UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Use of Mass Spectrometry to Characterize the Metabolic Output of Bacterial Pathogens

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Chemistry

by

David Gonzalez

Committee in charge:

Professor Pieter C. Dorrestein, Chair
Professor Simpson Joseph
Professor Elizabeth Komives
Professor Victor Nizet
Professor Kimberly Prather

2011

The dissertation of David Gonzalez is approved, and acceptable in quality and form for publication electronically and on microfilm:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2011

DEDICATION

This thesis is dedicated to my faith, family, and friends. **FAITH:** From the first day at Miracosta College to sitting here in front of my computer preparing for a Ph.D thesis, the “Our Father” has continued to give me the strength to surpass the many hurdles encountered in this long academic journey. **FAMILY:** Thank you to my beautiful wife of fifteen years Yolanda Gonzalez for being patient, understanding, and a companion in this journey. Thank you to my son David for his friendship, love, and continued support. Do you remember when I used to ride you on the handle bars of your bike coming back from elementary school? Thank you to my son Isaiah for his friendship, love, and continued support. Do you remember helping me edit papers and analyze data? Thank you to my Mom for teaching me the meaning of accountability and to fight for what you believe in no matter the consequences. Thank you to my Dad for his love, friendship and demonstrating that one can teach with compassion. Thank you to my sisters and brother for being there at times of need. **FRIENDS:** To all the “good people” I’ve met along the way that have helped or inspired me to reach my full potential as a student. Thank you to all the scientific collaborations and the many wonderful scientists involved in the projects. A special thanks to my friend and mentor Professor Pieter C. Dorrestein for taking a chance on me as his first Ph.D student. The guidance and scientific education received from Pieter have been one of a kind. Thank you for not letting me give up!

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Pieter C. Dorrestein for his support as the chair of my committee. Professors Kimberly Prather, Simpson Joseph, Victor Nizet, and Elizabeth Komives for being members of the committee panel. I would also like to acknowledge the Victor Nizet (Victor, Mary, Nina[2], Samira, Andrew), Jack E. Dixon (Jack, Shaun, Doug, Andrew), Kit Pogliano (Kit, Tinya) labs, without their support my research and scientific findings would have not progressed to a level suitable for a doctorate degree. Chapter 2, in full, is a reprint of the material as it appears in Lee et al. PNAS 2008 105(15):5879-84. The dissertation author was a collaborative researcher and co-author of this publication. Chapter 3, in full, is a reprint of the material as it appears in Gonzalez et al. JBC 2010 285(36):28220-8. The dissertation author was the primary investigator and author of this paper. Chapter 4, in part is currently being prepared for submission for publication of the material. The dissertation author was a collaborative researcher, co-author of this publication, and drafted the manuscript. Chapter 5, in full, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this paper.

VITA

Education

2006-2011 Ph.D. in Chemistry, University of California, San Diego
2006-2008 M.S. in Chemistry, University of California, San Diego
2003-2006 B.S. in Chemistry, California State University, San Marcos

Academic Honors and Awards

2011 A.P. Giannini Foundation Medical Research Postdoctoral Fellowship
2011 IRACDA Postdoctoral Fellowship
2010 Yale Bouchet Honor Society Inductee
2010 Annual Compact for Faculty Diversity Institute on Teaching and Mentoring
2010 NIH/NIDDK Heme and Blood Protein Training Fellowship
2009 NIH/NIDDK Heme and Blood Protein Training Fellowship
2008 NIH/NIDDK Heme and Blood Protein Training Fellowship
2007 GAAN Research Fellowship
2006 AGEProfessoriate First-year Doctorial Fellowship
2006 Competitive-edge Predoctoral Research Award
2005 Poster presentation award at the Annual Biomedical Research Conference
2005 UCSD Summer Training Academy for Research in the Sciences
2004 Selected as a CSUSM Research Initiative for Scientific Enhancement Scholar

Research Experience

2006-3/4/2011 Research Associate, Pieter Dorrestein Research Group, UCSD.

Design and develop experiments involving Fourier Transform Ion Cyclotron Mass Spectrometry (FTICR-MS), bottom-up mass spectrometry, molecular biology techniques, chromatographic separation, and matrix assisted laser desorption ionization mass spectrometry imaging . Research efforts have involved integrating these tools in order to investigate orphan gene clusters that are involved in a human pathogen's infectious cycle. In particular, my current project has focused on a gene cluster responsible for the formation of a hemolytic/cytolytic factor from *Clostridium botulinum*, the organism associated with botulism. Additionally, my studies have extended to investigate the pathogenicity associated with methicillin-resistant *Staphylococcus aureus*, focusing on epidemic strains.

2005-2006 Research Assistant, Simpson Joseph Research Group, UCSD.

Project 1 involved investigating the role of a Watson-Crick base pair between the tRNA acceptor arm and 23S rRNA. My main contribution to this project was constructing tRNA mutants and isolating the 30S ribosomal subunits (wild-type and mutant) which were used in kinetic translocation assays. Project 2 investigated the interaction of initiation factor III (IF3) with the 30S small ribosomal subunit. The mechanism of interaction was assayed by FRET between a fluorescently labeled IF3 and the 30S ribosomal.

Teaching Experience

2007-2008	TA, UCSD. General Chemistry, Molecular Biology Laboratory
2006-2007	TA, UCSD. Biochemistry Laboratory and Organic Chemistry
2005-2006	TA, CSUSM. Chemistry of Life, General Chemistry
2004-2011	Tutor, Independent. Algebra, G. Chem, O. Chem, Biochem, Biology

Oral and Poster Presentations

2011 UCSD Chemistry/Biochemistry PhD recruitment seminar, Poster
2010 Office of Biomedical and Research Training, CSUSM, Invited Presenter
2010 Pharmacology Research Seminar Series, UCSD, Invited Presenter
2010 Heme and Blood Protein Symposium, UCSD, Invited Presenter
2010 Natural Product Affinity Group Seminar Series, UCSD, Invited Presenter
2010 Yale University Bouchet Honor Society Seminar, Poster
2009 Office of Biomedical and Research Training Roundtable Discussion, CSUSM
2009 Heme and Blood Protein Symposium, CSUSM, Poster
2007 Office of Biomedical and Research Training Roundtable Discussion, CSUSM
2006 Office of Biomedical and Research Training Roundtable Discussion, CSUSM
2005 UCSD Summer Training Academy for Research in the Sciences, Poster
2005 Annual Biomedical Research Conference for Minority Students, ABRCMS, Poster

Publications

1. **Gonzalez D.**, Haste N.M, Hollands A., Fleming T., Hamby M., Pogliano K., Nizet V. and Dorrestein P.C. Insight Into Probiotic Effects by *Bacillus subtilis* Against *Staphylococcus aureus*. Submitted Microbiology Feb 2011

2. **Gonzalez D.**, Lee S.W., Hensler M.E., Mitchell D.A., Markley A.L., Nizet V., Dixon J.E., Dorrestein P.C. Clostridiolysin S: a post-translationally modified biotoxin from *Clostridium botulinum.* Journal of Biological Chemistry. 2010 Sep 3:285(36)

3. Lee S.W., Mitchell D.A., Markley A.L., Hensler M.E., **Gonzalez D.**, Wohlrab A., Dorrestein P.C., Nizet V., Dixon J.E. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci USA 2008 Apr 15;105(15):5879-84

4. Simmons, T. L., Coates R.C., Clark B. R., Engene N., **Gonzalez D.**, Esquenazi, E., Dorrestein P. C., Gerwick W. H. Biosynthetic Origin of Natural Products Isolated from Marine Microorganism-Invertebrate Assemblages. Proc. Natl. Acad. Sci. USA. 2008 Mar 25; 105(12): 4587-94

5. Esquenazi, E., Coates, C., Simmons, L., **Gonzalez, D.,** Gerwick, W.H., Dorrestein, P.C. Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging. Molecular BioSystems. 2008 Jun; 4(6): 562-70. ****Cover article**** **

6. Jones A., Gerwick L., **Gonzalez D.**, Dorrestein, P.C., Gerwick W. H. Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. BMC Microbiology. 9:247 Oct 2009

7. He C., Nora G., Schneider E., Kerr I.D. , Hansell E., Hirata K., **Gonzalez D.**, Sajid M., Boyd S.E., Hruz P., Cobo E., Le C., Liu W., Eckmann L., Dorrestein P.C., Houpt E.R., Brinen L.S., Craik C., Roush W.R., McKerrow J., Reed S.L. Upregulation of a Novel *Entamoeba histolytica* Cysteine Proteinase is Key for Invasive Amebiasis. Journal of Biological Chemistry. 2010 Jun 11; 285(24): 18516-27.

8. Liu WT., Yang Y., Xu Y., Lamsa A., Yang J., Ng J., Haste N., **Gonzalez D**., Nizet V., Straight P., Pavel P, Pogliano J., Pogliano K., Dorrestein P.C. Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. Proc Natl Acad Sci USA. 2010 107, 16286-16290

9. Udwary D.W, Gontang E.A., Jones A.C., Schultz A.W., Sorrels C.M., Winter J.M., Yang J.Y., Beauchemin N., Capson T.L, Clark B.R., Esquenazi E., Eustáquio A.S., Freel K., **Gonzalez D.**, Gerwick L., Gerwick W.H., Liu W.T., Malloy K.L, Maloney K.N. , Nett M., Nunnery J.K., Penn K., Prieto-Davo A., Simmons T.L., Weitz S., Wilson M.C., Tisa L., Dorrestein P.C., Moore B.S. Comparative genomic and proteomic analysis of the actinorhizal symbiont *Frankia* reveals significant natural product biosynthetic potential. Submitted Microbiology (Invited to revise Feb 2011)

10. Yuquan X.*, Yang Y.*, **Gonzalez D**., Esquenazi E., Liu WT, Edlund A., Du L., Molnár I. Gerwick W. H., Jensen P., Fischbach M., Liaw C., Straight P, Dorrestein P.C. Observing the invisible, a window into the metabolic output of microbial colonies. * Contributed equally **(In preparation)**

11. Van Sorge N., Gusarov I., **Gonzalez D.**, von Köckritz-Blickwede M., Sabina A., Borkowski A., Dorrestein P.C., Nudler E., and Nizet V. Loss of bacterial nitric oxide synthase sensitizes MRSA to host innate defenses and antibiotics. **(In preparation)**

References

Prof. Pieter C. Dorrestein: Professor Skaggs School of Pharmacy and Pharmaceutical SciencesDepartments of Pharmacology, Chemistry and Biochemistry, UCSD (pdorrestein@ucsd.edu)

Prof. Victor Nizet: Professor and Chief of Pharmacology and Drug Discovery, School of Medicine, Department of Pediatrics and Skaggs School of Pharmacy and Pharmaceutical Sciences , School of Medicine, Department of Pediatrics, UCSD (vnizet@ucsd.edu)

Prof. Jack E. Dixon: Professor of Pharmacology, Cellular and Molecular Medicine, and Chem/Biochem, UCSD, Member of the National Academy of Sciences, Vice President and Chief Science Officer, Howard Hughes Medical Institute (jedixon@ucsd.edu)

Prof. Elizabeth Komives: Professor of Chemistry and Biochemistry, Program Director for the Blood and Heme Program Training Program, UCSD (ekomives@ucsd.edu)

ABSTRACT OF THE DISSERTATION

The use of mass spectrometry to characterize the metabolic output of bacterial pathogens

by

David J. Gonzalez

Doctor of Philosophy

University of California, San Diego

Professor Pieter C. Dorrestein, Chair

The overall theme of this thesis is the use of mass spectrometry to answer questions of biological relevance. The thesis begins by informing the reader on the basics of mass spectrometry and associated tools with an introductory chapter. Chapter 2 describes collaborative work performed between three labs at UCSD, the Jack E. Dixon, Victor Nizet and Pieter C. Dorrestein labs. The project was initiated by the use of comparative genomics to discover a widely distribution toxin biosynthetic gene cluster. Molecular and biochemical analysis of two members of the conserved family, namely Group A Streptococcus and *Clostridium Botulinum*, demonstrated the organisms indeed contained similar cellular machinery for the biosynthesis of ribosomally encoded peptides.

Chapter 3 continues characterizing the widely distributed family of toxins specifically the gene cluster within the bacterium *C. Botulinum*. The work describes the use of mass spectrometry to obtain structural evidence, verifying the biosynthetic gene cluster indeed produces heterocycles as hypothesized by the genomic studies.

Additionally, the work uses biochemical and genetic approaches to characterize a novel β-hemolytic and cytolytic toxin in two Clostridia species.

Chapter 4 introduces the method of imaging mass spectrometry (IMS) of microbial colonies. Although IMS is not a novel technique, very few reports had used the technology to investigate the microbial kingdom. The work within chapter 4 describes an IMS survey performed on a vast number of differential microbes to visualize their associated metabolic output. The goal was to show that IMS is an amenable technique that can be applied to a diverse number of microbes. Chapter 5 goes beyond the proposal set forth in chapter 4, as IMS was used to visualize the interaction between two well characterized model bacterial systems, the probiotic bacterium *Bacillus subtilis* and the bona fide human pathogen methicillin resistant *Staphylococcus aureus* (MRSA). Imaging mass spectrometry showed a directional release of compounds by the probiotic organism *B. subtilis* that have antibiotic effects against MRSA giving insight into the potential mode of action for probiotic organisms.

Lastly, the thesis puts forth a set of experimental proposals (A-D) based on continued experiments that extend from Chapters 2-5. Proposals (A) and (B) focus on the developing alternative mass spectral techniques and characterization of the other genes within the conserved family of toxins. Proposals (C) and (D) extend the use of IMS to the field of bacterial pathogenesis. An introduction, preliminary experiments, conclusions and challenges are also provided in order to determine the potential next steps for each project.

Chapter 1

General Introduction

Summary

The main theme of the thesis presented herein is the use of mass spectrometry to answer biologically relevant questions. Therefore, a general introductory chapter is provided with the intention to introduce the reader with the fundamentals of mass spectrometry and specific mass spectral techniques used during the doctoral studies.

A. Mass spectrometry basics

The science of mass spectrometry has made great advances in the last few decades and has become an invaluable tool in basic science research as well as the pharmaceutical industry (1, 2). A mass spectrometer is considered a highly amenable analytical tool since sample introduction can be performed in the gaseous, solid, or liquid state. The field of biological mass spectrometry made great advances by the development of soft ionization methods such as electrospray ionization (ESI) and matrix-assisted desorption ionization (MALDI) (3). Soft ionization techniques allow for the analyte to enter the gas phase with minimal fragmentation; allowing researchers to use mass spectrometry for investigations involving proteins, peptides, carbohydrates, DNA, drugs, therapeutic targets, and other biologically relevant molecules (4, 5, 6, 7). The future is bright for the field of mass spectrometry, as technological advances broaden its applicability on a daily basis as new ionization techniques arise and elaborate ion separation techniques are being developed to examine crude samples which require limited to no sample preparation (8, 9).

A first step in understanding the science of mass spectrometry is to become familiar with the four basic components that comprise an instrument (Figure 1.1) (10,

11). Before exploring the basic components of a mass spectrometer keep in mind that an instrument, although highly technical, can be described as a scale. Simply put, a mass spectrometer weighs objects similar to a weight scale found at a gym or the grocery store. The difference between a common weight scale and a mass spectrometer is the size of the object being weighed. A mass spectrometer has the ability to weigh small molecules not visible to the human eye at very precise measurements (12).

Typically, the first instance a sample encounters the mass spectrometer is at the sample inlet. The sample inlet is located at the front end of the mass spectrometer (Figure 1.1). Sample introduction can be a continuous stream supplied by liquid chromatography or could be a solid sample mounted on a steel plate as the case with matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) (13,14). Once a sample is introduced into the mass spectrometer it is passed through the ionization source (component 2). Many different ionization sources are commercially available; electron impact (EI), electrospray ionization (ESI), MALDI, fast-atom bombardment (FAB), secondary ionization (SIMS) and others (15, 16, 17). All ionization sources function in converting the analyte to a gaseous ionized state that can be measured by the instrument. Thereafter, gas ions are guided by a series of voltages and filters to component 3 the mass analyzer (18). The function of the mass analyzer is to separate ions according to their *m/z* (mass to charge) value. A typical mass analyzer, referred to as time-of-flight (TOF), separates ions according to size, which is governed by the laws of mass as larger ions travel slower and smaller ions travel quicker down the flight tube. The type of mass analyzer used for a certain experiment depends upon

the desired resolution, mass range, scan rate and detection limits required for sample analysis (19). Component 4, the ion detector, recognizes and analyzes ions based upon their charge or momentum. The resulting output from passing a sample through the above mentioned components is a mass spectrum. A mass spectrum typically displays ion count on the y-axis and the m/z value for each individually detected ion on the x-axis.

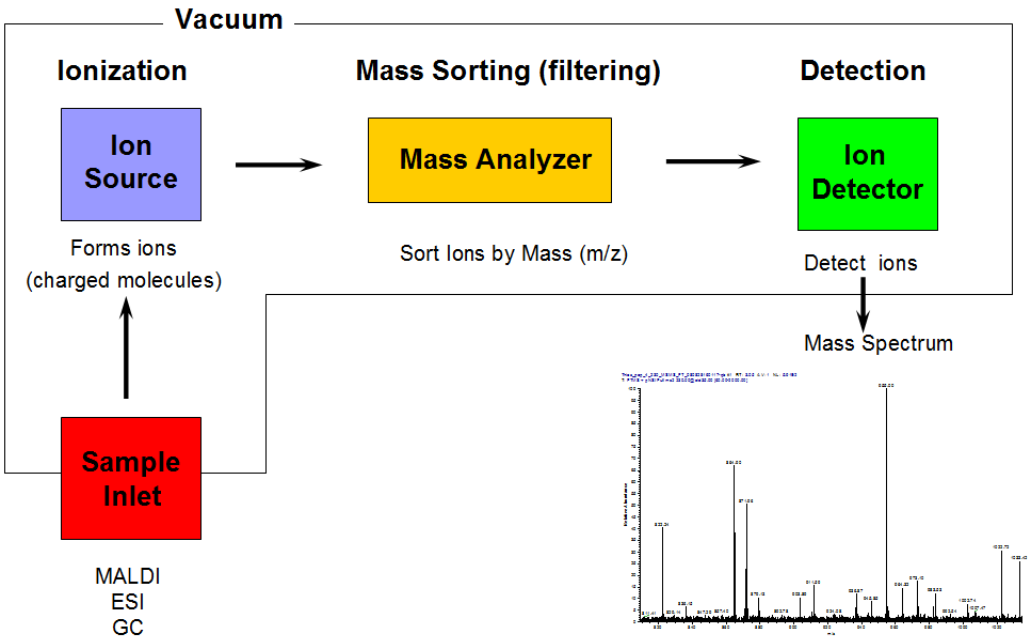

Figure 1.1 Basics of mass spectrometry components. Samples are introduced into the mass spectrometer via the sample inlet (red). Following introduction, samples are charged at the ion source (lavender**).** Thereafter, ions are separated according to m/z values at the mass analyzer (gold). Detection of the ions occur at the ion detector (green) resulting in a mass spectrum.

B. Bottom-up proteomics basics

An organism's complete set of proteins, the proteome, is fundamental to understanding biological processes as proteins are the functional unit of a cell (20). Characterizing an organism's proteome is complicated because a single gene can give rise to a number of different proteins through alternative splicing, RNA editing, and posttranslational modification (21). Proteomics entails the systematic study of proteins in order to obtain a thorough understanding of the structure, function and regulation of a given biological system (22).

Initially proteomics focused on the generation of protein maps using two-dimensional polyacrylamide gel electrophoresis as a means to numerically characterize a proteome. Through the years the field of proteomics has progressed as advancements in technology, particularly analytical instrumentation, have contributed to proteomic discovery capabilities. A powerful tool for defining components of a proteome has proven to be the use of mass spectrometry (23, 24). The use of mass spectrometry in the field of proteomics has expanded the study to include not only protein expression profiling, but the analysis of post-translational modifications and protein-protein interactions (25, 26). Specifically, liquid chromatography, interfaced at the mass spectrometer sample inlet (LCMS), has given the analysis the ability to analyze whole proteomic complex mixtures by separating components as they are introduced into the mass spectrometer. Methods that combine liquid chromatography (LC) with mass spectrometry (MS) such as nanoLC-ESI-MS/MS, and LC-MALDI-TOF/TOF-MS have the ability to identify hundreds to thousands of proteins in a given experiment (27, 28).

Through the progression of mass spectrometry in the field of proteomics, terms have been coined to describe commonly used techniques. One example is bottom-up proteomics, a method that uses the combination of proteolytic digestion, liquid chromatography and tandem mass spectrometry to identify the components of a complex sample. Tandem mass spectrometry in a simple sense uses two components within the mass spectrometer to selectively isolate an ion (precursor mass) and then activate the ion to induce ion fragmentation, typically by a technique termed collision induced dissociation (CID) (29). The second component, mass analyzer, separates the product ions resulting from the precursor ion according to their m/z value. Successful bottom-up experiments are highly dependent on sample preparation and separation at the front end before sample introduction at the inlet. Different types of separation methods at the sample inlet can be used according to the complexity of the analyte. For example, if the complex sample contains one to few proteins, 1-dimensional LC can be used. For a sample containing few to thousands of proteins, 2-dimensional LC can be used. If hundreds to thousands of proteins are analyzed, a multidimensional protein identification technology (MuDPIT) is typically appropriate (30, 31).

Full proteomic studies were not performed in the thesis herein. Strategies involving bottom-up proteomic techniques were adapted and applied to capture post-translationally modified peptide ions in the species *Clostridium Botulinum* and *Streptococcus pyogenes*. The developed strategies for identification of peptides containing specific modifications are described in Chapter 2. Small scale proteomic strategies are also described in Chapter 5. The small scale strategies were used to

determine the alteration of Staphylococcus aureus metabolic output in the presence of antibiotic derived from the probiotic organism *Bacillus subtilis*.

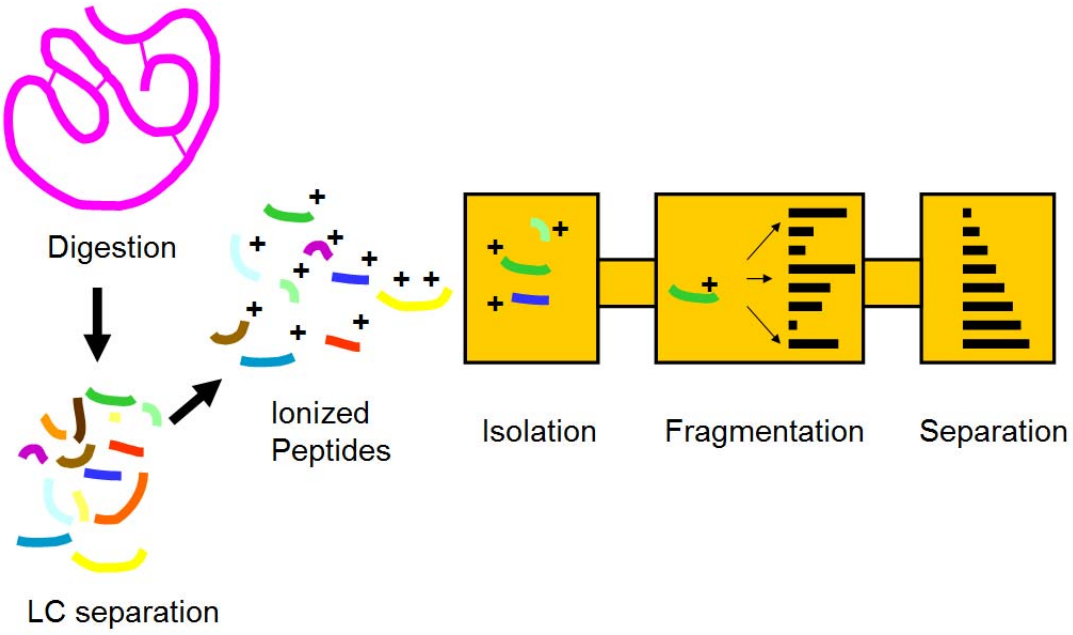

Figure 1.2 Bottom-up proteomics. A sample containing an array of proteins is digested with a protease (e.g. trypsin), resulting in numerous fragment peptides. The fragment peptides (complex sample) are separated by liquid chromatography pre-introduction into the mass spectrometer. Introduced peptides are ionized and then selectively isolated for fragmentation purposes. Resulting fragments are separated in the mass filter before be detected.

