable of providing enough binding information to retain SagC through the washing steps. Peptide 1 comprises residues 1-10 of SagA and not only contains the FxxxB motif but also the important adjacent residues (L2, K3, F4, I8 and L9). Peptide 5 contains I8 and L9 but only the first residue of the TQV site (T17). This peptide is insufficient at retaining SagC. Taken together, these findings demonstrate that separate sites within the N-terminal leader peptide synergize to provide SagC with a high affinity binding site.

Binding parameters for the SagA-WT, SagA-FIA (FxxxB double alanine mutant), and SagA-TQV (triple alanine mutant) leader peptides with SagC were then measured by surface plasmon resonance to quantify the contribution of each binding motif. Using a C-terminal hexahistidine tagged version of SagA-WT, -FIA, and -TQV, the rates of association (k_a) and dissociation (k_d) were first measured with SagC (Table 1). A typical sensorgram and curve-fitting analysis are provided in Fig. 3.6. As expected from the peptide array data, the effects of FxxxB and TQV mutation were substantial. Compared to SagA-WT, SagC bound to the SagA-FIA peptide ~16-fold less tightly. The reduction in affinity was accounted for by an increased k_d (k_a was unchanged). SagC affinity for SagA-TQV was also reduced (~24-fold). The major factor in the SagA-TQV mutant's reduction in affinity is an increased k_d; however, there is also a slight decrease in k_a (<2-fold) upon comparison to SagA-WT (Table 1).

We next determined if the rules of substrate recognition applied to other cyclodehydratases. ClosC and McbB, from *C. botulinum* and *E. coli* respectively,

were tested for SagA binding. Relative to SagC, binding to the SagA-WT leader peptide was reduced with both ClosC and McbB (~12-fold and ~130-fold, respectively, Table 1). This trend correlates with the degree of amino acid similarity to SagC (59% for ClosC; 28% for McbB). Mutation of the SagA FxxxB and TQV motifs led to additional losses of affinity in the same manner as for SagC binding (*i.e.* both the FxxxB and TQV mutants effect k_d; the k_a is only slightly affected by the mutation of TQV). This suggests that although each cyclodehydratase is likely fine-tuned for a single endogenous substrate, the substrate binding mechanism is conserved across species.

Mutational Analysis of Heterocyclizable Residues of SagA

Numerous attempts at identifying sites of heterocycle formation on SagA by mass spectrometry have failed. As a result, a site-directed mutagenesis approach was initiated to determine functionally important residues of SagA with the assumption that the lack of cytolytic activity by particular SagA point mutants is suggestive of heterocycle locations. To begin this study, every heterocyclizable residue (cysteines, serines, and threonines) located on the protoxin half of SagA (residues 24-53) was individually mutated. Cysteines and serines were mutated to alanines while threonines were mutated to valines, a closer structural mimic. The lytic activity of these proteins was assessed *in vitro* after treatment with purified SagBCD and through genetic complementation of *S. pyogenes DsagA*. Each mutation was classified as having no effect (mutant SLS was fully active), a contributing effect (reduced activity), or a critical effect (inactive). Using these two independent assays, key functional residues of SLS were identified. It is important to note that our method of testing the cytolytic

activity of SagA point mutants though genetic complementation measures activity and secretion in a simultaneous fashion.

Compiled results are overlaid on the amino acid sequence of SagA (Fig. 3.3*A*) and also provided in table form (Fig. 3.3*B*). This analysis revealed that not all sites of potential heterocycle formation were important for activity but most mutations had a measurable impact on function. Two mutations, C32A and S39A, completely abolished the cytolytic activity of SLS in both assays and are thus highly likely sites of thiazole/oxazole formation. Moreover, it was found that serine and cysteine could substitute for one another at these two critical sites (C32S and S39C) with a small reduction of *in vitro* activity. However, serine cannot be used to replace multiple adjacent cysteines as indicated by the lack of activity in C31S/C32S (2C/S), C26S/C27S/C28S (3C/S), and a total serine (7C/S) mutant (Fig. 3.3*B*). Similarly, valine cannot completely replace threonine (4T/V).

Overall, the activities measured by *in vitro* reconstitution and genetic complementation agreed. The one case were the assays did not concur was with the C24A mutation. We observed substantial lytic activity *in vitro* but no activity in genetically complemented bacteria (Fig. 3.3*B*). This residue is predicted to flank the leader peptide cleavage site. We have previously shown that leader peptide removal is not required for *in vitro* activity (20). Thus, a reduction of activity in genetic complementation experiments may result from reduced leader peptide processing and/or secretion. These results confirm that many residues within SagA are important for activity and predict that SLS is highly post-translationally modified.

Proline Can Substitute for Thiazoles and Oxazoles in the SagA Propeptide

We rationalized that since thiazoles and oxazoles are 5-membered, nitrogen-containing heterocycles that restrict peptide backbone conformation, proline may be capable of serving as a structural mimic. Therefore, the SagA-C32P and -S39P mutants were prepared and treated with purified SagBCD. These yielded substantial *in vitro* lytic activity (Fig. 3.3*B*) and encouraged the construction of SagA mutants that were cytolytic in the absence of SagBCD. Thus, proline was systematically incorporated into select heterocyclizable sites of SagA that were previously determined to be important for activity (positions 32, 34, 39, 46, 48). This proline-substituted SagA panel was then tested for lytic activity in the absence of SagBCD (Fig. 3.3*C*). Wild-type SagA and two single substituted proline mutations (C32P and S39P) did not exhibit intrinsic lytic activity. In contrast, two double proline substituted mutants showed dose-dependent activity without SagBCD treatment (C32P/S39P and S34P/S39P) (Fig. 3.3*C* and Fig. 3.7). A triple mutant, C32P/S34P/S39P, predicted to further enhance lytic activity, killed *E. coli* upon induction of protein expression and yielded no protein. Likewise, additional permutations revealed that if either C32 or S34 are substituted in combination with S39, S46, and S48, *E. coli* toxicity was encountered. To support the claim that toxicity arose due to SLS-like activity, a synthetic peptide comprised of the C-terminal 21 residues of SagA was synthesized with prolines at positions 34, 39, 46, and 48. This peptide, although devoid of cysteine, was intrinsically lytic in a dose-dependent manner (Fig. 3.3*C* and Fig. 3.7). The observation that select proline mutations endow SagA with intrinsic lytic activity bolsters the conclusion that the above sites are heterocyclized in SLS.

Key Heterocycles of SLS are Required for *S. pyogenes* Virulence in Mice

To determine the importance of heterocyclic conversion of SLS for disease progression, mice were subcutaneously injected with *S. pyogenes* M1 Δ*sagA* mutant complemented with either *sagA*-WT, *sagA*-S39A, or *sagA*-7C/S. Seven of eight mice infected with the *sagA*-WT complemented strain developed significant necrotic ulcers and swelling, while no change was observed in any infections of *sagA*-S39A (Fig. 3.4*A*) and *sagA*-7C/S (data not shown). Histological examination of tissues from WT and the mutant infections shows that WT infections exhibit more extensive tissue disruption and a higher infiltration of neutrophils (Fig. 3.4*B*). These results demonstrate that a single SagA point mutation predicted to interfere with heterocycle formation markedly reduces the virulence of *S. pyogenes* in a murine model of skin infection.

Discussion

In order for SagA to be converted into the active SLS cytolysin, the SagBCD synthetase complex must first recognize the SagA substrate. Heterocycles must then be installed at the proper locations in order to elicit biological activity. By using two independent assays, we analyzed the substrate permissivity of SagBCD using *in vitro* biochemical assays and genetic reconstitution. The substrate tolerance of SagBCD was initially assessed using substrates from other members of the firmicutes phylum that were suspected of producing SLS-like cytolysins. To begin, we chose the substrate from *Clostridium botulinum*, which harbors the most highly related *sag*-like genetic

cluster (*clos*A-I) known outside of the *Streptococcus* genus. SagBCD accepted ClosA as a non-cognate substrate *in vitro* and in genetically complemented *S. pyogenes DsagA* bacteria, but the activity was reduced relative to SagA (Fig. 3.1*A*). This could stem from a number of possibilities. The most likely explanation, in light of our binding data, is that SagBCD process ClosA more slowly due to a lowered substrate affinity. *In vitro,* ClosA treated with purified SagBCD resulted in activity equal to SagA. However, under the conditions employed, non-physiological concentrations of synthetase and substrate could easily result in an elevated level of substrate permissivity. Other explanations for reduced ClosA activity upon genetic reconstitution could be that on a molar basis, SLS is more potent or that the leader peptidase and/or export machinery process the toxin more slowly. If the latter was true, toxicity would be an expected outcome but analysis of growth curves indicated that this is not the case (data not shown).

Since ClosA and McbA were both known *in vitro* substrates of the SagBCD synthetase (20), we were initially surprised when SagBCD rejected potential substrates more similar to SagA than McbA (*i.e.* StaphA and ListA, Fig. 3.1*B*). However, chimeric substrates consisting of the SagA leader peptide coupled with the StaphA or ListA protoxin were fully active *in vitro* after treatment with recombinant SagBCD. This indicated that the SagA leader peptide conferred substrate recognition. Later experiments showed that a partially conserved motif in the SagA leader peptide, FxxxB, played a major role in efficient cytolytic conversion (Fig. 3.1*C*) and SagC affinity (Fig. 3.2 and Table 1). Additionally, the SagA leader peptide can be fused to properly designed artificial protoxins to yield cytolysins after treatment with SagBCD.

Peptide array binding experiments uncovered that in addition to the FxxxB motif and some adjacent residues, SagC also recognizes another site (TQV) on the SagA leader peptide (Fig. 3.2). Given that the K_D was measured to be in the low nM range, multiple binding sites on SagA would be presumed necessary to drive such an avid interaction (Table 1). Mutation of the SagA-TQV site led primarily to an increased k_d with a modest decrease of k_a relative to SagA-WT (Table 1). This is in contrast to the SagA-FIA mutant, which only led to an increased k_d. These binding trends also held for the SagA binding interaction with the cyclodehydratases from a closely related organism (*C. botulinum*, ClosC) and a distantly related organism (*E. coli*, McbB, Table 1). We hypothesize this relationship will be true of all *sag*-like clusters that harbor substrates with FxxxB motifs in the leader peptide.

Elucidation of the SagC substrate recognition requirements was undertaken to judge why certain substrates are accepted by the SagBCD synthetase and why others are not. Earlier studies have demonstrated that the SLS precursor from a fish pathogen, *Streptococcus iniae*, was able to complement the hemolytic activity of *S. pyogenes* Δ*sagA* (10, 23). Upon comparison to SagA from *S. pyogenes*, the *S. iniae* leader peptide has a single mutation in the leader sequence (K3Q). Thus, it should not be surprising that SagA from *S. iniae* was an efficient substrate of *S. pyogenes* SagBCD. We conclude that ClosA was accepted by recombinant and endogenous SagBCD due to a complete FxxxB site (LKF and VL) and a partial TQV site (V19). Previous results has also shown that SagBCD accepts McbA as a substrate (20). This is supported by the present work that demonstrates SagC directly binds to the McbA leader peptide (Fig. 3.2*A*). Relative to SagA, the SagC binding to McbA was greatly reduced but the

data provide a satisfactory explanation: the leader peptide of McbA contains partial SagC binding sites. The FxxxB sequence is present but the important surrounding residues are non-conservatively mutated. Also, the "Q" of the TQV site is present (Fig. 3.2*A*).

The identification of the TQV site helps explain why simple inclusion of FxxxB into the StaphA and ListA leader peptides did not result in high affinity SagC binding. Residues adjacent to the FxxxB motif (LKF and IL) also contribute to this interaction. Out of a total 23 residues in the predicted SagA leader peptide, eight are directly involved in driving substrate affinity. This implies that upon SagC binding SagA, nearly the entire length of the leader peptide will be associated, possibly buried, in a SagC binding site. Undoubtedly, the extended recognition capability of SagC acts to prevent the synthetase complex from processing unwanted peptides/proteins. Perhaps this affinity is also a protective measure employed by *S. pyogenes*. Our work has demonstrated that pro-SLS (Fig. 3.1*A*) is highly active and it is possible that a tightly bound synthetase complex will cage the activity until the metabolite is ready for export.

SLS has historically been recalcitrant to isolation and characterization. However, the early work of Bernheimer (1, 3, 4) and Koyama (17-19) shed important light on the nature and action of SLS. Using gel filtration analysis on partially purified SLS, Bernheimer predicted that the molecular weight of the polypeptide would be 2.8 kDa (2). This is quite remarkable, given that a modern calculation of 2.7 kDa is based on modifications to SagA, a peptide discovered over 30 years later. Our results suggest that in the mature structure of SLS, oxazoles will be formed from SagA residues S34,

S39, S46, and S48 (Fig. 3.3*B*). The work of Kolter and Walsh on microcin B17 provides a literature precedent for heterocycle formation when the preceding residue is glycine (S39, S36, S48) (16, 22, 33, 38, 39). Presumably, the increased flexibility in the peptide backbone at these locations facilitates the orbital alignment required for cyclodehydration. Oxazole formation at the abovementioned sites is further supported by the creation of proline substituted SagA mutants that exhibit intrinsic cytolytic activity (Fig. 3.3*C*).

Our mutagenesis data confirm a previous study identifying SagA-C24 and -C27 as residues important for the activity of SLS on blood agar plates (7). Despite this, the fate of the cysteine-rich region (C24-C32) remains speculative. Given the observation that C32S and C32P were cytolytic, we expect C32 to be converted to a thiazole. However, serine cannot simultaneously substitute for multiple adjacent cysteines. This implies that there is either a polythiazole-specific electronic contribution to cytolysis or that another type of modification is present in the mature structure of SLS. Our current data cannot distinguish between these two possibilities, which will require extensive methodology improvements. We hypothesize that it is unlikely free thiols, disulfides, or lanthionines will be present in the final SLS structure. This is because SLS is known to be oxygen-stable (11, 12) and cystine was not detected in an amino acid analysis study (17). Moreover, the *sagA-I* cluster, which is sufficient to confer SLS production to a heterologous bacterial species (28), does not harbor the appropriate dehydratases and cyclases to form lanthionine linkages (21, 28, 30).

The murine infection study supports our previous data that SLS heterocycle formation is required for the virulence of *S. pyogenes*. Furthermore, we show that a single amino acid change, SagA-S39A, renders this pathogen avirulent in a skin infection model (Fig. 3.4). Without active SLS, the bacteria exhibit undetectable pathogenicity and are efficiently cleared by the murine immune system. It is conceivable that in such a setting, antibodies could be raised against inactive SLS mutants. Whether this represents a viable strategy for vaccine development remains to be seen but with the tactical role SLS-like toxins play in pathogenesis, further study is warranted.

This study has elucidated the requisite features of SagBCD substrate recognition and SagA residues of functional importance. We conclude that although the C-terminus of SagA undergoes enzymatic conversion, only the N-terminal leader peptide substantially contributes to substrate recognition. Conceivably, these principles could be used to guide the design of other artificial toxins for antibiotic or anticancer applications. Additionally, our study has yielded a wealth of structure-activity data on a conserved family of bacterial virulence factors and lays the foundation for the development of a structure-based vaccine.

Acknowledgements

We thank CJ Allison (S. Taylor Lab, UCSD) for preparing the peptide arrays, and Jill Harrington (Kolodner Lab, UCSD) for assistance with the Biacore T100. We are grateful to members of the Dixon lab and Mary Hensler for helpful discussions.

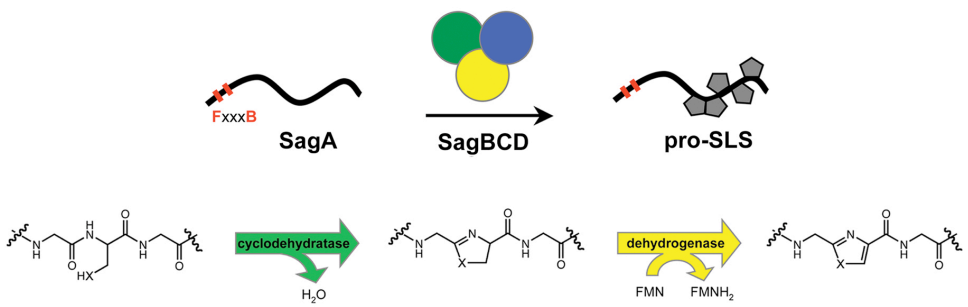
Table 3.1 Contribution of SagA interaction sites to cyclodehydratase binding kinetics

Ligand	Analyte	k_a ($\times 10^3$)	k_d ($\times 10^{-4}$)	K_D (nM)	χ^2
SagA-WT	SagC	3.7 ± 0.4	0.25 ± 0.03	6.7 ± 0.8	0.04
SagA-FIA	SagC	3.7 ± 0.3	3.5 ± 0.3	95 ± 7.0	0.04
SagA-TQV	SagC	2.1 ± 0.3	3.4 ± 0.8	160 ± 30	0.03
SagA-WT	ClosC	3.8 ± 0.3	3.1 ± 0.7	82 ± 14	0.10
SagA-FIA	ClosC	3.3 ± 0.5	7.2 ± 1.0	220 ± 30	0.07
SagA-TQV	ClosC	2.5 ± 0.3	9.6 ± 0.9	380 ± 40	0.03
SagA-WT	McbB	2.5 ± 0.3	22 ± 4.0	880 ± 140	0.58
SagA-FIA	McbB	2.3 ± 0.2	27 ± 1.0	1200 ± 100	0.25
SagA-TQV	McbB	2.0 ± 0.3	41 ± 5.2	2100 ± 310	0.04

SagA-WT	MLKFTSNILATSVAETTQVAPGGHHHHHH
SagA-FIA	MLKATSNALATSVAETTQVAPGGHHHHHH
SagA-TQV	MLKFTSNILATSVAETAAAAPGGHHHHHH

Surface plasmon resonance was performed using leader peptides of SagA with a C-terminal His_6-tag as the immobilized ligand. The sequence of each peptide is given in the lower section of the table. The red residues indicate a SagC binding motif. The analytes tested for binding were the cyclodehydratases from *S. pyogenes* (SagC), *C. botulinum* (ClosC), and *E. coli* (McbB). The association rates (k_a, $M^{-1}s^{-1}$) and dissociation rates (k_d, s^{-1}) were recorded in duplicate at two concentrations and used to calculate the dissociation constant (K_D). Error is reported as standard deviation of the mean. Chi-squared (c^2) values are given as an assessment of curve-fitting accuracy.

A



B

Substrate	Wild-type Sequence	Full-length Similarity (%)	Lytic Activity *In vitro*	Lytic Activity *In vivo*
SagA	MLKFTSNILATSVAETTQVAPGG*CCCCCTTCCFSIATGSGNSQGGSGSYTPGK	100	+++	+++
ClosA	MLKFNEHVLTTTNNSNNKVTVAPG*SCCCCSCCCCVSVSVGGGSASTGGGAAAGQGGN	62	+++	++
StaphA	MMKINNHTINGYSDINSSEAMQYAAG*CCSCSCSCSCSCSCSCTSASTAEQ	46	-	-
ListA	MNIKSQSSNGYSNNAVGSEAMNYAAG*CCSCSCSTCTCTCTCASSAATKM	44	-	-
McbA	MELKASEFGVVLSVDALKLSRQSPLG-VGIG$_{10}$SCGGQGGGCGGCSNGCSGGNGGSGGSGSHI	32	n.a.	n.a.

FxxxB Leader peptide — Protoxin

C

Substrate	Mutant Sequence	Protoxin Similarity (%)	Lytic Activity *In vitro*	Lytic Activity *In vivo*
SagA-FIA	MLKATSNALATSVAETTQVAPGG*CCCCCTTCCFSIATGSGNSQGGSGSYTPGK	100	+	-
SagA-StaphA	MLKFTSNILATSVAETTQVAPGG*CCSCSCSCSCSCSCSCTSASTAEQ	50	+++	+++
SagA-ListA	MLKFTSNILATSVAETTQVAPGG*CCSCSCSTCTCTCTCASSAATKM	45	+++	-
SagX	MLKFTSNILATSVAETTQVAPGG*CCCCGGIAGAINGGSGGSYTPNK	57	+++	-
SagA-inverse	MLKFTSNILATSVAETTQVAPGG*CCCCCTTCCTYSGSGGQSNGSGTAISFPGK	68	-	-
SagA-scramble	MLKFTSNILATSVAETTQVAPGG*CCCCCTTCCNSGGFGSATYSSQGGSTIPGK	60	++	-

Figure 3.1 SagBCD substrate recognition is provided by the SagA leader peptide
(A) SagA is converted into an active cytolysin, pro-streptolysin-S (pro-SLS), by the actions of SagBCD (a trimeric oxazole/thiazole synthetase). Heterocycles are schematically represented as shaded pentagons. A marginally conserved motif in the SagA leader peptide, FxxxB (where B is a branched-chain amino acid), is highlighted in red. Individual reactions catalyzed by SagC (cyclodehydratase) and SagB (FMN-dehydrogenase) are shown. (B) Representative amino acid sequences and cytolytic activity of SagA-like substrates. Shown in red are leader peptide residues that comprise the FxxxB motif. The putative leader peptide cleavage sites are shown as asterisks, except for McbA, where the site is known (hyphen). In blue are sites of potential heterocycle formation (for McbA, known sites are blue). Percent amino acid similarity to full-length SagA (as determined by ClustalW alignment) is given. The cytolytic activity was tested for these substrates *in vitro* using purified proteins and *in vivo* using the SLS-deficient strain, *S. pyogenes DsagA*, complemented with the desired substrate. Activity equal to wild-type SagA is designated as (+++); activity that is 30-70% of wild-type SagA is (++); detectable activity that is less than 30% of SagA is noted as (+); while non-detectable activity is (-). The activity for McbA is not applicable (n.a.) because this secondary metabolite is a DNA gyrase inhibitor, not a cytolysin. (C) Sequences and lytic activity of mutant substrates. All substrates contain the wild-type SagA leader peptide, except for the first entry (FxxxB mutant, SagA-FIA). The percent amino acid similarity to the protoxin half of SagA is shown. The second and third entries are SagA leader peptides fused to the protoxin of StaphA and ListA. SagX is an artificially designed toxin while the inverse and scrambled substrates manipulate the sequence of SagA between residues 33-50 (underlined).