C. Imaging Mass Spectrometry basics

Imaging mass spectrometry (IMS) is a powerful tool to monitor the spatial distribution of ions over a solid sample (32, 33). The ability to visualize were an ion is located within a sample can be invaluable information for an analyst. Early IMS experiments focused on localizing targets or drug delivery within tissue and analysis of murine body section (34). Recently, IMS has been extended to examine the microbial kingdom in a previously uncharacterized fashion. The metabolic outputs of bacterial colonies were monitored to characterize molecules associated with colony growth and morphology. Additionally, IMS was extended to visualize the molecules involved in the chemical communication resulting from a bacterial-bacterial interaction (35, 36). This introduction will focus on IMS of microbial colonies by introducing the method of thin-layer imaging mass spectrometry (TLA-IMS) and hurdles associated with IMS of microbial colonies.

Figure 1.3 displays a general overview of an imaging mass spectrometry (IMS) workflow. Bacterial cultures are grown in a Petri dish to a desired time point. Samples are then excised and mounted on the MALDI target plate surface (Step 1) and immediately followed by matrix application (Step 2 and 3). The mass spectrometer is then programmed to raster across the surface of the sample collecting spectra at specified sampling locations (Step 4). The spectra are then compiled into a single dataset (Step 5), where the occurrence of any single mass can be visualized as a scaled false color overlay depicting the relative intensity of the ion across all sampled locations (Step 6). This data can be further processed to improve its quality followed by merging with other data including optical images. Within this proposal, the

resulting data obtained from a co-culturing experiment will allow for the assessment of the metabolic output a given bacterial pathogen possesses when challenged with biotic factors. The acquired information will guide the isolation and characterization of the observed target molecule.

Although advancements in sample preparation have allowed for whole bacterial cell colonies to be directly introduced into the mass spectrometer, there are some pressing analytical challenges when performing IMS of microbial colonies. Currently, microbes are grown on media with minimal components in order to overcome any complications in ion detection. Because detection of ions within a mass spectrometer is a competitive process, high salts and small molecules content is controlled in order to not suppress a target ion. The use of minimal media limits the number and type of microbes that can be examined by IMS. A second and potentially more pressing hurdle is determining the identities of the many ions observed in the resulting mass spectrum. In other words, even if an ion derived form a microbial colony is observed that has a unique spatial distribution and relevant to the experiment, a bottleneck would be to determine if the ion is already known or if it is uncharacterized. Therefore, the future success of microbial IMS is highly dependent on strong collaborative efforts between mass spectrometrists and natural product chemists.

Nevertheless, IMS of bacterial colonies has immense potential to monitor microbial systems in vast range of disciplines and is predicated to positively impact the practice of everyday microbiologist as the technique becomes an everyday lab practice. The thesis herein introduces IMS in Chapter 4 with a study of work

performed on a large number of differential microbes with the goal of showing the diverse amenability of IMS. Chapter 5 use IMS as a hypothesis generating tool to guide successive experiments based on the phenotypes observed by IMS. *B. subtilis* and *S. aureus*, two well-characterized model systems, are used.

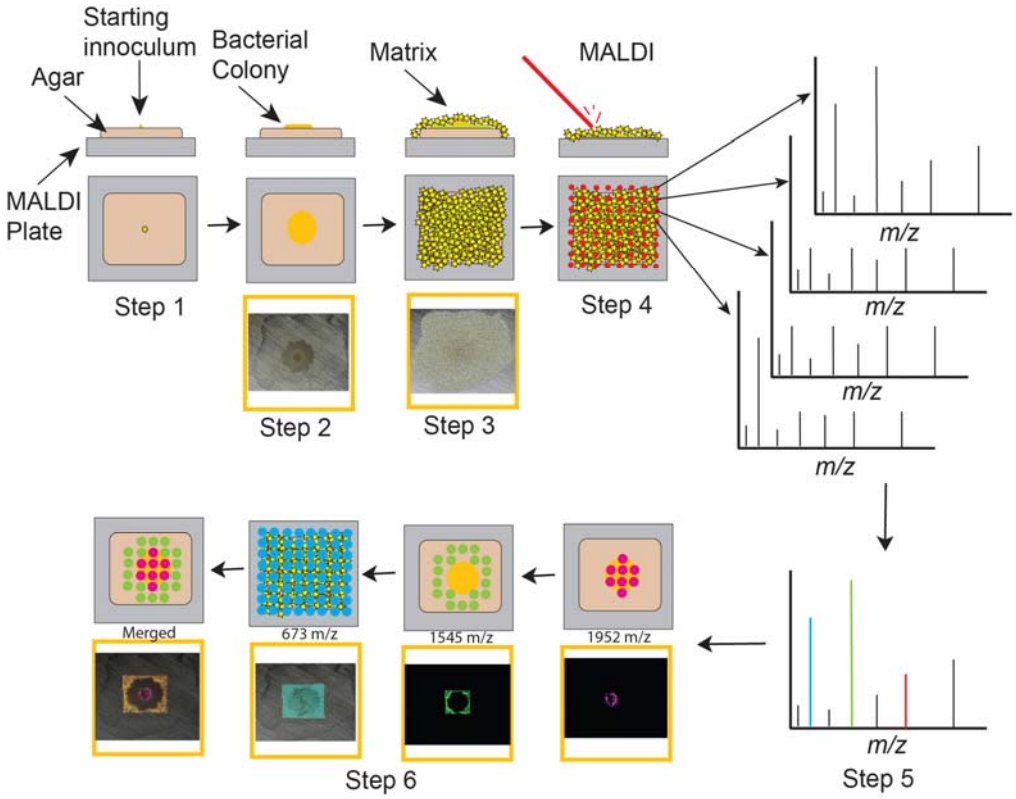

Figure 1.3 Imaging Mass Spectrometry. General work flow of the strategy of imaging mass spectrometry. Figure obtained form Yang et al. NCB, 2010 (35).

Acknowledgements

The introduction portion of this thesis uses figures 1.1-1.2 obtained from a presentation given by Toni Koller, Ph.D., Technical Director, Proteomics Center Stony Brook University Medical Center. Figure 1.3 was published by the Dorrestein lab (Yang et al. NCB, 2010).

Reference

1. Loo JA. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. Mass Spectrom Rev. 1997 Jan-Feb;16(1):1-23.

2. Cheng, X. & Hochlowski, J. Current application of mass spectrometry to combinatorial chemistry. Anal. Chem. 74, 2679–2690 (2002).

3. Hillenkamp F, Karas M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. Methods Enzymol. 1990;193:280-95

4. Hill JA, Annan RS, Biemann K. Matrix-assisted laser desorption ionization with a magnetic mass spectrometer. Rapid Commun Mass Spectrom. 1991 Sep;5(9):395-9.

5. Castro JA, Koster C, Wilkins C. Matrix-assisted laser desorption/ionization of high-mass molecules by Fourier-transform mass spectrometry. Rapid Commun Mass Spectrom. 1992 Apr;6(4):239-41.

6. Pieles U, Zürcher W, Schär M, Moser HE. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. Nucleic Acids Res. 1993 Jul 11;21(14):3191-6.

7. Geisow M. Electrospray ionization mass spectrometry--a powerful new analytical tool. Trends Biotechnol. 1990 Nov;8(11):301-3

8. Boyle JG, Whitehouse CM, Fenn JB. An ion-storage time-of-flight mass spectrometer for analysis of electrospray ions. Rapid Commun Mass Spectrom. 1991 Sep;5(9):400-5.

9. Henion J, Brewer E, Rule G. Sample preparation for LC/MS/MS: analyzing biological and environmental samples. Anal Chem. 1998 Oct 1;70(19):650A-656A

10. Gary L. Glish & Richard W. Vachet. The basics of mass spectrometry in the twenty-first century *Nature Reviews Drug Discovery* 2, 140-150

11. Goolsby, B. J. & Brodbelt, J. S. Characterization of β-lactams by photodissociation and collision-activated dissociation in a quadrupole ion trap. *J. Mass Spectrom.* **33**, 705–712 (1998).

12. Payne, A. H. & Glish, G. L. Thermally assisted infrared multiphoton photodissociation in a quadrupole ion trap. *Anal. Chem.* **73**, 3542–3548 (2001).

13. Watson E, Shah B, DePrince R, Hendren RW, Nelson R. Matrix-assisted laser desorption mass spectrometric analysis of a pegylated recombinant protein. Biotechniques. 1994 Feb; 16(2):278-81.

14. Fitzgerald MC, Parr GR, Smith LM. Basic matrices for the matrix-assisted laser desorption/ionization mass spectrometry of proteins and oligonucleotides. Anal Chem. 1993 Nov 15; 65(22):3204-11.

15. Dessort D, Mersel M, Lepage P, Van Dorsselaer A. Fast-heating mass spectrometry of phosphatidylcholine, lysophosphatidylcholine, phosphatidylethanolamine, and sphingomyelin. Anal Biochem. 1984 Oct;142(1):43-52.

16. McLuckey SA, Van Berkel GJ, Goeringer DE, Glish GL. Ion trap mass spectrometry. Using high-pressure ionization. Anal Chem. 1994 Jul 15; 66(14):737A-743A. Review.

17. Doroshenko VM, Cotter RJ. Injection of externally generated ions into an increasing trapping field of a quadrupole ion trap mass spectrometer. J Mass Spectrom. 1997 Jun; 32(6):602-15.

18. Andersen CA, Hinthorne JR. Ion microprobe mass analyzer. Science. 1972 Feb 25; 175(24):853-60.

19. James P. Protein identification in the post-genome era: the rapid rise of proteomics. Q Rev Biophys. 1997 Nov; 30(4):279-331.

20. Persidis A. Proteomics. Nat Biotechnol. 1998 Apr; 16(4):393-4.

21. Cash P. Proteomics in medical microbiology. Electrophoresis. 2000 Apr; 21(6):1187-201.

22. Neubauer G, Wilm M. From genome to function: analysis of multi-protein complexes and protein phosphorylation. Curr Opin Mol Ther. 1999 Dec; 1(6):695-701.

23. Clewley JP. Proteomics--discovering the protein constellation of the cell. Commun Dis Public Health. 2000 Jun;3(2):146-7.

24. Pandey A, Mann M. Proteomics to study genes and genomes. Nature. 2000 Jun 15; 405(6788):837-46.

25. Castoro JA, Nuwaysir LM, Ijames CF, Wilkins CL. Comparative study of photodissociation and surface-induced dissociation by laser desorption Fourier transform mass spectrometry. Anal Chem. 1992 Oct 1;64(19):2238-43.

26. Talluri S, Scheraga HA. Amide H/D exchange in the thermal transition of bovine pancreatic ribonuclease A. Biochem Biophys Res Commun. 1990 Oct 30;172(2):800-3.

27. Higgs RE, Knierman MD, Gelfanova V, Butler JP, Hale JE. Label-free LC-MS method for the identification of biomarkers. Methods Mol Biol. 2008;428:209-30.

28. Moore C, McKeown P. LCMS/MS and TOF-SIMS identification of the color bodies on the surface of a polymer. J Am Soc Mass Spectrom. 2005 Mar;16(3):295-301.

29. McFadden WH, Garteiz DA, Siegmund EG. Thermospray collisionally induced dissociation with single and multiple mass analyzers. J Chromatogr. 1987 May 8;394(1):101-8.

30. Kuksis A, Myher JJ. Application of tandem mass spectrometry for the analysis of long-chain carboxylic acids. J Chromatogr B Biomed Appl. 671(1-2):35-70.

31. Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001 73(23):5683-90.

32. Clerc J, Fourré C, Fragu P. SIMS microscopy: methodology, problems and perspectives in mapping drugs and nuclear medicine compounds. Cell Biol Int. 1997 Oct;21(10):619-33.

33. Weidner S, Knappe P, Panne U.MALDI-TOF imaging mass spectrometry of artifacts in "dried droplet" polymer samples. Anal Bioanal Chem. 2011 Feb 19

34. Gustafsson JO, Oehler MK, Ruszkiewicz A, McColl SR, Hoffmann P. MALDI Imaging Mass Spectrometry (MALDI-IMS)-Application of Spatial Proteomics for Ovarian Cancer Classification and Diagnosis. Int J Mol Sci. 2011 12(1):773-94.

35. Yang YL, Xu Y, Straight P, Dorrestein PC. Translating metabolic exchange with imaging mass spectrometry. Nat Chem Biol. 2009 Dec;5(12):885-7.

Chapter 2

Discovery of a widely distributed toxin biosynthetic gene cluster

Abstract

Bacteriocins represent a large family of ribosomally produced peptide antibiotics. Here we describe the discovery of a widely conserved biosynthetic gene cluster for the synthesis of thiazole and oxazole heterocycles on ribosomally produced peptides. These clusters encode a toxin precursor and all necessary proteins for toxin maturation and export. Using the toxin precursor peptide and heterocycle-forming synthetase proteins from the human pathogen Streptococcus pyogenes, we demonstrate the in vitro reconstitution of streptolysin S activity. We provide evidence that the synthetase enzymes, as predicted from our bioinformatics analysis, introduce heterocycles onto precursor peptides, thereby providing molecular insight into the chemical structure of streptolysin S. Furthermore, our studies reveal that the synthetase exhibits relaxed substrate specificity and modifies toxin precursors from both related and distant species. Given our findings, it is likely that the discovery of similar peptidic toxins will rapidly expand to existing and emerging genomes.

B. 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 (1, 2). *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 (1, 3). The molecular structure of SLS has remained elusive for more than a century; however, SLS is known to be an oxygen-stable, nonimmunogenic cytolysin (4). 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, 5–7). 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-aa protoxin (Fig. 2.1 *A*). *SagA* is followed by a set of three genes (*sagBCD*) exhibiting low sequence identity to three genes found in the *Escherichia coli mcb* gene cluster that produces the DNA gyrase inhibitor microcin B17 (MccB17) [supporting information (SI) Fig. S2.5 (5, 8). *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. S2.5–S2.9). The final genes in each cluster are dedicated ATP-binding cassette (ABC) transporters.

Analogous to the *sag* locus, the first ORF of the MccB17-producing gene cluster *mcbA* encodes a 69-aa precursor protein (9, 10). Walsh and colleagues (5) have shown that McbA is extensively posttranslationally modified by the McbBCD

synthetase complex. These modifications convert four serines and four cysteines into oxazole and thiazole heterocycles, respectively (Fig. 2.1 *B* and Fig. S2.5). Unmodified McbA exhibits no effect on DNA gyrase (11).

Results and Disscusion

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 used McbA, which is amenable to mass spectrometry, to detect and confirm heterocycle formation by SagBCD (12). After removal of the MBP tag, recombinant McbA has a calculated molecular mass of 6,293 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. 2.1 *B*). 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. 2.1 *C* and *D*). These masses are within error for linear mode 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 experimental evidence that SagBCD functions in a manner analogous to McbBCD.

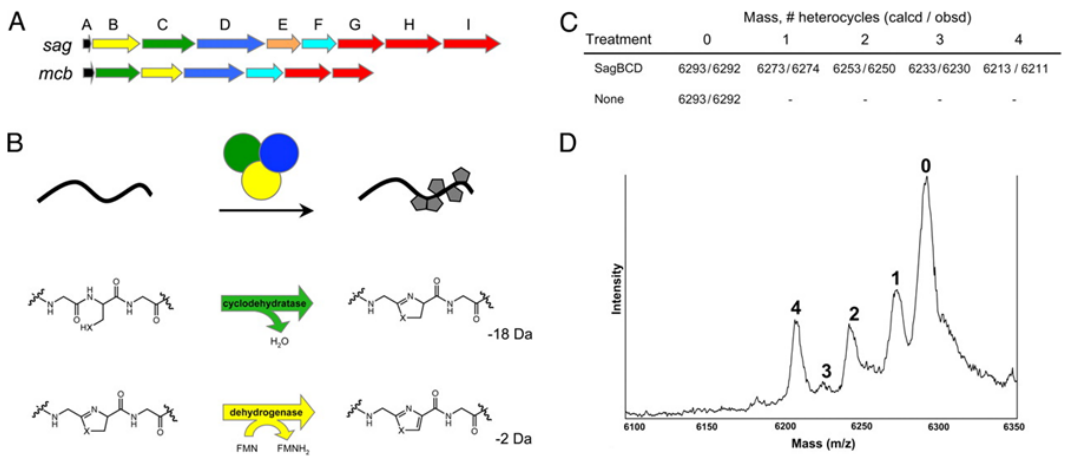

Figure 2.1 Conservation of toxin biosynthesis operons in *S. pyogenes* and *E. coli*. (*A*) Genetic organization of the streptolysin S-associated gene cluster (*sagA–I*) from *S. pyogenes* and the *E. coli* microcin B17 gene cluster (*mcbA–G*). (*B*) Through the action of a trimeric synthetase complex, oxazole and thiazole heterocycles are incorporated into a peptidic protoxin scaffold (black) and are active *in vitro*. Chemical transformations carried out by SagC/McbB (green, cyclodehydratase) and SagB/McbC (yellow, dehydrogenase) orthologs are shown. Molecular mass 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. Shown are 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. (*D*) Linear mode MALDI-TOF mass spectrum of McbA treated with SagBCD. The number of heterocycles on the protoxin peptide is indicated above the respective mass (singly charged species).

SagBCD Converts SagA into a Cytolysin.

Because 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. After 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. 2.2 *A*). 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. 2.2 *B*). 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. 2.2 *B*). Lactate dehydrogenase (LDH) release into the media provided quantitative confirmation that cytotoxicity required both the SagA substrate and the SagBCD complex (Fig. 2.2 *C*). 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 either that SagA-panC/S is 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.

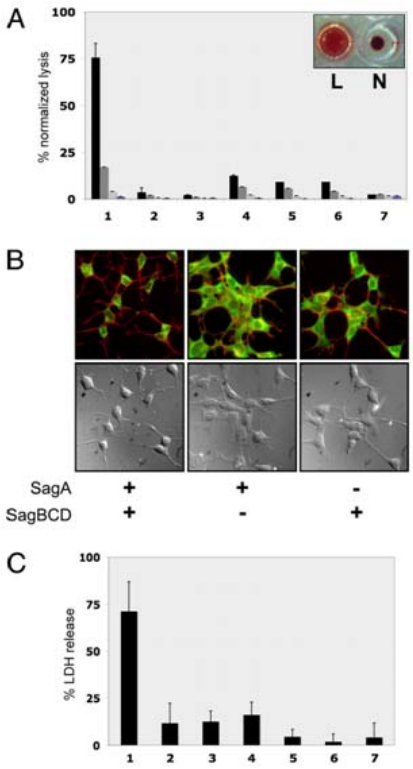

Figure 2.2 Cytolytic activity of *in vitro* synthetase reactions. (*A*) Hemolytic assays of SagA plus SagBCD synthetase reactions in microtiter wells containing defibrinated sheep blood. Bars indicate lysis normalized to a positive control (Triton X-100). Levels indicated are 1:1, 3:4, 1:2, and 1:4 ratios of synthetase reaction to blood (left to right) of 16-h reactions (n = 3). Lane 1, SagA plus SagBCD; lane 2, SagA alone; lane 3, SagBCD alone; lane 4, SagA plus SagBC; lane 5, SagA plus SagCD; lane 6, SagA plus SagBD; lane 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 *Upper*, and DIC images are in *Lower*. (*C*) LDH release assay of HEK293a cells treated with SagA plus SagBCD. Lysis is measured as A_{490} normalized to positive control (Triton X-100). Lane 1, SagA plus SagBCD; lane 2, SagA alone; lane 3, SagA plus SagBC; lane 4, SagA plus SagCD; lane 5, SagA plus SagBD; lane 6, SagBCD alone; lane 7, SagA-panC/S plus SagBCD.

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 (13). 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 (14). This search revealed that similar gene clusters are widespread among prokaryotes, most of which are not associated with a function (Fig. 2.3 *A*). The genes encoding the SagB-like dehydrogenase (Fig. S2.6), SagC-like cyclodehydratase (Fig. S2.7), and SagD-like protein (Fig. S2.8) 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. S2.9).

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. 2.3 *A*). Similar gene clusters are also found in distantly related prokaryotes, such as cyanobacteria (15, 16) 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 (13). *P. putida* colonizes the nutrient-rich rhizosphere of plants and induces plant growth while

secreting antibiotics to limit the growth of competing soil bacteria (17, 18).

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 (19). The conservation of this synthetase gene cluster across prokaryotes—in particular, among 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. 2.3 *B*). 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. 2.3 *B*). 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 (20), will likely produce toxins with acetylated and dehydrated amino acids (Fig. S2.5). 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 (21). Inclusion of these ancillary enzymes is expected to dramatically increase the chemical diversity of the toxins, perhaps for niche purposes. The cyanobacteria *Prochloron didemni* (15) and *Trichodesmium erythraeum* IMS101 (16) have been previously identified as sources of cyclic peptides with Ser/Thr/Cys-derived heterocycles (Fig. S2.5). The *sag*-like biosynthetic clusters of *T. erythraeum* and *P. didemni* harbor an additional gene encoding a putative serine protease (Fig. 2.3 *A*). 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. 2.3 *A*). Based on a similar

arrangement of genes, we hypothesize that *Rhodobacter sphaeroides*, *Lyngbya* sp. PCC 8106, and *Microcystis aeruginosa* will also form macrocyclized peptides.

Figure 2.3 A. 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. Numerous ancillary enzymes are included and increase the chemical diversity of the toxin family.

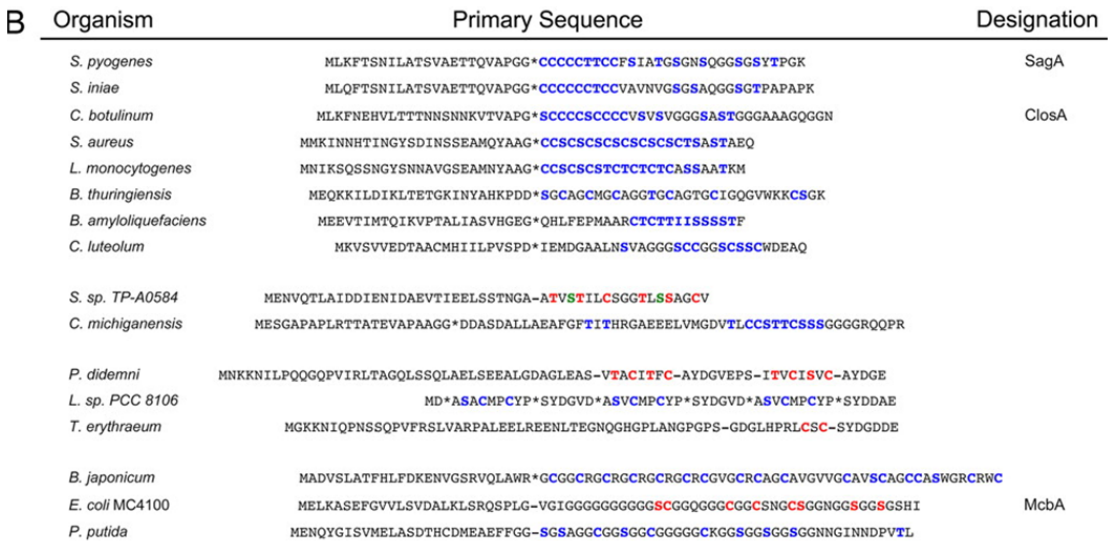

Figure 2.3 B. Select protoxin amino acid sequences. Predicted leader peptide cleavage sites are denoted with an asterisk. Hyphens indicate a known cleavage site. Potential sites of heterocyclization are indicated in blue, and known sites of heterocycle formation are indicated in red. Green text signifies conversion to dehydroalanine. Toxins similar to SagA (top) are predicted cytolysins.

The Heterocycle Synthetase Complex Displays Functional Promiscuity.

A gene cluster that is highly similar to the *sag* cluster was found in the pathogenic bacterium *C. botulinum* (Fig. 2.3 *A*). 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. 2.3 *B*). 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 observed only in samples containing both ClosA and the SagBCD synthetase complex (Fig. 2.4 *A*). Microscopy revealed that cytotoxicity in HEK cells treated with ClosA plus SagBCD was identical to cells treated with SagA plus SagBCD (Figs. 2.2 *B* and 2.4 *A*). These data demonstrate that 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. 2.4 *B*). 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 (22). 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 require further 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 *in vitro* reconstitution of 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.