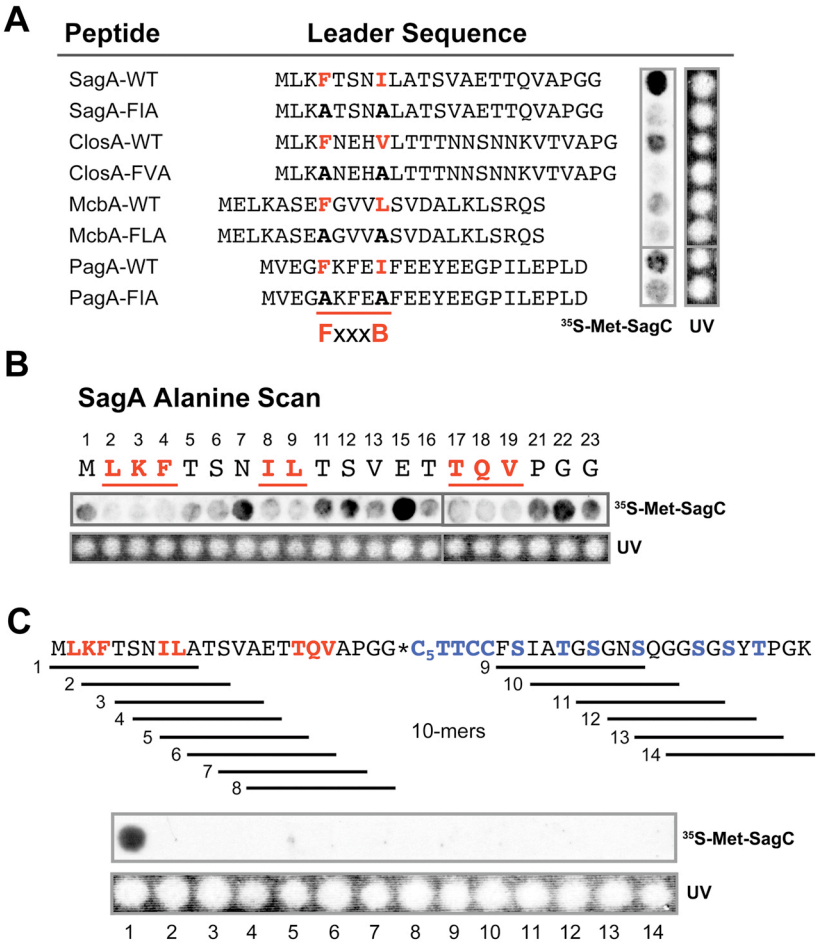
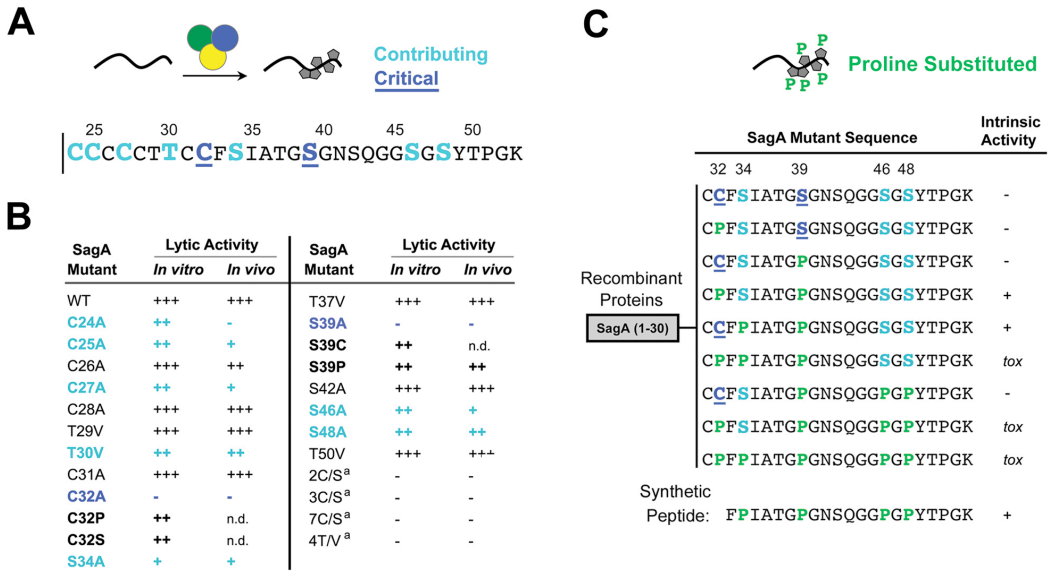


Figure 3.2 Elucidation of the SagC binding determinants

(A) Using a peptide array synthesizer, the given sequences were generated with C-terminal attachment to a cellulose membrane. The FxxxB motif was left intact (WT) and mutated to Ala for the substrates from *S. pyogenes* (SagA), *C. botulinum* (ClosA), *E. coli* (McbA), and *P. furiosus* (PagA). ^{35}S-Met-SagC was allowed to bind to the array before extensive washing and exposure to film. The relative binding of SagC to the array is indicated by the autoradiographic exposure. Equal peptide loading is confirmed by UV irradiation (254 nm). (B) The SagA leader peptide alanine scan array was synthesized to obtain positional contributions to SagC binding. The putative leader peptide of SagA contains three alanines (3/23); therefore 20 mutants of SagA were tested in this assay. The residue number, amino acid mutated to alanine, relative binding of ^{35}S-Met-SagC (autoradiography), and peptide loading control (UV) are aligned. Positions found to be important for binding are shown in red. Positions 2-4 and 8-9 contain the FxxxB motif; positions 17-19 comprise the TQV motif. (C) A peptide array comprised of fourteen 10-mer peptides of SagA was prepared to evaluate the binding contributions of each interaction site within the leader and the protoxin C-terminus. The entire length of SagA is represented, except for the middle cysteine-rich region. Relative binding and peptide loading are demonstrated as above.



SagA Mutant	Lytic Activity *In vitro*	Lytic Activity *In vivo*	SagA Mutant	Lytic Activity *In vitro*	Lytic Activity *In vivo*
WT	+++	+++	T37V	+++	+++
C24A	++	-	S39A	-	-
C25A	++	+	S39C	++	n.d.
C26A	+++	++	S39P	++	++
C27A	++	+	S42A	+++	+++
C28A	+++	+++	S46A	++	+
T29V	+++	+++	S48A	++	++
T30V	++	++	T50V	+++	+++
C31A	+++	+++	2C/S[a]	-	-
C32A	-	-	3C/S[a]	-	-
C32P	++	n.d.	7C/S[a]	-	-
C32S	++	n.d.	4T/V[a]	-	-
S34A	+	+			

Figure 3.3 Identification of SagA residues necessary for cytolytic activity
(A) Compilation of the cytolytic data for SagA point mutations after reaction with SagBCD. Residues that are C-terminal to the predicted leader peptide cleavage site (vertical line) of SagA (24-53) are shown. Positions that contribute to cytolytic activity are shown in light blue. Positions critical for cytolysis are dark blue and underlined. (B) The activity of each SagA mutant was tested by the erythrocyte lysis assay *in vitro* using purified proteins and *in vivo* by genetic complementation. Activity scoring is as in Fig. 3.1B. [a] Multiple sites were mutated. 2C/S, C31S/C32S; 3C/S, C26S/C27S/C28S; 7C/S, all seven cysteines mutated to serine; 4T/V, all four threonines mutated to valine. (C) The intrinsic activity of SagA substituted with proline was tested using purified protein. No SagBCD was used in this assay. Toxicity (*tox*) was observed for some SagA proline mutants, which prevented recombinant expression. A synthetic peptide was prepared to test proline substitution at each serine found to be important for SagA activity.

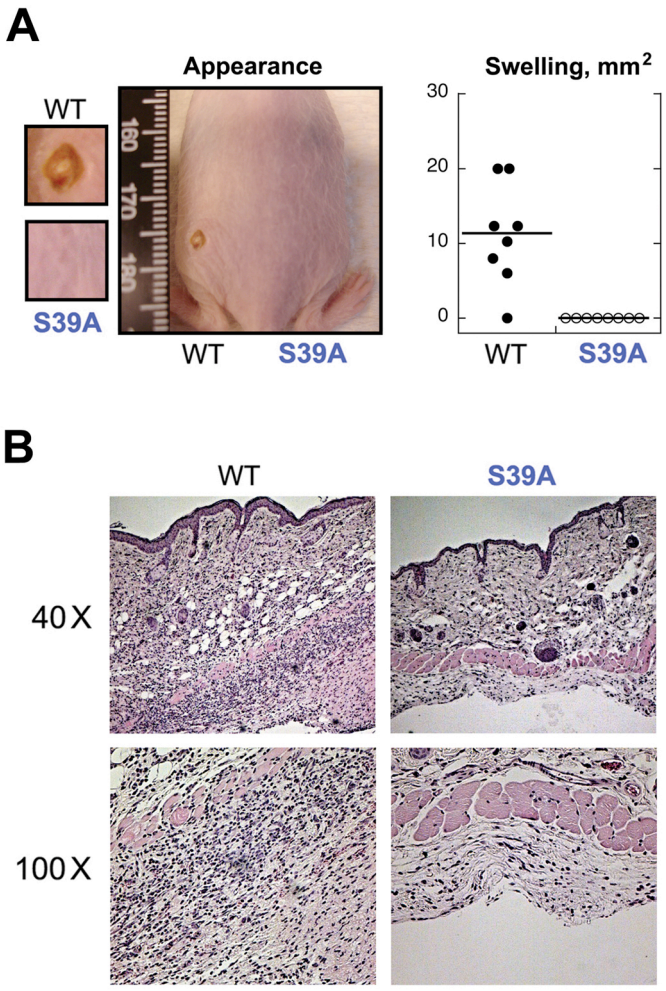


Figure 3.4 Ser39 of SagA is required for virulence in a mouse infection model
(A) Four week-old hairless crl:SKH1(hrhr)Br mice were challenged subcutaneously with 10^7 colony-forming units of *S. pyogenes* M1 Δ*sagA* complemented with *sagA*-WT (left flank) and *sagA*-S39A (right flank). Gross appearance of a typical mouse and lesion size (mm^2) are shown at 4 d post inoculation. Left insets show closer view of injection site. A necrotic ulcer is visible with *sagA*-WT. No lesions are visible with *sagA*-S39A. (B) Histological evaluation (hemotoxilyn/eosin staining) of excised tissue from site of injection. Magnification is given at 40X and 100X.

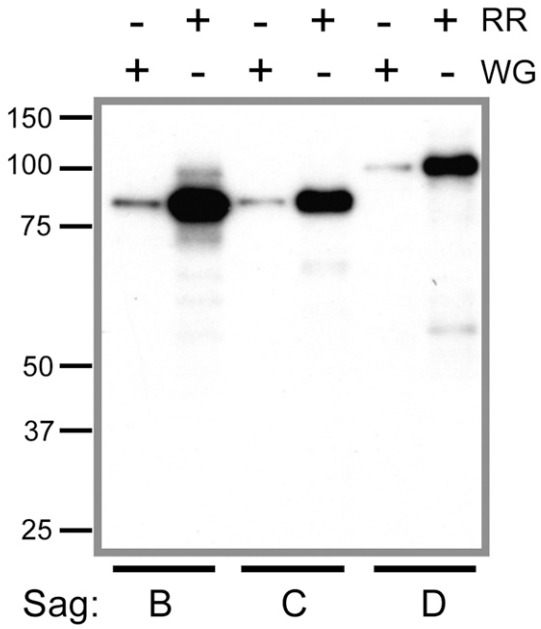


Figure 3.5 *In vitro* transcription/translation (TNT) of SagBCD
TNT reactions were carried out using maxiprepped plasmid for pET28-MBP-SagB, pET28-MBP-SagC, and pET28-MBP-SagD using kits available from Promega. Here, the results from using the extracts rabbit reticulocyte (RR) and wheat-germ (WG) are shown. S30 extract was also tested but led to formation of numerous truncation products. ^{35}S-Met was used to label these proteins during TNT and film was exposed using standard autoradiography techniques. The molecular weights, as indicated on the left side of the figure, are the expected size for each MBP-tagged protein. The protein produced from RR was used for the peptide array binding experiments.

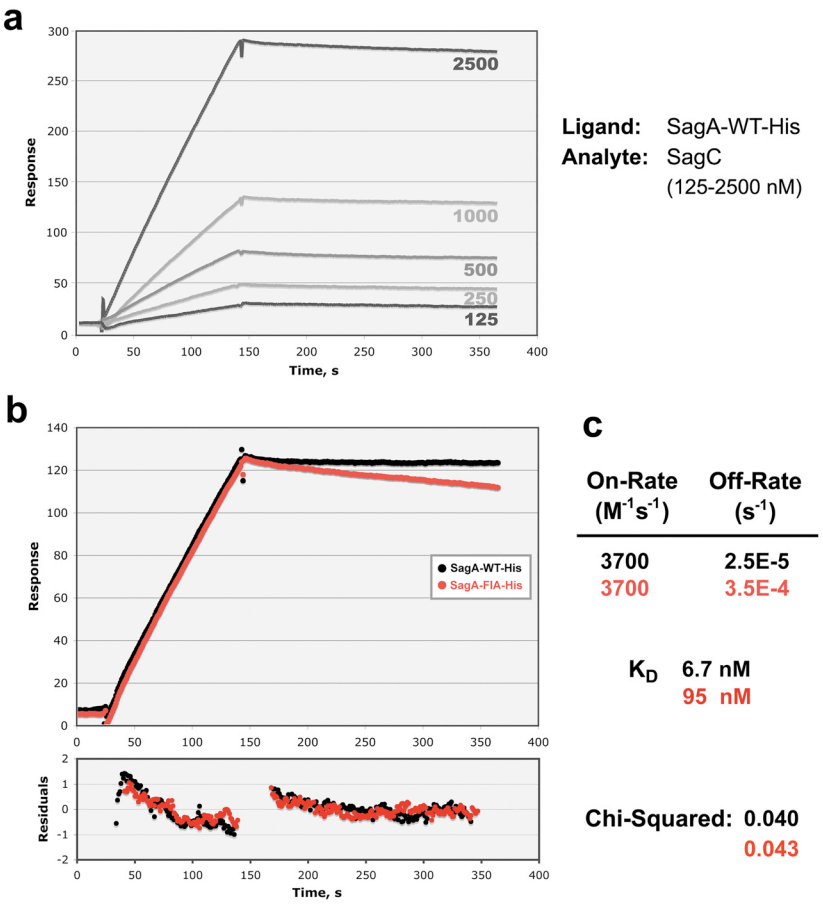


Figure 3.6 Quantification of Leader Peptide Binding to SagC

Surface plasmon resonance (Biacore T100, NTA sensor chip) was used to quantify the rate of association (on-rate) and rate of dissociation (off-rate) between various cyclodehydratases and hexahistidine-tagged SagA leader peptides. (A) Representative sensorgram for the interaction of SagA and SagC. SagA-WT-His ligand was immobilized (110 RU) on the NTA sensor chip and various concentrations of SagC analyte (125 – 2500 nM) were flowed over the surface. (B) Based on the magnitude of the response in panel a, 1000 nM SagC was chosen to evaluate the contribution of the FxxxB motif. SagA-WT-His (black) and SagA-FIA-His (red) were immobilized in separate runs, as described, and SagC binding was measured. The data were fit to a 1:1 binding model after accounting for a minor baseline drift (dissociation of Ni-NTA and the His-tagged ligand). Residuals are shown in the lower panel, along with the respective Chi-squared values, as calculated by the BiaEval software package. (C) Results of the curve fitting (BiaEval). As can been seen by visual inspection of the sensorgram, mutation of FxxxB affects only the dissociation rate (~16-fold). Values obtained with the SagA-TQV mutant and other cyclodehydratases (ClosC and McbB) are given in Table 1 of the main text.

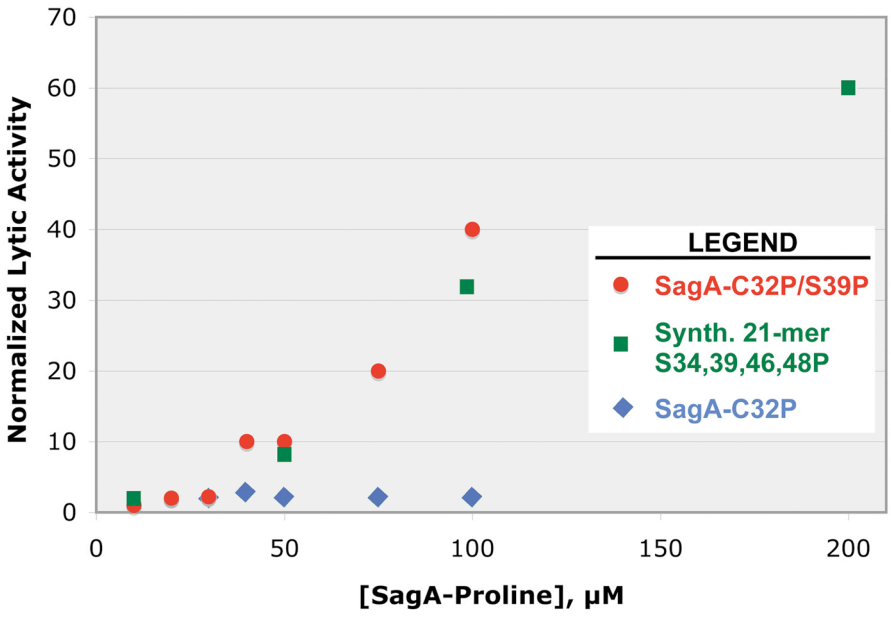


Figure 3.7 Dose-dependent activity of SagA-proline mutants
A panel of recombinant MBP-tagged SagA proline mutants was tested for intrinsic (SagBCD-independent) lytic activity (Fig. 3.3*C*). A single mutation at each critical residue (C32P and S39P) did not produce an intrinsic cytolysin. A dose-response for C32P is shown (blue diamonds). However, a double mutation (C32P/S39P) was cytolytic in the absence of SagBCD-treatment (red circles). As more sites of suspected heterocycle formation were replaced with proline, toxicity arose. A synthetic peptide comprised of the C-terminal 21 residues of SagA, with prolines replacing four serines (S34P/S39P/S46P/S48P), was also intrinsically active (green squares). No cysteines are present in this peptide.

References

1. **Bernheimer, A. W.** 1949. Formation of a bacterial toxin (streptolysin S) by resting cells. J Exp Med **90:**373-92.

2. **Bernheimer, A. W.** 1967. Physical behavior of streptolysin S. J Bacteriol **93:**2024-5.

3. **Bernheimer, A. W.** 1950. Stabilization of streptolysin S by potassium ions. J Exp Med **92:**129-32.

4. **Bernheimer, A. W., and L. L. Schwartz.** 1965. Lysis of Bacterial Protoplasts and Spheroplasts by Staphylococcal Alpha-Toxin and Streptolysin S. J Bacteriol **89:**1387-92.

5. **Betschel, S. D., S. M. Borgia, N. L. Barg, D. E. Low, and J. C. De Azavedo.** 1998. Reduced virulence of group A streptococcal Tn916 mutants that do not produce streptolysin S. Infect Immun **66:**1671-9.

6. **Cotter, P. D., L. A. Draper, E. M. Lawton, K. M. Daly, D. S. Groeger, P. G. Casey, R. P. Ross, and C. Hill.** 2008. Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I Listeria monocytogenes. PLoS Pathog **4:**e1000144.

7. **Datta, V., S. M. Myskowski, L. A. Kwinn, D. N. Chiem, N. Varki, R. G. Kansal, M. Kotb, and V. Nizet.** 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Mol Microbiol **56:**681-95.

8. **de Vos, W. M., O. P. Kuipers, J. R. van der Meer, and R. J. Siezen.** 1995. Maturation pathway of nisin and other lantibiotics: post-translationally modified antimicrobial peptides exported by gram-positive bacteria. Mol Microbiol **17:**427-37.

9. **Donia, M. S., J. Ravel, and E. W. Schmidt.** 2008. A global assembly line for cyanobactins. Nat Chem Biol **4:**341-3.

10. **Fuller, J. D., A. C. Camus, C. L. Duncan, V. Nizet, D. J. Bast, R. L. Thune, D. E. Low, and J. C. De Azavedo.** 2002. Identification of a streptolysin S-associated gene cluster and its role in the pathogenesis of Streptococcus iniae disease. Infect Immun **70:**5730-9.

11. **Ginsburg, I., Z. Bentwich, and T. N. Harris.** 1965. Oxygen-Stable Hemolysins of Group a Streptococci. 3. The Relationship of the Cell-Bound Homolysin to Streptolysin S. J Exp Med **121:**633-45.

12. **Ginsburg, I., T. N. Harris, and N. Grossowicz.** 1963. Oxygen-Stable Hemolysins of Group a Streptococci. I. The Role of Various Agents in the Production of the Hemolysins. J Exp Med **118:**905-17.

13. **Hryniewicz, W., and J. Pryjma.** 1977. Effect of streptolysin S on human and mouse T and B lymphocytes. Infect Immun **16:**730-3.

14. **Humar, D., V. Datta, D. J. Bast, B. Beall, J. C. De Azavedo, and V. Nizet.** 2002. Streptolysin S and necrotising infections produced by group G streptococcus. Lancet **359:**124-9.

15. **Igarashi, Y., Y. Kan, K. Fujii, T. Fujita, K. Harada, H. Naoki, H. Tabata, H. Onaka, and T. Furumai.** 2001. Goadsporin, a chemical substance which promotes secondary metabolism and Morphogenesis in streptomycetes. II. Structure determination. J Antibiot (Tokyo) **54:**1045-53.

16. **Kelleher, N. L., C. L. Hendrickson, and C. T. Walsh.** 1999. Posttranslational heterocyclization of cysteine and serine residues in the antibiotic microcin B17: distributivity and directionality. Biochemistry **38:**15623-30.

17. **Koyama, J.** 1963. Biochemical Studies on Streptolysin S'. Ii. Properties of a Polypeptide Component and Its Role in the Toxin Activity. J Biochem **54:**146-51.

18. **Koyama, J.** 1964. Biochemical Studies on Streptolysin S'. Iv. Properties of the Oligoribonucleotide Portion. J Biochem **56:**355-60.

19. **Koyama, J., and F. Egami.** 1963. Biochemical studies on streptolysin S' formed in the presence of yeast ribonucleic acid. I. The purification and some properties of the toxin. J Biochem **53:**147-54.

20. **Lee, S. W., D. A. Mitchell, A. L. Markley, M. E. Hensler, D. Gonzalez, A. Wohlrab, P. C. Dorrestein, V. Nizet, and J. E. Dixon.** 2008. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci U S A **105:**5879-84.

21. **Li, B., J. P. Yu, J. S. Brunzelle, G. N. Moll, W. A. van der Donk, and S. K. Nair.** 2006. Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. Science **311:**1464-7.

22. **Li, Y. M., J. C. Milne, L. L. Madison, R. Kolter, and C. T. Walsh.** 1996. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. Science **274:**1188-93.

23. **Locke, J. B., K. M. Colvin, N. Varki, M. R. Vicknair, V. Nizet, and J. T. Buchanan.** 2007. Streptococcus iniae beta-hemolysin streptolysin S is a virulence factor in fish infection. Dis Aquat Organ **76:**17-26.

24. **Madison, L. L., E. I. Vivas, Y. M. Li, C. T. Walsh, and R. Kolter.** 1997. The leader peptide is essential for the post-translational modification of the DNA-gyrase inhibitor microcin B17. Mol Microbiol **23:**161-8.

25. **Marmorek, A.** 1895. Ann. Inst. Pasteur **9:**593-620.

26. **Milne, J. C., R. S. Roy, A. C. Eliot, N. L. Kelleher, A. Wokhlu, B. Nickels, and C. T. Walsh.** 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:**4768-81.

27. **Nizet, V.** 2002. Streptococcal beta-hemolysins: genetics and role in disease pathogenesis. Trends Microbiol **10:**575-80.

28. **Nizet, V., B. Beall, D. J. Bast, V. Datta, L. Kilburn, D. E. Low, and J. C. De Azavedo.** 2000. Genetic locus for streptolysin S production by group A streptococcus. Infect Immun **68:**4245-54.

29. **Ofek, I., S. Bergner-Rabinowitz, and I. Ginsburg.** 1972. Oxygen-stable hemolysins of group A streptococci. 8. Leukotoxic and antiphagocytic effects of streptolysins S and O. Infect Immun **6:**459-64.

30. **Okeley, N. M., M. Paul, J. P. Stasser, N. Blackburn, and W. A. van der Donk.** 2003. SpaC and NisC, the cyclases involved in subtilin and nisin biosynthesis, are zinc proteins. Biochemistry **42:**13613-24.

31. **Onaka, H., H. Tabata, Y. Igarashi, Y. Sato, and T. Furumai.** 2001. Goadsporin, a chemical substance which promotes secondary metabolism and morphogenesis in streptomycetes. I. Purification and characterization. J Antibiot (Tokyo) **54:**1036-44.

32. **Robinson, J.** 1951. The nonantigenicity of streptolysin S. J Immunol **66:**661-5.

33. **Roy, R. S., O. Allen, and C. T. Walsh.** 1999. Expressed protein ligation to probe regiospecificity of heterocyclization in the peptide antibiotic microcin B17. Chem Biol **6:**789-99.

34. **Roy, R. S., A. M. Gehring, J. C. Milne, P. J. Belshaw, and C. T. Walsh.** 1999. Thiazole and oxazole peptides: biosynthesis and molecular machinery. Nat Prod Rep **16:**249-63.

35. **Roy, R. S., S. Kim, J. D. Baleja, and C. T. Walsh.** 1998. Role of the microcin B17 propeptide in substrate recognition: solution structure and mutational analysis of McbA1-26. Chem Biol **5:**217-28.

36. **Schmidt, E. W., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood, and J. Ravel.** 2005. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proc Natl Acad Sci U S A **102:**7315-20.

37. **Williams, A. B., and R. S. Jacobs.** 1993. A marine natural product, patellamide D, reverses multidrug resistance in a human leukemic cell line. Cancer Lett **71:**97-102.

38. **Yorgey, P., J. Davagnino, and R. Kolter.** 1993. The maturation pathway of microcin B17, a peptide inhibitor of DNA gyrase. Mol Microbiol **9:**897-905.

39. **Yorgey, P., J. Lee, J. Kordel, E. Vivas, P. Warner, D. Jebaratnam, and R. Kolter.** 1994. Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. Proc Natl Acad Sci U S A **91:**4519-23.

Chapter IV

Clostridiolysin S, a Post-translationally Modified Biotoxin from *Clostridium Botulinum*

Introduction

Microbial virulence and survival are often defined by metabolic output. Among the many molecular species used to give a competitive advantage are hydrogen peroxide, genetically encoded small molecules, siderophores, proteins, and bacteriocins (30) (26). Streptolysin S (SLS) is a well known hemolytic/cytolytic, ribosomally encoded bacteriocin and virulence factor group A *Streptococcus* (GAS) (4, 8, 18). To this day, the characteristic β-hemolytic phenotype observed on blood agar plates is used as a clinical diagnostic tool for GAS identification. GAS is best known as the agent of acute pharyngitis (strep throat) but also may cause invasive infections, including necrotizing fasciitis and toxic shock syndrome. In the 100-year history of our knowledge of streptococcal bacteria, the precise chemical structure of this toxin has remained elusive, although the discovery of its biosynthetic gene cluster more than a decade ago (18) has guided investigations into its post-translational modification and likely heterocyclic nature (12, 16).

Through recent comparative genomic analysis, an SLS-type gene cluster was identified in clostridia species including *Clostridium botulinum* and *Clostridium sporogenes*, two disease-causing bacteria known to endanger food supplies (6, 23, 25). Similar gene clusters were found in *Staphylococcus aureus* RF122, *Bacillus thuringiensis*, *Streptococcus iniae*, and *Listeria monocytogenes* 4b. The *L. monocytogenes* 4b strain is the primary serotype and causative agent for outbreaks of listeriosis (3). The *S. aureus* RF122 strain is responsible for bovine mastitis. *S. iniae* is

a cytotoxic fish pathogen. This suggests that a shared metabolic output of these pathogens including SLS-like toxins may contribute to their pathogenic potential (3).

Each of these SLS family gene clusters contains a similar set of genes. For instance, in *C. botulinum* and *C. sporogenes*, the *closA–I* genes are related by sequence to sagA–I from GAS (see Fig. 4.1A). Of these genes, ClosF is a protein of unknown function. ClosG, -H, and -I are ABC transporters and therefore are likely to be responsible for exporting the mature hemolytic product. ClosA is the prepropeptide (structural peptide) that is post-translationally modified to form the propeptide. ClosE has been annotated as an immunity protein but is similar to the Caa*X* protease superfamily in sequence. Thus, ClosE may be the enzyme responsible for proteolytic processing of the post-translationally modified propeptide to the mature biotoxin. ClosBCD resembles the McbBCD-modifying proteins found in the *Escherichia coli* microcin B17 system and the *Prochloron didemni patDG* genes of the patellamide biosynthetic gene cluster. This family of natural product toxins was recently shown by several groups to include the thiopeptides, which are nearly ubiquitous in soil-dwelling *Bacillus* and *Streptomyces* bacteria (11, 15, 17, 24, 29). In the microcin B17 system, McbB, -C, and -D were shown to be necessary and sufficient for the installation of heterocyclic moieties on C-terminal serine and cysteine residues of the McbA substrate. Chemical transformations installed on the McbA substrate were assayed by mass spectrometry and were shown to result in the loss of multiples of 20 Da (see Fig. 4.3A). Based on the comparative genomics with the above-mentioned clusters, ClosBCD also are likely involved in the post-translational conversion of serine, threonine, and cysteine residues of ClosA to their corresponding thiazole and

methyloxazole heterocycles to generate a modified propeptide. BLAST analysis indicates that ClosB is similar to a FMN-dependent oxidase, ClosC is similar to the ATP-dependent enzymes E1, ThiF, and MoeB, which contain a structural zinc that is not involved in catalysis (5, 10). The E1/ThiF/MoeB protein family utilizes ATP to activate the C-terminal end of ubiquitin or an ubiquitin-like protein. ClosD belongs to a biochemically uncharacterized protein family YcaO (DUF181).

In this paper, we characterize the SLS-type toxin from *C. botulinum* and *C. sporogenes*. We demonstrate that the gene clusters in clostridia are functionally equivalent to the SLS biosynthetic pathway by complementing targeted GAS SLS-operon knock-outs with the corresponding *Clostridium* genes. Once the genome became available (10), plasmid integrational mutants of the *closA* and *closC* gene products also were generated in *C. sporogenes* to verify that the clostridiolysin S gene cluster was indeed responsible for a hemolytic phenotype. Hemolytic activity of this gene cluster was reconstituted *in vitro*, allowing the direct mass spectral confirmation that the clostridial toxin is post-translationally modified to contain heterocyclic moieties and that this process required structural zinc bound to the putative cyclodehydratase, ClosC. Finally, we demonstrate that the synthetase enzymes install heterocyles on ClosA using nanocapillary LC-MS/MS in conjunction with the mass spectral data analysis programs InSpecT (27) and Spectral Networks (1), to establish that the nontoxic precursor peptide (ClosA) is converted into the final post-translationally modified toxin, clostridiolysin S.

Results

Comparative Bioinformatics of Clos and Sag Locus

Previously, we and others (12) (25) (24) have identified a large, conserved family of bacteriocin gene clusters present across six phyla of microorganisms. These were identified by similarity to the microcin B17 (14, 31, 32) biosynthetic gene cluster and include the important human pathogens GAS, *L. monocytogenes*, *S. aureus*, and *C. botulinum*. Comparison of the gene clusters between the SLS biosynthetic operon in GAS and its ortholog in *C. botulinum*, displayed high levels of similarity both in gene content and organization (Fig. 4.1A). The SLS gene cluster contains the gene encoding the SLS precursor peptide, designated SagA, as well as genes encoding the SLS thiazole/oxazole synthetase (SagB–D), a putative immunity protein (SagE), a protein of unknown function (SagF), and three proteins containing homology to ATP-dependent binding cassette (ABC) transporters (SagG–I). All of these genes are present in the gene cluster of *C. botulinum* (ATCC 3502) in identical gene order and orientation to the SLS gene cluster, which we hereby refer to as the clostridiolysin S gene cluster.

ClosA–D Genetic Complementation Studies in Group A Streptococcus

To determine whether the clostridiolysin S genes are biochemically equivalent to the *sag* gene cluster in GAS, genetic complementation studies were performed using the *sagA–D* deletion strains of GAS NZ131 (M49 serotype) (Δ*sagA*, Δ*sagB*, Δ*sagC*, and Δ*sagD*) (4). We transformed these strains with a plasmid containing the

corresponding wild-type copies of the *clos* genes (*closA*, *closB*, *closC*, and *closD*), to determine the extent of genetic tolerance between GAS and *C. botulinum.* The ability of the complemented strains to lyse erythrocytes was quantified by measuring the amount of hemoglobin released into the supernatant by measuring the absorbance at 450 nm (Fig. 4.1B). These were compared with the hemolytic activity of wild-type GAS, Δ*sagA* complemented with pDCerm-*sagA*, and Δ*sagA–D* complemented with the pDCerm plasmid alone. The closA, closC, and closD complemented strains were able to lyse erythrocytes, whereas complementation of the GAS Δ*sagB* mutant with a plasmid containing the *closB* gene was not sufficient to produce a hemolytic phenotype, even though ClosB is 60% similar to SagB. Of the genes that were able to reconstitute the hemolytic phenotype, *closC* partially complemented, whereas the *closA* and *closD* complemented strains restored, near wild-type levels of hemolysis. This experiment produced the initial indication that the orphan clostridium gene cluster indeed is functionally equivalent to the SLS cluster. During our investigations into the function of the SLS-like gene cluster from *C. botulinum*, the genome sequence of *C. sporogenes* became available (25). *C. sporogenes*, an opportunistic foodborne disease-causing bacteria, also contains the SLS-like gene cluster. In *C. sporogenes*, this gene cluster is 85–98% identical to the gene cluster found in *C. botulinum*. Because *C. sporogenes* is a BSL1 level organism, it was possible to generate integrated plasmid mutants of this gene cluster within our laboratories, enabling the direct assessment of the involvement of this gene cluster in the generation of an SLS-type toxin. Indeed, when *closA* and *closC* were inactivated, *C. sporogenes* no longer retained its hemolytic activity (Fig. 4.1C). Results of both the complementation of

GAS with the *C. botulinum* genes and the targeted mutagenesis in *C. sporogenes* confirm that the SLS-like gene clusters from clostridia are involved in the biogenesis of a hemolysin.

In Vitro Reconstitution of Clostridiolysin S Activity

Because the above genetic complementation experiments indicated that the *C. botulinum* gene cluster is functionally equivalent to the SLS gene cluster, we attempted to reconstitute hemolytic activity *in vitro*. To determine whether the precursor protein from *C. botulinum* ClosA could be transformed into an active cytolysin in a synthetase-dependent manner, ClosA was produced as an MBP-fusion protein for solubility and then subjected to post-translational modification by the addition of ClosBCD. Incubating ClosA with purified ClosBCD did not produce a lytic product (data not shown). The individual enzymes were analyzed and ClosB, the flavin mononucleotide (FMN)-dependent oxidoreductase, did not exhibit a yellow color upon purification. Several attempts to purify an active ClosB (yellow) failed under different affinity purification systems. The anticipated yellow color is indicative of FMN-binding proteins and adding FMN to the assay also failed to reconstitute the hemolytic activity. We surmised that the ClosB, as purified, was in an inactive apoprotein and hypothesized that this enzyme may undergo an additional processing event that could not be reconstituted in an *E. coli* system. Based on genetic complementation studies and insertional mutants, ClosB is active *in vivo*. We hypothesize that *in vitro*, despite an overall similarly score of 60%, SagB and ClosB

diverge enough that one is purified as an active holoprotein, whereas the other is purified in an inactive apoform.

Previous studies showed ClosA was converted readily to an active hemolytic toxin *in vitro* by the SagBCD-modifying enzymes (12). SagB is 60% similar to ClosB, and additionally, recombinant expression of SagB yielded holoprotein, as indicated by a bright yellow color (12). Therefore, to overcome the inability to purify holo-ClosB, we substituted the ClosB ortholog SagB into the reaction. When ClosA was incubated with SagB, ClosC, and ClosD, robust hemolysis was observed. Omitting ClosA, SagB, ClosC, and ClosD, in any combination, resulted in the abolishment of hemolysis, demonstrating the *in vitro* reconstitution of a hemolytic SLS-like biotoxin from *C. botulinum*. To gauge the relative *in vitro* hemolytic activity of the CLS system, reconstitution of SLS activity was assayed in parallel. Based on *in vitro* hemolysis, CLS and SLS have comparable bioactivities (Fig. 4.2A). This observation is supported by the genetic complementation studies as the *closD* and *closA* genes produced a natural product with potency at near wild-type levels. To provide additional confirmation of the *in vitro*-reconstituted reaction, we set out to characterize ClosC in further detail..

Sequence alignments of McbB, E1, MoeB, ThiF, and MccB (Fig. 4.3B), suggest that ClosC is related to the ThiF/MoeB/E1 superfamily of proteins (13, 21, 22). MoeB, ThiF, and E1 are proteins that adenylate the C-terminus of ubiquitin-like proteins, whereas MccB is responsible for activating the C-terminal glutamine end of microcin C7. In clostriodiolysin S, an amide carbonyl undergoes activation, as is the

case for microcin B17, patellamide, goadsporin, and the recently described thiopeptide family (9, 15, 29). However, sequence alignments of ClosC with the E1-type family of proteins, displays that ClosC contains two conserved C*xx*C motifs that typically bind Zn^{2+} (Fig. 4.3B). In the microcin system, Zn^{2+} was originally proposed to serve as a Lewis acid, activating the amide carbonyl for cyclodehydration (31). However, that mechanism was later amended to have a general acid attack the carbonyl to form a leaving group for the final elimination reaction (19). The crystal structures of ThiF, MoeB, and MccB support this second mechanism, showing a tertracysteine coordinated Zn^{2+} 20 Å away from the active site pocket where ATP binds (13, 21). In addition to the C*xx*C motifs, the crystal structures of ThiF and MoeB resolved the amino acid residues critical for ATP binding. Of these residues, 43% are conserved in ClosC, indicating that it requires ATP for catalysis, which is in agreement with the requirement of ATP in the *in vitro* reconstitution assay (12). Based on the structural arguments, we predict that the tetracysteine coordinated Zn^{2+} in ClosC plays a structural role rather than a catalytic role. We therefore anticipated that the addition of EDTA could disrupt the structure of ClosC. The addition of EDTA did not affect the structure of ClosC as judged by CD nor did it kill the hemolytic activity (Fig. 4.6), similar to the observations by the homolog McbB in the microcin B17 pathway (31). Therefore, we determined the effect of the cysteine mutations on the Zn^{2+} content, *in vitro* reconstitution, and structural integrity of ClosC.

Point mutations in the C*xx*C motifs resulted in the abolishment of hemolytic activity and minimal levels of Zn^{2+} binding compared with wild-type ClosC (Fig. 4.2B

and Fig. 4.7). CD measurements indicated that wild-type MBP-ClosC had two distinct negative minima at 215 and 225 nm, indicative of α-helical content. Based on the molar ellipticity at 222 nm, the helical content of MBP-ClosC was determined to be 49%, in close agreement with the predicted content of 38–42%. The ClosC C*xx*C mutants, lacking the ability to bind Zn^{2+}, showed the loss of the negative minima at 215 and 225 nm (Fig. 4.2C). The resulting CD spectra obtained by the C*xx*C mutants, indicating the striking loss of α-helical integrity, were unexpected results because the dramatic loss of α-helical integrity was caused by a single amino acid point mutation on the ClosC portion of the MBP-ClosC construct. The amylose resin purification of the MBP-ClosC C*xx*C mutant proteins was verified by SDS-PAGE (Fig. 4.8). Experiments were performed to rescue hemolytic activity by zinc titration of the C*xx*C proteins. Addition of zinc did not restore a hemolytic phenotype. Zinc titration experiments of the C*xx*C proteins were also assayed by CD. The CD showed a minor restoration of α-helical content, but the extent was not sufficient for hemolytic reconstitution (Fig. 4.9). These findings indicate that the tetracysteine coordination of Zn^{2+} is required for a properly folded ClosC, which can function *in vitro* and further solidify the *in vitro* reconstitution of a hemolytic toxin.

Detection of Heterocyclized ClosA Peptides via Off-line Separation and LC-MS/MS Methods

From the first reports of SLS activity in the early 1900s, structural characterization attempts have failed, and all attempts to characterize other hemolytic biotoxins from *Listeria*, *Streptococci*, or *Clostridia* have met the same fate (2, 3, 12,

16, 28). Therefore, any direct structural insight into a SLS-type toxin would be a major advancement. For the past three years, we also were unsuccessful when attempting to analyze this toxin by traditional approaches. Using intact protein mass spectrometry, observation of the ClosA protein by Fourier transform-ion cyclotron mass spectrometry was successful. However, upon addition of the synthetase enzymes (SagB, ClosC, and ClosD), the ClosA protein only could be observed as a heterogeneous species of overlapping ions in the Fourier transform-ion cyclotron mass spectrum, which provided little insight into the modified structure (Fig. 4.10). One of the limitations of intact protein mass spectrometry, especially when there are many forms of one protein, is that the signal is dispersed over a large number of different and overlapping masses, which are difficult, if not impossible, to characterize. Tandem mass spectrometry also consistently failed to identify the toxin portion of this protein. For other proteins of this size or larger such as BSA, GFP, a freestanding AT domain on the Group B *Streptococcus* hemolytic pathway, it was possible to observe the intact proteins with relative ease, indicating that our problem most likely was due to the troublesome biophysical properties of SLS-like biotoxins. In the case of clostridiolysin S, the overlap of heterologous isotopic distributions hindered the assignment of an accurate mass, rendering the top-down data uninterpretable. Therefore, we set out to capture the modifications on ClosA by a bottom-up mass spectrometry strategy.

Traditional proteomic approaches, such as iodoacetamide alkylation, trypsin digestions, and capillary LC-MS/MS analysis also failed to recover any of the peptides covering the C-terminal end of the protoxin. We assumed that the modified peptides

(ClosA and SagA) were hydrophobic, and therefore, we applied approaches usually reserved for membrane proteins, such as dissolving in 75% formic or acetic acid. This too failed to recover any of these peptides. Because ClosA and SagA have an array of contiguous cysteines within the primary sequence (CCCCSCCCC), another approach was taken based on some classic enzymological studies carried out on acetoacetate decarboxylase (7). This particular enzyme decarboxylates acetoacetate via the formation of an imine with an active site lysine. One approach used to determine that the lysine was critical for the activity of acetoacetate decarboxylase was to replace it with a cysteine. This rendered the protein inactive. However, upon 2-bromoethylamine treatment, the cysteine is converted to *S*-aminoethyl cysteine, an unnatural side chain with sterics and electronics complimentary to lysine. To observe heterocycles on ClosA and SagA, alkylation of the cysteines with BrEA was performed. BrEA alkylation has three significant advantages: it 1) improves water solubility; 2) improves the ionization by introducing an amine, which is positively charged under our conditions; and 3) enables tryptic digestion after the modified cysteine, generating a ladder of peptides. Because the SLS proteins were available in large quantities, we first optimized this protocol using the *in vitro* reconstitution hemolytic activity of SagA with SagBCD. The *in vitro*-reconstituted sample was alkylated with 2-bromoethylamine, then trypsin-digested, and analyzed by nanocapillary LC-MS/MS. The resulting data were subjected to the tandem mass spectral processing algorithms, InSpecT, and Spectral Networks (1, 27). Examination of the reconstituted system, consisting of SagA and SagBCD, uncovered a peptide containing two oxazole moieties observed at positions Ser^{46} and Ser^{48} (see Fig. 4.5B, peptide 2), which were

absent in the control samples that lacked SagBCD or a reaction mixture that excluded one of the three synthetase proteins. Indeed, as demonstrated in previous mutational analyses in the SLS system, these identified sites of heterocyclic residues contribute to the hemolytic properties of the toxin *in vitro* (16).

Similarly, the *in vitro*-reconstituted system containing ClosA, digested, and 2-bromoethylamine-alkylated, was subjected to nanocapillary LC-MS/MS analysis on an LTQ instrument. Spectral Networks processed 576,339 total spectra and identified 133,435 clusters. Of the identified clusters, 42 were observed in the protoxin-encoding, C-terminal half of ClosA (CCVSVS region; the N-terminus is a leader peptide). Using this approach, Spectral Networks and InSpecT identified several reoccurring peptides as containing −20 Da losses that were absent in the control assays. The identification of −20 Da species in both the Sag and Clos *in vitro*-reconstituted samples represents the first direct information that SLS-type protoxin substrates accept the installation of heterocycles by the synthetase enzymes (Fig. 4.4). In the clostridiolysin S-reconstituted system, peptides with modifications at Thr^{11}, Thr^{12}, and Thr^{46} (Fig. 4.5B, peptide 1; Figs. 4.11 and 4.12) were identified as a heterogeneous mixture, in agreement with top-down mass spectrometry data. Although the data do not provide a complete structure, the post-translationally modified peptides confirm that the clostridiolysin S precursor undergoes extensive post-translational modification *in vitro* and contains a series of methyloxazole motifs. Because a heterocycle was identified at Thr^{46} (Fig. 4.5B, peptide 1) within ClosA, which follows the classical paradigm in the microcin B17 system of C-terminal amino

acid residue heterocyclization, a mutant bearing an individual point mutation was generated to assess the importance of heterocyclic conversion at this residue on the overall cytolytic properties of clostridiolysin S.

In vitro synthetase reactions using the purified ClosA point mutant (T46A) and the synthetase enzymes were assessed for hemolytic activity. Substitution of an alanine at Thr^{46} had a dramatic effect on the *in vitro* hemolytic phenotype (Fig. 4.5A). In a similar fashion to *in vitro* reconstitution experiments in microcin B17 system, where an additional heterocycle was produced relative to the known structure (32), we believe heterocycles at positions Thr^{11} and Thr^{12} on the leader peptide are results of excessive *in vitro* processing because they are found on the proposed leader peptide of the protoxin (20).

Discussion

Based on the striking genetic and functional similarities with the key GAS virulence factor SLS, we demonstrate that two clostridial species produce a related hemolytic factor. This article has provided evidence that pathogenic bacteria utilize synthetase enzymes to produce chemical changes in ribosomally encoded precursors to biosynthesize biologically active toxins, which gain activity from the installation of heterocyclic moieties. Given the functional similarities and equivalent potencies, it is likely that the metabolic output of clostridiolysin S by *C. botulinum* can play a role in host tissue injury and contribute to clostridial virulence. While recognizing its broader classification as a bacteriocin-like peptide, it also could be fruitful to explore whether

clostridiolysin S may aid *C. sporogenes*, a known component of the human gut microbiome, to establish a competitive niche within the vast number of microbes comprising the gut microflora. The mass spectrometry approach outlined herein is the first step toward the characterization of this class of hemolytic/cytolytic toxins. Preliminary results on the structural characterization of the SLS biosynthetic pathway have further solidified the described mass spectral method to be significant and reproducible for this family of toxins. Our findings on the structure/function of the ClosC subunit are likely to extend to other cyclodehydratases from this family of bacteriocin synthetase enzyme complexes. Because targeting virulence factors is becoming an attractive anti-infective strategy, selective inhibition of this family of enzymes may well become an attractive therapeutic target or could be used as an approach to prevent future food spoilage.

Experimental Procedures

Plasmid Integrational Mutagenesis of closA and closC

PCR was used to amplify the central regions of C. sporogenes closA and closC genes. PCR products were recovered by T-A cloning in the vector pCR2.1-TOPO (Invitrogen) and then subcloned into the temperature-sensitive erythromycin (Erm^R) plasmid, pHY304. The knock-out plasmids were introduced by electroporation into C. sporogenes, and transformants were recovered on THA-Erm 500 μg/ml. Single recombination events were identified by shifting to the nonpermissive temperature (37 °C) while maintaining Erm selection and were later confirmed by PCR. Intragenic

plasmid integrational mutants were streaked onto blood agar plates (5% sheep blood, Hardy Diagnostics) to assess hemolytic activity.

Cloning and Protein Purification of closA, closB, closC, closD, and sagB

Single synthetic gene copies of closA–D subcloned into pGS-21a KpnI/HindIII restriction enzyme sites were purchased from GenScript Corp. The pGS-21a vectors harboring the closA–D genes contain a GST tag and His_6 tag at the N-terminus of each clos gene for affinity purification purposes, followed by an enterokinase cleavage site. Using the closA–D/pGS-21a vectors as templates, closA–D were PCR-amplified with BamHI/NotI flanking sequences. The closA–D PCR amplicons were verified by agarose (1%) gel electrophoresis. Positively identified closA–D amplicons were then excised and purified using the Qiagen QIAquick PCR purification kit. The purified amplicons and the pET28 vector containing an N-terminal maltose-binding protein (MBP) tag were then sequentially digested by allowing the NotI digest to proceed overnight followed by a 2-h BamHI digestion at 37 °C (New England Biolabs). Following the BamHI/NotI digest of the pET28-MBP-modified vector, a calf intestine phosphatase reaction was performed (New England Biolabs). The digested amplicons and vector were then separated by agarose gel electrophoresis, excised, and purified with a Qiaquick gel extraction kit. Purified closA–D amplicons were then ligated into the similarly digested vector using a rapid ligation kit (Roche Applied Science). Ligation products of MBP-closA–D were then transformed into one-shot Top10 DH5a chemically competent cells (Invitrogen) and plated on Kan^{50} Luria Broth (LB) plates. Single colonies of each transformant were picked and grown overnight in preparation

for plasmid recovery. The recovered closA–D plasmids were then verified by PCR amplification and sequenced by Eton Biosciences (La Jolla, CA). The validated plasmids were then transformed into Invitrogen E. coli BL21(DE3) cells by heat shock and recovered in super-optimal broth with catabolic repression (SOC) medium for 1 h (Invitrogen). Cells were harvested by centrifugation at 13,000 rpm for 1 min and then plated on Kan^{50} LB agar plates at 37 °C overnight. Single colonies were picked and verified for the presence of the closA–D genes by colony PCR. Colonies that contained the closA–D genes were grown and then stored at −80 °C until used in the hemolytic/cytolyic assays. Preparation of the MBP-SagB inducible construct was described previously (12).

Protein Purification of MBP-ClosA–D and MBP-SagA–D

All MBP-tagged proteins used in the in vitro hemolytic assays were purified as described previously (12). Upon purification, all proteins were immediately stored at −80 °C until further use.

***In Vitro* Synthetase Reactions to Test Cytolytic Activity**

Protein preparation and synthetase reactions employing MBP-tagged substrate ClosA and the BCD synthetases were performed as described previously for the in vitro reconstitution of SLS activity (12). In every case, omission of the substrate or the synthetase resulted in no detectable hemolytic activity. All assays were internally normalized and baseline-adjusted using two positive controls (Triton X-100 and wild-

type SagA treated with SagBCD) and two negative controls (substrate and SagBCD alone).