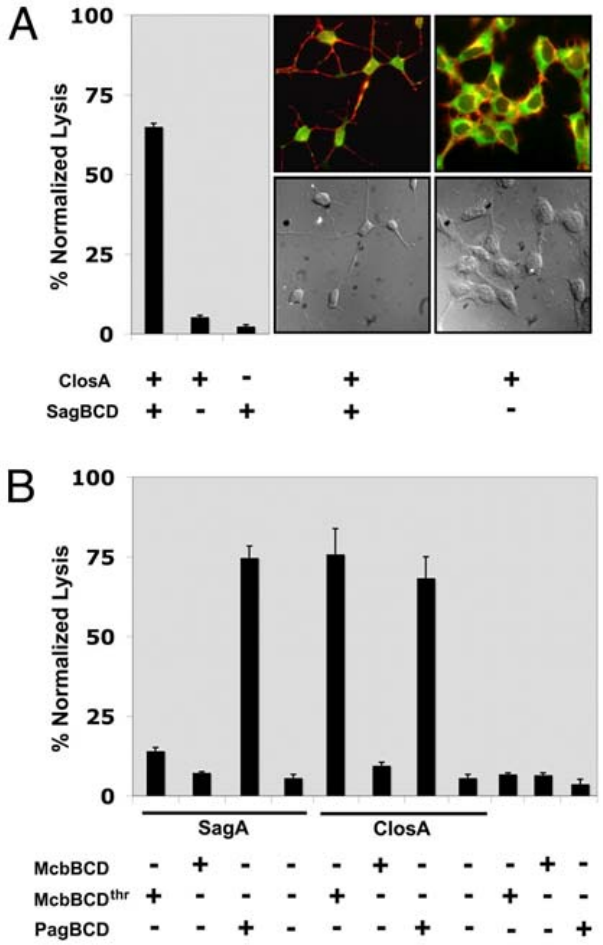

Figure 2.4 Heterocycle synthetases accept substrates from related and distant prokaryotes. (*A*) The *S. pyogenes* synthetase SagBCD accepts the *C. botulinum* protoxin and produces a cytolytic toxin. (*Left*) Hemolytic assay. (*Center and Right*) Fluorescence microscopy and DIC images of HEK293a cells treated as indicated. (*B*) Hemolytic assay of SagA and ClosA 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.

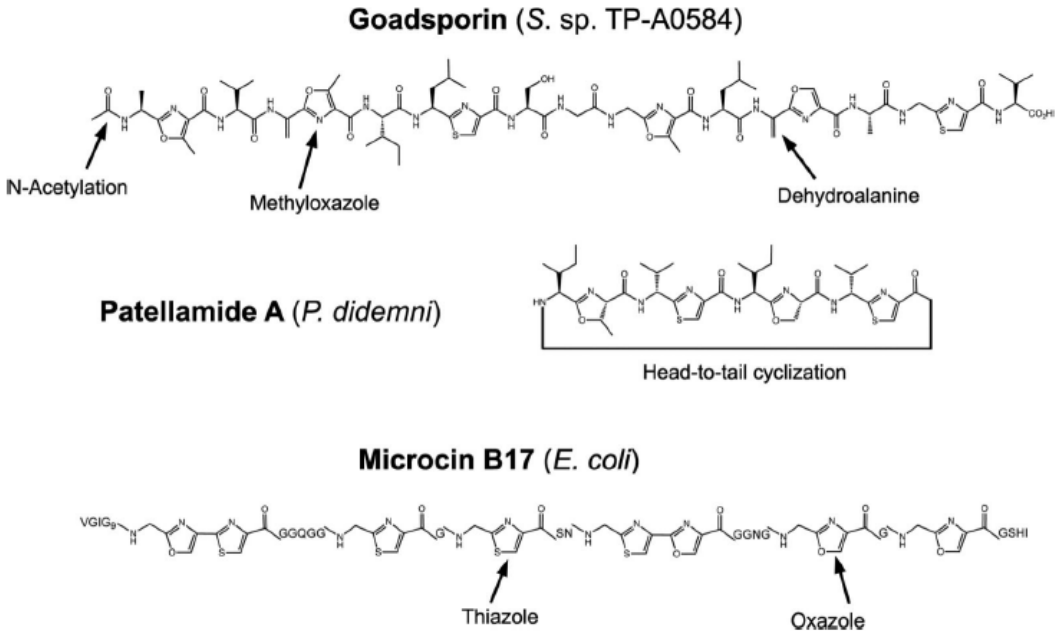

Figure S2.5 Molecular structures of compounds synthesized by a *sag*-like gene cluster. Posttranslational modifications are indicated. Dehydroalanines and thiazoles are derived from cysteine, oxazole from serine, and methyloxazole from threonine. Patellamide A contains two reduced heterocycles: a methyloxazoline and oxazoline. These are derived from threonine and serine, respectively, and are not labeled for clarity.

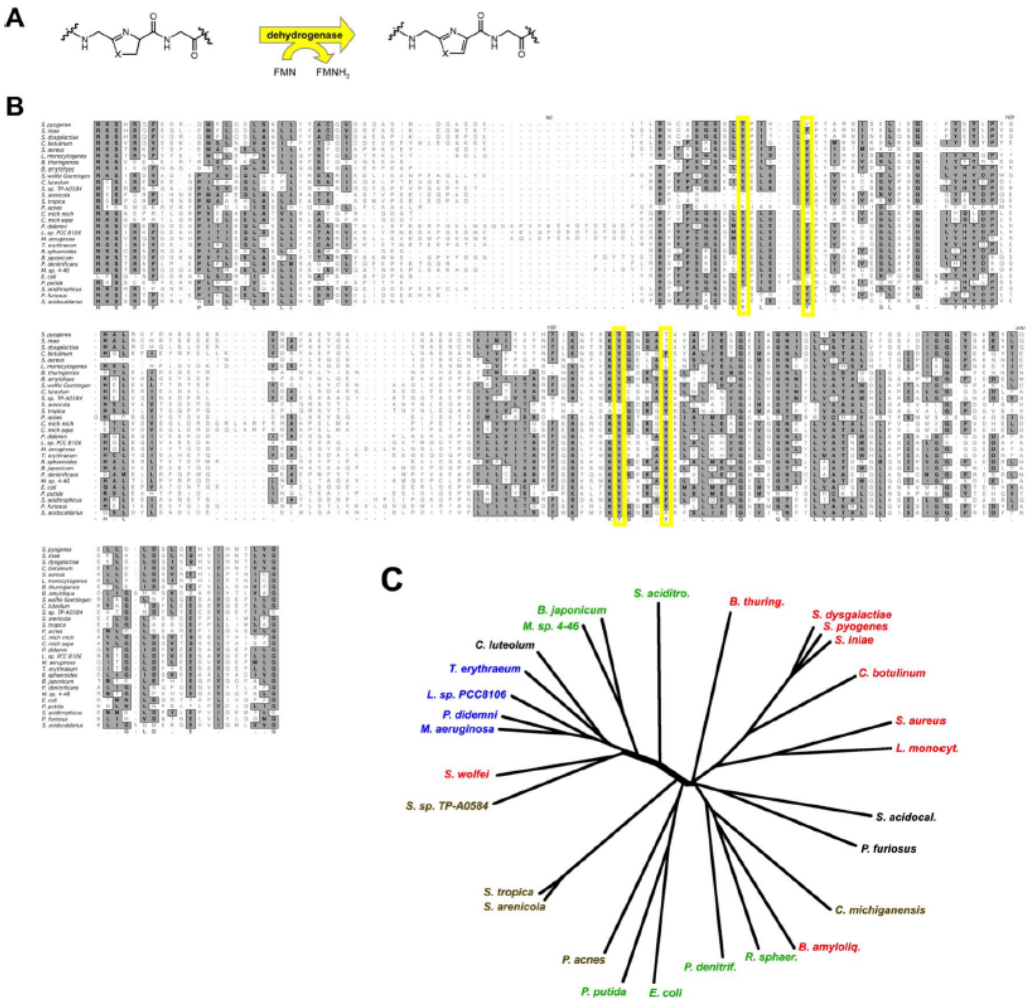

Figure S2.6 Multiple sequence alignment and phylogenetic analysis of the dehydrogenase. (*A*) The chemical transformation carried out by SagB orthologs. Oxazoline (X _ O) and thiazoline (X _ S) heterocycles are oxidized by two electrons to oxazole and thiazole, respectively. During this oxidative aromatization reaction, oxidized flavin mononucleotide (FMN) is reduced by two electrons to FMNH2. 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 _-stacked between Y orW. 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 coloredas follows: red, firmicutes; green, proteobacteria; blue, cyanobacteria; brown, actinobacteria; black, other.

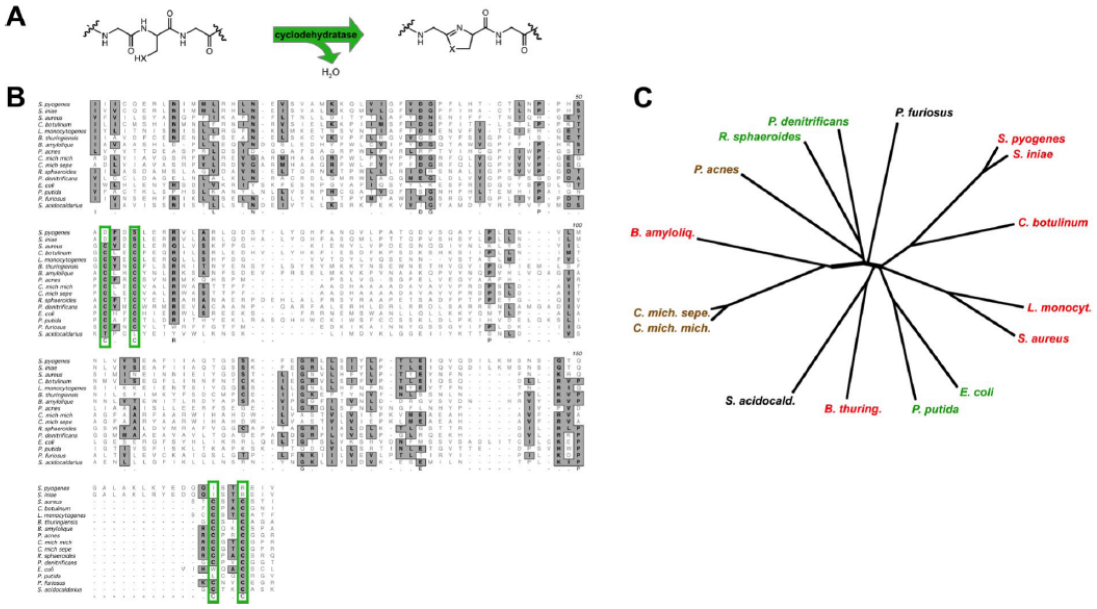

Figure S2.7 Multiple sequence alignment and phylogenetic analysis of the cyclodehydratase. (*A*) The chemical transformation carried out by SagC orthologs. Thecyclodehydratase catalyzes the conversion of serine to oxazoline (X_O), cysteine to thiazoline (X_S), and threonine to methyloxazoline with loss 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 theNandCtermini 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 containtwo nearly invariant CXXC motifs (green boxes) that likely serve to tetrahedrally coordinate a structural Zn2_. 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.

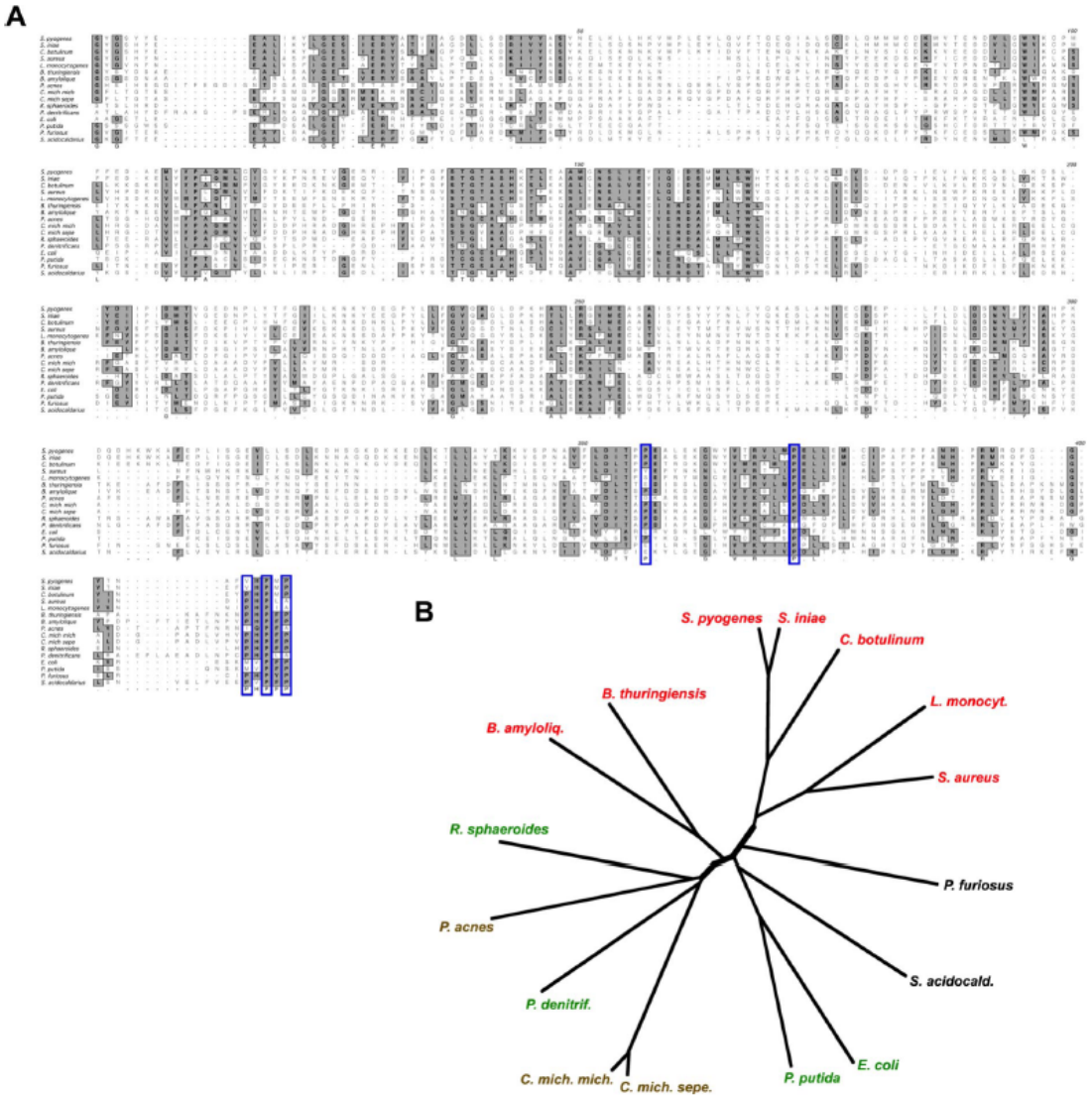

Figure S2.8 Multiple sequence alignment and phylogenetic analysis for the docking protein. (*A*) ClustalW alignment of 16 SagD orthologs (the others are encoded as a fusion protein to SagC). Because of divergence at the N terminus, the _400 C-terminal residues are shown in this alignment. This region 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.

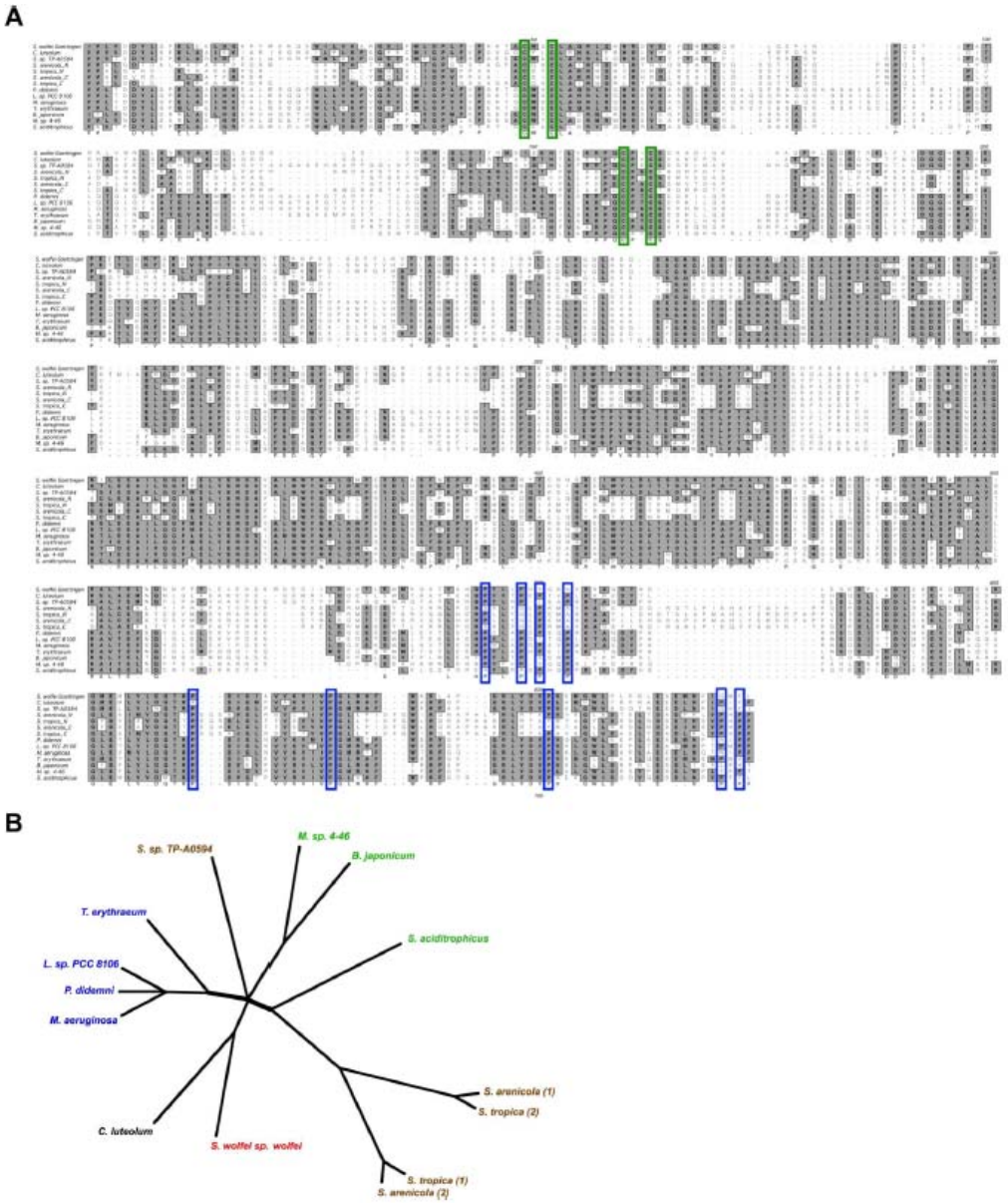

Figure S2.9 Multiple sequence alignment and phylogenetic analysis for the fused cyclodehydratase/docking protein. (*A*) ClustalW alignment of 14 fused SagCD orthologs (the *Salinospora* clusters contain two copies of this protein). Because of divergence at the N terminus, the _700 C-terminal residues 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 theend of the operon. The labels are sorted by phylum and are colored as follows: brown, actinobacteria; green, proteobacteria; blue, cyanobacteria; red, firmicutes;black, bacteriodetes.

Materials and Methods

Bioinformatics.

The amino acid sequences of SagBCD from *S. pyogenes* (2) and McbBCD from *E. coli* (23) were initially used as a BLAST query against the nonredundant protein sequence database of the National Center for Biotechnology Information (14). 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 by 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 by using ClustalW alignment to known toxin structural genes from *S. pyogenes* and *E. coli*. *Streptococcus iniae* has been shown to contain a gene cluster nearly identical to *S. pyogenes* and was therefore not used for bioinformatics searching (24, 25). Furthermore, another *Streptococcus* species, *Streptococcus dysgalactiae* sp. *equisimilis*, contains a SagA and SagB homolog (3). Because of an incomplete genome sequence, we cannot confirm the presence of SagC–I at this time.

Wider subsequent searches of the nonredundant protein sequence database (National Center for Biotechnology Information) 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. S2.5), *Streptomyces* sp. TP-A0584 (goadsporin) (26), *P. didemni* (patellamide A/C) (15), and *T. erythraeum* IMS101 (trichamide) (16) 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 two to three times with 100 μl of deionized water 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.

MALDI-TOF mass spectra were acquired on a Voyager DE-STR instrument (Applied Biosystems) in linear positive mode. Specific settings were typically as follows: 2,000–2,100 laser power, 93% grid, 0.12% guide wire, 400 nsec delay, mass window 3,000–18,000 Da.

***In Vitro* Synthetase Assay.**

Synthetase reactions using MBP-tagged SagA and SagBCD and other fusion proteins were performed in a manner described earlier (27, 28). Reaction mixtures consisted of MBP-SagA (10 μM) and MBP-SagBCD (2 μM each) in a total volume of 100 μl of 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. Whole blood was washed three times in PBS and diluted to a final concentration of 1:25 vol/vol in PBS. Prepared whole blood (100 μl) was then placed into individual wells of a flat-bottom 96-well microtiter plate (Costar, Corning). 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. After incubation, plates were centrifuged at 500 × *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).

For the sake of simplicity, all gene identifiers for this paper use the following designations unless otherwise stated: A, protoxin; B, dehydrogenase; C, cyclodehydratase; D, docking protein. This family of bacteriocin is defined as a ribosomally produced toxin with conserved machinery (the BCD enzymes) to introduce thiazole, thiazoline, (methyl)-oxazole, and (methyl)-oxazoline heterocycles, which are derived from cysteine, threonine, and serine residues, into the peptidic backbone provided by a structural protoxin protein. 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* genomicDNAand 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 Nterminal 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 DH5_. 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 AGGTAA C; PagB_rev, AAGCGG 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; PagD_fwd, 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 _g/ml kanamycin (or 100 _g/ml ampicillin) and 35 _g/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-_-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 [50mMTris, 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 8ml 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 over expressed 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 (6 M 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) supple-mented with 5% FBS, penicillin (100 units/ml), and streptomycin (100 _g/ml).

Cytotoxicity Assay

Lactate dehydrogenase (LDH) levels were determined by using a colorimetric assay (Roche). HEK293a cells were grown to 70–80% confluence 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 min 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

Author contributions: S.W.L. and D.A.M. contributed equally to this work; S.W.L., D.A.M., M.E.H., P.C.D., and J.E.D. designed research; S.W.L., D.A.M., A.L.M., M.E.H., D.G., A.W., and P.C.D. performed research; S.W.L., D.A.M., A.L.M., and M.E.H. contributed new reagents/analytic tools; S.W.L., D.A.M., A.L.M., M.E.H., D.G., A.W., P.C.D., V.N., and J.E.D. analyzed data; and S.W.L., D.A.M., and J.E.D. wrote the paper.

Reference

1. Datta V, et al. (2005) Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Mol Microbiol 56:681–695.

2. Nizet V, et al. (2000) Genetic locus for streptolysin S production by group A streptococcus. Infect Immun 68:4245–4254.

3. Humar D, et al. (2002) Streptolysin S and necrotising infections produced by group G streptococcus. Lancet 359:124–129.

4. Nizet V, (2002) Streptococcal beta-hemolysins: Genetics and role in disease pathogenesis. Trends Microbiol 10:575–580.

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

6. Yorgey P, Davagnino J, Kolter R (1993) The maturation pathway of microcin B17, a peptide inhibitor of DNA gyrase. Mol Microbiol 9:897–905.

7. Betschel SD, et al. (1998) Reduced virulence of group A streptococcal Tn916 mutants that do not produce streptolysin S. Infect Immun 66:1671–1679.

8. San Millan JL , Kolter R , Moreno F (1985) Plasmid genes required for microcin B17 production. J Bacteriol 163:1016–1020.

9. Davagnino J, et al. (1986) The DNA replication inhibitor microcin B17 is a forty-three-amino-acid protein containing sixty percent glycine. Proteins 1:230–238.

10. Yorgey P, et al. (1994) Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. Proc Natl Acad Sci USA 91:4519–4523.

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

12. Kelleher NL , Belshaw PJ , Walsh CT (1998) Regioselectivity and chemoselectivity analysis of oxazole and thiazole ring formation by the peptide-heterocyclizing microcin B17 synthetase using high-resolution MS/MS. J Am Chem Soc 120:9716–9717.

13. Severinov K, et al. (2007) Low-molecular-weight post-translationally modified microcins. Mol Microbiol 65:1380–1394.

14. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

15. Schmidt EW, et al. (2005) Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proc Natl Acad Sci USA 102:7315–7320.

16. Sudek S, Haygood MG, Youssef DT, Schmidt EW (2006) Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium Trichodesmium erythraeum, predicted from the genome sequence. Appl Environ Microbiol 72:4382–4387.