Preparation of GAS *ΔsagA–D* Strains complemented with *closA–D*

The construction of GAS ΔsagA, ΔsagB, ΔsagC, and ΔsagD allelic exchange mutants in GAS strain NZ131 (M49 serotype) was described previously (4). To introduce the individual closA–D genes in trans to the corresponding sagA–D allelic exchange mutants, closA–D were individually PCR-amplified with XbaI/BamHI flanking sequences from pGS-21a closA–D gene containing vectors (GenScript). Each PCR amplicon was verified on a 1.0% gel, and the positively identified closA–D amplicons were then excised and purified using a Promega PCR purification wizard kit. Sequential restriction enzyme digests, BglII/XbaI for the pDCerm plasmid (4) and BamHI/XbaI for the closA–D amplicons followed (New England Biolabs). Restriction enzyme products were separated on a 1.0% agarose gel, excised, and purified by the use of a Promega gel extraction wizard kit. Ligation was performed by the use of a 3:1 insert to vector ratio with a ligation kit (Roche Applied Science). The resulting ligation product was then transformed into E. coli MC1061 competent cells via electroporation and then plated onto $LAerm^{500}$ plates. Single colonies of each transformant were then picked and grown overnight in preparation for plasmid recovery. The recovered pDCerm-closA–D plasmid was then verified by PCR amplification and sequenced by Eton Biosciences. The validated pDCerm-closA–D plasmids were then transformed into the corresponding GAS ΔsagA–D strains via electroporation and plated on $THBerm^{5}$ plates (Todd Hewitt Broth). Single colonies

were picked and verified for the presence of the closA–D genes by colony PCR. Complementation of the constructed GAS ΔsagA–D harvesting pDCerm-closA–D was assessed by quantification of SLS-like hemolytic activity.

Growth and Quantification of GAS *ΔsagA–D* + *closA–D* pDCerm Strains

Extracts containing BSA-stabilized SLS or SLS-like products were prepared in the following manner. Overnight cultures (10 ml) of Streptococcus pyogenes M1 mutants containing the appropriate pDCerm plasmids for complementation were grown to A_{600} ~ 0.6 in THB containing 2 μg/ml erythromycin. Cultures were treated with BSA (10 mg/ml) for 1 h at 37 °C and then centrifuged (6000 × g for 10 min) before passing the supernatant through a 0.2-μm acrodisc syringe filter (Pall Corp.). These samples were centrifuged again (6000 × g for 10 min), and the supernatants were assayed for hemolytic activity by addition to defribinated sheep blood (in V-bottom microtiter plates at 1:25 and 1:50 dilutions). The blood was treated for 2–4 h before assessing hemolytic activity as reported previously (12). Genetic complementation was assayed in both the Dixon and Dorrestein laboratories on multiple occasions and resulted in similar, consistent results.

Site-directed Mutagenesis of ClosC

Preparation of ClosC CxxC Mutants

A Stratagene QuikChange kit was used to create the closC CxxC mutants using pET28b-MBP closC plasmid as a template for PCR. The PCR product was then mixed

with 1 μl of DpnI (Stratagene) and allowed to digest at 37 °C for 1 h. Following digestion of the parental template, 1 μl of each PCR product was transformed into Invitrogen One-shot DH5α chemically competent cells by heat shock at 37 °C. Following the transformation, the DH5α cells were SOC recovered at 37 °C for 1.5 h then plated on Kan^{50} plates and incubated overnight at 37 °C. Single colonies were picked and inoculated into 8 ml of LB Kan^{50} for 16 h. The plasmid was recovered using a Qiaquick miniprep kit. Each CxxC plasmid was then sequenced by Eton Biosciences (La Jolla, CA). Once the sequences were correctly verified, 20 ng of plasmid for each CxxC mutant was transformed into BL21 competent cells (Invitrogen).

Protein Purification of CxxC Mutants

All ClosC CxxC proteins were MBP-tagged and purified as previously described (12). Upon purification, all proteins were immediately stored at −80 °C until needed.

ICP-MS Analysis of wt ClosC and ClosC CxxC Mutants

30 μm of ClosC, ClosC_C242A, ClosC_C245A, ClosC_C332A, ClosC_C335A, and ClosC_C242A_C245A were prepared and sent to Ted Huston (University of Michigan Department of Geological Sciences, W. M. Keck Elemental Geochemistry Laboratory) for metal content analysis. ICP-MS experiments on MBP-ClosC and CxxC mutants were performed in duplicate with two different protein preparations.

Circular Dichroism of ClosC CxxC Mutants

Circular dichroism measurements were performed on a Jasco J-815 spectropolarimeter. The temperature in the sample container was maintained using a Jasco PFD temperature stabilizer. Freshly purified MBP-ClosC and MBP-ClosC CxxC mutants were buffer-exchanged into 10 mm phosphate buffer using equilibrated PD-10 columns (GE Healthcare). Thereafter, MBP-ClosC and all MBP-CxxC mutants were prepared to a final protein concentration of 1.2 mg/ml. Before running the samples on a CD spectrometer, the machine was flushed with nitrogen gas for 1 h. Once the CD spectrometer was equilibrated, each sample was scanned 10× within a wavelength range of 185–300 nm. Data were processed by applying baseline subtraction, deconvolution, and noise reduction using Spectra Manager® (version 2). A xenon lamp was used as the light source. Sensitivity of the machine was set at 100 millidegrees. The scanning method was set to step fashion with a response time of 2 s. Bandwidth was set at 1, and two open channels, CD and dynode voltage, were operated. CD data were verified by repeating each experiment with different protein preparations, which resulted in consistent data. The CD results were expressed as a molar ellipticity ([θ]) in degrees cm^2 $dmol^{-1}$, as defined: $[\theta] = \theta_{obs}/10nlC_p$, where θ_{obs} is the CD in millidegrees, n is the number of amino acids, l is the path length of the cell (cm), and C_p is the mole fraction. The [θ] value obtained at θ_{222} was used to calculate the helical content using the following equation: $\% \text{ helix} = ([\theta]_{222} - 4000)/(33{,}000 - 4000) \times 100$. The following secondary structure prediction programs were used to

obtain a theoretical value of helical content for comparative purposes: PSIPRED, SOPMA, and nnPredict.

Detection of Heterocyclized ClosA and SagA Peptides via Bottom-up Mass Spectrometry Analysis

Sample Preparation

In vitro ClosA and synthetase samples, with the ability to lyse red blood cells, first were chemically derivatized by the addition of 50 mm 2-bromoethylamine (BrEA) in 200 mm Tris buffer (pH 8.8). The BrEA-lytic sample reaction was allowed to proceed for 16 h, at which time the samples were desalted by passing 70 µl of reaction mixture through a Bio-Rad Spin6 column. 50 µl of the flow-through was then trypsin-digested for 10 min, 20 min, 1 h, 4 h, and overnight by the use of a trypsin singles proteomic grade kit (Sigma).

HPLC

Tryptic fragments were separated off-line by the use of an Agilent 1200 HPLC system equipped with an analytical C4 resin column. HPLC grade acetonitrile/0.1% TFA (Fisher) and water/0.1%TFA were used as solvents in the 70-min gradient that ran from 10–85% acetonitrile. Thirty-five 1.5-ml fractions were selectively collected starting at the 10 min time point within the LC run and then lyophilized overnight. Fractions 1–8, 9–12, 13–25, and 26–35 were pooled into three separate tubes and then

stored at −80 °C until mass spectral analysis. Pooled fractions 1–8 and 9–12 contained the identified heterogeneous mixture of modified and unmodified protoxins.

Nanocapillary LC-MS/MS

Nanocapillary columns were prepared by drawing 360-μm outer diameter, 100-μm inner diameter deactivated, fused silica tubing (Agilent) with a Model P-2000 laser puller (Sutter Instruments) (heat: 330, 325, 320; velocity, 45; delay, 125) and were packed at 600 psi to a length of ~10 cm with C18 reverse-phase resin suspended in methanol. The column was equilibrated with 90% of solvent A (water, 0.1% AcOH) and loaded with 10 μl of trypsin digested lytic sample (ClosA, SagB, ClosC, and ClosD) by flowing 90% of solvent A and 10% of solvent B (CH_3CN, 0.1% AcOH) at 2 μl/min for 5 min, 7 μl/min for 3 min, and 10 μl/min for 12 min. At 20 min, the flow rate was increased to 200 μl/min and infused into a split-flow so that ~200–500 nL/min went through the capillary column, whereas the remainder of the flow was diverted to waste. A gradient for eluting trypsin-digested peptides was established with a time-varying solvent mixture and directly electrosprayed into a calibrated Thermo Finnigan LTQ-XL (source voltage, 2.0 kV; capillary temperature, 200 °C). Note that in all occasions, for heterocycle identification, LTQ-MS was tuned and calibrated to achieve a background signal NL < 4.5E3.

LC-MS/MS Acquisition

Seven different MS/MS acquisitions methods were used to compile >500,000 spectra of the ClosA and synthetase lytic samples. Methods M1 and M2 use data-

dependent acquisition with varying cut-offs on the exclusion list capacity and time. Methods M3–M7 select the 1–25th most abundant ions for fragmentation by collision-induced dissociation using data-dependent acquisition. The identified peptides, in a reoccurring fashion, for the SLS and CLS systems were observed in M1 (fragmentation of the first through fifth most abundant ions), M2 (data-dependent list capacity cut) and M3 (fragmentation of the first through tenth most abundant ions).

MS/MS Data Processing

All collected data files (RAW) were processed with a DOS command line version of InSpecT software (University of California, San Diego). Relevant input search parameters included the following: (i) post-translational modification search, −18 Da (Cys, Ser, and Thr for cyclodehydration), −20 Da (Cys, Ser, and Thr for cyclodehydration and dehydrogenation), and +43 Da (Cys, Ser, and Thr, for BrEA); (ii) database, Clostridium genome (The Sanger Institute), Clostridium genome "reversed" or "phony database," and common contaminants database; (iii) PTM allowed per peptide = 6; (iv) Δm b- and y-ion tolerance = 0.5 Da; and (v) parent mass tolerance = 1.5 Da.

The acquired RAW files were converted to mzXML files and then processed by Spectral Networks. Peptides that InSpecT and Spectral Networks identified as having PTM of −20 Da were further processed manually by de novo sequencing of the spectrum. Specifically, the mass error of the parental ion fragmented and the b- and y-ion series were verified.

Point Mutation Analysis of Identified Heterocyclized Site on Clostridiolysin S

Preparation of ClosA Ala Mutant at Identified Heterocycle

A QuikChange kit was used to create the ClosA mutants using pET28b-MBP closA plasmid as a template for PCR. Further processing of the construct follows the standard preparation described for CxxC mutants.

Protein Purification of MBP-ClosA Ala Mutants

ClosA mutant was MBP-tagged and purified as described previously (12). Upon purification, the protein was immediately stored at −80 °C until needed for assays.

Additional description of Mass Spectrometry

To observe the C-terminal end of ClosA, the digested 2-bromoethylamine alkylated samples were subjected to nanocapillary–LCMS/MS analysis on an LTQ instrument. Because of the heterologous nature of the converted ClosA substrate observed in the top down analysis, it was expected that similar heterologous mixtures of modified and unmodified peptides would result in the bottom-up approach. Therefore the data was subjected to the tandem mass spectral processing program Spectral Networks. Spectral Networks processed 576,339 total spectra and identified 133,435 clusters. Of the indentified clusters 42 were observed in the C-terminal portion of ClosA (CCVSVS region). In conjunction with Spectral Networks, the

tandem MS/MS data was analyzed by InSpecT. The resulting MS/MS data analysis of ClosA outputted a series of sequence tags, which covered the C-terminal ends in a ladder type fashion that had a -20 and/or +43 Da modification. All residues in the ClosA primary sequence were annotated except C26 and C27. Each of the peptides identified to contain a 20 Da loss were manually confirmed. With exception of an -18 Da modification on T11 and the expected +43 Da modifications on Cys, Ser, Thr that were found, no other modifications were recovered in this search. We believe that the 18 Da loss found in the ClosA sample was due to a low level of source fragmentation as this ion is found to be nearly undetectable species in the broadband scan.

Figure 4.4 shows the tandem mass spectrum of a CLS peptide with a (methyl)-oxazole at position T46. Manual inspection of the PTM peptide shows the parental mass fragmented has a Δm= 0.42 Da (230 ppm mass error) relative to the theoretical mass. The b and y-ion series ppm error range from 5 ppm-200 ppm. Within the b and y-ion series 14/22 y-ions were annotated with 7 contingent y-ions capturing the -20 Da modification. The b-ion series annotated 7/22 with 5 contingent b-ions capturing the -20 Da modification. Because the y12/b11 ions were not annotated and y13/b10 were annotated, we infer the 20 Da loss is localized at T46 and thus represents methyloxazole formation. Our conclusion is supported by the chemistry which takes place upon methyloxazole formation, since the amide bond between S45 and A44 would have been lost upon heterocyclization and therefore y12/b11 would be nonexistent. In addition to the PTM peptide, the unmodified peptide was also captured with a Δm= 0.52 (288 ppm mass error) and b and y-ion series ppm error range from 5

ppm-300 ppm. Furthermore, the analysis of the tandem mass spectrometry data identified several other peptides with multiple 20 Da losses. The first localizes heterocyclized residues at positions: T11 and T12 (Peptide 3 Fig. 4.11). The parent mass of the fragmented peptide with heterocycles at T11 and T12 had a mass error of 260 ppm mass error and a b and y-ion series with ppm error ranging from 10 ppm-250 ppm. The identified peptide displayed 9 contingent y-ions all of which contained the modification. The b-ion series annotated 7/14 ions. Peptide 4 is a similar peptide as Peptide 3 but has one less heterocycle. The mass error of this peptide was 333 ppm. Our method was able to identify an SLS peptide with a 40 Da loss relative to the unmodified peptide (installation of two oxazoles by SagBCD enzymes). The identified peptide (Peptide 2) had a mass error of 71 ppm.

Primers used in studies

Intergrational mutagenesis of closA and closC

(i) ClosA_For_*EcoRI*

GCGGAATTCTAATGAACACGTACTGACAACTAC)

(ii) ClosA_Rev_*BamHI*

GCCGGATTCGCTGCTCCGCCCACCTGTT .

ClosC_For_*EcoRI*

GCGGAATTCGCGGAATTCGCCCCAAAAAGTATTATTTTATATCAG

(iii) ClosC_Rev_*BamHI*

GCCGGATTCCACTCTAGACATCCCGTTTTAGG

Cloning and protein purification of ClosA, ClosB, ClosC, ClosD, SagB

(i) ClosA(ForBamHI) AAG GAT CCA TGC TGA AAT TTA ACG AAC ATG TGC TGA CC ClosA(RevNotI) AAG CGG CCG CTT AGT TGC CGC CCT GAC CCG C (ii) ClosB(ForBamHI) AAG GAT CCA TGC TGC TGA AAA ACC TGA AAA AAC AGA AAG TC ClosB(RevNotI) AAG CGG CCG CTT AAT CAC CCT GTT CCC AGC CCA CG (iii) ClosC(ForBamHI) AAA GGA TCC ATG AAA AAT AAT ACC ATC TAT CGT CTG AGC ClosC(RevNotI) AAA GCG GCC GCT TAG CTA TTA ATA TCT TCC AGA ATA CCA TCG (iv) ClosD(ForBamHI) AAA GGA TCC ATG ATC AAA TTT AGT CCG TCA TTT AAT AAT ATT CTG G ClosD(RevNotI) AAA GCG GCC GCT TAT GGC ATC GGA TGC GGG TAT TC.

In vivo Lytic activity: Clos single-gene complementation into Group A Streptococcus (GAS) ΔsagA-D strains

(i) ClosA (ForXbaI) TAA TCT AGA TAA ATG CTG AAA TTT AAC GAA CAT GTG CT ClosA(RevBamHI) TAA GGA TCC TAA TTA GTT GCC GCC CTG ACC CGC CGC CGC A (ii) ClosB (ForXbaI) TAA TCT AGA TAA ATG CTG CTG AAA AAC CTG AAA AAA CAG A ClosB(RevBamHI) TAA GGA TCC TAA TTA ATC ACC CTG TTC CCA GCC CAC GAC T (iii) ClosC (ForXbaI) TAA TCT AGA TAA ATG AAA AAT AAT ACC ATC TAT CGT CTG A ClosC(RevBamHI) TAA

GGA TCC TAA TTA GCT ATT AAT ATC TTC CAG AAT ACC A (iv) ClosD (ForXbaI) TAA TCT AGA TAA ATG ATC AAA TTT AGT CCG TCA TTT AAT ClosD (RevBamHI) TAA GGA TCC TAA TTA TGG CAT CGG ATG CGG GTA TTC GTT.

Sited Directed Mutagenesis of ClosC CxxC mutants

(i) ClosC_C332A TTC CGT TTT GTC CGG CGG CCG GCA ATA TTA GCA AAG C ClosC_C332A GCT TTG CTA ATA TTG CCG GCC GCC GGA CAA AAC GGA A (ii) ClosC_C329A TCT GCT GCG CGT TCC GTT TGC TCC GGC GTG C ClosC_C329A GCA CGC CGG AGC AAA CGG AAC GCG CAG CAG A (iii) ClosC_C245A GAA AAC CGG GTG CCT GGA AGC TTT TGA GCA GCG TAT TCT G ClosC_C245A CAG AAT ACG CTG CTC AAA AGC TTC CAG GCA CCC GGT TTT C (iv) ClosC_C242A GCC GCC GAA AAC CGG GGC CCT GGA ATG TTT TGA G ClosC_C242A CTC AAA ACA TTC CAG GGC CCC GGT TTT CGG CGG C ClosC_C242A_C245A GCC GCC GAA AAC CGG GGC CCT GGA AGC TTT TGA GCA GCG TAT T ClosC_C242A_C245A AAT ACG CTG CTC AAA AGC TTC CAG GGC CCC GGT TTT CGG CGG C (v) ClosC_C329A_C332A GAT CTG CTG CGC GTT CCG TTT GCT CCG GCG GCC GGC AAT ATT A ClosC_C329A_C332A TAA TAT TGC CGG CCG CCG GAG CAA ACG GAA CGC GCA GCA GAT C.

Point mutation analysis of identified heterocyclized sites on clostriodiolysin S

(i)T46A_fwd GCA GCG CGA GCG CCG GTG GTG GT T46A_rev ACC ACC ACC GGC GCT CGC GCT GC

Acknowledgements

We greatly appreciate: (i) Jane Yang for critically reading the manuscript (ii) Ted Huston for performing ICP-MS analysis (iii) Theresa Smith and Gary Xie for performing comparative sequence alignments of the clos cluster against their available clostridia genomes (iv) Richard Klemke and Mark Lortie (UCSD Pathology Biomarker and Diagnostic Discovery Center) for use of the LTQ-MS in the validation experiments (v) Support by the Beckman Foundation, the NIH Hemoglobin and Blood Protein Chemistry Training Program (5T32DK007233-34) and the NIH Molecular Biophysics Training Program (GM08326).

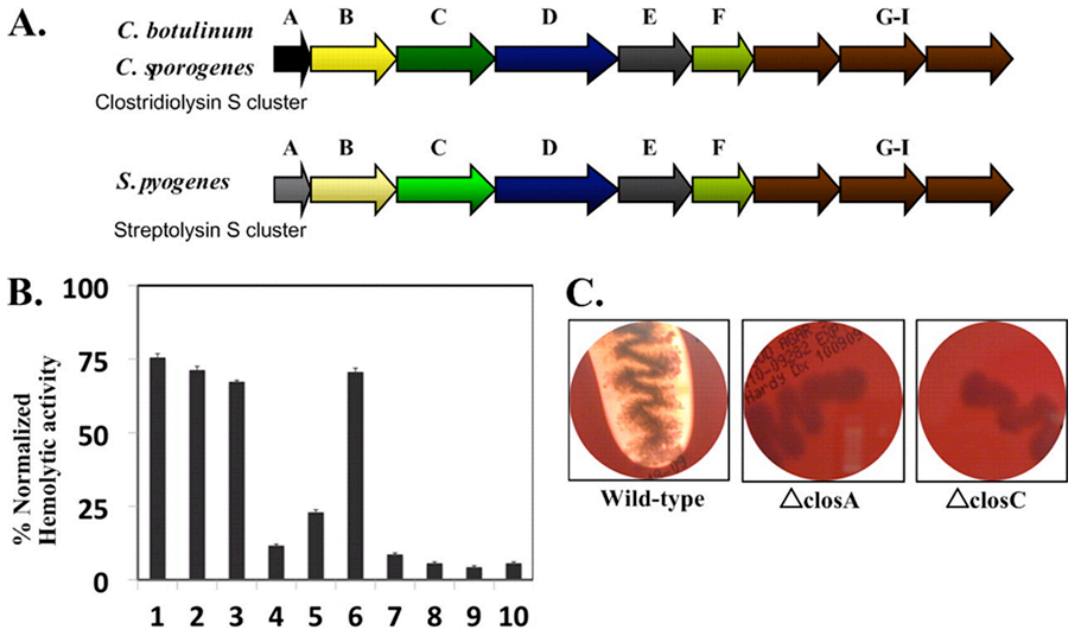


Figure 4.1 Comparison of the streptolysin S-associated genetic cluster (Sag) with the SLS-like clostridiolysin S genetic cluster in *Clostridium botulinum*

(A) the Sag cluster in GAS contains a SLS precursor sequence (SagA), which is modified by the SagBCD synthetase complex. SagE and -F are potentially involved in immunity function. SagG–I have homology to ABC transporters. The clostridiolysin S gene cluster contains genes similar to the Sag locus, in identical order, and is present in *C. botulinum* ATCC 3502, and *C. sporogenes* ATCC 15579. *B*, shown is the genetic complementation of Clos genes in GAS M1 Sag mutants. BSA-stabilized extracts were assayed for lytic activity. *Bar 1*, M1wtGAS; *bar 2*, M1 GAS Δ*sagA* and *wtsagA*; *bar 3*, M1 GAS Δ*sagA* and *wtclosA*; *bar 4*, M1 GAS Δ*sagB* and *wtclosB*; *bar 5*, M1 GAS Δ*sagC* and *wtclosC*; *bar 6*, M1 GAS Δ*sagD* and *wtclosD*; *bar 7*, M1 GAS Δ*sagA* and plasmid alone; *bar 8*, M1 GAS Δ*sagB* and plasmid alone; *bar 9*, M1 GAS ΔsagC and plasmid alone; and *bar 10*, M1 GAS Δ*sagD* and plasmid alone. Hemolytic activity was normalized against a Triton X-100 positive control. *C*, *C. sporogenes* exhibits a strong β-hemolytic phenotype when grown on blood agar plates. Δ*closA* and Δ*closC C. sporogenes* allelic exchange mutants lose the ability to lyse erythrocytes. Bacteria were grown for 36 h at 37 °C under anaerobic conditions.

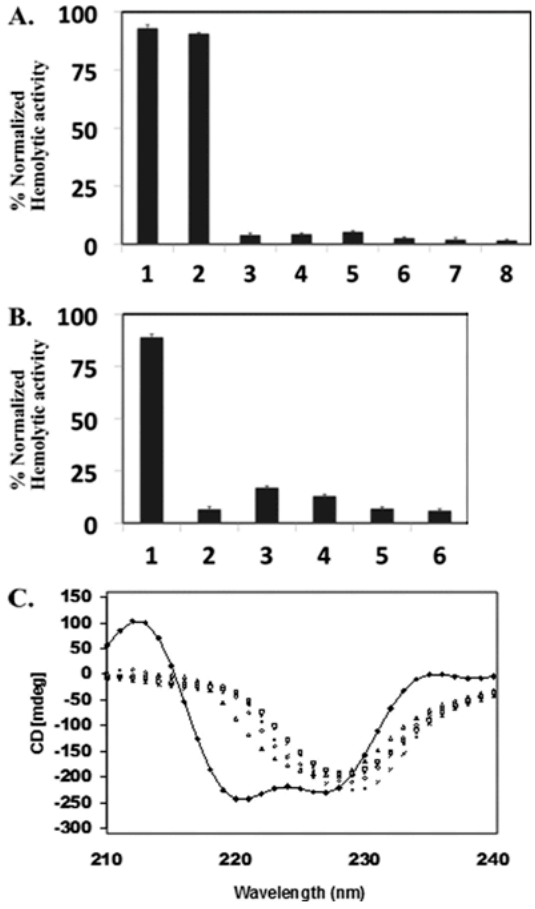


Figure 4.2 Hemolytic activity of clostridiolysin is detectable *in vitro*
(A) in vitro reconstitution of ClosA hemolytic activity. Synthetase reactions using MBP-ClosA and SagB/ClosC/ClosD produce a hemolytic toxin. Hemolytic activity was normalized against a Triton X-100 positive control. Bar 1, wt SagA and SagBCD; bar 2, wt ClosA and SagB/ClosC/ClosD; bar 3, wt ClosA only; bar 4, SagB/ClosC/ClosD only; bar 5, wt ClosA and SagB/ClosC; bar 6, wt ClosA and ClosC/ClosD; bar 7, wt ClosA and SagB/ClosD; bar 8, SagA alone. (B) in vitro reconstitution of ClosA hemolytic activity using CxxC mutants as wild-type ClosC surrogates. Bar 1, wt ClosA and SagB/ClosC/ClosD; bar 2, wt ClosA and SagB/ALEC/ClosD wt; bar 3, ClosA and SagB/CLEA/ClosD; bar 4, wt ClosA and SagB/APAA/ClosD wt; bar 5, ClosA and SagB/APAC/ClosD; bar 6, wt ClosA and SagB/CPAA/ClosD. Hemolytic activity was normalized against a Triton X-100 positive control. (C) CD spectra of MBP-ClosC and MBP-tagged CxxC mutants. ■, CD spectra of MBP-ClosC (a negative minimum at 215 and 225 nm confirms α-helical integrity); □, CD for CPAA mutant; ▵, CD for CLEA mutant; ×, CD for APAC mutant; *, CD for APAA mutant; ○, CD for ALEC mutant. CD was repeated three times with different protein purifications and resulted in consistent results.