17. Dos Santos VA, et al. (2004) Insights into the genomic basis of niche specificity of Pseudomonas putida KT2440. Environ Microbiol 6:1264–1286.

18. Nelson KE, et al. (2002) Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. Environ Microbiol 4:799–808.

19. Broderick NA , Raffa KF , Handelsman J (2006) Midgut bacteria required for Bacillus thuringiensis insecticidal activity. Proc Natl Acad Sci USA 103:15196–15199.

20. Onaka H, et al. (2005) Cloning and characterization of the goadsporin biosynthetic gene cluster from Streptomyces sp. TP-A0584. Microbiology 151:3923–3933.

21. Sofia HJ, et al. (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 29:1097–1106.

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

23. Roy RS, et al. (1999) Thiazole and oxazole peptides: Biosynthesis and molecular machinery. Nat Prod Rep 16:249–263.

24. Locke JB, et al. (2007) Streptococcus iniae beta-hemolysin streptolysin S is a virulence factor in fish infection. Dis Aquat Organ 76:17–26.

25. Fuller JD, et al. (2002) Identification of a streptolysin S-associated gene cluster and its role in the pathogenesis of Streptococcus iniae disease. Infect Immun 70:5730–5739.

26. Igarashi Y, et al. (2001) Goadsporin, a chemical substance which promotes secondary metabolism and morphogenesis in streptomycetes. II. Structure determination. J Antibiot (Tokyo) 54:1045–1053.

27. Sinha Roy R, Belshaw PJ, Walsh CT (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 37:4125–4136.

28. Sinha Roy R, Kelleher NL, Milne JC, Walsh CT (1999) In vivo processing and antibiotic activity of microcin B17 analogs with varying ring content and altered bisheterocyclic sites. Chem Biol 6:305–318.

Chapter 3

Clostridiolysin S, a Post-translationally modified biotoxin from *Clostridium Botulinum*

Abstract

Through elaboration of its botulinum toxins, Clostridium botulinum produces clinical syndromes of infant botulism, wound botulism, and other invasive infections. Using comparative genomic analysis, an orphan nine-gene cluster was identified in C. botulinum and the related foodborne pathogen Clostridium sporogenes that resembled the biosynthetic machinery for streptolysin S, a key virulence factor from group A Streptococcus responsible for its hallmark beta-hemolytic phenotype. Genetic complementation, in vitro reconstitution, mass spectral analysis, and plasmid intergrational mutagenesis demonstrate that the streptolysin S-like gene cluster from Clostridium sp. is responsible for the biogenesis of a novel post-translationally modified hemolytic toxin, clostridiolysin S.

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 (1, 2). Streptolysin S (SLS) is a well known hemolytic/cytolytic, ribosomally encoded bacteriocin and virulence factor group A *Streptococcus* (GAS) (3,–,5). 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 (4) has guided investigations into its post-translational modification and likely heterocyclic nature (6, 7).

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 (8,–,10). 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 (11). 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 (11).

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. 3.1*A*). 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 (12,–,16). 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. 3.3*A*). 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 (17, 18). 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 (19) and Spectral Networks (20), to establish that the nontoxic precursor (ClosA) is converted into the final post-translationally modified toxin, clostridiolysin S.

Results

Comparative Bioinformatics of Clos and Sag Locus

Previously, we and others (6, 10, 12) 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 (21,–,23) 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. 3.1*A*). 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*) (3). 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. 3.1*B*). 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 (10). *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. 3.1*C*). 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.

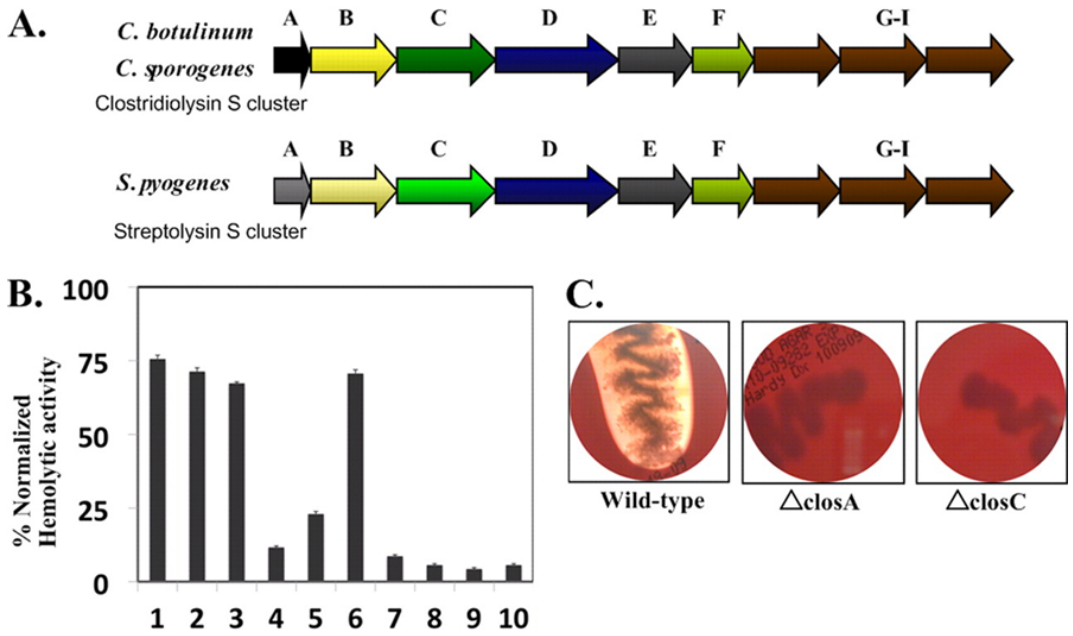

Figure 3.1 Comparison of the streptolysin S-associated genetic cluster (Sag) with the SLS-like clostridiolysin S genetic cluster in *Clostridium botulinum*. *A*, the Sag cluster in GAS contains a SLS precursor sequence (SagA), which is modified by the SagBCD synthetase complex. SagE and -F are potentially 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.

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 (6). SagB is 60% similar to ClosB,

and additionally, recombinant expression of SagB yielded holoprotein, as indicated by a bright yellow color (6). 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. 3.2*A*). 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.

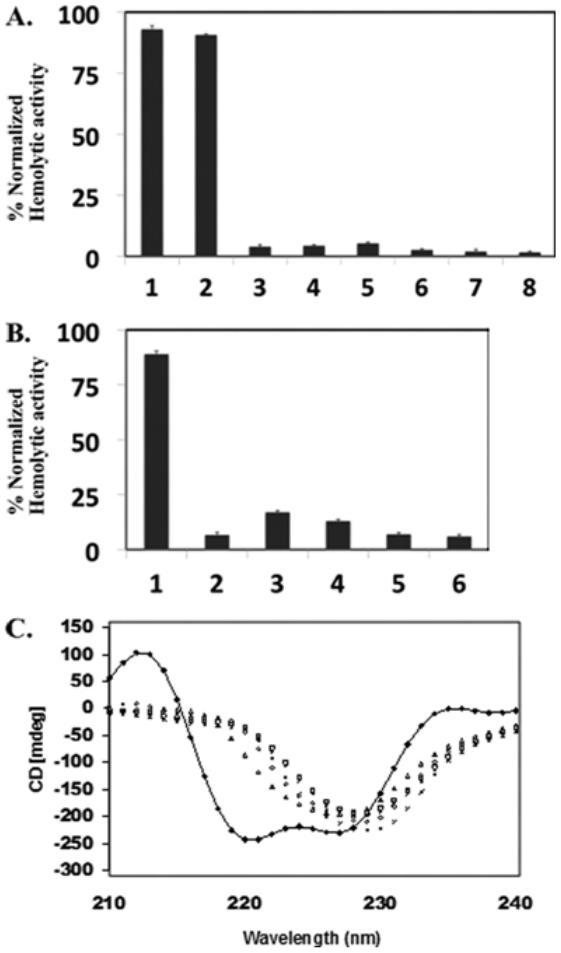

Figure 3.2 Hemolytic activity of clostridiolysin is detectable in vitro**.** A, in vitro reconstitution of ClosA hemolytic activity. Synthetase reactions using MBP-ClosA and SagB/ClosC/ClosD produce a hemolytic toxin. Hemolytic activity was normalized against a Triton X-100 positive control. Bar 1, wt 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.

Sequence alignments of McbB, E1, MoeB, ThiF, and MccB (Fig. 3.3*B*), suggest that ClosC is related to the ThiF/MoeB/E1 superfamily of proteins (24, 25, 26). 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 this regard, the chemistry described by these systems is very different. In clostriodiolysin S, an amide carbonyl undergoes activation, as is the case for microcin B17, patellamide, goadsporin, and the recently described thiopeptide family (13,–,16, 27). Because ClosA contains a C-terminal asparagine, in addition to several Ser, Thr, and Cys residues, we cannot predict the biochemical reaction catalyzed by ClosC (supplemental Fig. 3.6). 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. 3.3*B*). In the microcin system, Zn^{2+} was originally proposed to serve as a Lewis acid, activating the amide carbonyl for cyclodehydration (22). Although that mechanism is certainly plausible, the crystal structures of ThiF, MoeB, and MccB have become available and show that the tertracysteine coordinated Zn^{2+} is 20 Å away from the active site pocket where ATP binds (25, 26). 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 (6). 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 (supplemental Fig. 3.7), similar to the observations by the homolog McbB in the microcin B17 pathway (22). 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. 3.2*B* and supplemental Fig. 3.8). 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. 3.2*C*). 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 (supplemental Fig. 3.9). 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 (supplemental Fig. 3.10). 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.

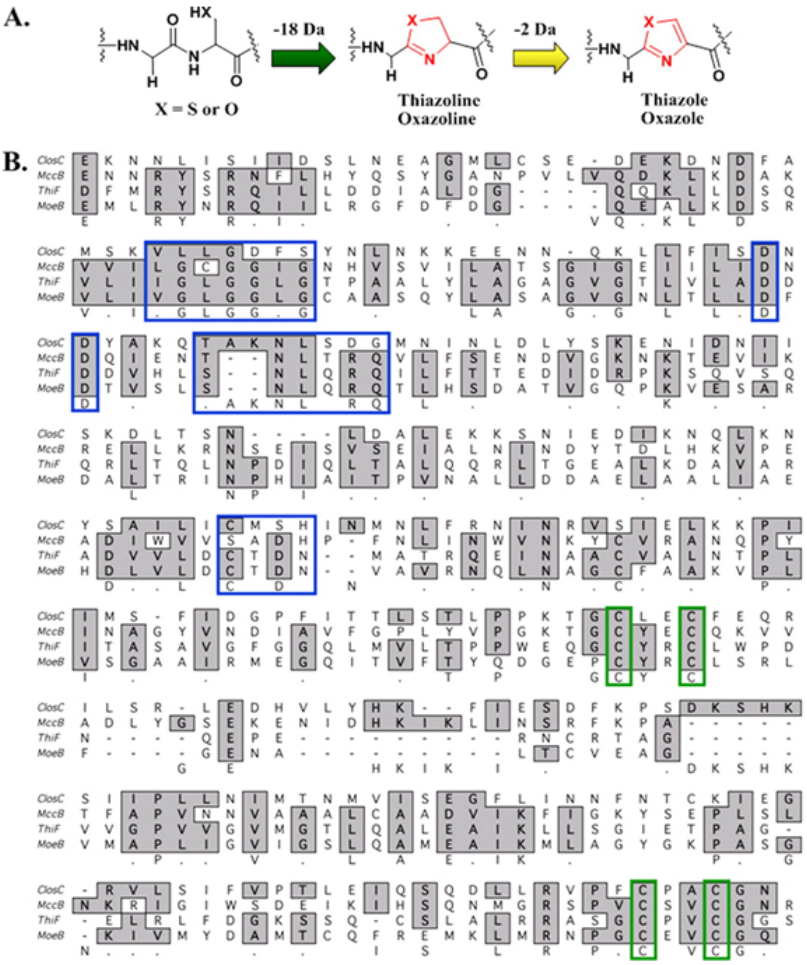

Figure 3.3 Alignment of ClosC in C. botulinum ATCC 3503 and MoeB/ThiF and proposed mechanisms. A, mechanisms for the formation of thiazole or methyloxazole. B, sequence alignments of ClosC, MccB, E. coli MoeB, and E. coli ThiF. Blue highlighted boxes show conserved ATP binding residues determined by the crystal structures of ThiF and MoeB. Green highlighted boxes show the residues involved in zinc-tetrathiolate formation.

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 (6, 7, 11, 28, 29). 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 (supplemental Fig. 3.11). 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 (30). 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 (19, 20). 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. 3.5*B*, 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* (7).

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. 3.4).

In the clostridiolysin S-reconstituted system, peptides with modifications at Thr^{11}, Thr^{12}, and Thr^{46} (Fig. 3.5*B*, peptide 1; supplemental Figs. 3.12 and 3.13) 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. 3.5*B*, 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. 3.5*A*). In a similar fashion to *in vitro* reconstitution experiments in microcin B17 system, where an additional heterocycle was produced relative to the known structure (23), 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 (31).

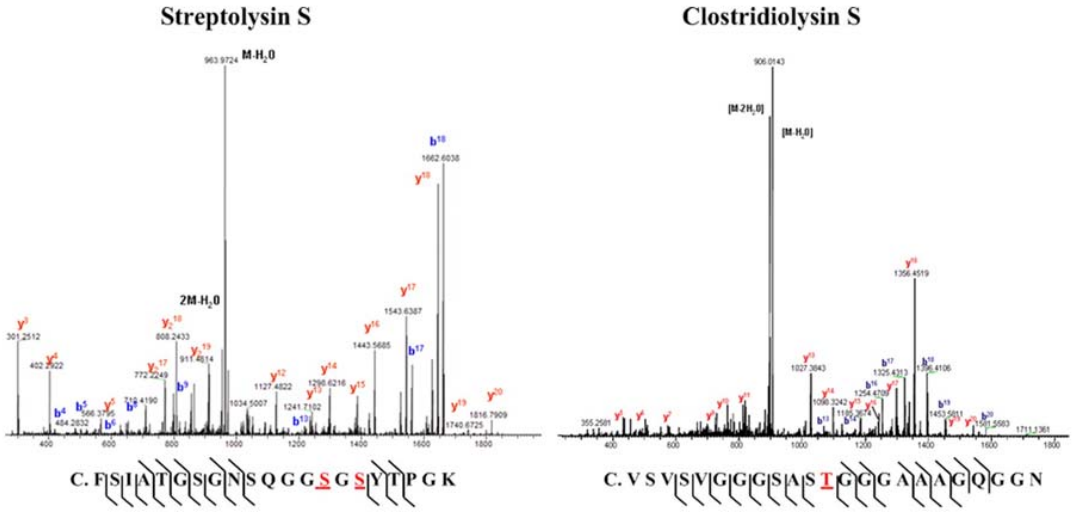

Figure 3.4 Detection of heterocyclized ClosA peptides via nanocapillary LC-MS/MS method. Mass spectra of C-terminal peptides identified as being heterocyclized in the SLS and CLS in vitro systems. A detailed description of mass spectral annotations is supplied in the supplemental data.

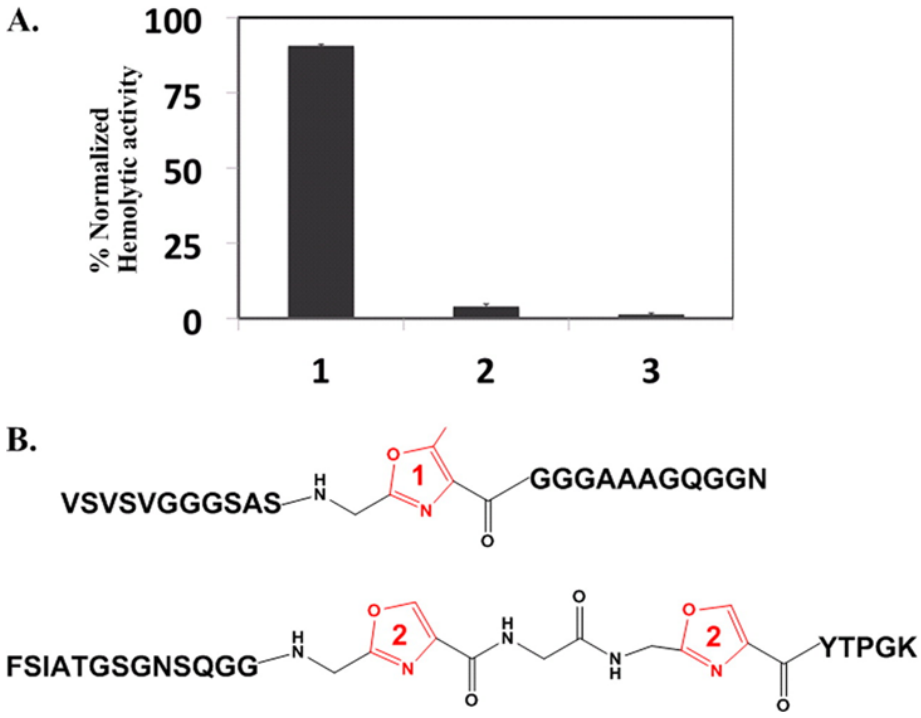

Figure 3.5 Hemolytic activity of MBP-ClosA T46A and structures of identified PTM peptides. *A*, synthetase reactions using MBP-ClosA T46A and SagB/ClosC/ClosD. Hemolytic activity was normalized against a Triton X-100 positive control. *Lane 1*, wt ClosA and SagB/ClosC/ClosD; *lane 2*, SagB/ClosC/ClosD 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}.

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

Cloning of closA, closC, and closD

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 (6).

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 (6). Upon purification, all proteins were immediately stored at −80 °C until further use.

Cytolytic Activity Assay

In Vitro Synthetase Reactions

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 (6). 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).

Clos Single Gene Complementation into Group A Streptococcus ΔsagA–D Strains

Preparation of GAS ΔsagA–D Strains Harvesting the closA–D Genes

The construction of GAS ΔsagA, ΔsagB, ΔsagC, and ΔsagD allelic exchange mutants in GAS strain NZ131 (M49 serotype) was described previously (3). 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 (3) 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 Strains Harvesting the closA–D Genes

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 (6). 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 (6). Upon purification, all proteins were immediately stored at −80 °C until needed.

Inductively Coupled Plasma Mass Spectrometry (ICP-MS) Analysis of ClosC CxxC Mutants

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 (6). 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 3.4 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 Supplemental Fig. 3.12). 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

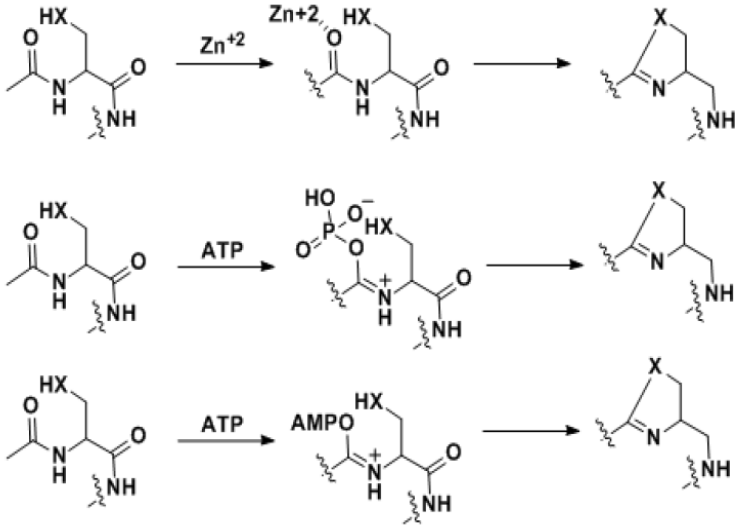

Figure S3.6 Alternative mechanisms for the formation of thiazolines or (methyl)-oxazolines using Zn+2 or ATP as Lewis acid for carbonyl destabilization.

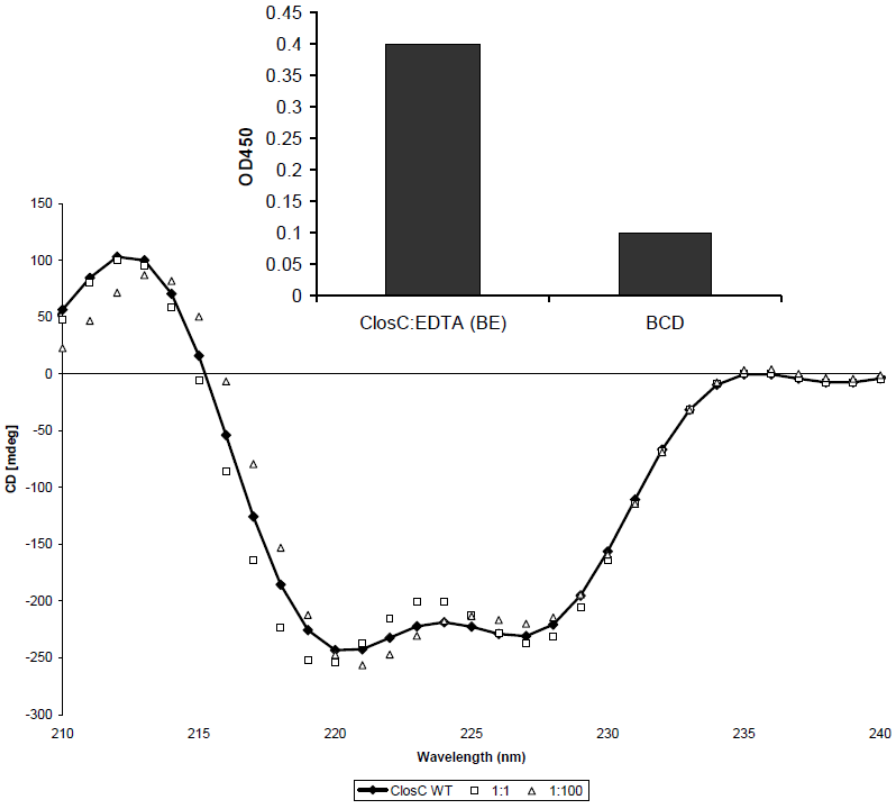

Figure S3.7 MBP-ClosC was incubated with 1:1 and 1:100 protein-to-EDTA molar ration then buffer exchanged (BE). The CD shows the MBP-ClosC remains unperturbed by the EDTA and is active in the bioassays (insert, y-axis shows OD450 monitoring heme release upon erythrocyte lysis).

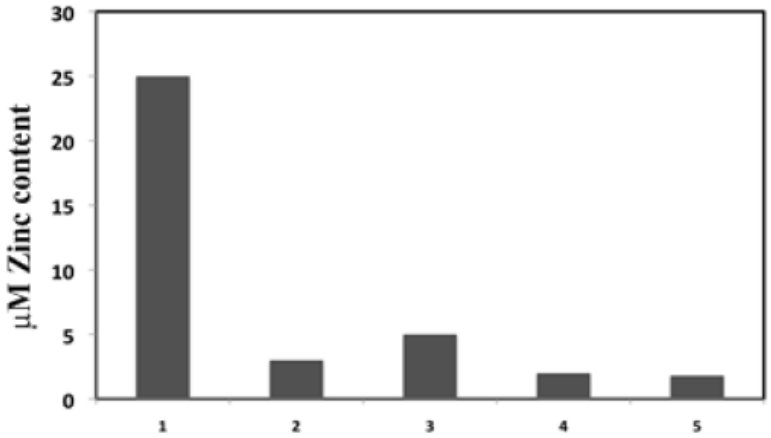

Figure S3.8 ICP-MS data of wt ClosC versus CxxC mutants. In order: 1. ClosC 2. ALEC 3. CLEA 4. APAC 5. CPAA. ICP-MS experiments were performed in duplicate with two different protein preparations.

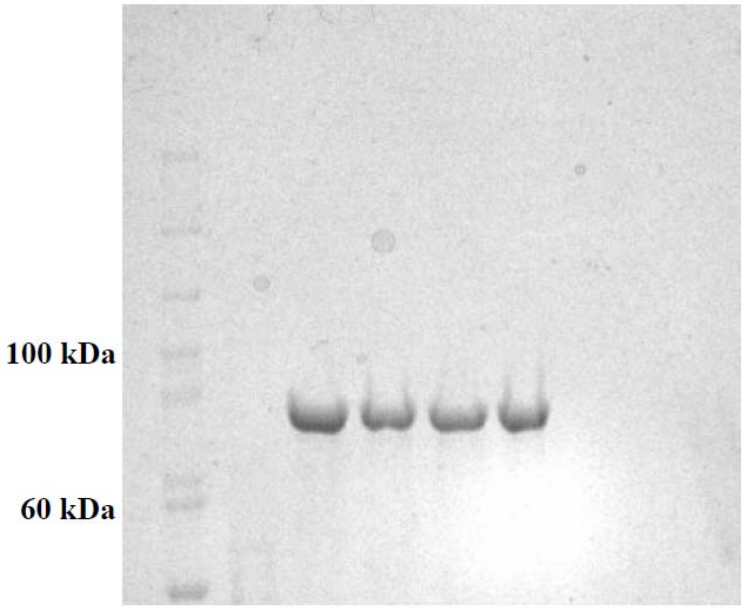

Figure S3.9 SDS-PAGE of the proteins used for CD analysis. Proteins were concentrated (40kDa cut-off) and buffer exchanged. Molecular weight of MBP-ClosC is ~89kDa. 1. ALEC 2. CLEA 3. APAC 4. CPAA.