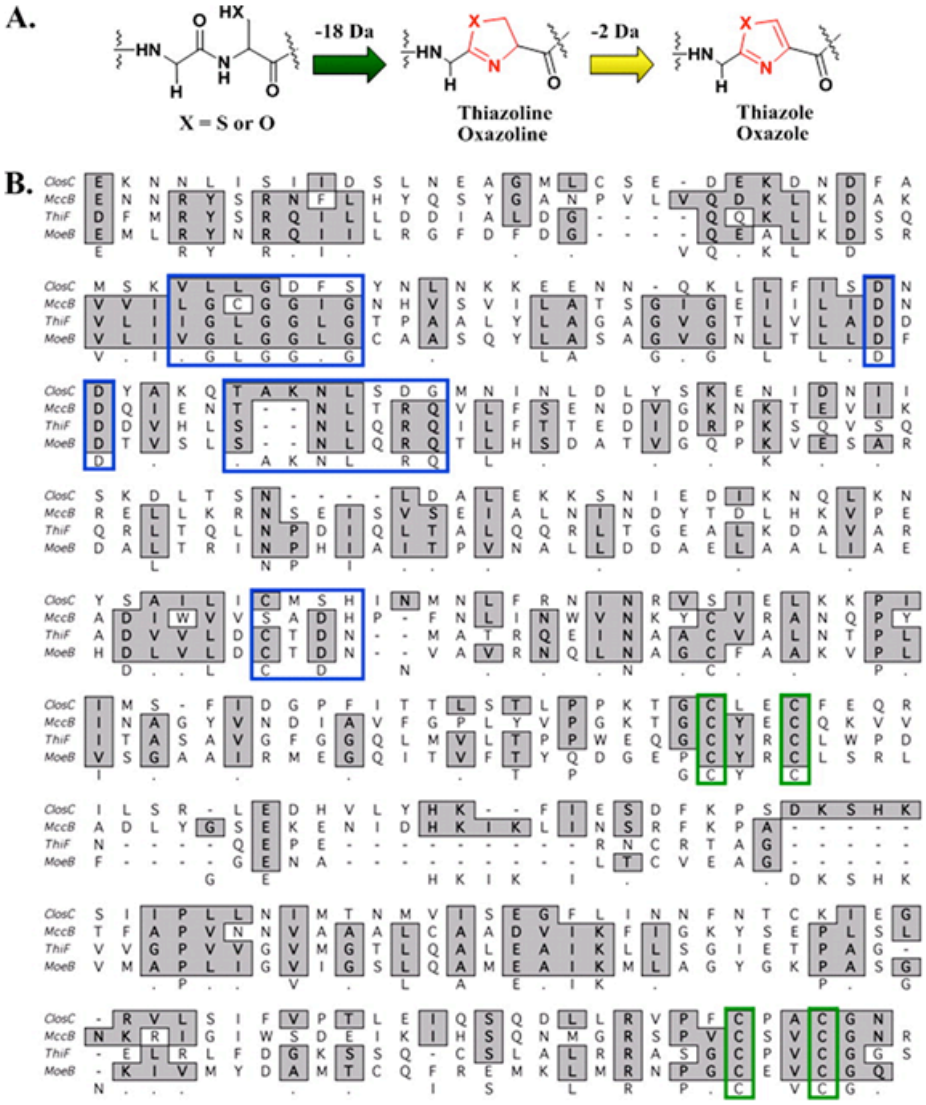


Figure 4.3 Alignment of ClosC in *C. botulinum* ATCC 3503 and MoeB/ThiF and proposed mechanisms

A mechanisms for the formation of thiazole or methyloxazole. B, sequence alignments of ClosC, MccB, E. coli MoeB, and E. coli ThiF. Blue highlighted boxes show conserved ATP binding residues determined by the crystal structures of ThiF and MoeB. Green highlighted boxes show the residues involved in zinc-tetrathiolate formation.

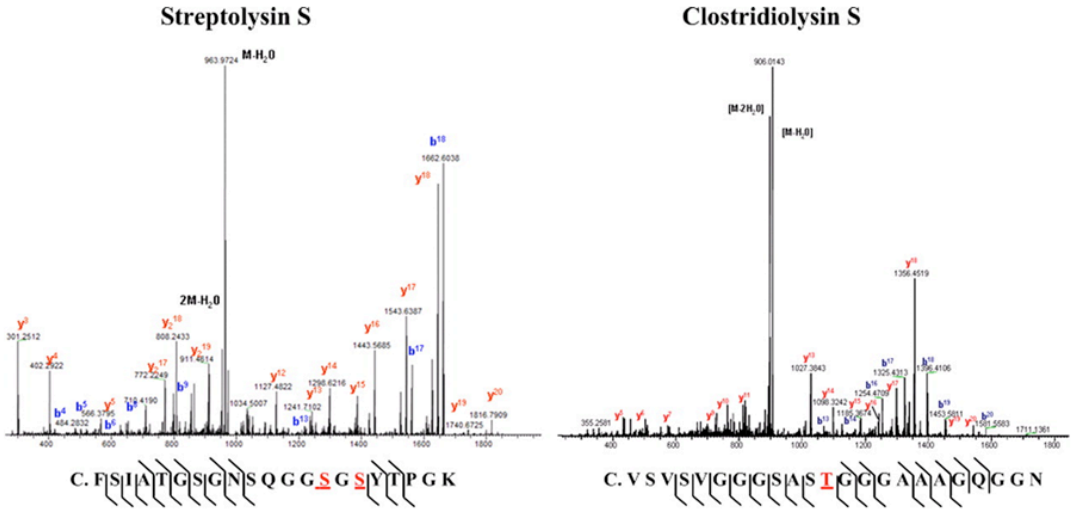


Figure 4.4 Detection of heterocyclized ClosA peptides via nanocapillary LC-MS/MS method

Mass spectra of C-terminal peptides identified as being heterocyclized in the SLS and CLS in vitro systems. A detailed description of mass spectral annotations is supplied in the materials and methods section.

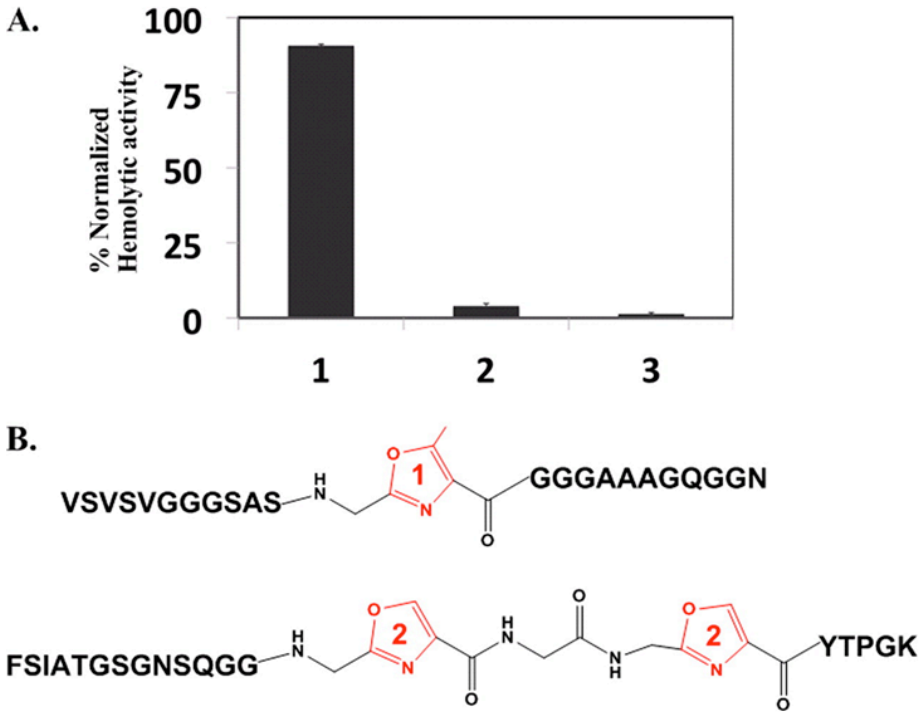


Figure 4.5 Hemolytic activity of MBP-ClosA T46A and structures of identified PTM peptides

(A) synthetase reactions using MBP-ClosA T46A and SagB/ClosC/ClosD. Hemolytic activity was normalized against a Triton X-100 positive control. *Lane 1*, wt ClosA and SagB/ClosC/ClosD; *lane 2*, SagB/ClosC/ClosD only; *lane 3*, ClosA T46A and SagB/ClosC/ClosD. (B) peptide 1, CLS peptide containing a methyloxazole at position Thr^{46}; peptide 2, SLS peptide containing two oxazoles at positions Ser^{46} and Ser^{48}.

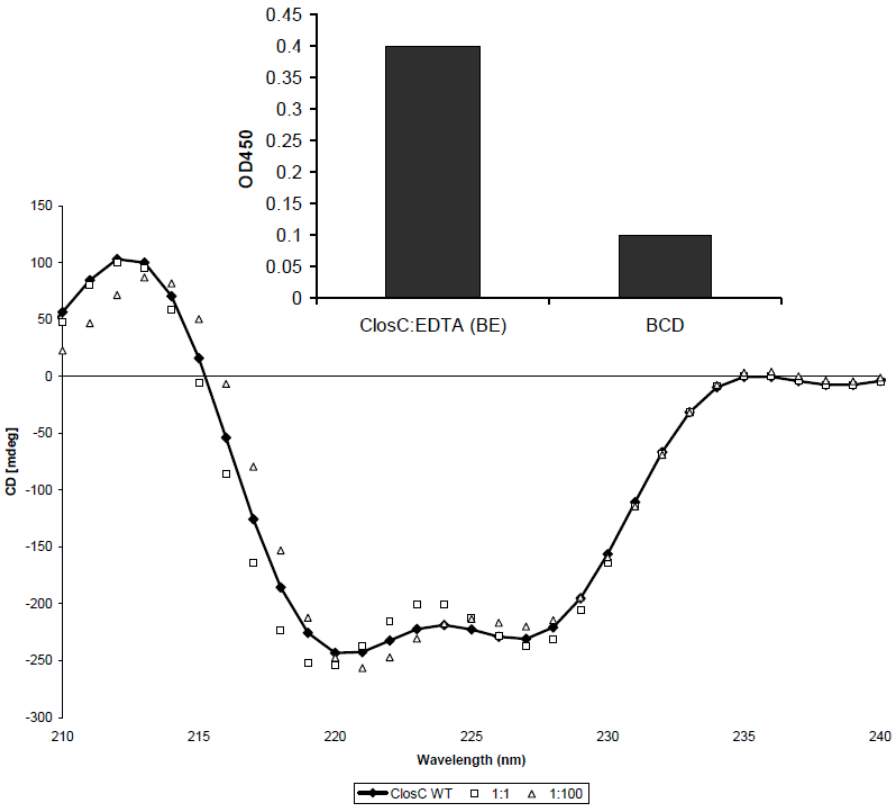


Figure 4.6 Effect of zinc on ClosC circular dichroism
MBP-ClosC was incubated with 1:1 and 1:100 protein-to-EDTA molar ration then buffer exchanged (BE). The CD shows the MBP-ClosC remains unperturbed by the EDTA and is active in the bioassays (insert, y-axis shows OD450 monitoring heme release upon erythrocyte lysis).

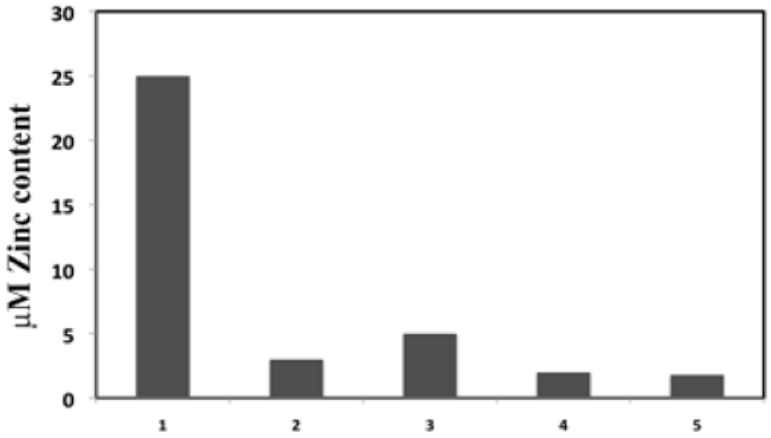


Figure 4.7 ICP-MS data of wt ClosC versus CxxC mutants
ICP-MS data is reported in order: 1. ClosC 2. ALEC 3. CLEA 4. APAC 5. CPAA. ICP-MS experiments were performed in duplicate with two separate protein preparations.

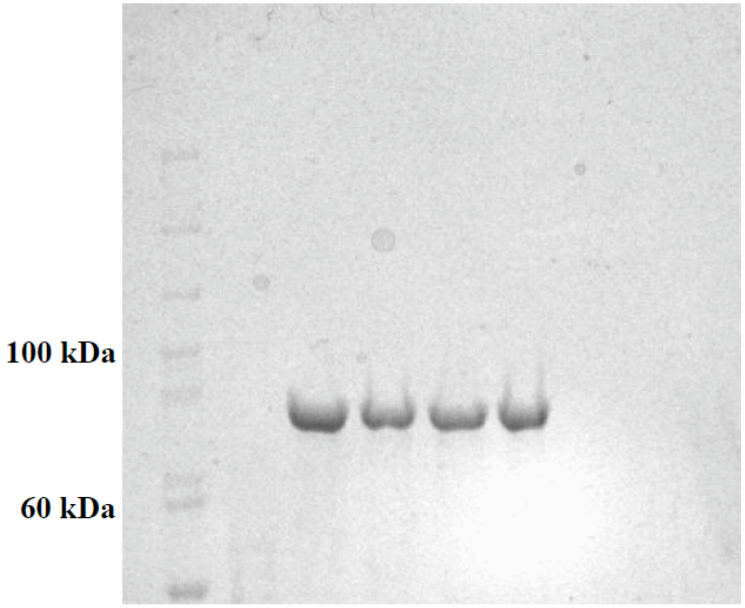


Figure 4.8 SDS-PAGE of the proteins used for CD analysis
Proteins were concentrated (40kDa cut-off) and buffer exchanged. Molecular weight of MBP-ClosC is ~89kDa. 1. ALEC 2. CLEA 3. APAC 4. CPAA.

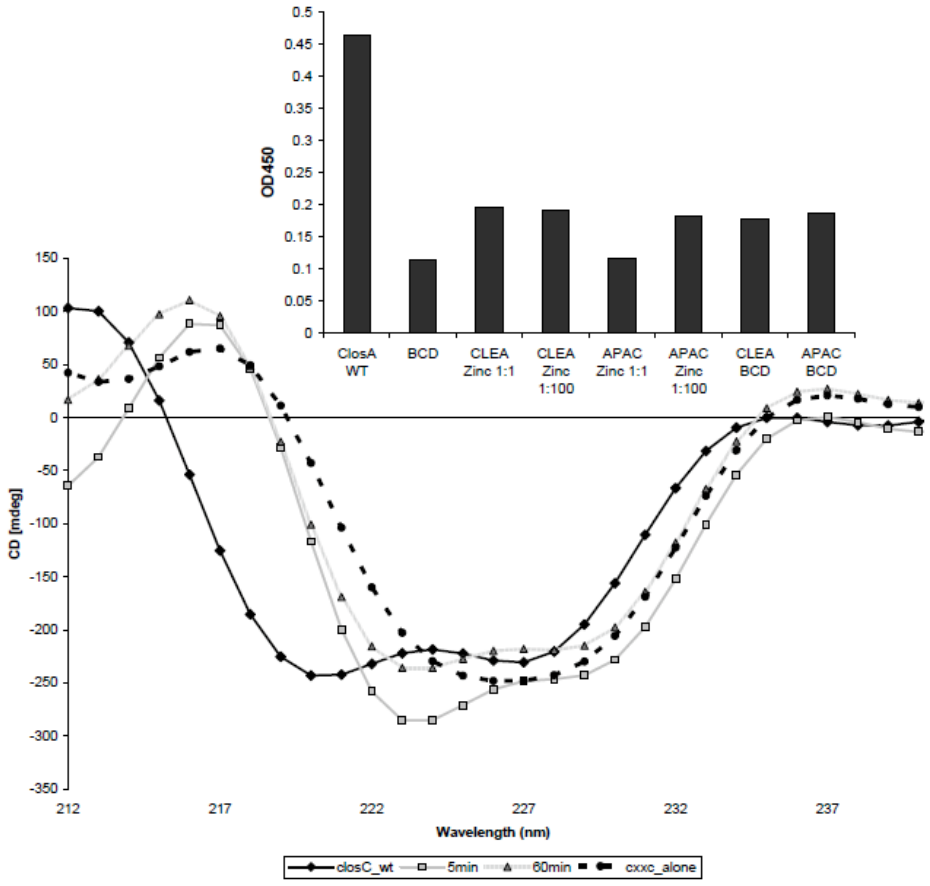


Figure 4.9 Effect of zinc on ClosC circular dichroism
CxxC proteins CLEA and APAC were incubated with zinc at a 1:1 and 1:100 protein-to-zinc molar ratio then buffer exchanged (BE). The CD shows the zinc could not restore alpha-helical content comparable to wt. MBP-ClosC. Insert shows bioactivity of the CxxC proteins upon zinc incubation (y-axis shows OD450 monitoring heme release upon erythrocyte lysis).

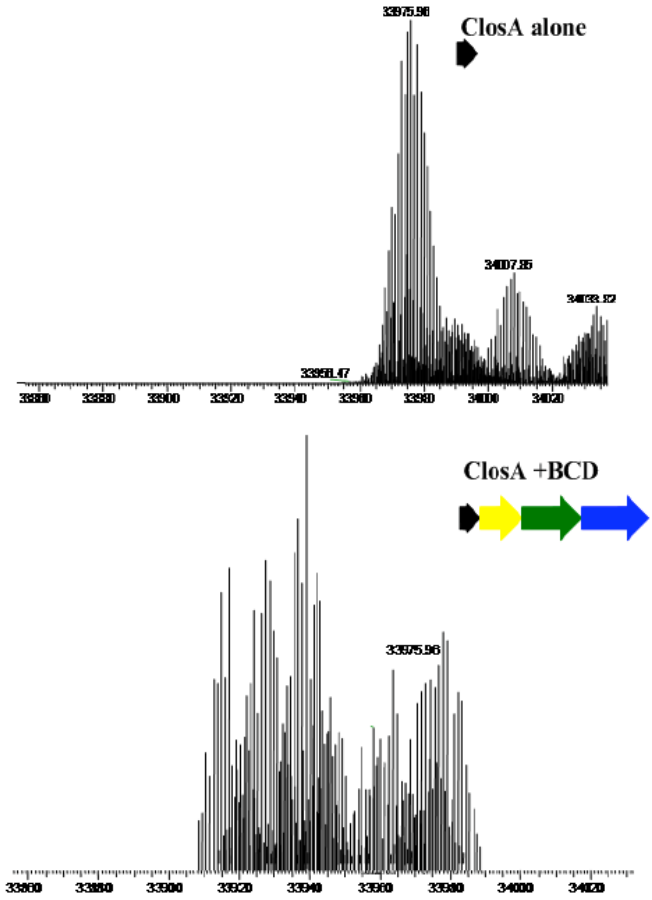


Figure 4.10 HPLC of modified ClosA
HPLC purified recombinantly expressed ClosA was introduced into the FTICR-MS and visualized at 21ppm. Incubating ClosA with the BCD synthetase complex, followed by HPLC produced a series of overlapping ions. Manual annotation of the resulting spectra produced ambiguous mass assignments; therefore a bottom-up approach was taken to visualize the PTM occurring on ClosA.

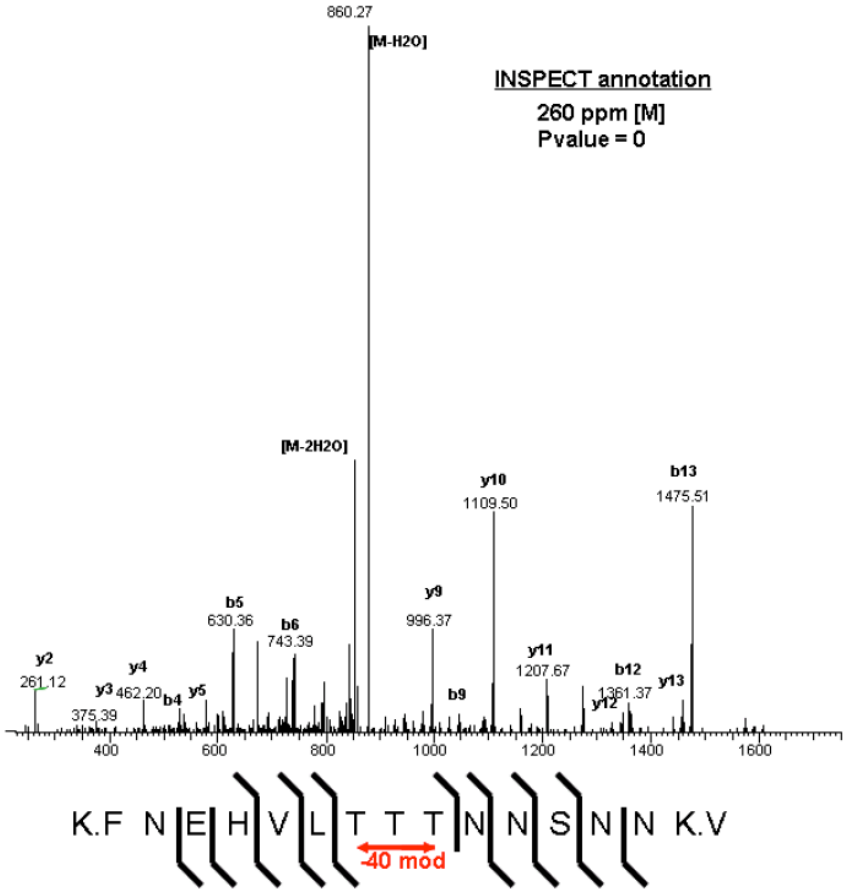


Figure 4.11 Tandem mass spectrometry of T11/T12 heterocyclic peptide

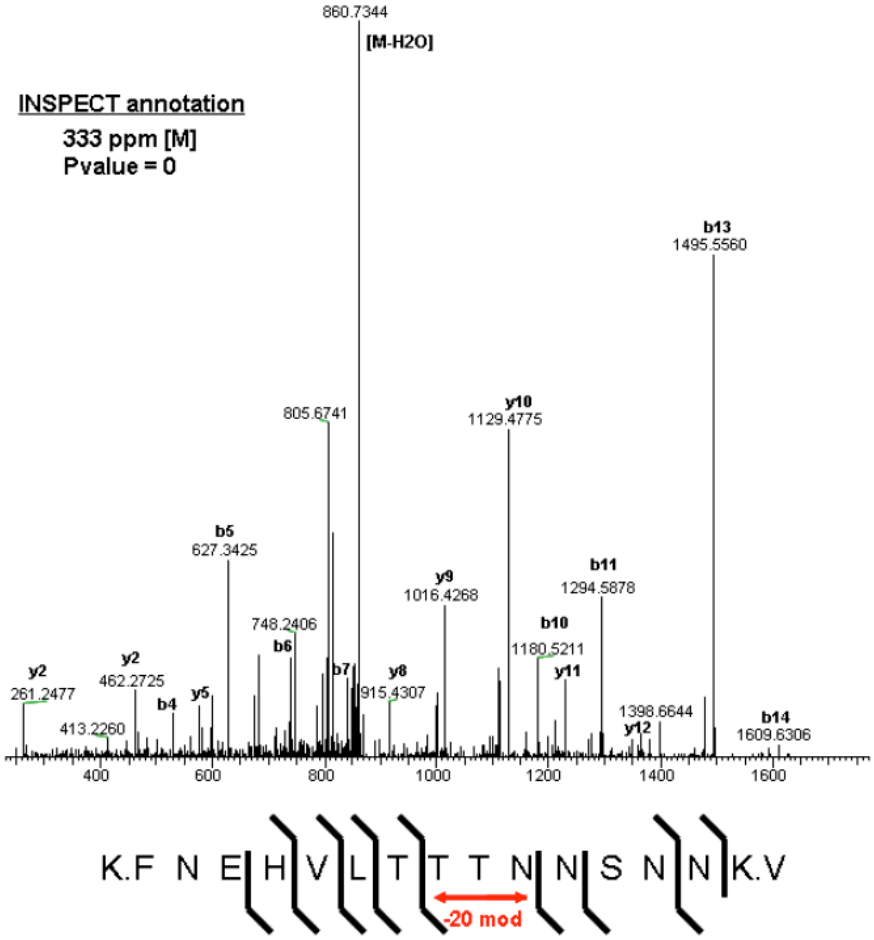


Figure 4.12 Tandem mass spectrometry of T12 heterocyclic peptide

E. References

1. **Bandeira, N., D. Tsur, A. Frank, and P. A. Pevzner.** 2007. Protein identification by spectral networks analysis. Proc Natl Acad Sci U S A **104:**6140-5.