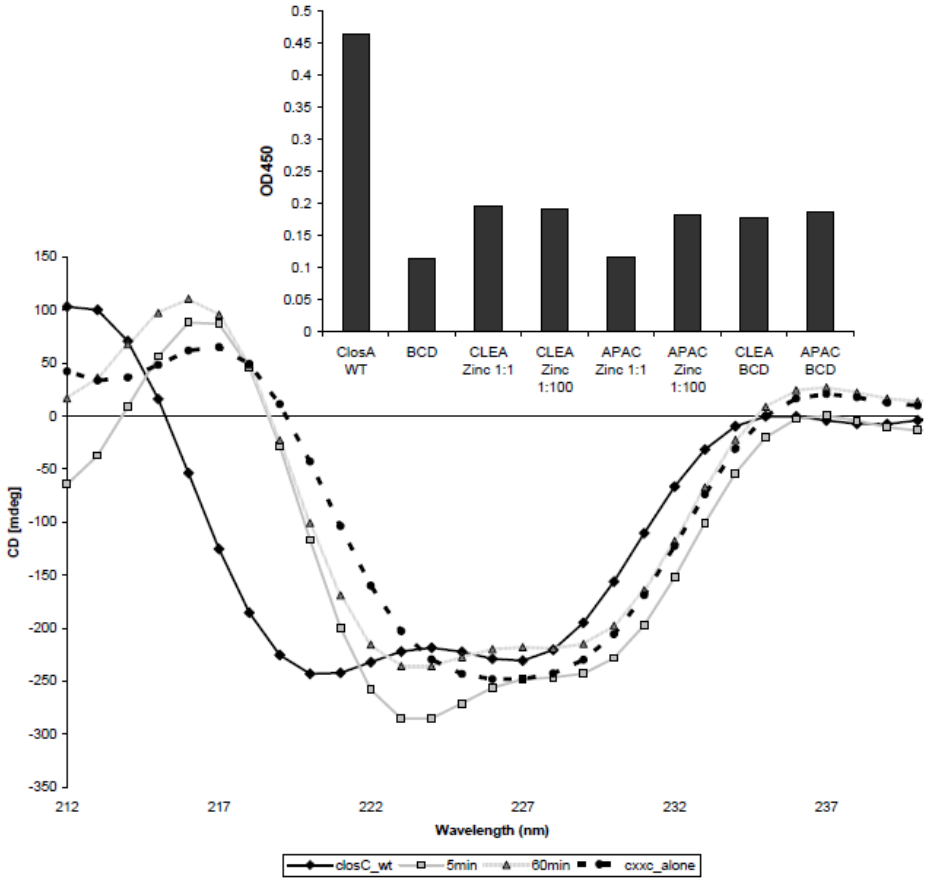

Figure S3.10 CxxC proteins CLEA and APAC were incubated with zinc at a 1:1 and 1:100 protein-to-zinc molar ratio then buffer exchanged (BE). The CD shows the zinc could not restore alpha-helical content comparable to wt. MBP-ClosC. Insert shows bioactivity of the CxxC proteins upon zinc incubation (y-axis shows OD450 monitoring heme release upon erythrocyte lysis).

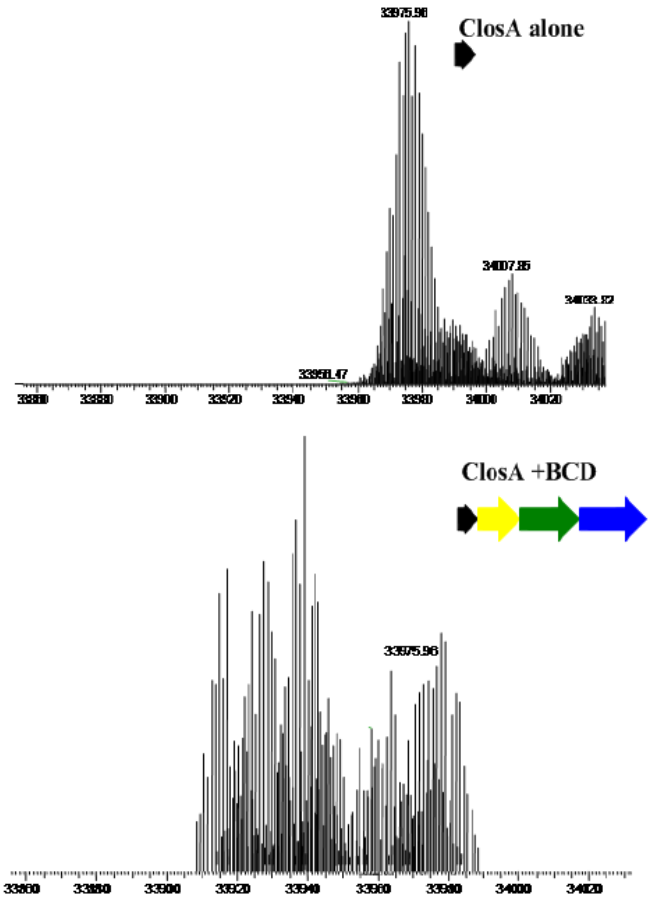

Figure S3.11 HPLC purified recombinantly expressed ClosA was introduced into the FTICR-MS and visualized at 21ppm. Incubating ClosA with the BCD synthetase complex, followed by HPLC produced a series of overlapping ions. Manual annotation of the resulting spectra produced ambiguous mass assignments; therefore a bottom-up approach was taken to visualize the PTM occurring on ClosA.

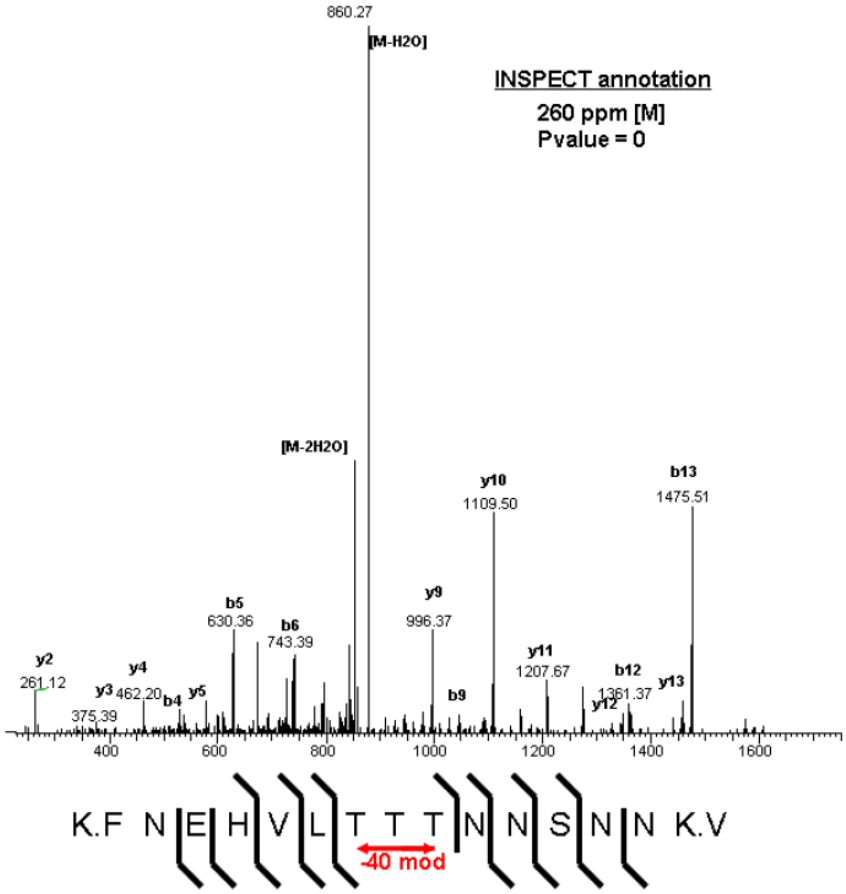

Figure S3.12 Tandem mass spectrometry of T11/T12 heterocyclic peptide.

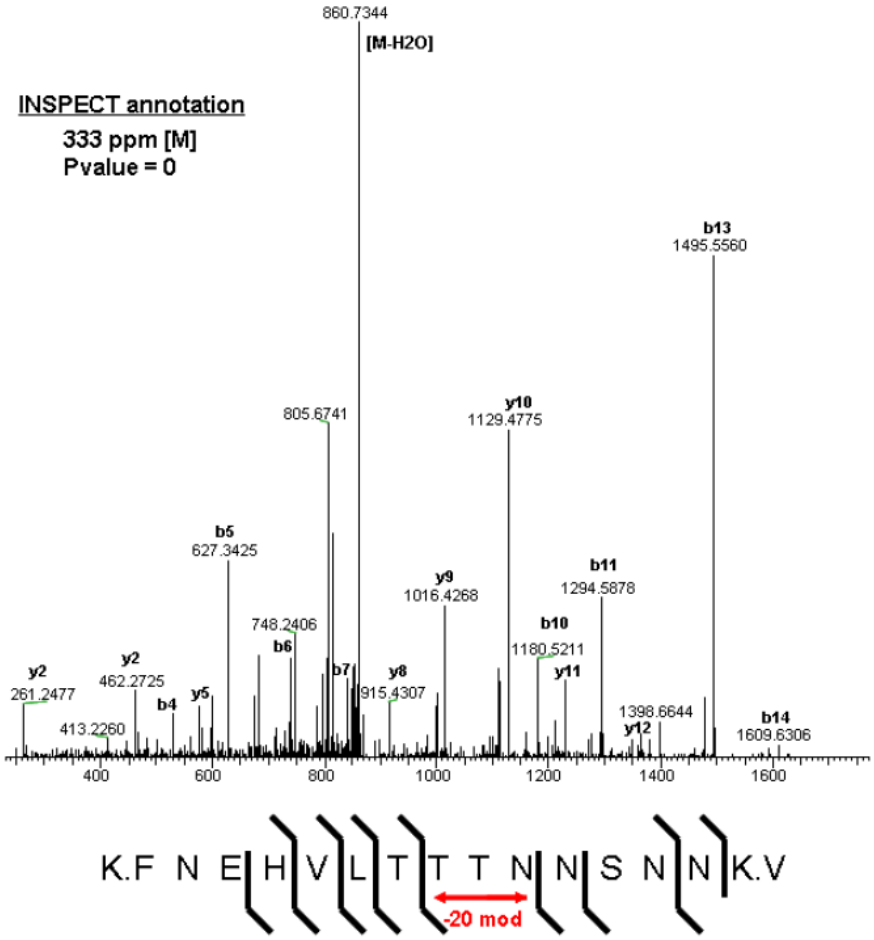

Figure S3.13 Tandem mass spectrometry of T12 heterocyclic peptide.

Acknowledgements

Author contributions: D.G, S.W.L., D.A.M., M.E.H., A.L.M, V.N., P.C.D., designed research; S.W.L., D.A.M., A.L.M., M.E.H., D.G., A.W., J.E.D., V.N., and P.C.D. performed research; N.B., S.D., S.W.L., D.A.M., A.L.M., and D.G. contributed new reagents/analytic tools; D.G., A.L.M., P.C.D., V.N. analyzed data; and D.G., P.D. wrote the paper.

References

1. Yang, Y. L., Xu, Y., Straight, P., and Dorrestein, P. C. (2009) *Nat. Chem. Biol.* 5, 885–887

2. Severinov, K., Semenova, E., Kazakov, A., Kazakov, T., and Gelfand, M. S. (2007) *Mol. Microbiol.* 65, 1380–1394

3. Datta, V., Myskowski, S. M., Kwinn, L. A., Chiem, D. N., Varki, N., Kansal, R. G., Kotb, M., and Nizet, V. (2005) *Mol. Microbiol.* 56, 681–695

4. Nizet, V., Beall, B., Bast, D. J., Datta, V., Kilburn, L., Low, D. E., and De Azavedo, J. C. (2000) *Infect. Immun.* 68, 4245–4254

5. Humar, D., Datta, V., Bast, D. J., Beall, B., De Azavedo, J. C., and Nizet, V. (2002) *Lancet* 359, 124–129

6. Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V., and Dixon, J. E. (2008) *Proc. Natl. Acad. Sci. U.S.A.* 105, 5879–5884

7. Mitchell, D. A., Lee, S. W., Pence, M. A., Markley, A. L., Limm, J. D., Nizet, V., and Dixon, J. E. (2009) *J. Biol. Chem.* 284, 13004–13012

8. Hara-Kudo, Y., Ogura, A., Noguchi, Y., Terao, K., and Kumagai, S. (1997) *Microb. Pathog.* 22, 31–38

9. Savani, J., and Harris, N. D. (1978) *J. Food Sci.* 43, 222–224

10. Sebaihia, M., Peck, M. W., Minton, N. P., Thomson, N. R., Holden, M. T., Mitchell, W. J., Carter, A. T., Bentley, S. D., Mason, D. R., Crossman, L., Paul, C. J., Ivens, A., Wells-Bennik, M. H., Davis, I. J., Cerden˜o-Ta´rraga, A. M., Churcher, C., Quail, M. A., Chillingworth, T., Feltwell, T., Fraser, A., Goodhead, I., Hance, Z., Jagels, K., Larke, N., Maddison, M., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., White, B., Whithead, S., and Parkhill, J. (2007) *Genome Res.* 17, 1082–1092

11. Cotter, P. D., Draper, L. A., Lawton, E. M., Daly, K. M., Groeger, D. S., Casey, P. G., Ross, R. P., and Hill, C. (2008) *PLoS Pathog.* 4, e1000144

12. Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., and Ravel, J. (2005) *Proc. Natl. Acad. Sci. U.S.A.* 102, 7315–7320

13. Liao, R., Duan, L., Lei, C., Pan, H., Ding, Y., Zhang, Q., Chen, D., Shen, B., Yu, Y., and Liu, W. (2009) *Chem. Biol.* 16, 141–147

14. Kelly, W. L., Pan, L., and Li, C. (2009) *J. Am. Chem. Soc.* 131, 4327–4334

15. Morris, R. P., Leeds, J. A., Naegeli, H. U., Oberer, L., Memmert, K., Weber, E., LaMarche, M. J., Parker, C. N., Burrer, N., Esterow, S., Hein, A. E., Schmitt, E. K., and Krastel, P. (2009) *J. Am. Chem. Soc.* 131, 5946–5955

16. Wieland Brown, L. C., Acker, M. G., Clardy, J., Walsh, C. T., and Fischbach, M. A. (2009) *Proc. Natl. Acad. Sci. U.S.A.* 106, 2549–2553

17. Duda, D. M., Walden, H., Sfondouris, J., and Schulman, B. A. (2005) *J. Mol. Biol.* 349, 774–786

18. Jurgenson, C. T., Burns, K. E., Begley, T. P., and Ealick, S. E. (2008) *Biochemistry* 47, 10354–10364

19. Tanner, S., Shu, H., Frank, A., Wang, L. C., Zandi, E., Mumby, M., Pevzner, P. A., and Bafna, V. (2005) *Anal. Chem.* 77, 4626–4639

20. Bandeira, N., Tsur, D., Frank, A., and Pevzner, P. A. (2007) *Proc. Natl. Acad. Sci. U.S.A.* 104, 6140–6145

21. Li, Y. M., Milne, J. C., Madison, L. L., Kolter, R., and Walsh, C. T. (1996) *Science* 274, 1188–1193

22. Zamble, D. B., McClure, C. P., Penner-Hahn, J. E., and Walsh, C. T. (2000) *Biochemistry* 39, 16190–16199

23. Zamble, D. B., Miller, D. A., Heddle, J. G., Maxwell, A., Walsh, C. T., and Hollfelder, F. (2001) *Proc. Natl. Acad. Sci. U.S.A.* 98, 7712–7717

24. Roush, R. F., Nolan, E. M., Lo¨hr, F., and Walsh, C. T. (2008) *J. Am. Chem. Soc.* 130, 3603–3609

25. Regni, C. A., Roush, R. F., Miller, D. J., Nourse, A., Walsh, C. T., and Schulman, B. A. (2009) *EMBO J.* 28, 1953–1964

26. Lehmann, C., Begley, T. P., and Ealick, S. E. (2006) *Biochemistry* 45, 11–19

27. Igarashi, Y., Kan, Y., Fujii, K., Fujita, T., Harada, K., Naoki, H., Tabata, H., Onaka, H., and Furumai, T. (2001) *J. Antibiot.* 54, 1045–1053

28. Cinader, B., and Pillemer, L. (1950) *J. Exp. Med.* 92, 219–237

29. Tsaihong, J. C., and Wennerstrom, D. E. (1983) *Curr. Microbiol.* 9, 333–338

30. Hopkins, C. E., O'Connor, P. B., Allen, K. N., Costello, C. E., and Tolan, D. R. (2002) *Protein Sci.* 11, 1591–1599

31. Oman, T. J., and van der Donk, W. A. (2010) *Nat. Chem. Biol.* 6, 9–18

Chapter 4

Observing the Invisible, a Window into the Metabolic Output of Microbial Colonies

Abstract

The emerging tool of thin-layer-agar MALDI-imaging was applied in the characterization of microbial colonies, a technique that can be adapted to observe and identify the metabolic output of a diverse array of microbial species. Organisms characterized by TLA-IMS included fungi, themophillic fungi, cyanobacteria, marine as well as terrestrial actinobacteria, pathogenic bacteria, and human microbiome derived bacteria. Spatial distributions of the observed and identified metabolic factors were characterized, which are related to their biological and ecological functions.

Introduction

On a daily basis microbiologist harvest and culture microbial organisms for experimental purposes. Lost within the norm of microbial culturing is the invaluable information associated with colony growth, adaptation, and sustainment. The ability to characterize such factors has significant challenges given the chemistries of such molecules are enormously diverse. The molecules range from polyketides (e.g. erythromycin), non-ribosomal peptides (e.g. penicillin), isoprenoids (e.g. artemisinin), fatty acids (e.g octanoic acid), microcins (e.g. Nisin), peptides, to poly-nucleotides and proteins. Because of this chemical diversity, most of these molecules are studied one at a time and out of their spatial and morphological context of a growing microbial colony. Furthermore, traditionally such molecules are extracted and analyzed resulting in limited available options to study the metabolic output of colonies in a multiplexed fashion.

Historically, advancements in technology have been the catalyst for scientific renovation. The emerging tool of matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) imaging mass spectrometry (IMS) is a powerful technique for investigating the spatial distribution of the multiple different intact biological molecules without the need for a nonnative tag or radioactive marker (1, 2). Every detectable target compound can be measured and visualized simultaneously within the mass spectrometry output window. This technology has been wildly used in biomedical research, as pioneer experiments have focused on the topics of disease pathology and pharmaceutical research (3, 4). The systems studied derived from animal tissues such as brain, liver (5) and plant tissue (6). In these methods, timely

preparation methods involving tissue slicing, sample freezing or thawing are required before mass spectral analysis (7).

Recently, our lab has advanced the capabilities of MALDI-TOF imaging by adapting the tool to investigate microbes, namely bacteria, in a previously uncharacterized fashion. Preliminary efforts involved visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria (8). Experimental efforts were aimed at deciphering the producer of specific natural products in order to aid in the process of isolation. The method involved isolation of a cyanobacteria filament harvested from a marine assemblage, followed by anchoring the filament onto a target plate, and then coating the sample with matrix for MALDI-TOF analysis (9). Thereafter, thin-layer-agar MALDI-imaging (TLA-IMS) was developed to visualize the secondary metabolites involved in the chemical communication between two well characterized soil-dwelling bacteria, *Bacillus subtilis* and *Streptomyces coelicolor*. The development of TLA-IMS minimized sample handling and preparation by growing the microbial colonies directly on top of the MALDI plate. The study enabled the visualization of the spatial and temporal characterization of metabolic output during growth of microbial colonies in a competitive environment (10). Furthermore, TLA-IMS allowed for the differentiation of secreted metabolites and metabolites that are associated with the microbial colony. Figure 4.1 displays the overview of TLA-IMS of microorganisms.

The work herein describes advances in the application of TLA-IMS towards the investigation of microbial metabolic output and shows the amenability of IMS to be diverse. Through media optimization, the use of TLA-IMS was expanded to

investigations involving a variety of differential microbial species, establishing MALDI based TLA-IMS as a general, highly applicable method. Using the optimized growth media based TLA-IMS, it was possible to visualize that the sampled microbes display a form of metabolic release into the environment leading credence to the hypothesis that microbes interact with their surroundings toward niche establishment. Additionally, the work shows that a large number of molecular entities can be found from heterogeneous mixtures of organisms isolated from different ecological niches (soil and marine sediment) underscoring the importance of metabolic release involved for the co-existence of multiple microbes in a community. Co-operated with the FT-ICR-MS or MALDI-TOFTOF the work displays a general scheme for the observed ions to be partially or completely identified. This method opens a new window into the metabolic output of microbial colonies, with the potential to spearhead novel lines of investigations within the microbial kingdom.

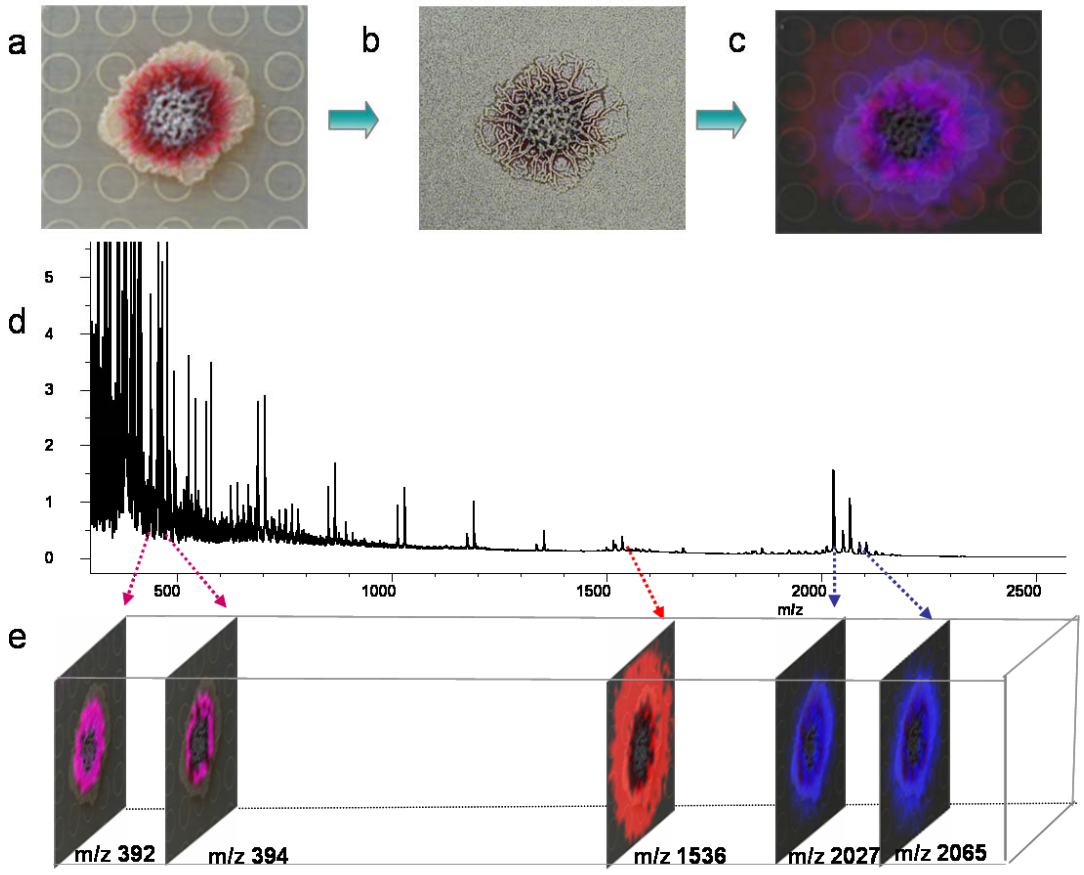

Figure 4.1 Overview of TLA-IMS of microorganism. A. Photograph of the colony grown on the top of the MALDI plate. B. The colony with the agar coated with universal matrix. C. The merged signals with an optical image of the colony. D. The average signal from *m/z* 300-2300 of all the spectra obtained in the imaging runs. E. The separate single *m/z* ion images generated by Bruker Daltonics FlexImaging 2.0.