2. **Cinader, B., and L. Pillemer.** 1950. The purification and properties of streptolysin S. J Exp Med **92:**219-37.

3. **Cotter, P., L. A. Draper, E. M. Lawton, K. M. Daly, D. S. Groeger, P. G. Casey, R. P. Ross, and C. Hill.** 2008. Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I Listeria monocytogenes. PLoS pathogens **4:**e1000144.

4. **Datta, V., S. Myskowski, L. Kwinn, D. Chiem, N. Varki, R. Kansal, M. Kotb, and V. Nizet.** 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Molecular Microbiology **56:**681-695.

5. **Duda, D. M., H. Walden, J. Sfondouris, and B. A. Schulman.** 2005. Structural analysis of Escherichia coli ThiF. Journal of Molecular Biology **349:**774-86.

6. **Hara-Kudo, Y., A. Ogura, Y. Noguchi, K. Terao, and S. Kumagai.** 1997. Effect of hemorrhagic toxin produced by Clostridium sporogenes on rabbit ligated intestinal loop. Microb Pathog **22:**31-8.

7. **Hopkins, C. E., P. B. O'Connor, K. N. Allen, C. E. Costello, and D. R. Tolan.** 2002. Chemical-modification rescue assessed by mass spectrometry demonstrates that gamma-thia-lysine yields the same activity as lysine in aldolase. Protein Sci **11:**1591-9.

8. **Humar, D., V. Datta, D. J. Bast, B. Beall, J. C. S. De Azavedo, and V. Nizet.** 2002. Streptolysin S and necrotising infections produced by group G streptococcus. Lancet **359:**124-9.

9. **Igarashi, Y., Y. Kan, K. Fujii, T. Fujita, K. Harada, H. Naoki, H. Tabata, H. Onaka, and T. Furumai.** 2001. Goadsporin, a chemical substance which promotes secondary metabolism and Morphogenesis in streptomycetes. II. Structure determination. J Antibiot **54:**1045-53.

10. **Jurgenson, C. T., K. E. Burns, T. P. Begley, and S. E. Ealick.** 2008. Crystal structure of a sulfur carrier protein complex found in the cysteine biosynthetic pathway of Mycobacterium tuberculosis. Biochemistry **47:**10354-64.

11. **Kelly, W., L. Pan, and C. Li.** 2009. Thiostrepton Biosynthesis: Prototype for a New Family of Bacteriocins. J Am Chem Soc.

12. **Lee, S. W., D. A. Mitchell, A. L. Markley, M. E. Hensler, D. Gonzalez, A. Wohlrab, P. C. Dorrestein, V. Nizet, and J. E. Dixon.** 2008. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci USA **105:**5879-84.

13. **Lehmann, C., T. P. Begley, and S. E. Ealick.** 2006. Structure of the Escherichia coli ThiS-ThiF complex, a key component of the sulfur transfer system in thiamin biosynthesis. Biochemistry **45:**11-9.

14. **Li, Y. M., J. C. Milne, L. L. Madison, R. Kolter, and C. T. Walsh.** 1996. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. Science **274:**1188-93.

15. **Liao, R., L. Duan, C. Lei, H. Pan, Y. Ding, Q. Zhang, D. Chen, B. Shen, Y. Yu, and W. Liu.** 2009. Thiopeptide biosynthesis featuring ribosomally synthesized precursor peptides and conserved posttranslational modifications. Chem Biol **16:**141-7.

16. **Mitchell, D. A., S. W. Lee, M. A. Pence, A. L. Markley, J. D. Limm, V. Nizet, and J. E. Dixon.** 2009. Structural and functional dissection of the heterocyclic Peptide cytotoxin streptolysin s. J Biol Chem **284:**13004-12.

17. **Morris, R., J. Leeds, H. Naegeli, L. Oberer, K. Memmert, E. Weber, M. Lamarche, C. Parker, N. Burrer, S. Esterow, A. Hein, E. Schmitt, and P. Krastel.** 2009. Ribosomally Synthesized Thiopeptide Antibiotics Targeting Elongation Factor Tu. J Am Chem Soc.

18. **Nizet, V., B. Beall, D. J. Bast, V. Datta, L. Kilburn, D. E. Low, and J. C. De Azavedo.** 2000. Genetic locus for streptolysin S production by group A streptococcus. Infect Immun **68:**4245-54.

19. **Nolan, E. M., and C. T. Walsh.** 2008. How Nature Morphs Peptide Scaffolds into Antibiotics. ChemBioChem.

20. **Oman, T. J., and W. A. van der Donk.** 2010. Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nature Chemical Biology **6:**9-18.

21. **Regni, C. A., R. F. Roush, D. J. Miller, A. Nourse, C. T. Walsh, and B. A. Schulman.** 2009. How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic. EMBO J**:**1-12.

22. **Roush, R. F., E. M. Nolan, F. Löhr, and C. T. Walsh.** 2008. Maturation of an Escherichia coli ribosomal peptide antibiotic by ATP-consuming N-P bond formation in microcin C7. J Am Chem Soc **130:**3603-9.

23. **Savani, J., N. Harris, and W. Gould.** 1978. Survival of Clostridium Sporogenes PA 3679 in Home-Canned Tomatoes. Journal of Food Science **43:**222-224.

24. **Schmidt, E., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. Haygood, and J. Ravel.** 2005. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proceedings of the National Academy of Sciences of the United States of America **102:**7315-20.

25. **Sebaihia, M., M. W. Peck, N. P. Minton, N. R. Thomson, M. T. G. Holden, W. J. Mitchell, A. T. Carter, S. D. Bentley, D. R. Mason, L. Crossman, C. J. Paul, A. Ivens, M. H. J. Wells-Bennik, I. J. Davis, A. M. Cerdeño-Tárraga, C. Churcher, M. A. Quail, T. Chillingworth, T. Feltwell, A. Fraser, I. Goodhead, Z. Hance, K. Jagels, N. Larke, M. Maddison, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M. Simmonds, B. White, S. Whithead, and J. Parkhill.** 2007. Genome sequence of a proteolytic (Group I) Clostridium botulinum strain Hall A and comparative analysis of the clostridial genomes. Genome Research **17:**1082-92.

26. **Severinov, K., E. Semenova, A. Kazakov, T. Kazakov, and M. S. Gelfand.** 2007. Low-molecular-weight post-translationally modified microcins. Mol Microbiol **65:**1380-94.

27. **Tanner, S., H. Shu, A. Frank, L.-C. Wang, E. Zandi, M. Mumby, P. A. Pevzner, and V. Bafna.** 2005. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal Chem **77:**4626-39.

28. **Tsaihong, J. C., and D. E. Wennerstrom.** 1983. Effect of carrier molecules on production and properties of extracellular hemolysin produced byStreptococcus agalactiae. Curr Microbiol **9:**333-338.

29. **Wieland Brown, L. C., M. G. Acker, J. Clardy, C. T. Walsh, and M. A. Fischbach.** 2009. Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. Proc Natl Acad Sci USA **106:**2549-53.

30. **Yang, Y.-L., Y. Xu, P. Straight, and P. C. Dorrestein.** 2009. Translating metabolic exchange with imaging mass spectrometry. Nature Chemical Biology**:**1-3.

31. **Zamble, D. B., C. P. McClure, J. E. Penner-Hahn, and C. T. Walsh.** 2000. The McbB component of microcin B17 synthetase is a zinc metalloprotein. Biochemistry **39:**16190-9.

32. **Zamble, D. B., D. A. Miller, J. G. Heddle, A. Maxwell, C. T. Walsh, and F. Hollfelder.** 2001. In vitro characterization of DNA gyrase inhibition by microcin B17 analogs with altered bisheterocyclic sites. Proc Natl Acad Sci USA **98:**7712-7.

Chapter V

An *E. coli*-based Bioengineering Strategy to Study Streptolysin S Biosynthesis

Introduction

Group A *Streptococcus pyogenes* (GAS) is a leading human pathogen that can progress to dramatic invasive syndromes such as necrotizing fasciitis and toxic shock syndrome (2). A hallmark of GAS infection is a clear zone of hemolysis surrounding colonies grown on blood agar media. The ability of GAS to lyse blood cells is due to a powerful cytolysin known as streptolysin S (SLS) (1, 10). SLS production by GAS results in infections with invasive and severe outcomes (5). Furthermore, recent studies have shown that streptolysin S can inhibit neutrophil recruitment and impair immunity (7, 9). SLS has also been implicated in facilitating transepithelial spread of GAS (15). These findings demonstrate that SLS is likely to contribute to the pathogenesis of GAS infections by multiple strategies, making it a critical virulence factor for study. Genetic analyses have shown that SLS biosynthesis is mediated by a nine-gene cluster (streptolysin S-associated gene, *sag*) in the GAS chromosomal genome (11).

Though previous attempts to identify the streptolysin S structure have not been successful, we have demonstrated the first *in vitro* reconstitution of the activity of SLS activity (6). Using the purified forms of the precursor peptide and modifying complex from *S. pyogenes*, we have demonstrated that SLS is a small, ribosomally produced bacteriocin-like toxin in the thiazole-oxazole modified microcins (TOMMs) family. An inactive peptide precursor is expressed by GAS that then undergoes heterocyclic addition at specific residues to confer hemolytic activity. Our *in vitro* studies revealed that three candidate enzymes, SagBCD, are necessary and sufficient to convert the

SLS precursor Sag A into a hemolytic molecule. The biosynthetic gene cluster that is responsible for the synthesis of SLS is highly conserved and widespread, and its presence spans multiple phyla, including Cyanobacteria and Euryarchaeota (6).

SLS is closely related to other hemolytic TOMM natural products from firmicutes phyla organisms including *Clostridium botulinum, Listeria monocytogenes* and *Staphylococcus aureus*, however other TOMM natural products such as microcin B17 have unique bacterial targets (4, 17). Thus, many of these undiscovered heterocyclic peptide compounds, classified as bacteriocins, are likely to be important and powerful antibiotics that bacteria use to establish and maintain their environmental niche. It is likely that the discovery of similar peptidic antibiotics will rapidly expand as more genomes are sequenced. Importantly, these new peptide antibiotics are produced ribosomally, and thus are amenable to genetic engineering strategies. Finally, many gene clusters resembling Microcin B17 and SLS are present in microorganisms that are difficult to isolate, maintain, and study (12, 14). Therefore, heterologous approaches using model organisms such as *E. coli* would offer a marked advantage in studying the biosynthetic process behind many of these novel bacteriocin-like compounds. Gene complementation studies using *Streptococcus pyogenes* have demonstrated that insertional inactivation of each of the nine genes located in the *sag* gene cluster can be complemented *in trans* using plasmid-based single gene complementation methods (3). However, previous data have shown that not all of the complemented genes can restore wild type levels of toxin activity using this approach.

The limitations inherent in GAS recomplementation experiments make it

difficult to perform detailed studies, such as assessing the role of particular domains in overall function of the SLS synthetases. Heterologous expression trials with various components of the SLS biosynthetic gene cluster have been attempted. Nizet et al. cloned the entire nine-gene region of the Sag gene cluster and introduced a plasmid containing this region into the non-hemolytic Gram-positive coccus *Lactococcus lactis* (11). Using the pSagLocus plasmid, Nizet demonstrated that the nine-gene *sag* cluster was sufficient to confer a significant amount of hemolytic activity when introduced heterologously into *Lactococcus lactis.* However, a major limitation of this method is the inability to control the expression of the individual genes that are introduced into the non-native host. Indeed, introduction of the pSagLocus plasmid into the Gram-negative host *E. coli* did not result in hemolytic activity as reported by the same authors. This is likely due in part to the incompatibility of promoters between the Gram-positive Group A *Streptococcus* and the Gram-negative *E. coli.* To overcome the existing limitations with heterologous expression of SLS and other bacteriocins from divergent microorganisms, we describe herein an *E. coli*-based multigene expression system for the biosynthesis of the SLS toxin. This system allows for all of the genes necessary in the natural product biosynthesis pathway to be individually added as independent open reading frames into the expression host plasmid. All of the genes necessary for the production of the natural product can be added or removed by means of simple cloning procedures into robust *E. coli* expression sites. Our expression system also allows mutagenesis studies to be rapidly performed by PCR-based methods, because the genes are located on easily isolated plasmids. Finally, our heterologous expression system can be utilized to rapidly insert similar genes from

other similarly produced bacteriocins, such as those from the TOMM family (8). Importantly, our *E. coli* based heterologous expression method will be especially applicable for investigating the biosynthesis of ribosomally derived compounds from microorganisms for which genomes have been sequenced, but for organisms that have been difficult to culture in sufficient yields for study.

Materials and Methods

Plasmid Construction

The expression plasmids, listed in table 5.1, were all constructed using the Duet family of co-expression vectors (EMD4Biosciences) in a sequential fashion by adding to the first MCS then the second. SagB was subcloned from MBP-pET-28b SagB (the construction of which we have previously reported {Lee, 2008 #107}) into the first MCS of pET-DUET-1 using EcoR I/Not I restriction sites. Cloning was done in this fashion to create a TEV/thrombin protease cleavable His 6x fusion tag on the amino terminus of SagB . Next, SagC was amplified by PCR using primers in table 5.2 (IDT) and ligated into the second MCS position using Nde I/Xho I restriction sites. SagB was inserted into pET-Duet using the BamHI/NotI restriction sites to create an amino terminal His 6x tag fusion. SagDA pACYCDuet-1 was constructed by first amplifying SagC by PCR then ligating it into the first MCS position of pACYCDuet-1 using Nco I/Not I restriction enzymes. Next, SagA was inserted into the second MCS position an amino terminal maltose binding protein tag by subcloning the *MBP-SagA* gene from the previously created pET-28b MBP vector (PNAS). All inserts and

vectors were double restriction digested using NEB enzymes as per the manufacturer's instructions. The digested DNA was then purified using a 1% agarose gel, excised, gel extracted (Qiagen), and ligated using T4 DNA ligase (NEB). The resulting constructs were transformed into chemically competent DH5α *E. coli* cells for amplification, purified using the Invitrogen miniprep kit and validated by DNA sequencing (ETON) using the pETUpstream, DuetDOWN1, DuetUP2, and t7 terminator primers (table 5.2).

Co-expression and Hemolytic tests

The two combinations of SagABCD expression vectors were co-transformed into BL21-DE3 cells (Invitrogen) and plated on LB agar plates with 100 μg/mL ampicillin and 35 μg/mL chloramphenicol. The resulting colonies were then tested for their ability to produce a hemolytic molecule. 10 mL LB cultures with 100 μg/mL ampicillin and 35 μg/mL chloramphenicol and grown overnight at 37°C with shaking. 3 10 mL LB cultures were inoculated with each of these cultures and grown to OD_{600} 0.7. Two of the cultures were induced with 0.8 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and BSA was added to one of the induced cultures. An aliquot was removed at 2, 4, and 6 hours for hemolytic assay testing.

Results

We developed a multi-plasmid based expression strategy as a means to successfully reconstitute SLS activity by heterologous expression in lab strains of *E.*

coli using a pETDUET based protein expression system. This is a pET-derived *E. coli* expression system that contains two unique cloning sites for the simultaneous expression of two genes off of a single plasmid. The pETDUET expression system can be extended such that up to four expression plasmids can be introduced into a single bacterium, each with an independent origin of replication and two cloning sites. In this manner, up to eight proteins can be expressed simultaneously, under individual inducible control. This system has been used previously in other systems to characterize protein complexes in eukaryotic systems (13, 16). The overall organization of the plasmids used for heterologous reconstitution of SLS in *E. coli* is depicted in **Figure 5.1**. Our previous *in vitro* reconstitution studies demonstrated that the precursor protein SagA and the synthetase complex SagBCD are necessary and sufficient to convert the SagA precursor into an active hemolysin. Thus we reasoned that cloning of these four genes into pETDUET vectors would drive sufficient expression of the minimal products necessary to produce a functional SLS-like toxin in *E. coli.* These plasmids, designated pETDUET-1-SagBC and pACYCDuet-1-SagDA were transformed into *E. coli* expression strain BL21-DE3 (Invitrogen), and positive transformants were selected by plating on LB agar plates containing appropriate antibiotics. As a safety measure, antibiotic selection pressure ensures that SLS production by *E. coli* is reversed when antibiotics are not present in growth media. To test the ability of pETDUET-SLS *E. coli* to produce a hemolytic molecule, we induced *E. coli* containing SagABCD by 0.8 mM isopropyl-β-D-thiogalactopyranoside (IPTG); BSA was added to one of the induced cultures as a stable carrier protein for active SLS. Supernatants were collected and processed for

SLS activity measurement by a microtiter hemolysis assay. **Figure 5.2A** demonstrates levels of hemolytic activity present after protein induction and SLS production by *E. coli.* SLS activity from *E. coli* supernatants was also strictly concentration dependent, as characteristic of an active hemolysin. Taken together, these findings represent the first heterologous expression of SLS activity using *E. coli*-based bioengineering methods. One substantial advantage of our bioengineering strategy is the ability to rapidly screen enzyme mutants for activity in a high-throughput fashion. We performed mutagenesis studies to examine the importance of putative conserved residues present in the amino acid sequence of one of the candidate synthetase enzymes, SagC. Based on sequence homology of SagC in *Streptococcus pyogenes* to SagC-like genes in the closely related pathogens *Streptococcus iniae, Staphylococcus aureus,* and *Clostridium botulinum,* we probed the functional importance of several candidate residues using our pETDUET SLS system. Quikchange site-directed PCR mutagenesis was used to generate point mutants in SagC (Invitrogen). **Figure 5.2B** shows a partial alignment of the SagC amino acid sequences from *Streptococcus pyogenes, Streptococcus iniae, Staphylococcus aureus,* and *Clostridium botulinum,* and depicts the selected conserved residues that were probed for function. Hemolytic assays performed on supernatants of pETDUET SLS SagC mutants demonstrated that mutation of many of these conserved residues greatly lowered SLS-toxin production **(Figure 5.2C).** Importantly, these initial findings demonstrate that our pETDUET-SLS system can be used to quickly probe structure/function relationships that can be investigated further by biochemical approaches.

Discussion

We have reported here the first successful multi-plasmid system for the expression of the SLS biosynthetic gene cluster in *E. coli.* This method holds promise for investigating the biosynthesis of ribosomally-derived compounds from microorganisms that are difficult to culture in sufficient yields for natural product extraction and identification. We are currently using this system to characterize several other TOMM family members. Finally, it has already been demonstrated that natural products derived from posttranslationally modified bacteriocins such as TOMMs can have a range of medically useful activities (8). This system will allow researchers to probe a vast chemical space in a high-throughput manner by producing libraries of enzyme combinations and unnatural precursor peptides. We are currently using this system to explore the potential of the pETDUET *E. coli* system to produce a library of novel heterocyclic bacteriocins to screen for medically relevant, active compounds.

Acknowledgements

We thank Clayton Thomas and the members of the Dixon Lab for helpful comments. Research was supported by 5R01GM 903 28-07. ALM was supported by the Ruth L. Kirschstein National Research Service Award NIH/NCI T32 CA009523.

Table 5.1 pETDUET Construct Information

Name	Vector	Genes Inserted
SagBC	pET-DUET1	His-SagB, Untagged SagC
SagDA	pACYCDuet-1	Untagged SagD, MBP-SagA

Table 5.2 List of oligonucleotide primers used in this study

Name	Sequence 5' to 3'
SagC_for	AAAAAAAACATATGAAATATCAACTTAATAGTAATG
SagC_rev	AAAACTCGAGTCGATTACTCGTCAAGGAG
SagD_for	AAAAACCATGGGATCCATGTTATACTATTATCCTTCTTTTAATCATATTTTTGATG
SagD_rev	AAAAGCGGCCGCGAATTCTAAGGCATTGG
pet Upstream	ATGCGTCCGGCGTAGA
DuetDown1	GATTATGCGGCCGTGTACAA
DuetUP2	TTGTACACGGCCGCATAATC
T7 Terminator	GCTAGTTATTGCTCAGCGGT

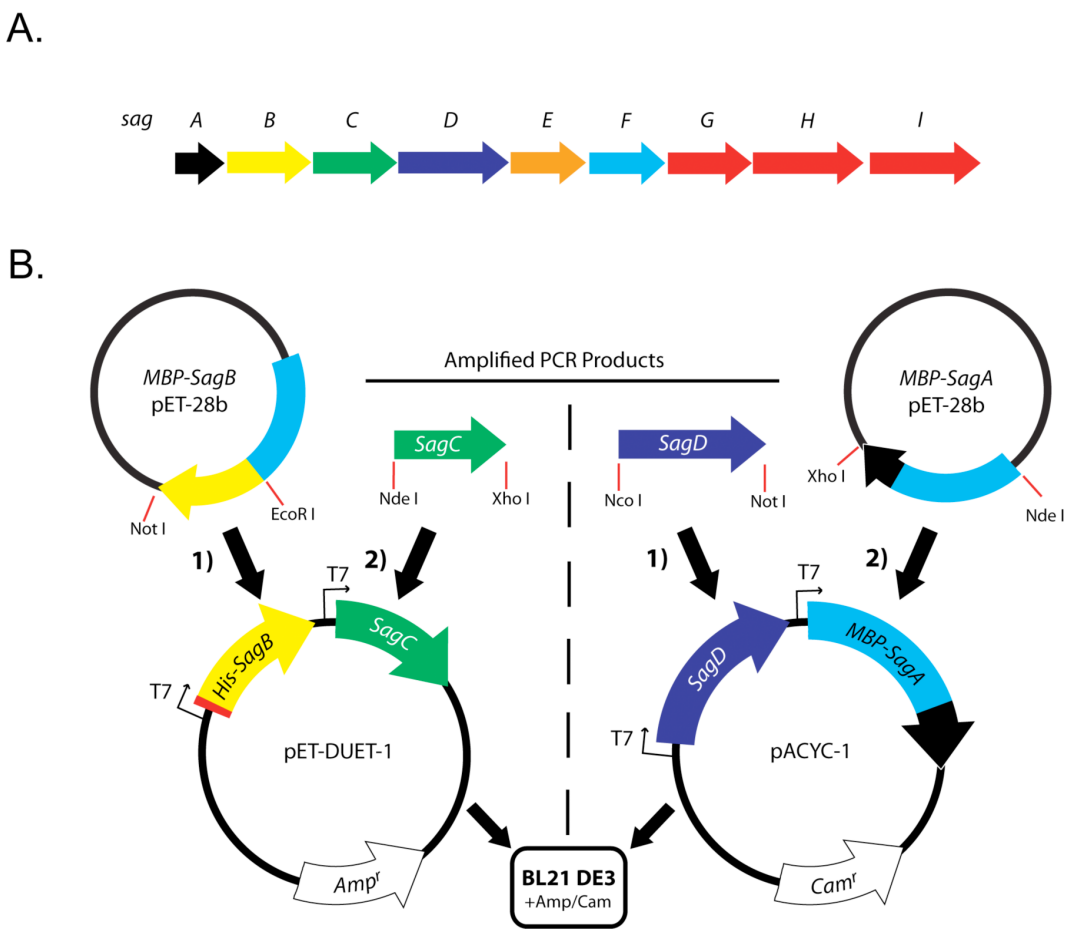


Figure 5.1 Overview of the SLS expression cluster and expression strategy.
(A) Architecture of the Group A *Streptococcus* SLS expression operon. *sagA* encodes the substrate peptide SLS precursor, *sagB, sagC, and sagD* encode the heterocyclic modification enzymes for SLS maturation. (B) Flowchart depicting cloning strategy for pETDUET-SLS plasmid construction and gene expression. Overall organization of the plasmids used for heterologous reconstitution of SLS in *E. coli* is indicated. The genes were inserted into the pETDUET expression plasmid sequentially either by subcloning (*SagA* and *SagB*) or PCR amplification (*SagC* and *SagD*). Once the clones were made they were both transformed into BL21 DE3 cells under double selection for protein expression using the T7 promoter system.

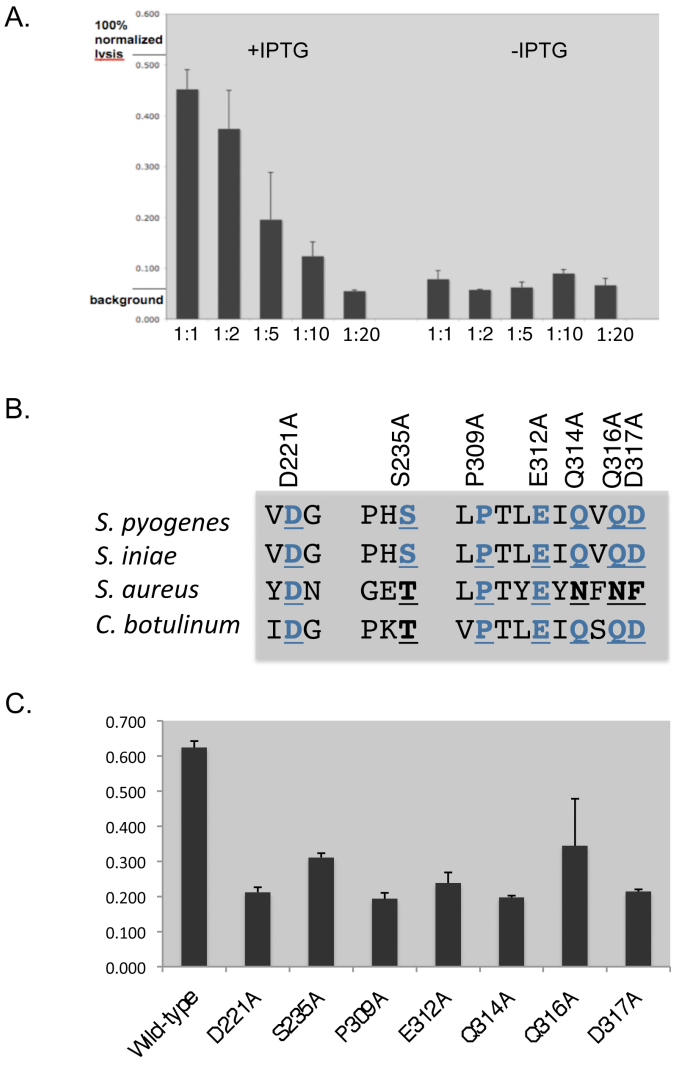


Figure 5.2 Hemolytic activity of the *E. coli* pETDUET-SLS system.
(A) Hemolytic activity of the produced toxin is concentration dependent (Left bars). Dilution indicates ratio of induced culture supernatant to PBS containing defibrinated sheep blood. Hemolytic activity at 1:1 dilution is almost identical to activity produced by wt GAS as indicated by the line indicating normalized 100% lysis. Uninduced cultures (right bars, -IPTG) do not express genes needed to biosynthesize toxin, and do not produce hemolytic toxin (levels of uninduced cultures are near background levels). (B) Partial alignment of SagC amino acid sequences from *S. pyogenes, S. iniae, S. aureus, C. botulinum*, showing candidates used for mutagenesis studies. Each residue was mutated to alanine in the inserted pETDUET-SLS genes using Quikchange mutagenesis kit. **C.** Hemolytic activity of the wt pETDUET-SLS toxin and SagC mutants for comparative study. Bars indicate OD600 values as relative measures of hemolysis.

References

1. **Bernheimer, A. W.** 1967. Physical behavior of streptolysin S. Journal of Bacteriology **93:**2024-5.

2. **Carapetis, J. R., A. C. Steer, E. K. Mulholland, and M. Weber.** 2005. The global burden of group A streptococcal diseases. Lancet Infect Dis **5:**685-94.

3. **Datta, V., S. Myskowski, L. Kwinn, D. Chiem, N. Varki, R. Kansal, M. Kotb, and V. Nizet.** 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Molecular Microbiology **56:**681-695.

4. **Heddle, J. G., S. J. Blance, D. B. Zamble, F. Hollfelder, D. A. Miller, L. M. Wentzell, C. T. Walsh, and A. Maxwell.** 2001. The antibiotic microcin B17 is a DNA gyrase poison: characterisation of the mode of inhibition. Journal of Molecular Biology **307:**1223-34.

5. **Humar, D., V. Datta, D. J. Bast, B. Beall, J. C. S. De Azavedo, and V. Nizet.** 2002. Streptolysin S and necrotising infections produced by group G streptococcus. Lancet **359:**124-9.

6. **Lee, S., D. Mitchell, A. Markley, M. Hensler, D. Gonzalez, A. Wohlrab, P. Dorrestein, V. Nizet, and J. Dixon.** 2008. Discovery of a widely distributed toxin biosynthetic gene cluster. Proceedings of the National Academy of Sciences **105:**5879.

7. **Lin, A., J. Loughman, B. Zinselmeyer, M. Miller, and M. Caparon.** 2009. Streptolysin S Inhibits Neutrophil Recruitment During the Early Stages of Streptococcus pyogenes Infection. Infection and Immunity.

8. **Melby, J. O., N. J. Nard, and D. A. Mitchell.** 2011. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. Current opinion in chemical biology.

9. **Miyoshi-Akiyama, T., D. Takamatsu, M. Koyanagi, J. Zhao, K. i. Imanishi, and T. Uchiyama.** 2005. Cytocidal effect of Streptococcus pyogenes on mouse neutrophils in vivo and the critical role of streptolysin S. J Infect Dis **192:**107-16.

10. **Nizet, V.** 2002. Streptococcal beta-hemolysins: genetics and role in disease pathogenesis. Trends Microbiol **10:**575-80.

11. **Nizet, V., B. Beall, D. J. Bast, V. Datta, L. Kilburn, D. E. Low, and J. C. De Azavedo.** 2000. Genetic locus for streptolysin S production by group A streptococcus. Infect Immun **68:**4245-54.

12. **Schmidt, E., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. Haygood, and J. Ravel.** 2005. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proceedings of the National Academy of Sciences of the United States of America **102:**7315-20.

13. **Small, E., A. Eggler, and A. D. Mesecar.** 2010. Development of an efficient E. coli expression and purification system for a catalytically active, human Cullin3-RINGBox1 protein complex and elucidation of its quaternary structure with Keap1. Biochemical and Biophysical Research Communications **400:**471-5.

14. **Sudek, S., M. Haygood, D. Youssef, and E. Schmidt.** 2006. Structure of Trichamide, a Cyclic Peptide from the Bloom-Forming Cyanobacterium Trichodesmium erythraeum, Predicted from the Genome Sequence. Applied and Environmental Microbiology **72:**4382.

15. **Sumitomo, T., M. Nakata, M. Higashino, Y. Jin, Y. Terao, Y. Fujinaga, and S. Kawabata.** 2011. Streptolysin S contributes to group A streptococcal translocation across an epithelial barrier. J Biol Chem **286:**2750-61.

16. **Tolia, N. H., and L. Joshua-Tor.** 2006. Strategies for protein coexpression in Escherichia coli. Nat Methods **3:**55-64.

17. **Yorgey, P., J. Lee, J. Kördel, E. Vivas, P. Warner, D. Jebaratnam, and R. Kolter.** 1994. Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. Proc Natl Acad Sci USA **91:**4519-23.

Chapter VI

Validation of small, unannotated ORFs as encoded substrates for thiazole/oxazole modified microcin synthetase complexes

Introduction

Previous bioinformatic efforts have identified over 250 new thiazole/oxazole-modified microcin (TOMM) natural product gene clusters widely distributed among prokaryotes (7). By definition, all TOMM gene clusters have a precursor gene, referred to as gene (A), as well as genes encoding a cyclodehydratase (B), oxidoreductase (C), and, in most cases, a scaffolding protein (D). The TOMM modifying enzymes bind to the N-terminus of the precursor peptide and modify cysteine, serine, and threonine residues on its C-terminus (11). The substrate binding region is then cleaved off by a protease leaving the modified, active C-terminal natural product (8). Compared to the surrounding gene cluster, the C-terminus of the precursor gene constitutes a hyper-variable region, which codes for in a diverse set of substrate peptides (4). Because of its small size and variable sequence, the precursor gene was incorrectly annotated in approximately one-third of the gene clusters and had to be identified through manual searches of the genomic DNA (Fig. 2.1) (7). Precursor peptide genes were identified by searching for small open reading frames encoding between 50 and 100 amino acids with an α-helical N-terminus and Cys/Ser/Thr/Gly-rich C-terminus (12, 15). The peptide genes normally are found within 5 kb of the cyclodehydratase, although some TOMMs precursor genes have been found more than a megabase away (6, 7).

This study seeks to validate the expression of these ORFs and show that the peptide products are TOMM substrates. Five diverse TOMM gene clusters were chosen, each with an unannotated ORF. Reverse transcriptase PCR (RT-PCR) analysis

of *Bacillus amyloliquefaciens*, *Pseudomonas putida*, KT 2440 and *Bradyrhizobium japonicum* USDA 110 demonstrated transcript expression. A fifth organism, the thermophile *Pyrococcus furiosus*, could not be cultured because of its extreme and stringent growth conditions. A knockout of the substrate gene of *Pseudomonas putida* was then constructed as a control and to aid in the search for the TOMM natural product.

Materials and Methods

Growth and reverse transcription

Cultures of *B. amyloliquefaciens* and *P. putida* were grown overnight in 10 mL LB media cultures of at 37°C with shaking. *B. japonicum* was grown in AG media at 30°C for 72 hours. The optical density of the cultures was then taken at 600 nm, and 1 volume of cells at OD_{600} = 1 was added to 2 volumes of Bacterial Protect solution (Qiagen). The solution was incubated at room temperature for 5 min and then spun down at 5000 × *g*. A Qiagen RNA extraction kit was used in extracting the transcript RNA from the cell pellet according to the manufacturer's protocol for bacterial cells. Before elution, the remaining DNA was removed by using a Qiagen RNAse Free DNAse Set. The purified RNA was then eluted in 30 μL RNAse free water. A Superscript III kit (Invitrogen) was used to carry out reverse transcription according to the manufacturer's protocols.

RT-PCR

The expression of every gene in each of the three TOMM natural product gene clusters was tested by RT-PCR reactions. The primers used for these PCR reactions are shown in Table 6.1. Each 25 μL PCR reaction included 10 ng of the bacterial cDNA, Ex Taq buffer (TaKaRa), 8% DMSO, 400 nM primers, 400 μM dNTP, and 0.25 μl Taq polymerase. 30 cycles of RT-PCR were carried out under the following conditions: denaturation (94 °C, 30 s); annealing (52 °C, 60 s); elongation (72 °C, 45 s).

Insertional inactivation of *P. putida* precursor peptide

The pJQ200sk suicide vector system was used for the allelic exchange (13). The strains constructed for these experiments are listed in Table 6.2. Rifampicin-resistant isolates of *P. putida* were produced by sequentially plating *P. putida* KT2440 on LB agar plates containing rifampicin concentrations increasing from 10 μg/mL to 50 μg/mL (2). In each plating round, individual colonies were picked, PCR tested for the presence of the *P. putida* TOMM genes, and then streaked on a plate containing a higher concentration of rifampicin. The PutidaA pJQ200sk suicide vector was assembled by amplifying 1-kb fragments of the *P. putida* genome immediately upstream and downstream of the *putidaA* gene using the PutidaA upstream and PutidaA downstream primer pairs shown in Table 6.1. The first and last 15 base pairs of *putidaA* were included in the upstream and downstream allelic exchange regions, respectively. Not I/BamHI endonucleases were used to insert the PutidaA upstream amplified region into plasmid pJQ200sk; then the BamHI/Xho I endonuclease pair was used to insert the PutidaA downstream amplified region into the plasmid. Finally,

the *kanamycin* resistance gene was amplified and inserted (primers, Table 6.1) into the BamHI restriction site. Each cloning step was validated by DNA sequencing (ETON Biosciences). The resulting suicide vector was transformed into *E. coli* DH5α (called ppuApJQ200sk). Rif^R *P. putida*, *E. coli* ppuApJQ200sk, and *E. coli* SM10 helper cells were grown overnight 37 °C with shaking in LB media containing Rif, Kan, and Kan antibiotics, respectively. Aliquots from these cultures were then sequentially spread on LB agar plates without selection and allowed to grow overnight. The plates were scraped, and the bacteria were resuspended in 1 mL of LB. 25 μL of this mixed bacterial culture was then spread onto each of 40 plates containing 25 μg/mL kanamycin and incubated overnight at 37 °C. Single crossover colonies were selected with putidaoutside, putidaA and kantoputidaB primer pairs (Table 6.1). Positive colonies were then grown up in LB with 25 μg/mL kan (overnight, 37 °C) and then spread on LB agar plates with 25 μg/mL Kan and 5% sucrose. Colonies harboring double crossover recombinations were identified by PCR tests and DNA sequencing (ETON Biosciences).

Recomplementation of putidaA on plasmids

PutidaA was amplified from *P, putida* genomic DNA by using the PutidaAwt and PutidaAFLAG primer pairs shown in Table 6.1 and inserted into the pBBRMCS5 expression plasmid (9) by means of Kpn I/Xho I endonucleases. The plasmids were then transformed into *P. putida* Rif^R Δ*ppuA* by electroporation and grown on LB agar plates with 25 μg/mL kan. 5% sucrose and 10 μg/μL gentamycin. RT-PCR experiments were carried out on *P. putida* Rif^R Δ*ppuA, P. putida* Rif^R Δ*ppuA* +*ppuA* and *P. putida* Rif^R Δ*ppuA* + *ppuA-FLAG* as above.

PagA cloning and expression

PagA_for and PagA_rev primers (Table 6.1) were used to amplify the *Pyrococcus furiosus* gene *pagA* from genomic DNA; then BamHI/Not I restriction endonucleases were used to insert the amplified gene into the pET-28b MBP expression plasmid. The resultant ligated plasmids were transformed into *E. coli* DH5α cells and plated on LB agar with 50 μg/μL kanamycin. Several colonies were picked and grown overnight in LB with 50 μg/μL kanamycin before the plasmids were isolated by miniprep (Qiagen). The plasmids were checked by DNA sequencing (ETON Biosciences). MBP-PagA protein was expressed and purified as described previously for MBP-SagA (7).

Hemolytic assays

MBP-PagB, C and D, as well as SagA, B, C and D, were expressed and purified as previously described (7). *In vitro* synthetase assays were performed on PagA+PagBCD, PagA+SagBCD, SagA+PagBCD, and SagA+SagBCD. PagA, SagA, PagBCD, and SagBCD were each added to the synthetase reaction solution individually as a negative control for hemolytic activity as previously described (7). After the 16 h reaction, an aliquot of each sample was added to defibrinated sheep blood (Hemostat Laboratories) and placed at 37 °C for 8 h (7).

Results

RT-PCR experiments showed that the transcripts for the *B. amyloliquefaciens, P. putida* and *B. japonicum* precursor peptide genes were all amplified (Figure, 6.2). In order to better understand the architecture of the TOMM operons, all genes in the *B. amyloliquefaciens* gene cluster were tested for transcript expression (Figure 6.3 A). Primer pairs were also chosen spanning pairs of sequential genes to test whether the *B. amyloliquefaciens* gene cluster was expressed as polycystronic mRNA (Figure 6.3 B). As expected, all genes in the cluster were expressed. It was also determined that the processing and export genes of the cluster were expressed as two polycystronic mRNA transcripts, F−I and J−E. Despite repeated experiments, attempts to amplify a cDNA containing portions of the I and J genes were unsuccessful.

The *P. putida* TOMM gene cluster was chosen for validation in this study due to the fact that its architecture is very similar to the microcin B17 gene cluster. However, with the exception of *ppuD*, none of the genes in the cluster had been annotated. RT-PCR experiments identified transcript mRNAs for all genes in the cluster (Figure 6.4 A). A knockout of the *ppuA* TOMM precursor peptide gene in *P. putida* was created in an attempt to identify a corresponding phenotype. As expected, the *P. putida ΔppuA* strain failed to show precursor peptide expression by RT-PCR (Figure 6.4 B). The *ppuA* knockout did not substantially change transcript expression for the other genes in the cluster (Figure 6.4 B and data not shown). Finally, the *ppuA* gene recomplemented into the *ΔppuA* strain on a plasmid both in a wild type form and with a C-terminal Flag tag. RT-PCR experiments on these bacteria showed *ppuA*

expression in each. Western blot experiments probing of *P.putida ΔppuA + ppuA-Flag* with α-Flag antibodies did not yield any signal in the correct size range for ppuA-Flag (data not shown).

Finally, the unannotated *P. furiosus* TOMM precursor peptide, PagA, was shown to have robust hemolytic activity, similar to the SLS precursor peptide, SagA (Figure 6.5). These data complement our previous work showing that PagBCD can modify SagA to a hemolytic molecule (7).

Discussion

All unannotated TOMM substrates tested were transcribed

TOMM substrate gene mRNA transcripts were identified in the three organisms tested in this study. These data suggest that our methodology for identifying TOMM precursor peptide genes is correct. All genes in the *B. amyloliquefaciens* and *P. putida* clusters were shown to be expressed by RT-PCR, despite the fact that none of genes in the *P. putida* TOMM cluster had been previously annotated, except for *ppuD.* The *B. amyloliquefaciens* TOMM gene cluster was found to contain two distinct polycystronic regions. The *P. putida* cluster was not analyzed for polycystronic regions. As expected, knockouts of *ppuA* led to loss of the mRNA signal. The transcript signal was restored when *ppuA* was recomplemented on a plasmid.

The *P. putida* and *B. japonicum* TOMM natural products have not yet been identified.

Repeated efforts to identify the *P. putida* and *B. japonicum* TOMM natural

products were unsuccessful. Extraction efforts were made more difficult by the fact that no activity could be traced to the TOMM natural products in either organism. I centered my identification efforts on the *P. putida* TOMM because of its similarity to microcin B17 and the relatively low number of cysteines in its precursor peptide sequence compared to the SLS precursor peptide. I carried out multiple natural product extractions on the wildtype and *ΔppuA P. putida* under conditions used to extract microcin B17 (5, 14) and used differential mass spectrometry to identify potential TOMM ions. Ions identified in these experiments were either shown not to be TOMM natural products or were not found in subsequent extractions (data not shown). A Flag tag fusion peptide gene was constructed and transformed into the *ΔppuA* strain in an attempt to find the resultant natural product by western blot using Flag antibodies, however no signal was found (data not shown). This could be because the additional amino acid residues in the Flag-tag blocked peptide acceptance by the posttranslational modification enzymes or transporters. Walsh and others found that certain C-terminal modifications of the microcin B17 precursor peptide abrogated its ability to be exported (10). The remote possibility remains that the *P. putida* and *B. japonicum* TOMM precursor peptide genes are transcribed but not translated. Western blots using antibodies specific for these peptides would be the best way to test for translation.

The *Pyrococcus furiosus* TOMM has hemolytic activity

Repeated attempts to grow *P. furiosus* in the lab were unsuccessful, probably because it is an obligate anaerobe that only grows at temperatures above 80 °C (1). As a result, I was unable to obtain RT-PCR data showing PagA expression. Because the

stretch of consecutive cysteines followed by a downstream heterocyclizable residue in PagA fit our model for SLS like natural products, I decided to investigate the gene cluster further. Indeed, despite the evolutionary divide between *S. pyogenes* and *P. furiosus*, both PagBCD and SagBCD can convert PagA to a hemolytic molecule. This result complements our previous data showing that PagBCD can modify SagA to a hemolytic molecule (7).

Recent supporting studies

During the time in which these experiments were undertaken two studies have been published adding more evidence that our previous bioinformatic searches identified the correct substrate genes. Cotter et. al. (3) identified the unannotated *Listeria monocytogenes* TOMM substrate gene as the precursor for listeriolysin S, a hemolytic natural product similar to Streptolysin S in backbone sequence and activity. More recently Scholz et. al. (16) linked a previously unknown *B. amyloliquefaciens* toxin, called plantazolicin, to the TOMM gene cluster studied in this report. The mechanism of toxicity was determined, but plantazolicin was shown to have an inhibitory effect on most Gram positive *Bacilli* surveyed. This type of preferential activity toward related species is common among TOMM natural products. The authors were unable to determine the structure of plantazolicin by mass spectrometry, however MALDI-TOF analysis of cell surface extracts showed that plantazolicin has a molecular mass of 1336 Da corresponding to a 14 amino acid processed peptide (RCTCTTIISSSSTF) (16).

Conclusion

This study provides support for our methodology for identifying TOMM substrate genes, regardless of whether they have been shown to be genes by annotation programs. The largest goals of this study, linking the *P. putida* and *B. amyloliquefaciens* TOMM substrates to new natural products, have not yet been accomplished. The fact that two of the five TOMM clusters investigated in this study have been assigned to new natural products indicates that our previous bioinformatic searches have identified the true TOMM substrate peptides. Furthermore, the *P. furiosus* gene cluster was shown to be capable of modifying the unannotated precursor peptide, PagA, into a hemolytic molecule *in vitro*. Further testing needs to be done on the *P. putida* and yet unconstructed *B. japonicum* wild type and knockout strain pairs in order to identify phenotype differences and isolate compounds for characterization by mass spectrometry. In order to prove that the *B. japonicum* and *P. putida* TOMM enzymes can install heterocycles on their corresponding substrate peptides, I have constructed BradyA-I and PutidaA-F petDUET expression plasmid sets. Fusion tagged BradyA and PpuA will be affinity purified from the *E. coli* BL21 expression cells for analysis by GC-MS-MS. As was the case with the vast majority of previous natural product characterizations, the listeriolysin S and plantazolicin identifications reported on above both had the luxury of starting with a known active compound and linking that activity to the TOMM gene cluster. It is hoped that my work will help lead the

way in being able to identify transient or poorly expressing natural products from a biosynthetic gene cluster identified through bioinformatic searches.

Acknowledgements

We thank Seema Mattoo for providing several bacterial strains and her input throughout the experimental studies. We also wish to thank Douglas Mitchell, Shaun Lee and many members of the Dixon lab for their helpful comments. Some data in Figure 6.3 was collected by Douglas Mitchell and included with permission. ALM was supported by the NIH Molecular Biophysics Training Program (GM08326) and the Ruth L. Kirschstein National Research Service Award NIH/NCI T32 CA009523.