RESULTS AND DISCUSSION

The growth conditions and imaging parameters for 41 different organisms were optimized (SI 1). The resulting table of mass-to-charge ratios (*m/z)* and corresponding images obtained with TLA-IMS are provided in the supplementary information (SI 2). The organisms subjected to TLA-IMS ranged from fungi (Figure 4.3A), bacteria and cyanobacteria. These microorganisms were isolated from a grand range of inhabits such as the human microbiome, marine sediment, terrestrial soil, insects and plants. Most organisms grew on ISP-2 agar except for five marine derived organisms tested. Alternative media types were pursued for the organisms that did not readily grow on ISP-2. The organisms included *Salinispora* bacteria grown on A1, cyanobacterium *Nostoc* sp. grown on BG-11 agar, and fungi grown on PDA agar. At the end, colonies from 41 different organisms were successfully harvested on ISP2, PDA, A1, or BG-11 and subjected to MALDI-imaging. Resulting data showed on average there are 11.27 significant signals observed within the mass spectral window range that was applied (Figure 4.2).

The most challenging colonies to image were the microbes grown on seawater media (A1), a result from the high salt content obstructing mass spectral analysis. To improve the mass spectral imaging output the A1 media was serially diluted to find a suitable condition. The A1 salt water content reduced to 25 % was used and with the optimized growth conditions the compound etamycin was readily identified from actinomycete strain CNS 575 (Figure 4.2B), a marine bacterial strain isolated from a sediment sample collected from the Nasese shoreline, Viti Levu, Fiji (11). In case of

the cyanobacterium *Nostoc* the majority of the observed compounds were associated with the colony. *Nostoc* cyanobacterium inhabit an aquatic environment therefore it would not make sense to observe release of its essential natural products such as pheophytin A (*m/z* 871 $[M+H]^+$) (4), a trypsin (used to digest food) and thrombin (responsible for blood clotting) inhibitor, that may be involved as a feeding deterrents. In agreement with TLA-IMS data on the cyanobacterium *Nostoc*, is that there is no evidence for release of metabolites from filamentous cyanobacteria except from dying filaments (12).

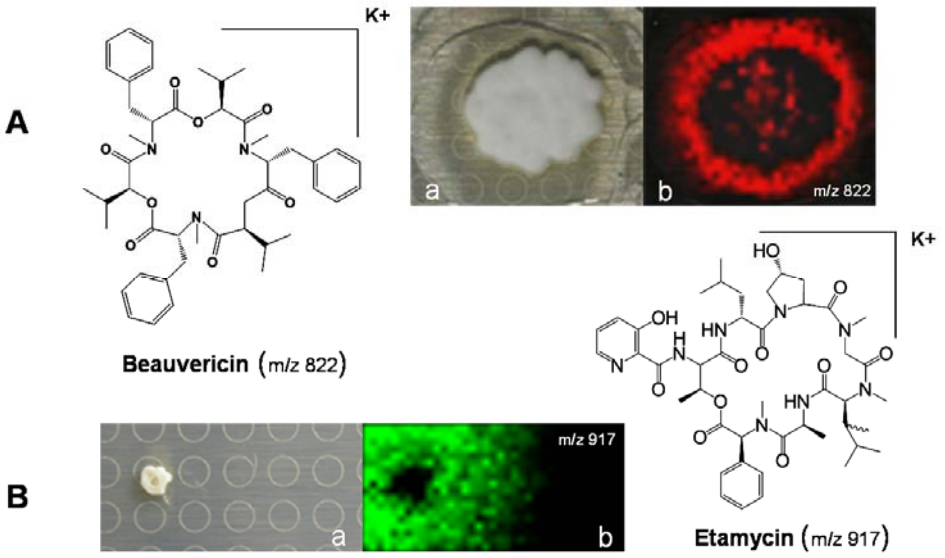

Figure 4.2 The TLA-IMS of *Beauveria bassiana* ATCC7159 , *and Actenomyce* CNS 575 a is the photograph of the colonies on thin layer agar, b is the merge of colonies and *m/z* signals**.** The numbers are the *m/z* signals observed by TLA-IMS. A TLA-IMS of *Beauveria bassiana* ATCC7159, *m/z* 822 is Beauvericin $([M+K]^+$. B TLA-IMS of *Actenomyce* CNS 575, *m/z* 917 is Etamycin ($[M+K]^+$.

Next we sought to examine terrestrial bacteria, which represent commonly found microorganisms in soil. On average the actinomycetes analyzed by TLA-IMS secreted 7.5 compounds while an average of 1.5 metabolites were associated with the colony on ISP2 medium (Figure 4.2). The secreted molecules from soil derived microbes are associated with their ecological roles and have been also explored for therapeutic potential (13, 14). For example, cinnamycin (*m/z* 2041 $[M+H]^+$, *m/z* 2063 $[M+Na]^+$,*m/z* 2079 $[M+K]^+$) of *Streptoverticillium griseoverticillatum* ATCC 31499 (15), CDA (*m/z* 1536 $[M+H]^+$) of *Streptomyces ceolicolor* A3(16) , surfactin (*m/z* 1075 $[M+K]^+$) and plipastatin (*m/z* 1545 $[M+K]^+$) of *Bacillus subtillis* 3610 (17), are antibiotics, maltophilin (*m/z* 511 $[M+H]^+$) and Dihydromaltophilin (*m/z* 513 $[M+H]^+$) of *Lysobacter enzymogenes* C3 (18) are antifungal factors. From the fifteen streptomyces that were investigated by TLA-IMS, eleven of the streptomyces displayed ions above 1800 *m/z*. Masses within this range are commonly associated with thioether, thiazole or oxazole post-translationally modified ribosomal modified peptides. One of the observed masses was dereplicated to a known thioether containing peptides, the lantibiotic SapB (*m/z* 2027 [M+H]+), a factor required for hyphae formation and sporulation, in *S. coelicolor* A3(19). Ten significant *m/z* signals greater than 1800 Da were observed in *S. albus* J1074 (*m/z* 2282), *Streptomyces sp*. SPB78 (*m/z* 2268), *S. ghanaensis* 14672 (*m/z* 2014), *S. clavuligerus* ATCC 53653 (*m/z* 2197), *S. viridochromogenes* 40736 (*m/z* 2169), *S. pristinispiralis* ATCC25486 (*m/z* 2014), *Streptomyces sp*. Mg1 (*m/z* 2003), *S. roseosporus* NRRL11379 (*m/z* 2252), and *S. roseosporus* NRRL15998 (*m/z* 2252). Based on fragmentation patterns obtained by tandem mass spectrometry, the ten signals greater than 1800 Da may be

postranslationally modified peptides (20, 21), a large class of compounds missed in traditional natural products isolation procedures. Given most therapeutics have been mined from natural products derived from microbes, we predict TLA-IMS can play an important role in the pipeline for therapeutic discovery.

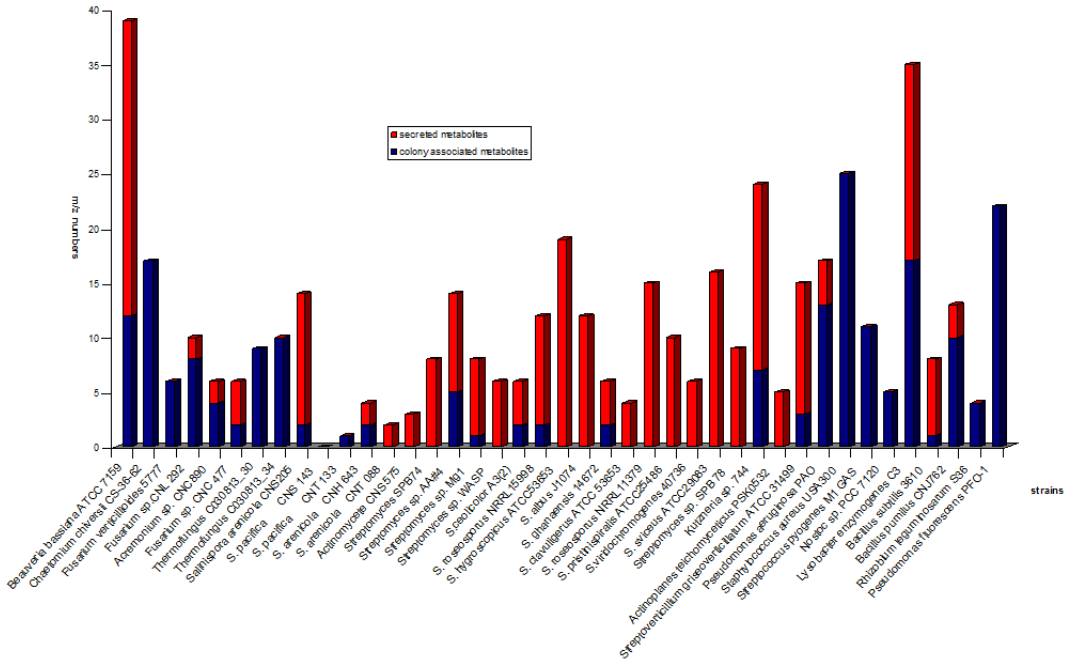

Figure 4.3 The ion numbers observed by TLA-IMS of the tested strains. The ion numbers for *Salinispora arenicola* CNS205, *S. pacifi*ca CNS 143, *S. pacifica* CNT 133, *S. arenicola* CNH 643, and *S. arenicola* CNT 088 are that these strains were grown on A1 agar, and *Nostoc* sp. PCC 7120 was grown on BG-11 agar. The ion numbers for others are that the microorganisms were grown on ISP-2 agar.

Of high biomedical relevance, because of their ability to cause infection in a host, is another class of microorganisms categorized as pathogens. Humans are continuously exposed to numerous pathogens that are the principal source of infections worldwide and cause significant morbidity and mortality. One such type, bacterial pathogens, produce and release a variety of molecular factors in a multiplexed fashion that assist in a pathogen's infectious cycle. A large percentage of these molecular species, which assist in pathogenicity, remain unknown. For example, traditionally *Staphylococcus aureus* infections have been hospital-associated and isolated to individuals with predisposed risks and neonates. Recently, the epidemiology of *S. aureus* has shifted, as infections have progressed to healthy adults and children. Such strains have been defined as community-associated *S. aureus* and are epidemic in some countries (22). Different degrees of pathogenecity have a likely correlation with different levels of metabolic output potential. For this reason, TLA-IMS was applied to study the metabolic output of three distinct pathogenic bacteria with the goal of identifying differential metabolic outputs between isolates. The following pathogens were studied: (i) epidemic methicillin-resistant *Staphylococcus aureues* (EMRSA) ST59 (ii) microbiome (nasal) derived *S. aureus* (MSSA), and the (iii) opportunistic pathogen *Staphylococcus epidermidis*. TLA-IMS of the EMRSA isolate showed 42 ions were present either cell bound or secreted. Of the 42 ions, the recently discovered PSMα1, PSMα2, PSMα3, PSMα4 and delta-toxin (PSMλ) were observed (23). The microbiome derived S. aureus strain indentified by 16S rRNA sequencing to be a MSSA isolate, showed the presence of PSMα3, PSMλ, and 9 other ions. Metabolic output for *S. epidermidis* showed the presence of PSMα and 7 other

ions (Figure 4.4). As predicted TLA-IMS shows each pathogen tested exhibited a different metabolic potential as the EMRSA isolate had the most robust metabolic output followed by the MSSA isolate then the *S. epidermidis*. The correlation between virulence and metabolic out is an avenue of research that shows promise in capturing and identifying molecular species pertaining to bacterial pathogenesis. The ability to visualize molecular species in a multiplex fashion can advance basic biomedical research by addressing the role each molecule plays in pathogenesis individually or in a synergistic fashion.

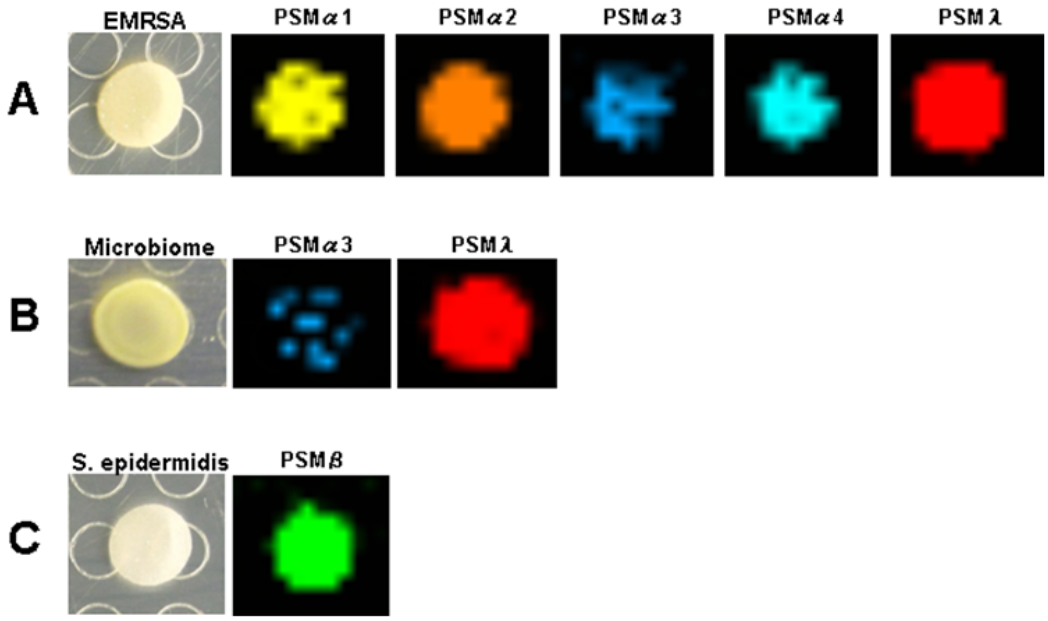

Figure 4.4 Metabolic output of distinct bacterial pathogens. A. The spatial distribution of PSMα1-4 and delta toxin (PSMλ) visualized by TLA-IMS. B. The microbiome derived MSSA isolate showed the presence of PSMα4 and PSMλ C. TLA-IMS of the opportunistic pathogen *S. epidermidis* showed the presence PSMβ. All other molecular species observed for each isolate are supplied in supplemental information strain list.

Finally, research efforts were aimed at capturing metabolic output within a community of bacteria. Every terrestrial or oceanic environment, including on or within each human being, an array of microbial communities exist. Therefore, TLA-IMS was performed to observe the metabolic output from a community of bacterial species. Since one gram of terrestrial or oceanic matter potentially contains hundreds to thousands of unique organisms, samples were collected from a soil garden located in Los Angeles, Ca and marine sediment from La Jolla, Ca. Each sample was then plated on solid agar media, which resulted in a complex mixture of differential microorganisms with distinct morphological phenotypes. Figure 4.5 shows the resulting metabolic output for each sample harvested. Within each sample a variety of different molecular species were observed to be produced by each organism within the community of microorganism. Given colony development is governed by the vast number of different molecular factors, the metabolic exchange visualized can contribute to the understanding of niche establishment in a complex microbial system. Furthermore, because metabolic exchange of a complex microbial system can be visualized by TLA-IMS we hypothesize the method can be extended to look at other complex systems in previously uncharacterized fashions (e.g. human microbiome).

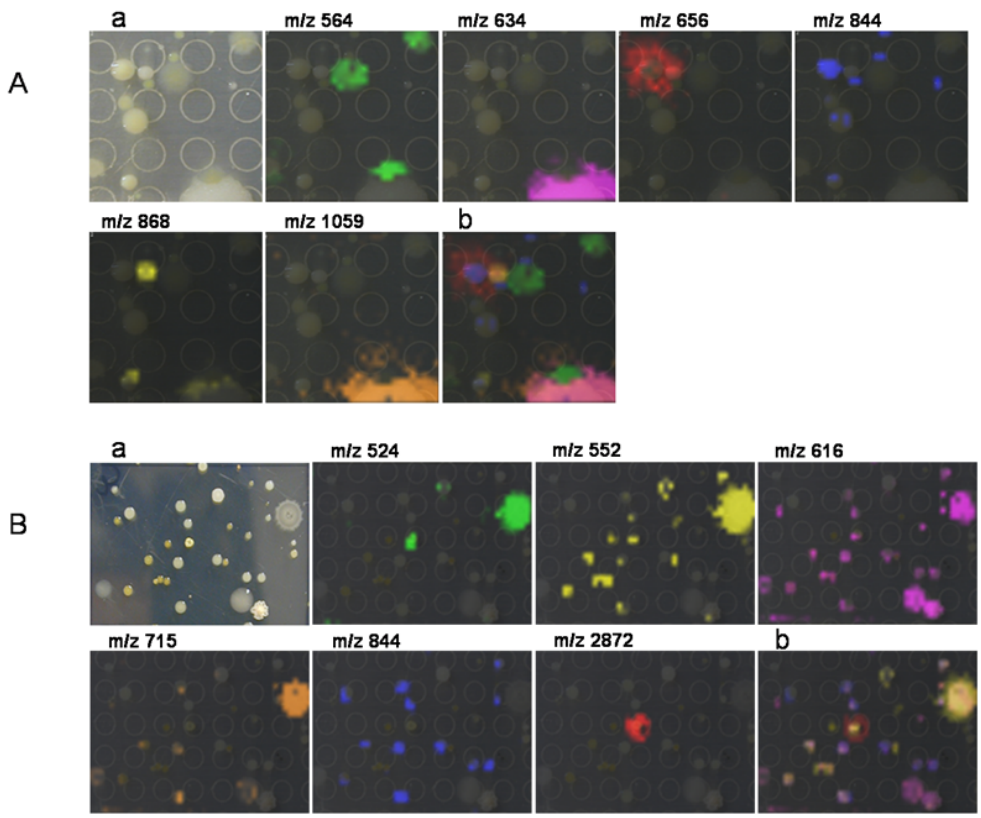

Figure 4.5 The TLA-IMS of multiple complex samples. A is the LA garden soil and B is the La Jolla marine sediment. a is the photograph of the colonies on thin layer agar, and b is the merge of colonies and *m/z* signals. The different colors present some of the *m/z* signals observed

Conclusions

TLA-IMS proves to be a powerful discovery tool that gives insight to the molecular factors, secreted or cell bound, associated with a specific organism during colony formation. Many unknown ions were observed by TLA-IMS (SI 1, SI 2), which indicates that traditional extraction-purification methods inevitably miss many important compounds. During the course of the experiments it was observed that metabolic output differs from culture conditions and medium components. Additionally, mass spectrometry parameters such as ionization and detection limits, play a factor in the type of metabolites visualized. Therefore, some known compounds for the specific strains investigated may not be detected in this study. Nevertheless, this method advances the capabilities of understanding the once hidden microbial world by illuminating the spatial distribution of multiple compounds derived from colonies grown directly on the nutrition agar. We hypothesize that TLA-IMS has the ability to spearhead many avenues of research as the methodology progresses to be a standardized lab practice.

MATERIALS AND METHODS

Strains, Media, and culture conditions. The Bacteria and Fungi used in this study were listed in supplementary information 1 (SI 1). ISP-2 Medium (Yeast extract 4.0 g, Malt extract 10.0 g, Dextrose 4.0 g, Agar 20.0 g, Distilled water 1000.0 mL), A1 medium (starch 10.0 g, Peptone 4.0 g, Yeast extract 2.0 g, Calcium carbonate 1.0 g, Agar 18.0 g, Distilled seawater 1000 mL), PDA medium (DificoTM Potato Dextrose Agar, Becton, Dickinson and Company), LB medium (Bacto-Tryptone 10 g,

Bacto-yeast extract 5 g, NaCl 10 g, Distilled water 1000.0 mL), and BG-11 medium added to 1% agar (Bg-11 is composed of $NaNO_3$ 1.5 g, K_2HPO_4 0.04 g, $MgSO_4 \cdot 7H_2O$, 0.075 g, $CaCl_2 \cdot 2H_2O$ 0.036 g, Citric acid 0.006 g, Ferric ammonium citrate 0.006 g, EDTA (disodium salt) 0.001 g, $NaCO_3$ 0.02 g, Trace metal mix A5 1.0 mL, Agar 10.0 g, Distilled water 1000 ml; Trace metal mix A5: H_3BO_3 2.86 g, $MnCl_2 \cdot 4H_2O$ 1.81 g, $ZnSO_4 \cdot 7H_2O$ 0.222 g, $NaMoO_4 \cdot 2H_2O$ 0.39 g, $CuSO_4 \cdot 5H_2O$ 0.079 g, $Co(NO_3)_2 \cdot 6H_2O$ 49.4 mg, Distilled water 1000 mL) were used for different species. All the strains were grown on the thin layer agar medium at 28 ^oC except *Staphylococcus aureus* USA300 and *Streptococcus pyogenes* M1 GAS at 37 ^oC.

Preparation of the samples. 13 mL of the selected agar medium was poured on the 100 × 15 mm Petri dish plate, and, the autoclaved MSP 96 MALDI target plate was mergered into the un-solid agar, then, stay the plate in bio-safety hood for 5 minutes with the lid open to form the thin-layer agar (TLA) MSP 96 MALDI target plate, at last, inoculate the microorganism. Bacteria *Lysobacter enzymogenes* C3, *Bacillus subtilis* 3610, *Bacillus pumilus* CNJ762, *Pseudomonas aeruginosa* PAO, and *Pseudomonas fluorescens* PFO-1 were inoculated into LB liquid medium to turn turbid. For *Nostoc sp*. PCC 7120, a 10 µl inoculums was pipetted onto a solidified 1% agar BG-11 mixture that had been poured on top of a square piece (4 cm X 4 cm) of aluminum foil contained in a Petri dish. After 7 days of growth, the foil was removed from the Petri dish and the culture-containing agar was peeled off and placed on the Bruker MSP 96 MALDI target plate. The fungi and the spore-producing bacteria were inoculated on the ISP-2 or A1 agar to produced spores. The spores were collected and

diluted into water. 0.2μL either the liquid bacteria or the 1 μL diluted spores were inoculated on the surface of the TLA MSP 96 MALDI target plates. The up side down inoculated Petri dish sealed with parafilm was incubated at 28 ^{o}C for several days. The inoculation times are listed in the SI1. When the colony of microorganism was grown well for analysis, the selected agar with colony was excised by razor blade and transferred to the top of MALDI plate. The colony with agar on the MALDI plate was taken an image with "Close Up" and "no flashing" model of the digital camera, and was evenly covered with Fluka Universal MALDI Matrix (puriss. P.a. matrix substance for MALDI-MS, CHCA:DHB = 1:1) by the sieve method, and then was dried up in 37 ^{o}C oven to generate a uniform distribution of crystalline matrix on the agar. Parameters for TLA-IMS. MALDI time-of-flight (TOF) (Bruker-microflex) was used for TLA-IMS analysis. The sample was subjected to TLA-IMS with an *m/z* range of 200-3600 difference from different species (SI1), and laser intervals in both the X and Y direction in 200 μm increments resulting in the collection of ~1300 spectra.

TLA-IMS Data analysis and the image generation.

The IMS data was analyzed using the FlexImaging software. The masses distributed in the specific area related to the colony were assigned to red colors.

MALDI-TOFTOF and FT-ICR-MS analysis

The microorganism inoculated at the same conditions. The microorganism cultures were extracted by n-butanol and methanol and the extracts were subjected to

FT-ICR-MS (Thermo) or MALDI TOFTOF (ABI 4800 TOFTOF) to observe high resolution mass spectra for molecular formula and MS2 for structure annotation.

Acknowledgments

The following thesis chapter is a work in progress. The chapter was written by the thesis author. The following individuals are the primary authors and have granted permission to add the chapter to the thesis herein: Yuquan Xu and Yu-Liang Yang. The following individuals contributed reagents or have performed research: Yuquan Xu, Yu-Liang Yang, David Gonzalez, Eduardo Esquenazi, Wei-Ting Liu, Anna Edlund, Liangcheng Du, István Molnár, William H. Gerwick, Paul R. Jensen, Michael Fischbach, Chih-Chuang Liaw, Paul Straight, and Pieter C. Dorrestein

Reference

1. Acquadro, E., C. Cabella, S. Ghiani, L. Miragoli, E.M. Bucci, and D. Corpillo. 2009. Matrix-assisted laser desorption ionization imaging mass spectrometry detection of a magnetic resonance imaging contrast agent in mouse liver. Anal. Chem. 81:2779-2784.