Table 6.1 List of oligonucleotide primers used in this study

Primer Name	Primer Sequence 5’ to 3’	Primer
5' PutidaA RT	GGAATTGGCTTCCGATACAC	RT
3' PutidaA RT	TACCGGGTCGTTGTTAATGC	RT
5' PutidaB RT	CGCCTGCACTAAACTTAGCC	RT
3' PutidaB RT	GGTGACTCCAGTCTCCAGGT	RT
5' PutidaA Outside RT	ATTCTTCGGAGCATGCTGAT	RT
3' PutidaA Outside RT	GTGTATCGGAAGCCAATTCC	RT
5' PutidaC RT	CACACCCGCTACGAAGAAAT	RT
3' PutidaC RT	CTCAAGCGGTAGCCTCGATA	RT
5' PutidaD RT	GACCCAGCGCAGTAAATCAT	RT
3' PutidaD RT	GCAACCGCATGTGTCTCTAA	RT
5' PutidaE RT	CCTCGACCTATACCGCTTCA	RT
3' PutidaE RT	GTGTATCCGGAGACCAGCAT	RT
5' PutidaF RT	AAGCGGGCAAGCTTATCAC	RT
3' PutidaF RT	CCATGAACTGTCCACCACTG	RT
5' Putida AB RT	TGGTTGTGGTGGTAGTGGTG	RT
3' Putida AB RT	AAGTTCATAGCGCTGGAAGG	RT
5' PutidaA Top Long	AAAAAGCGGCCGCGACCCGGTCAAGGCCCTG	RT
3' PutidaA Top	AAAAGGATCCATATTGATTTTCCATATAGGTGTCTCCTTG	RT
5' PutidaA Bottom	AAAAGGATCCCCGGTAACCCTGTAACAAAAAGCG	RT
3' PutidaA Bottom Long	AAAACTCGAGAAAGCCAATGTCTCATCCATGACCTC	RT
pJQ200-sk Upstream	CAAAAGCTGGAGCTCCAC	Seq
pJQ200-sk Downstream	CGCGTAATACGACTCAC	Seq
Kan Sequen Forw	ACAAGGGGTGTTATGAGCC	Seq
Kan Sequen Rev	TGGCTCATAACACCCCTTG	Seq
5' Putida A Kpn I +1	AAAAGGTACCAATGGAAAATCAATATGGTATTAGCGTC	Cloning
3' Putida A Xho I	AAAACTCGAGTTACAGGGTTACCGGGTCGTTG	Cloning
3' Putida A FLAG Xho I	AAAACTCGAGTTATTTATCGTCATCGTCCTTGTAGTCCAGGGTTACCGGGTCGTTG	Cloning
PagA_for	AAAGGATCCATGGTGGAAGGATTTAAATTTGAGATATTTGAG	Cloning
PagA_rev	AAAGCGGCCGCCTAAGGATTGGGTTTAGTTGAGAAGATTTTCC	Cloning

Table 6.2 List of bacterial strains and plasmids used in this study

Bacterial strain or plasmid name	Description	Reference or source
Bacterial strain		
B. japonicum USDA 110		ATCC
B. amyloliquefaciens FZB42		ATCC
P. putida		
KT2440		ATCC
KT2442	KT2440 isolates that are Rif^R	This study
ΔppuA	KT2442 with *ppuA* replaced with the kanamycin resistance gene	This study
E. coli		
SM10	K-*12 thi thr leu tonA lacY supE recA*::RP4-2Tc::Mu Km^R *λpir*	(17)
DH5α	*recA1*	Laboratory stock
Plasmids		
pJQ200sk	Gm^R, *sacB*, suicide vector	
pJQ200sk-ppuA	pJQ200sk derivative with 1kb regions flanking *ppuA* Kan^R	This study
pBBRMCS5	Gm^R broad-host-range vector	[Kovach, 1995]
pBBRppuA	pBBRMCS5 with ppuA wild type	This study
pBBRppuA-Flag	pBBRMCS5 with ppuA C-terminal Flag tag	This study

Organism	Precursor Peptide Sequence
E. coli	MELKASEFGVVLSVDALKLSRQSPLGVGIGGGGGGGGGGSCGGQGGGCGGCSNGCSGGNGGSGGSGSHI
B. japonicum	MADVSLATFHLFDKENVGSRVQLAWRGCG(GCR)$_5$CGVGCRCAGCAVGVVGCAVSCAGCCASWGRCRWC
P. putida	MENQYGISVMELASDTHCDMEAEFFGGSGSAGGCGGSGGCGGGGGCKGGSGGSGGSGGNNGINNDPVTL
P. syringae	MENDYISEFGEVVAVDGQRSSFDRPHTSLGGSCGGQGGGCGGCGGGGCSGGNGGSGGSGPSAPDHV
S. pyogenes	MLKFTSNILATSVAETTQVAPGGCCCCCTTCCFSIATGSGNSQGGSGSYTPGK
C. botulinum	MLKFNEHVLTTTNNSNNKVTVAPGSCCCCSCCCCVSVSVGGGSASTGGGAAAGQGGNS
S. aureus	MMKINNHTINGYSDINSSEAMQYAAGCCSCSCSCSCSCSCSCTSASTAEQ
L. monocytogenes	MNIKSQSSNGYSNNAVGSEAMNYAAGCCSCSCSTCTCTCTCASSAATKM
B. amyloliquefaciens	MEEVTIMTQIKVPTALIASVBGEGQHLFEPMAARCTCTTIISSSSTF
S. acidocaldarius	MRAYYTVRRANLIYQAYSEGRVLFSWFGCWFGCWFGGGGRVSG
P. furiosus	MVEGFKFEIFEEYEEGPILEPLDDDDPIIIIACCCCGGGGDNWKIFSTKPNP

Figure 6.1 Anatomy of annotated and unannotated TOMM substrate sequences. TOMM precursor peptides have an α-helical N-terminus and a C-terminus rich in glycine and heterocyclizable residues. Amino acid residues shown to be heterocyclized are colored red; in precursor peptides where the natural product structure is not known, possible sites of heterocyclization are colored blue. About one-third of all TOMM clusters found to date had an unannotated precursor peptide gene (underlined); we identified these on the basis of their similarity to known precursor peptides and proximity to the modifying enzymes.

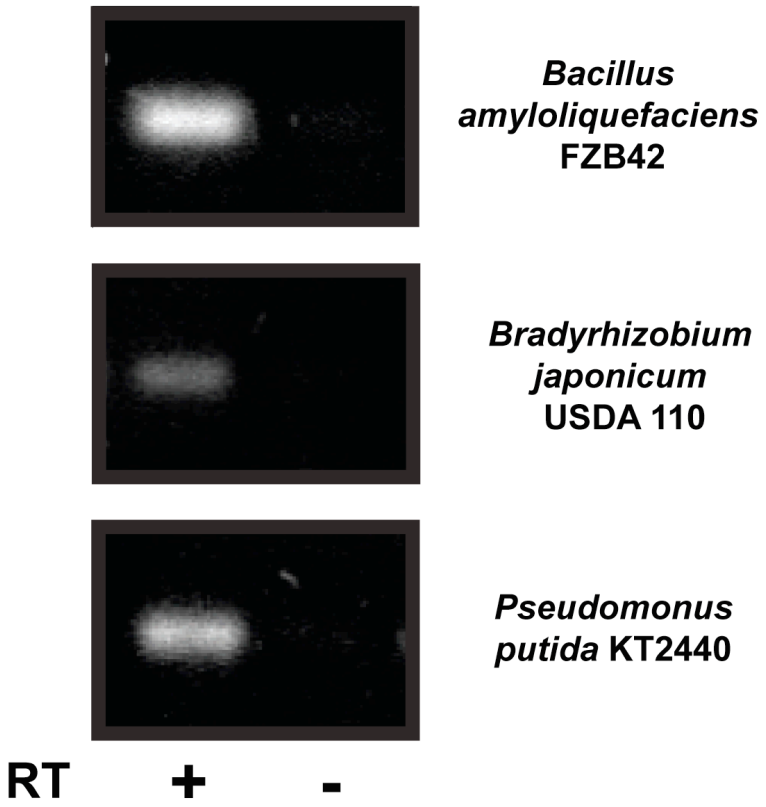


Figure 6.2 TOMM precursor peptide RT-PCR data.
RT-PCR data showing mRNA transcript expression of *B. amyloliquefaciens, B. japonicum and P. putida* unannotated precursor peptide genes. Genomic DNA contamination was assessed from the reverse transcriptase (-) control.

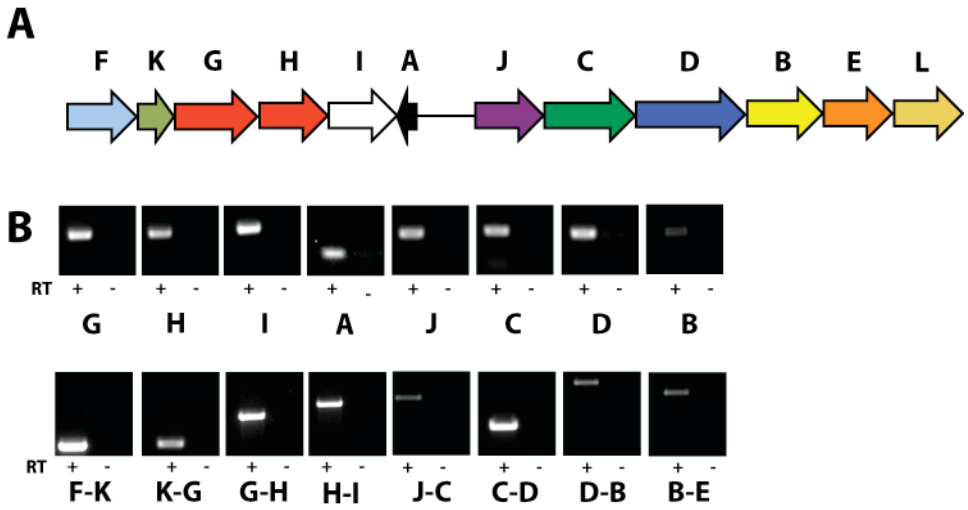


Figure 6.3 *Bacillus amyloliquefaciens* TOMM cluster and RT-PCR data.
(A) Genetic architecture of the *B. amyloliquefaciens* TOMM gene cluster. Notable genes in the cluster include: the precursor peptide, A; oxidoreductase, B; cyclodehydratase, C; docking scaffold, D; zinc protease, E; methyltransferase, L; ABC transporters, GH; and a putative immunity gene, F. (B) Top: RT-PCR showing transcription of genes in the *B. amyloliquefaciens* TOMM gene cluster. Bottom: RT-PCR data showing that portions of the *B. amyloliquefaciens* TOMM gene cluster are transcribed in polycystronic groups. The genes F−I share an mRNA transcript as do J−E.

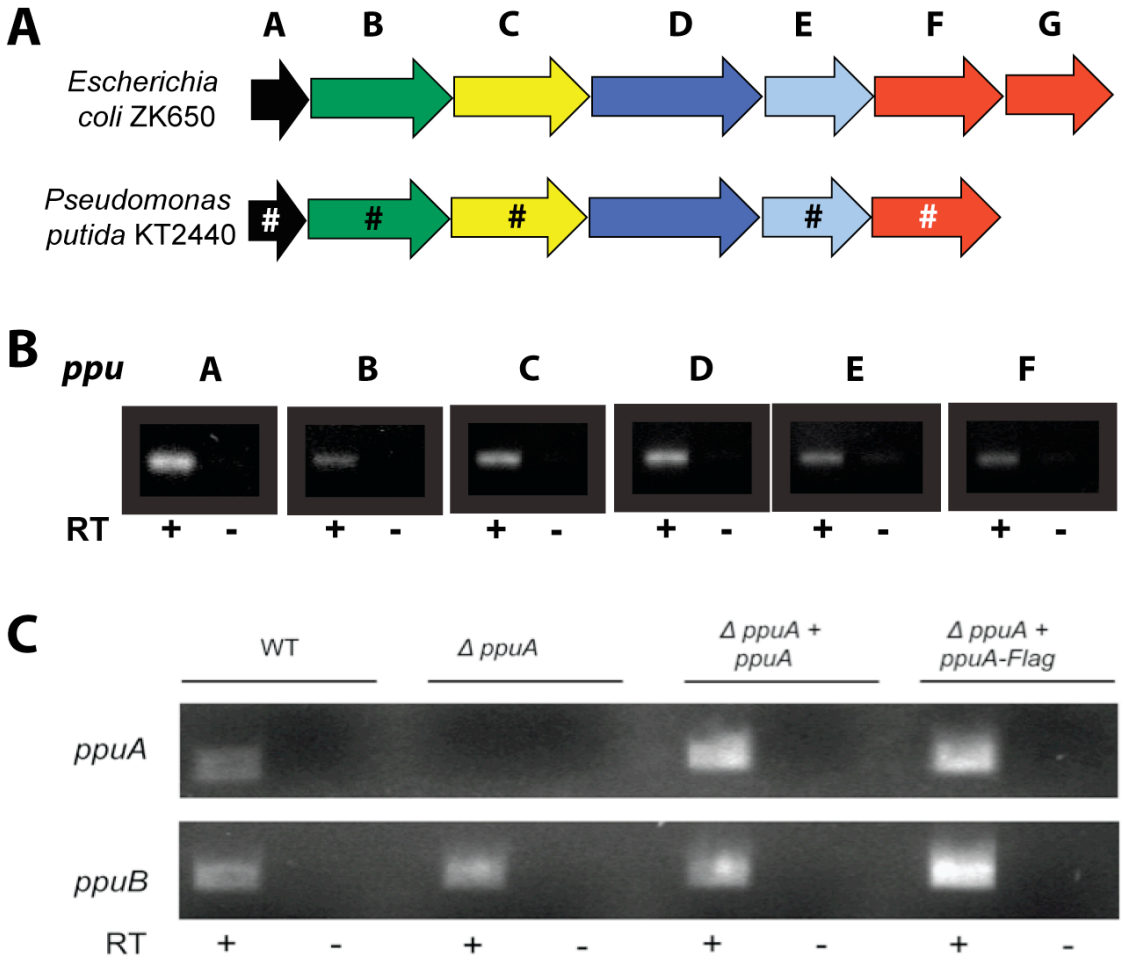


Figure 6.4 *Pseudomonas putida* gene cluster, RT-PCR data, and knockouts.
(A) Relative architecture of the microcin B17 and *P. putida* TOMM gene cluster. The color convention is maintained throughout the dissertation: black, precursor peptide; green, cyclodehydratase; yellow, oxidoreductase; blue, scaffolding; light blue, immunity; red, export. The (#) symbols denote genes that were previously unannotated. (B) RT-PCR analysis of the *P. putida* TOMM gene cluster. (C) Top: RT-PCR analysis of the *ppuA* WT and knockout *P. putida* as well as WT and *ppuA-Flag* recomplemented. Bottom: *PpuB* continued to be expressed in the *ΔppuA* knockout.

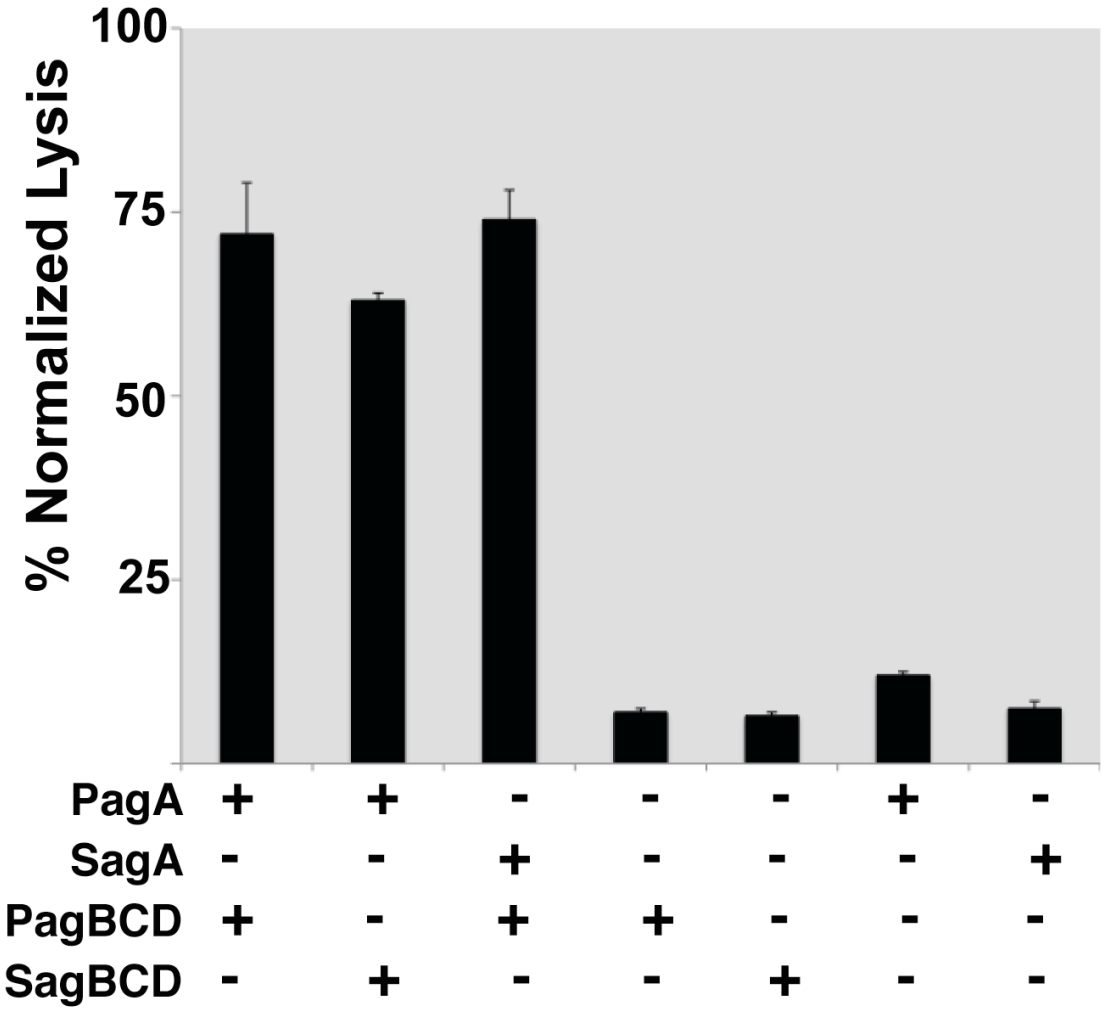


Figure 6.5 *P. furiosus* TOMM is a hemolytic natural product that can also be modified by *S. pyogenes* TOMM enzymes.
The *P. furiosus* TOMM (PagA) was recombinantly produced and tested for hemolytic activity after modification by *P. furiosus* TOMM enzymes (PagBCD) and *S. pyogenes* TOMM enzymes (SagBCD). SagA+SagBCD was used as a positive control, and the precursor peptides and enzymes alone were tested for hemolytic activity as a negative control.

References

1. **Adams, M. W., J. F. Holden, A. L. Menon, G. J. Schut, A. M. Grunden, C. Hou, A. M. Hutchins, F. E. Jenney, Jr., C. Kim, K. Ma, G. Pan, R. Roy, R. Sapra, S. V. Story, and M. F. Verhagen.** 2001. Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon Pyrococcus furiosus. J Bacteriol **183:**716-24.

2. **Boguslawska, J., J. Zycka-Krzesinska, A. Wilcks, and J. Bardowski.** 2009. Intra- and interspecies conjugal transfer of Tn916-like elements from Lactococcus lactis in vitro and in vivo. Applied and Environmental Microbiology **75:**6352-60.

3. **Cotter, P., L. A. Draper, E. M. Lawton, K. M. Daly, D. S. Groeger, P. G. Casey, R. P. Ross, and C. Hill.** 2008. Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I Listeria monocytogenes. PLoS pathogens **4:**e1000144.

4. **Donia, M., J. Ravel, and E. Schmidt.** 2008. A global assembly line for cyanobactins. Nature Chemical Biology **4:**341-343.

5. **Gao, Q., A. Fang, and A. L. Demain.** 2001. Induction of microcin B17 formation in Escherichia coli ZK650 by limitation of oxygen and glucose is independent of glucose consumption rate. J Ind Microbiol Biotechnol **26:**341-4.

6. **Haft, D. H., M. K. Basu, and D. A. Mitchell.** 2010. Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. BMC biology **8:**70.

7. **Lee, S. W., D. A. Mitchell, A. L. Markley, M. E. Hensler, D. Gonzalez, A. Wohlrab, P. C. Dorrestein, V. Nizet, and J. E. Dixon.** 2008. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci USA **105:**5879-84.

8. **Li, Y. M., J. C. Milne, L. L. Madison, R. Kolter, and C. T. Walsh.** 1996. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. Science **274:**1188-93.

9. **Lynch, M. D., and R. T. Gill.** 2006. Broad host range vectors for stable genomic library construction. Biotechnol. Bioeng. **94:**151-8.

10. **Madison, L. L., E. I. Vivas, Y. M. Li, C. T. Walsh, and R. Kolter.** 1997. The leader peptide is essential for the post-translational modification of the DNA-gyrase inhibitor microcin B17. Molecular Microbiology **23:**161-8.

11. **Milne, J. C., R. S. Roy, A. C. Eliot, N. L. Kelleher, A. Wokhlu, B. Nickels, and C. T. Walsh.** 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:**4768-81.

12. **Mitchell, D. A., S. W. Lee, M. A. Pence, A. L. Markley, J. D. Limm, V. Nizet, and J. E. Dixon.** 2009. Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. J Biol Chem **284:**13004-12.

13. **Ramos, H., E. Souza, J. Soaresramos, and F. Pedrosa.** 2006. Suicide vectors for the introduction of genetic markers into Bradyrhizobium strains by site-directed chromosomal integration between repeated sequences (RS-Œ±). Soil Biology and Biochemistry **38:**2481-2486.

14. **Roy, R. S., N. L. Kelleher, J. C. Milne, and C. T. Walsh.** 1999. In vivo processing and antibiotic activity of microcin B17 analogs with varying ring content and altered bisheterocyclic sites. Chem Biol **6:**305-318.

15. **Roy, R. S., S. Kim, J. D. Baleja, and C. T. Walsh.** 1998. Role of the microcin B17 propeptide in substrate recognition: solution structure and mutational analysis of McbA1-26. Chemistry & biology **5:**217-28.

16. **Scholz, R., K. J. Molohon, J. Nachtigall, J. Vater, A. L. Markley, R. D. Süssmuth, D. A. Mitchell, and R. Borriss.** 2011. Plantazolicin, a novel microcin B17/streptolysin S-like natural product from Bacillus amyloliquefaciens FZB42. Journal of Bacteriology **193:**215-24.

17. **Simon, R., U. Priefer, and A. Pühler.** 1983. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. Nat Biotechnol **1:**784-791.

Chapter VII

Conclusion and Future Directions

Conclusion:

I reported the discovery of a new class of ribosomally encoded natural products, with diverse activities, produced by disparate bacterial species. My colleagues and I showed that the important GAS virulence factor, streptolysin S, is a heterocycle containing thiazole oxazole modified microcin (TOMM) peptide and that it has several closely related virulence factors in other prevalent human pathogens such as *Clostridium botulinum, Listeria monocytogenes* and *Staphylococcus aureus.* Important motifs within the SagA precursor peptide were characterized, and the substrate binding site and residues critical for hemolytic activity were identified. After several years of performing mass spectrometry studies on *in vitro* modified SagA and ClosA, structural evidence for heterocycle formation on these peptides was discovered. I then reported the development of a method for engineering *E. coli* to produce modified streptolysin S precursor peptide with hemolytic activity. This method should prove useful for biosynthesizing TOMM natural product libraries in a high throughput manner. Finally, I validated our previous bioinformatic analysis of the TOMM gene cluster family by showing that the unannotated ORFs we identified as TOMM precursor peptide genes are expressed. Though these natural products are difficult to purify and analyze, my colleagues and I have discovered many new natural products and made substantial progress towards developing methods for characterizing them. This research is a stepping-stone to being able to reliably identify novel ribosomally encoded natural products through genome mining.

Future Directions:

Perhaps the most exciting part of the TOMM natural product project is the potential of TOMMs for use in the development of synthetic natural product libraries. We have built on previous research in the microcin B17 biosynthetic system to show that the peptides only need an amino-terminal enzyme binding site and downstream cysteines, serines, and threonines preceded by glycines in order for heterocyclization to occur (JBC, 2009). We have also shown that a fully synthetic peptide sequence that follows these rules can be heterocyclized *in vitro* (unpublished data). Finally, as part of the characterization of the streptolysin S biosynthesis cluster, I developed an *E. coli* expression system where all of proteins required for the production of the natural product were expressed of separate T7 promoters. Upon induction, the *E. coli* expressed the streptolysin S natural product and became highly hemolytic (unpublished data). Using this system, a library of peptide precursor genes could be quickly assembled and expressed, producing a corresponding library of novel, extensively posttranslationally modified natural products for use in activity assays. I have cloned the genes from *P. putida* and *B. japonicum* into *E. coli* in the same fashion to see whether my system is robust enough to work on other natural product biosynthesis systems. If successful, I will use site directed mutagenesis to make a small library of mutants in these precursor peptides as a proof of principal and test them for anti-cancer and anti-microbial activity.

This is an exciting time for research on ribosomally encoded natural products. As the pace of genome sequencing accelerates, I predict the TOMM natural product

family will grow at a similar pace. I believe that our knowledge of this pathway will expand to a point where the final structure of a newly discovered TOMM peptide can be determined solely from the precursor peptide sequence and gene cluster architecture, greatly accelerating the rate of natural product discovery. The diversity of species, precursor peptide sequence, tailoring enzymes and known biological activities that TOMM natural products represent leads me to believe that genomics based natural product discovery will only grow in importance.

Acknowledgements

I thank Douglas Mitchell and Vincent Tagliabracci for helpful discussions during the development of this chapter.