2. Anderson, D.M.G., V.A. Carolan, S. Crosland, K.R. Sharples, and M.R. Clench. 2009. Examination of the distribution of nicosulfuron in sunflower plants by matrix-assisted laser desorption/ionisation mass spectrometry imaging. Rapid Commun.Mass Sp. 23 :1321-1327.

3. Cornett, D.S., M.L. Reyzer, P. Chaurand, and R.M. Caprioli. 2007. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. Nat. Methods. 2 :828-833 .

4. Dembitsky, V.M., and T. Rezanka. 2005. Metabolites produced by nitrogen-fixing Nostoc species. Folia microbio. 50:363-391

5. Esquenaz, E., C. Coates, L. Simmons, D. Gonzalez , W.H. Gerwick , and P.C Dorrestein . 2008. Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging. Mol. Biosyst. 4:562-570.

6. Francese, S., F.R. Dani, P. Traldi, G. Mastrobuoni, G. Pieraccini, and G.Moneti. 2009. MALDI mass spectrometry imaging, from its origins up to today: the state of the art. Comb. Chem. High T. SCR. 1 2:156-174.

7. Haste, N.M., V.R. Perera , K.N. Maloney, D.N. Tran, P. Jensen, W. Fenical, V. Nizet, and M.E. Hensler. 2010. Activity of the streptogramin antibiotic etamycin against methicillin-resistant *Staphylococcus aureus* .J. Antibiot. 63 :219-224

8. Kaletta, C., K.D. Entian K, and G. Jung. 1991. Prepeptide sequence of cinnamycin (Ro 09-0198): the first structural gene of a duramycin-type lantibiotic. Eur. J. Biochem. 199 :411-415.

9. Rosengren K.J., R.J. Clark , N.L. Daly , U. Göoransson, A. Jones, and D.J. Craik. 2003. Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. J. Am. Chem. Soc. 125 :12464–12474.

10. Sauer, S., B.M. Lange, J. Gobom, L. Nyarsik, H. Seitz, and H. Lehrach. 2005. Miniaturization in functional genomics and proteomics. Nature reviews genetics. Nat. Rev. Genet. 6:465-476.

11. Schlievert, P.M., K.L. Strandberg, Y.C. Lin, M.L. Peterson, D. Pharm, and D.Y.M. Leung. 2010. Secreted virulence factor comparison between methicillin-resistant and methicillin-sensitive *Staphylococcus aureus*, and its relevance to atopic dermatitis. J. Allergy Clin. Immun. 125: 39-49.

12. Willey, J.M., and W.A. van der Donk. 2007. Lantibiotics: peptides of diverse structure and function. Annu. Rev. Microbiol. 61:477-501 .

13. Xu, Y., R. Orozco, E.M. Wijeratne, A.A. Gunatilaka, S.P. Stock, and I. Molnár. 2008. Biosynthesis of the Cyclooligomer Depsipeptide Beauvericin, a Virulence Factor of the Entomopathogenic Fungus *Beauveria bassiana*. Chem. Biol. 15 :898-907.

14. Xu, Y., R. Orozco, E.M. Wijeratne, P. Espinosa-Artiles, A.A. Gunatilaka , S.P. Stock, and I. Molnár. 2009. Biosynthesis of the cyclooligomer depsipeptide bassianolide, an insecticidal virulence factor of *Beauveria bassiana*. Fungal Genet. and Biol. 46 :353-364.

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

16. Yu, F., K. Zaleta-Rivera , X. Zhu , J. Huffman, J.C. Millet, S.D. Harris, G. Yuen, X. Li, and L. Du . 2007. Structure and biosynthesis of heat-stable antifungal factor (HSAF), a broad-spectrum antimycotic with a novel mode of action. Antimicrob. Agents Ch. 51 :64–72.

Chapter 5

Insight into Probiotic Effects by *Bacillus subtilis* Against *Staphylococcus aureus*

Summary

Probiotic action derives from microbial niche establishment through the elaboration of molecules with antibiotic benefit. The molecular actions that govern probiotic action leading to the clearing of toxin producing bacteria (e.g. pathogens) remain undetermined. The emerging technology of imaging mass spectrometry (IMS) enabled the observation that there is directionality in the metabolic output of the probiotic organism *B. subtilis* when co-cultured with *S. aureus* (MRSA). The directionally released antibiotic alters MRSA virulence factor production and colonization. Therefore, IMS provides insight into the largely hidden nature of competitive microbial encounters, niche establishment, and probiotic action.

Introduction

Microbial species exist in perpetual competition with one another for suitable ecological niches to support their survival and growth. Diverse niches exist in the general environment, such as soil or aquatic habitats, upon or within unicellular or multi-cellular eukaryotic life forms, including protozoa, plants and animals. On human skin and mucosal surfaces, the outcome of microbial competitions determines our normal flora, essential for health and immune homeostasis, or colonization with potential pathogens, the initial step in the pathogenesis of most infectious diseases (Grice et al., 2009; Hoffman et al., 2006). Recently microbial competition has been taking advantage of for health benefits by the administration of live microbes defined as Probiotics. The mechanisms and molecular species that govern probiotic effects, such as inhibiting the growth or effectiveness of toxic bacteria (e.g. pathogen) are not clearly understood.

Efficiency of nutrient acquisition and strategies for surface attachment are essential factors for microbial niche survival, while elaboration of compounds that kill or limit the growth of competing strains or species, can promote niche monopolization with probiotic benefits. The released compounds include secondary metabolite antibiotics (e.g. penicillin, chloramphenicol, tetracycline), bacteriocin peptides, or low-molecular weight toxic molecules such as hydrogen peroxide, each coupled to mechanisms for intrinsic resistance/immunity by the producing strain (Gonzalez et al., 2010; Li et al., 2010; Queck et al., 2009).

An attractive hypothesis holds that probiotic microbes would regulate and optimize their production of such molecules to kill, limit the growth, or modulate the

metabolism of toxic bacteria for maximal advantage. However, there have been no studies that address this problem in a geometric and multiplexed fashion, in part, due to the lack of available tools. The present work applies the emerging technology of imaging mass spectrometry (IMS) to study such interactions, using as models two well-characterized Gram-positive bacterial species, the probiotic bacterium *Bacillus subtilis*, and an epidemic *Staphylococcus aureus* isolate.

B. subtilis is an avirulent bacterium found on skin, in the digestive tract, in epithelial wounds, extremities of the human body, livestock, and soil (Ara et al., 2006; Earl et al., 2008). Because *B. subtilis* is ubiquitous, it has developed adaptive strategies to subsist in diverse environments via the production and secretion of a large number of genetically encoded molecules that control the growth of neighboring organisms (Stein, 2005; Liu et al., 2010). For this reason, *B. subtilis* is commercially sold as a skin care product, food ingredient for human consumption, animal feed, fertilizer, or an antibiotic substitute. The molecular mechanisms by which these products work are poorly understood, in particular how probiotic benefits manifest themselves. Similarly, *S. aureus* is found on human skin, digestive tracts, nares, livestock, and surgical instrumentation (Iwase, 2010; Otto, 2010; Roberson, 1994). At minimum, 30% of the world population is colonized with *S. aureus*, a bona fide pathogen that has developed significant resistance against a variety of antibiotics and is the cause of more fatalities in the United States than HIV/AIDS (Enright, 2010; Klevens et al., 2007).

Although *S. aureus* colonizes a large segment of the world inhabitants, it only produces clinical infection in a subset of this population. One attractive hypothesis is

that neighboring organisms occupying the same environmental niche (e.g. skin), particularly those with probiotic benefits, respond by secreting an array of antibiotic-type molecules to control *S. aureus* developmental phenotypes and thereby alter its ability to proliferate in the host. While secreted metabolites of neighboring organisms are likely to alter the growth of pathogens, there have been no studies that address this problem in a geometric and multiplexed fashion, in part, due to the lack of available tools. The study herein reveals *B. subtilis* can inhibit the growth of an epidemic *S. aureus* isolate, surprising directional release of a molecule with antimicrobial and metabolism-altering properties, and indicates a utility for IMS in dynamic analysis of interspecies metabolic exchange and perhaps future discovery of new antibiotics or probiotic candidates.

Results and Discussion

Imaging Mass Spectrometry of the interaction between *B. subtilis* and *S. aureus*

When *B. subtilis* was spotted on top of a lawn of methicillin-resistant *S. aureus* skin isolate ST59-MRSA-IV (Geng et al., 2010), a zone of clearing was observed. To provide insight into the molecular details of this interaction, the sample was subjected to IMS (Caprioli et al., 1997; Schwamborn et al., 2010; Yang et al., 2009). The imaging data identified the presence of two *B. subtilis* metabolites, the nonribosomal peptide synthetase-derived lipopeptide antibiotics surfactin and plipastatin (Stein, 2005), as well as delta-toxin (aka phenol-soluble modulin λ or PSM λ), a membrane disruptive peptide produced by *S. aureus* (Otto, 2010) (Fig. 5.1a). Although this

approach enabled the characterization of their interactions within a traditional experimental set up and put forth potential bioactive agents against MRSA, it is not representative of an encounter of the two organisms in nature as the experiment lacks spatial direction.

In order to provide a spatial dimension to the interacting organisms, *B. subtilis* isolated from soil was inoculated with *S. aureus* in a T-shaped configuration and assessed by IMS. Once again, the T-shape interaction demonstrated *B. subtilis* could inhibit *S. aureus* at the bacterial interface (Fig. 5.1b). Mapping the chemistry associated with the observed phenotypes by IMS resulted in an unexpected observation--the amount of surfactin produced by *B. subtilis* was increased at the bacterial interface (Fig. 5.1b and 5.1c), a directional phenomenon that did not require direct contact between the two bacterial species. This phenotype was not observed in the controls where each organism was cultured individually (Fig. 5.1d and 5.1e), nor was it a result of membrane damage in *B. subtilis* (Fig. 5.2a). Based on co-spots of a known amount of surfactin (Fig. 5.2b), we estimated that the highest concentration is 140 μg/mL, a concentration that is in agreement with previous findings (Coutte et al., 2010). Similar results were obtained with interactions of *S. aureus* and two additional *B. subtilis* isolates from soil or human skin (Fig. 5.3a).

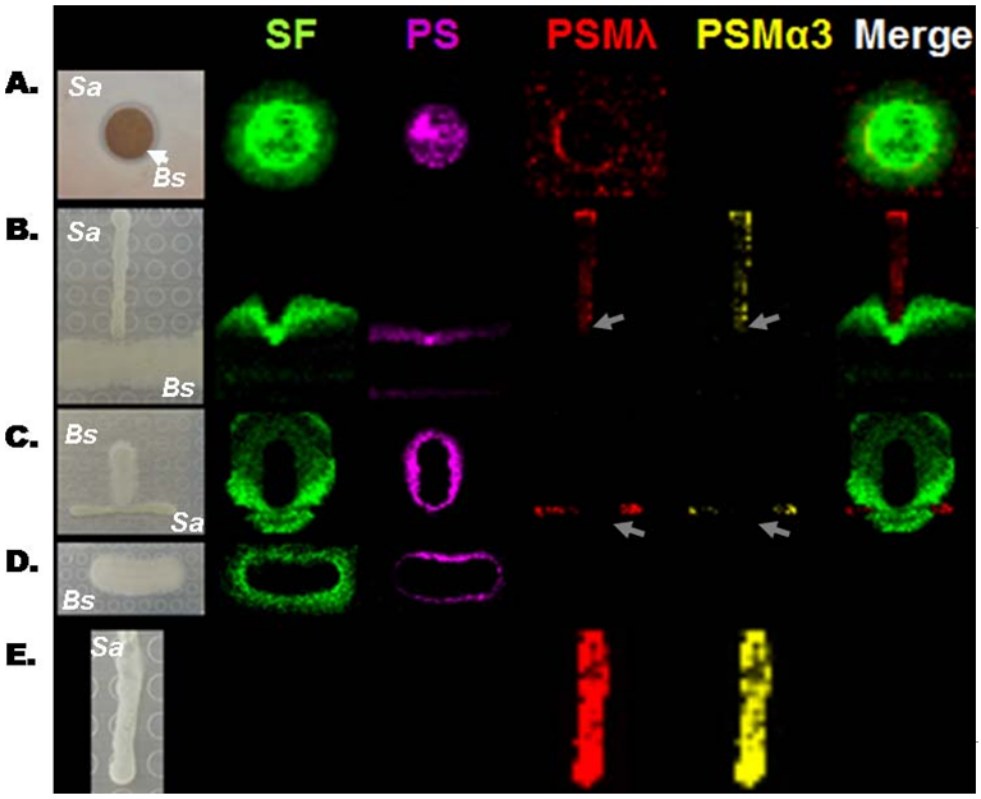

Figure 5.1 IMS of the interaction between *B. subtilis (Bs)* and *S. aureus (As)*. (A) Zone of inhibition (B) T-shape experiment (C) Converse of (B) (D) *B. subtilis* alone (E) *S. aureus* alone. Ions distributions are represented by color: surfactin (green), plipastatin (magenta), PSMλ (red) and PSMα3 (yellow). Grey arrows indicate areas of toxin suppression.

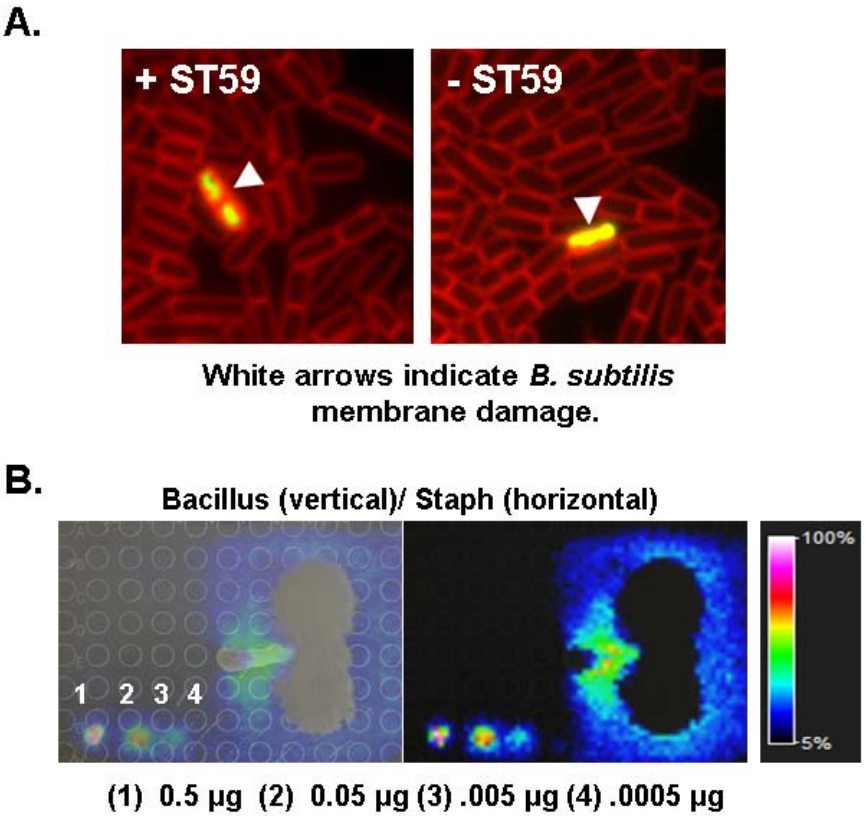

Figure 5.2 A. Viability assessment of *B. subtilis* cells by fluorescence microscopy. *B. subtilis* on the (+/-) MRSA sides were stained with Sytox green and FM 4-64. Viable cell counts were 97% on the (-) side and 98% on the (-) side. **B.** Imaging mass spectrometry of spot assays (0.5-0.0005µg). Ion intensity color scaling indicates the highest naturally released surfactin is located at the bacterial interface.

Surfactin and Plipastain Bioactivity against S. aureus

Based on the ring of inhibition and T-shaped IMS data, *B. su*btilis has the ability to prevent MRSA growth, and furthermore suggests that surfactin or plipastatin play a role in the inhibitory phenotype. To test the inhibitory hypothesis, both surfactin and plipastatin were purified from *B. subtilis* and tested for bioactivity. Both purified antibiotics suppressed *S. aureus* growth on solid agar; surfactin had the more pronounced effect (Fig. 5.3b).

Moreover, the IMS analysis suggested beyond growth suppression, *B. subtilis* influenced the level of production of PSMλ and related phenol-soluble modulins (PSM) in *S. aureus* (Fig. 5.2b and 5.2c). In the controls, without *B. subtilis,* robust production of PSMλ and PSM were observed (Fig. 5.2e). Given the recently discovered PSM have been attributed, at least in part, to the emergence of community-associated methicillin-resistant S. *aureus* (Wang et al., 2007), isolates with the ability to infect healthy adults and children, we attempted to reconstitute the PSM suppression phenotypes. When purified surfactin was spotted on an agar plate next to a colony of growing *S. aureus*, IMS of this sample showed that surfactin indeed inhibited the production of PSM (Fig. 5.4a).

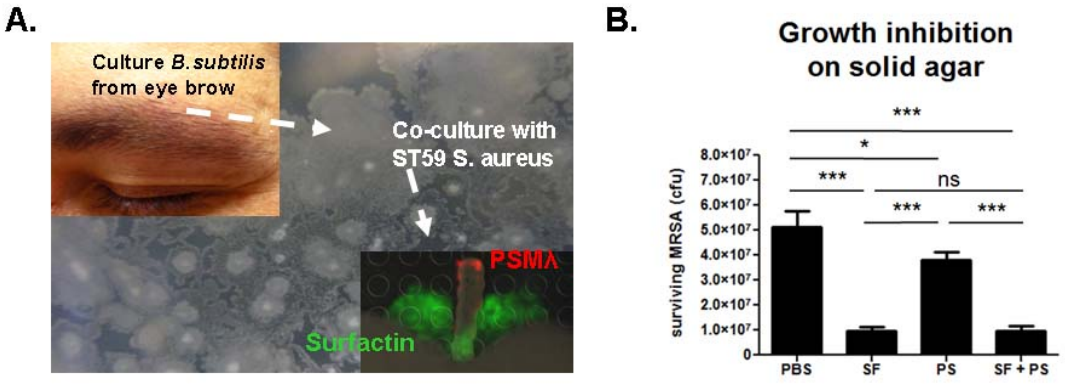

Figure 5.3 A. *B. subtilis* isolated from a human eye brow showed similar result to strain 3610. The identity of this *B. subtilis* isolate was validated by 16S rRNA sequencing. **B.** Inhibition of *S. aureus* growth on solid agar. Purified surfactin and plipastatin were resuspended in 70% ethanol alone or in combination and 10 µl of solution was pipetted onto agar plates. 2 x 10^5 cfu of *S. aureus* strain ST59 was pipetted onto agar at the location of compounds. Plates were incubated at 37°C for 4 hrs. Samples were serially diluted in PBS and plated on Todd Hewitt agar for enumeration. Inhibition of bacterial growth was calculated as a percentage of initial inoculums.

To further investigate PSM suppression, *S. aureus* was grown in the presence and absence of surfactin and subjected LC-MS/MS based spectral counting (Liu et al., 2004; Carvalho et al., 2008). In the presence of 10 μg/ml of surfactin, a concentration of surfactin that does not lyse *S. aureus*, at least a five-fold suppression of PSMλ, PSMα1, PSMβ2, and α-hemolysin (DeLeo, 2009) was observed though *S. aureus* reached similar cell densities and colony forming unit counts compared to the untreated control (Fig. 5.4b).

To validate surfactin mediated suppression of PSM, the detection limit of PSMλ was monitored in a concentration dependent manner by the use of intact cell MALDI-TOF analysis. Monitoring the PSMλ mass range (~3000 Da), surfactin was titrated at varies concentrations into inoculums of MRSA. In the absence of surfactin a robust ion signal was observed within the monitored mass range and lost when the highest titrated concentration of surfactin was monitored (10 μg/ml). Indeed, the ion detection of surfactin was restored in a concentration dependant manner as the subsequent lower titration ranges were monitored by intact cell MALDI-TOF (Fig. 5.4c).

Murine skin colonization models

Given the displayed activity of growth inhibition and PSM suppression activities observed *in vitro*, we sought to test if bioactivity could be reconstituted testing murine skin colonization models. To this end, the flanks of CD1 mice were shaved and individually treated with surfactin, plipastatin, or surfactin/plipastatin, then inoculated with *S. aureus.* We found that surfactin and plipastatin could restrict growth of *S. aureus* on mammalian skin. The data showed after 4 h, a significant

decrease in cfu count was detected with both natural lipopeptides, consistent with their involvement in suppressing *S. aureus* colonization as indicated by the imaging mass spectrometry data (Fig. 5.4D).

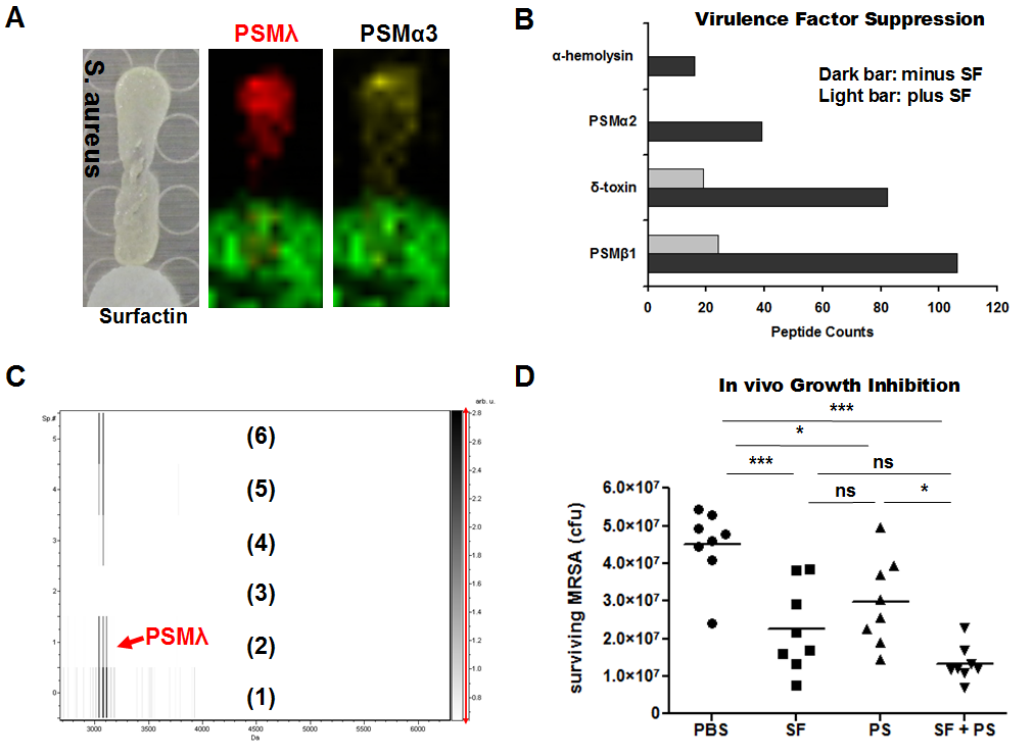

Figure 5.4 A. MRSA cultured adjacent to a filter disk containing surfactin showed suppression of PSMλ and PSMα3 when monitored by IMS. **B.** Peptide spectral counts of virulence factors identified in supernatant via nano-capillary LCMS/MS. Dark bars represent minus surfactin and light bars plus surfactin. **C.** Purified surfactin spotted next to a ST59 intact cell MALDI-TOF analysis of surfactin (SF) titration: (1) minus surfactin (2) DMSO (3) 10 µg/mL SF (4) 1.0 µg/mL SF (5) 0.01 µg/mL SF (6) 0.0001 µg/mL SF. Data shows PSMλ suppression and restoration as surfactin is titrated. **D.** Growth inhibition of MRSA on mammalian skin after 4 hrs: surfactin (SF), plipastatin (PS). * indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$, ns indicates no significant difference.

Conclusions

Because *B. subtilis* possess a nonpathogenic nature in healthy individuals, strains have been explored as potential probiotic agents of negligible consumer risk, in scenarios such as periodontitis (Tsubura et al., 2009) or *Helicobacter* gastritis (Park et al., 2007) or as an alternative to feed antibiotics in agricultural production of poultry (Li et al., 2006), cattle (Jenny et al., 2006), swine (Guo et al., 2006) or seafood (Liu et al., 2009). Since *B. subtilis* inhibited *S. aureus* growth and virulence factor expression through elaboration of its natural cyclic lipopeptide antibiotics, we hypothesized this competitive advantage could be harnessed to depopulate *S. aureus* from a specific niche on the host. Rapid advancement in DNA sequencing technologies is allowing metagenomic approaches to determine the composition of microbial niches (e.g. skin). However, companion technologies that can determine the molecular interactions between organisms in a multiplexed manner are largely absent. This work highlights IMS as a hypothesis-generating tool to reveal the molecular interplay that determines the outcome of competitive microbial encounters, potentially yielding novel candidate antimicrobials or optimizing strain selection for probiotic applications. The surprising finding of directional release of an inhibitory molecule by one bacterial strain toward a neighboring species such as the human pathogen, *S. aureus*, suggest that other secreted factors (e.g. cytotoxins, quorom sensing factors, nutrient scavenger proteins) could play important roles in health and disease and that these molecular events might be examined by IMS technology for more sophisticated modes of spatial deployment.

Experimental Procedures

Imaging mass spectrometry

ISP2 media was prepared with yeast extract 2 g, malt extract 4 g, dextrose 2 g, and agar 10 g, by addition to 500 mL of deionized water and autoclaved. 10 mL of autoclaved agar was poured into sterile Petri dishes under flame. Colony growth was initated by streaking 3 uL each of overnight growths of *S. aureus and B. subtilis* at equal cell densities on the prepared thin-layer agar ISP2 plates. The bacterial colonies were allowed to grow for 48 hrs at 30°C before transferring the co-culturing experiment to a Bruker MSP 96 MALDI anchor plate. Thereafter, the target plate was uniformly covered with Sigma universal matrix (α-cyano-4-hydroxycinnamic acid and 2, 5-dihydroxybenzoic acid) by the use of a 50 μm sieve. Once that sample was completely covered with matrix, it was baked in a 37 °C oven for 3-4 hours until it was deemed dried at which point it was subjected to imaging mass spectrometry. The Bruker MSP 96 anchor plate containing the sample was inserted into a Microflex Bruker Daltonics mass spectrometer outfitted with Compass 1.2 software suite (Consists of FlexImaging 2.0, FlexControl 3.0, and FlexAnalysis 3.0). The sample was run in positive mode, with 200-350 μm laser intervals in XY and 50–54% laser power. A photomicrograph of the colonies to be imaged by mass spectrometry was loaded onto the Fleximaging command window. Three teach points were selected in order to align the background image with the sample target plate. After the target plate calibration was complete, the AutoXecute command was used to analyze the samples. The method setting under the FlexControl panel was as ImaingRPpepmix. Which consist of the following settings: Laser: Fuzzy Control—On, Weight 1.00; Laser

Power varied between 50–54%; Matrix Blaster 0. Evaluation: Peak Selection Masses from m/z 400–4000, mass control list—Off. Peak Exclusion—Off. Peak Evaluation Processing Method Default, Smoothing—Off, Baseline Subtraction—On, Peak—Resolution higher than 100. Accumulation: Parent Mode: On, Sum up to 20 satisfactory shots in 20 shots, Dynamic Termination—off. Movement: Random Walk—2 shots at raster spot. Quit sample after—2 subsequent failed attempts. Processing: Flex analysis Method—none, Boito's MS method—none. Sample Carrier: nothing Spectrometer: On, Ion Source 1—19.00 mV, Ion Source 2—16.40 mV, Lens—9.45 mV, Reflector 20.00, Pulsed Ion Extraction—190 ns, Polarity—Positive. Matrix Suppression: Deflection, Suppress up to: m/z 400. Detector Gain—Reflector 4.1. Sample Rate–2.00 GS/s, Mode–low range, Electronic Gain—Enhanced, 100 mV. Real time Smooth—off. Spectrometer, Size: 81040, Delay 42968. Processing Method: Factory method RP_2465. Setup: Mass Range—Low. Laser Frequency—20 Hz, Auto teaching—Off. Instrument Specific Settings: Digitizer—Trigger Level— 2000 mV, Digital off Linear—127 cant, Digital off Reflector—127 cant. Detector Gain Voltage Offset, Linear—1300 V, Reflector—1400 V. Laser Attenuator, Offset—12%, Range—30%, Electronic Gain Button Definitions, Regular: 100 me (offset line) 100 mV (offset ref) 200 mV/full scale. End: 51 mV (offset line), 51 mV (offset ref) 100 mV/full scale. Highest: 25 mV (offsetting) 25 mV (offset ref) 50 mV/full scale. Calibration: Calibration was accomplished using a Bruker Daltonics Pepmix 4 as an external standard. Zoom Range 1.0% Peak Assignment Tolerance- User Defined-500 pap. After data acquisition, the data was analyzed using the FlexImaging software. The resulting mass spectrum was filtered manually in 0.5–3.0 Ad increments with

individual colors assigned to the specific masses. Ions of interest were identified by the use of tandem mass spectrometry. IMS of over twelve interactions between *S. aureus* and *B. subtilis* should similar, consistent phenotypes**.**

Visualization of membrane damage

B. subtilis cells were harvested from a colony that was either adjacent to that of *S. aureus* strain ST59 or grown in isolation. Cells were resuspended in T-base, stained with a final concentration of 30 μg/ml FM 4-64, which stains the membranes red, and 2.5 uM Sytox Green (Invitrogen), which brightly stains only cells with membrane damage. The cells were immobilized with poly-L-Lysine and visualized with an Applied Precision Spectris microscope. Experiments performed on different inoculums and microscope should consistent data.

Nanocapillary LC-MS/MS

Samples were prepared by spiking *S. aureus* isolate ST59 inoculums with 10 ug/mL of purified surfactin and allowed to grow for 16 hrs. Thereafter, samples were extracted with 1-butanol, lyophilized, and resuspended for LC-MS/MS. The LC-MS/MS apparatus was prepared by preparing nanocapillary columns. The columns were prepared by drawing 360-m outer diameter, 100-m 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 ST59-MRSA-IV extracts (+/-) 10 μg/mL surfactin by flowing 90% of solvent A and 10% of solvent B (CH3CN, 0.1% AcOH) at l/min for 5 min, l/min for 3 min, and l/min for 12

min. At 20 min, the flow rate was increased to 200 nL/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 extract contents 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 LTQ-MS was tuned and calibrated to achieve a background signal NL: 4.5E3. *LC-MS/MS Acquisition*. Two different MS/MS acquisitions methods were used to compile spectra of the surfactin spiked *S. aureus* ST59 prepared extracts. Methods M1 and M2 use data-dependent acquisition with varying cut-offs on the exclusion list capacity and time, collecting fragmentation data for the 1-10th most abundant ions. *Data Processing*. All collected data files (RAW) were processed with a DOS command line version of InSpecT software (Tanner et al., 2005). Relevant input search parameters included the following: (i) post-translational modification search, +28 Da (formylation) (ii) database, modified USA300 genome, USA300 genome "reversed" or "phony database," and common contaminants database (iii) PTM allowed per peptide = 3 (iv) *b*- and y-ion mass off set tolerance = 0.5 Da (v) parent mass tolerance = 1.5 Da. Experiments were reproducible between multiple runs.

***In vivo* inhibition of *S. aureus* growth**

Flanks of 8-10 week old female CD1 mice were shaved and naired. The following day, mice were anaesthetised with ketamine/xylazine and placed on heating pads to maintain body temperature. Purified surfactin and plipastatin were resuspended in 70% ethanol alone or in combination and 10 µl of solution was

pipetted onto pre-marked spots on the mouse flank and allowed to dry. Four spots per mouse were prepared as follows: negative control (70% ethanol), surfactin (20 μg), plipastatin (20 μg), or surfactin and plipastatin together (20 μg each). *S. aureus* strain ST59 was grown to mid-logarithmic phase (OD600=0.5), washed with sterile PBS and diluted to 2 x 10^7 cfu/ml. 10ul (2 x 10^5 cfu) of bacteria was spotted onto agar plates and allowed to dry. Once the spots had dried, agar discs were excised using a 6mm biopsy punch. One agar disc containing 2 x 10^5 cfu of *S. aureus* ST59 was placed onto each spot (4 per mouse) and affixed to the mouse with Tegaderm transparent wound dressing. After 4h, each spot was excised and skin together with agar disc was placed into 2 ml screw cap tubes containing 1 mm silica/zirconia beads in 1 ml PBS. The samples were homogenized in a mini-BeadBeater-8 for 1 min at full speed twice, placing the tubes on ice in between. Homogenized samples were serially diluted in sterile PBS and plated on Todd Hewitt agar plates for enumeration. Inhibition of bacterial growth by purified compounds was calculated as a percentage of initial innoculum. *Ethics approval.* Permission to undertake animal experiments was obtained from University of California, San Diego animal subject ethics committee.

Acknowledgments

Performed experiments: David Gonzalez, Nina Haste, Andrew Hollands, Tinya Fleming, Matthew Hamby. Designed experiments: David Gonzalez, Nina Haste, Kit Pogliano, Victor Nizet and Pieter Dorrestein. Wrote the paper: David Gonzalez, Victor Nizet, and Pieter Dorrestein

Reference

1. Ara, K., Hama, M., Akiba, S., Koike, K., Okisaka, K., Hagura, T., Kamiya, T., Tomita, F.(2006). Foot odor due to microbial metabolism and its control. *Can J Microbiol*. 52, 357-64

2. Caprioli, R.M., Farmer, T.B., Gile, J. (1997). Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem*. 69, 4751-4760

3. Carvalho, P.C., Hewel, J., Barbosa, V.C., Yates 3rd, J.R. (2008). Identifying differences in protein expression levels by spectral counting and feature selection. *Genet Mol Res*. 2, 342-356

4. Coutte, F., Leclere, V., Bechet, M., Gues, J.S., Lecouturier, D., Chollet-Imbert, M., Dhulster, P., Jacques, P. (2010). Effect of pps disruption and constitutive expression of srfA on surfactin productivity, spreading and antagonistic properties of Bacillus subtilis 168 derivatives. *J Appl Microbiol*. 2, 480-491

5. Deleo, F.R., Chambers, H.F. (2009). Reemergence of antibiotic-resistant Staphylococcus aureus in the genomics era. *J Clin Invest*. 119, 2464–2474
Earl, A.M, Losick, R., Kolter, R. (2008). Ecology and genomics of Bacillus subtilis. *Trends Microbiol*. 16, 269-275

6. Enright, M.C., Robinson, D.A., Randle, G., Feil, E.J., Grundmann, H., Spratt, B.G. (2010). The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). *Proc. Natl. Acad. Sci.* U.S.A. 99, 7687-7692

7. Geng, W. et al. (2010). Molecular characteristics of community-acquired, methicillin-resistant Staphylococcus aureus isolated from Chinese children. *FEMS Immunol Med Microbiol*. 58, 356-362

8. Gonzalez et al. (2010). Clostridiolysin S, a Post-translationally modified Biotoxin from Clostridium botulinum. *J. Biol. Chem*. 36, 28220-28228

9. Grice, E.A. et al. (2009). Topographical and Temporal Diversity of the Human Skin Microbiome. *Science*. 324, 1190-1192

10. Guo, X., Li, D., Lu, W., Piao, X., Chen, X. (2006). Screening of Bacillus strains as potential probiotics and subsequent confirmation of the in vivo effectiveness of Bacillus subtilis MA139 in pigs. *Antonie Van Leeuwenhoek*. 2, 139-146

11. Hoffman, L.R., Deziel, E., D'Argenio, D.A., Lepine, F., Emerson, J., McNamara, S., Gibson, R.L., Ramsey B.W., Miller, S.I. (2006). Selection for Staphylococcus

aureus small-colony variants due to growth in the presence of Pseudomonas aeruginosa. *Proc Natl Acad Sci U.S.A.* 52, 19890-19895

12. Iwase, T., Uehara, Y., Shinji, H., Tajima, A., Seo, H., Takada, K., Agata, T., Mizunoe, Y. (2010). Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization. *Nature*. 465, 346-349

13. Klevens, R.M. et al. (2007). Invasive Methicillin-Resistant Staphylococcus aureus Infections in the United States, *J. Am. Med.* Assoc. 298, 1763–1771

14. Jenny, B.F., Vandijk, H.J., Collins, J.A. (1991). Performance and fecal flora of calves fed a Bacillus subtilis concentrate. *J Dairy Sci.* 6,1968-1973.

15. Klevens, R.M. et al. (2007). Invasive Methicillin-Resistant Staphylococcus aureus Infections in the United States, *J. Am. Med. Assoc*. 298, 1763–1771.

16. Li, L., Xu, C.L., Ji, C., Ma, Q., Hao, K., Jin, Z.Y., Li, K. (2006). Effects of a dried Bacillus subtilis culture on egg quality. *Poult Sci*. 2, 364-368.

17.Li, B., Walsh, C.T. (2010). Identification of the gene cluster for the dithiolopyrrolone holomycin in Streptomyces clavuligerus. *Proc Natl Acad Sci* U.S.A. 48, 19731-19735

18. Liu, C.H., Chiu, C.S., Ho, P.L., Wang, S.W. (2009). Improvement in the growth performance of white shrimp, Litopenaeus vannamei, by a protease-producing probiotic, Bacillus subtilis E20, from natto. *J Appl Microbiol*. 3, 1031-41.

19. Liu, H., Sadygov, R.G., Yates 3rd, J.R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem*. 76, 4193-4201.

20. Liu, W. et al. (2010). Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of Bacillus subtilis. *Proc. Natl. Acad. Sci*. U.S.A. 107, 16286-16290.

21. Otto, M. (2010). Basis of Virulence in Community-Associated Methicillin-Resistant Staphylococcus aureus. *Annu Rev Microbiol* 64, 143-162.

22. Park, S.K. et al. (2007). The effect of probiotics on Helicobacter pylori eradication. *Hepatogastroenterology* 79, 2032-2036.

23. Queck, S.Y., Khan, B.A., Wang, R., Bach, T.H., Kretschmer, D., Chen, L., Kreiswirth, B.N., Peschel, A., Deleo, F.R., Otto, M. (2009). Mobile Genetic Element-Encoded Cytolysin Connects Virulence to Methicillin Resistance in MRSA. *PLoS Pathog*. 7, e1000533

24. Roberson, J.R., Fox, L.K., Hancock, D.D., Gay, J.M., Besser, T.E. (1994). Ecology of Staphylococcus aureus Isolated from Various Sites on Dairy Farms. *J. Dairy Sci* 77, 3354-3364.

25. Stein, T. (2005). Bacillus subtilis antibiotics: structures, syntheses and specific functions. *Mol Microbiol*. 56, 845-857.

26. Schwamborn, K., Caprioli, R.M. (2010). Molecular imaging by mass spectrometry — looking beyond classical histology. *Nat Reviews Cancer*. 10, 639-646.

27. Tanner, S., Shu, H., Frank, A., Wang, L.C., Zandi, E., Mumby, M., Pevzner, P.A., Bafna, V. (2005). InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem*. 14, 4626-39.

28. Tsubura, S. et al. (2009). The effect of Bacillus subtilis mouth rinsing in patients with periodontitis. *Eur J Clin Microbiol Infect Dis*. 11, 1353-1356.

29. Wang, R. et al. (2007). Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nat Med*. 12, 1510-1514.

30. Yang, Y.L., Xu, Y., Straight, P., Dorrstein, P.C. (2009). Translating metabolic exchange with imaging mass spectrometry. *Nat Chem Biol*. 12, 885-887

Chapter 6

Future Directions

Summary

The Future Directions chapter of this thesis is presented as experimental proposals. Proposals (A) and (B) focus on the further characterization of the Streptolysin S family of toxins by mass spectrometry and biochemical methods for protein functional characterization. Proposals (C) and (D) extend the experimental capabilities of imaging mass spectrometry of microbial species, with a particular focus on the field of infectious disease. Preliminary data associated with the proposals are provided as well as experimental observations, challenges and potential next steps.

A. Hypothesis Driven Identification of Heterocyclized Peptides from *in vitro* Reconstituted Streptolysin S by Spectral Networks Analysis

Introduction

Streptolysin S (SLS) is a ribosomally encoded, post-translationally modified virulent factor that endows Group A Streptococcus (GAS) with the ability to cause β-hemolysis (1). Genetic studies performed in *S. Pyogenes* determined a nine gene locus, termed SagA-I, was sufficient and required for SLS activity (2, 3). The contribution of SLS to the pathogenesis of streptococci was assayed in murine model of necrotizing fasciitis. In this model the wild-type *S. pyogenes* had the ability to elicit ulcers with bacterial proliferation, neutrophilic inflammation, histopathologic evidence of vascular injury, and tissue necrosis. In contrast, isogenic SLS-negative Sag gene mutants were hindered in their virulence capability, as they did not display similar phenotypes to the wild-type strain. Comparative bioinformatics demonstrates the Sag locus contains many features reminiscent of a bacteriocin-type biosynthetic operon. Of these genes, SagF is a protein of unknown function. SagG, H, and I are ABC transporters and therefore are likely to be responsible for exporting the hemolytic molecule. SagE was annotated as an immunity protein but has been found to be similar to CaaX proteases. SagB, C and D are similar to the McbB, C, D proteins found on the Microcin B17 system, a peptide antibiotic from *E. coli*. McbB, C and D are responsible for the cyclization of serines and cysteines on the B17 prepropetide (McbA) to thiazoles and oxazoles. In fact, our earlier studies demonstrated the SagB-D enzymes are highly promiscuous, as they readily accepted and installed heterocycles on the McbA

substrate of the Microcin B17 system (4). In the SLS-gene locus, the substrate that is modified is SagA, composed of 53 amino acids and is designated a prepropeptide, as it is hypothesized to undergo extensive post-translational modifications upon maturation (5).

In the 100 years of the known existence of SLS, no structural information has been obtained on the full length maturated toxin (6). To date, no experimental evidence has been presented verifying the purification of SLS from *S. Pyogenes* extracts; therefore its structural characterization has remained elusive. Hurdles in isolating SLS stem from its unstable nature once secreted. Early isolation attempts determined an increase in SLS activity from bacterial extracts upon supplementing the media with stabilizing carrier molecules such as RNA core or BSA (7). Although supplementing the media resulted in an improvement of SLS activity in the biological assays, this method adds a level of complexity into the system and thus difficulty in obtaining structural information by standard analytical methods (e.g. mass spectrometry, NMR).

Recently, we have been able to reconstitute SLS activity *in vitro* by the use of recombinately expressed SagA-D proteins. Using the recombinantly expressed proteins, it was shown that the combination of all four proteins produced an agent with the ability to lyse erythrocytes and in addition cause cytolytic effects on HEK cells (5). Therefore, attempts were made to experimentally characterize the structural changes occurring to the substrate SagA *in vitro* once incubated with the SagB-D modifying enzymes. Several constructs containing an N-terminal fusion tag were generated on the SagA substrate in attempt to analyze the structural changes upon modification by

top-down mass spectrometry (8). Because of experimental difficulties, thus far, no construct has shown to be conducive to the top-down mass spectral approach. These experimental difficulties led the mass spectrometry efforts to pursue an alternative approach, a method incorporating off-line chromatographic separation, chemical derivatization, and classical bottom-up proteomics. Confirmed hemolytic samples of the reaction containing SagA + SagBCD were subjected to the developed bottom-up mass spectrometry approach. The generated LCMS/MS data was analyzed with the tandem mass spectrometry processing program InSpecT, which searches against specified parameters and databases (9). Using this program, a peptide containing two post-translationally derived oxazoles at positions S46 and S48 of the modified SagA peptide was consistently indentified from successive LCMS/MS experiments (10). Although, post-translationally modified SagA peptides were captured, the substrate has several other serine, threonine, and cysteine that have been predicted to be modified by SagBCD (5). Therefore, given a limited number of post-translationally modified peptides were identified from the vast number of spectra collected, we surmised that a more elaborate method for tandem mass spectrometry identification of heterocyclized peptides would need to be established.

Spectral Networking is a tandem mass spectrometry processing program that identifies sequence tags of a given peptide by taking advantage of spectral pairs of similar overlapping peptides (11). Using Spectral Networks, we developed a hypothesis-driven approach for the identification of heterocyclized peptides on the SagA substrate. The developed method takes advantage of the Spectral Network premise—spectral pairing. The program works by first identifying spectral pairs, the

program will then propagate towards other similar modified or unmodified peptides as a means for identification of other spectral partners. The rationale behind the hypothesis-driven method proposed is that if Spectral Networks is given "half" of a spectral pair, originating from a peptide known to contain oxazoles or thiazoles, then Spectral Networks will propagate from this peptide and match spectral pairs based on the known modified peptide. Additionally, we provided Spectral Networks with LCMS/MS data sets originating from SagA alone, truncated SagA alone (21mer), SagA 21mer + BCD, and SagA full length + BCD. Selected reconstitution reactions were quenched at staggered time points with the aim of capturing intermediate peptides in the process of being post-translationally modified to contain oxazoles or thiazoles. In theory, this method should allow for the discovery of previously unidentified heterocyclized peptides (Fig 6.1).

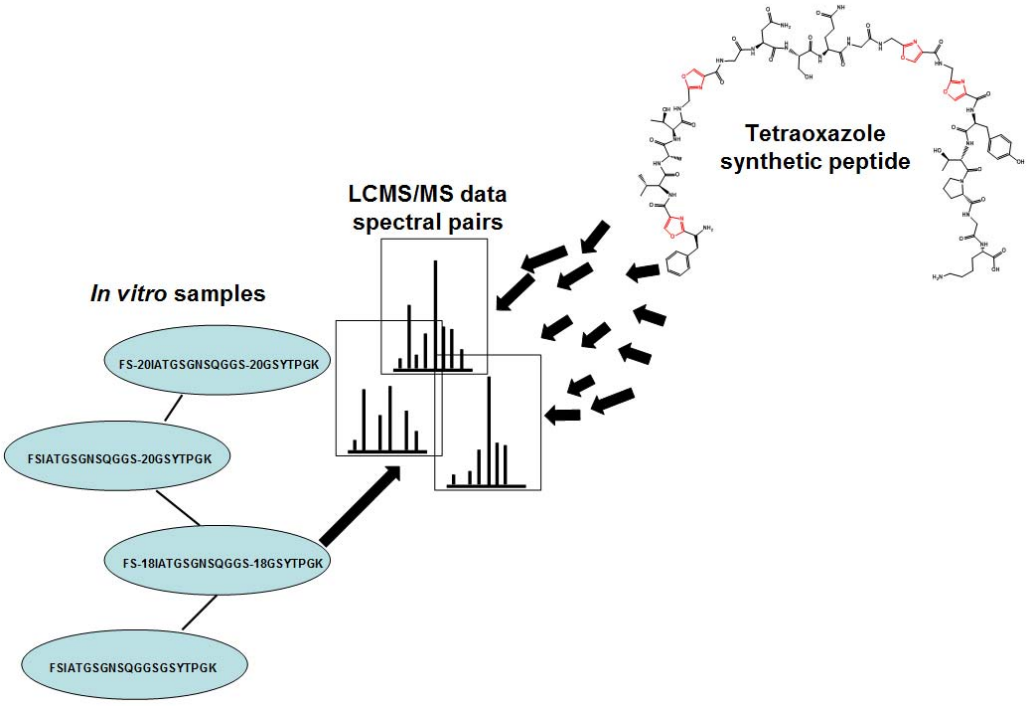

Figure 6.1 Schematic for the overall process of matching spectral pairs. A synthetic peptide containing four oxazoles will be fragmented by LCMS/MS and the data set will be used to propagate towards data sets obtained from *in vitro* reconstituted reactions prepared in varying manners. The goal of the experiment is to match similar overlapping peptides as a means to identify post-translationally modified peptides from SLS *in vitro* reconstituted reactions.

Method

A truncated SagA peptide (F33-K53) was synthesized to contain four oxazole heterocycles (tetraoxazole-SagA) at positions: S34, S39, S46, S48. Once synthesized, the tetraoxazole-SagA peptide was analyzed by FTICR-MS in order to obtain a single molecular formula. Thereafter, the peptide was fragmented using a linear ion trap mass spectrometer establishing amino acid connectivity, molecular weight, and the incorporation of the four oxazoles. The FTICR-MS showed the tetraoxazole-SagA had a mass accuracy of 1.1 ppm. Tandem mass spectrometry confirmed the location of the four oxazoles. Over one hundred forty seven *in vitro* reconstituted reactions of the SagA+BCD system were processed in different manners in or to obtain intermediate or fully post-translationally modified peptides. LCMS/MS of the varying samples was performed as previously described. To test our hypothesis, the data acquired from the fragmented tetraoxazole-SagA and the varying SagA + BCD samples were introduced as one data set and analyzed by spectral Networks analysis.

-

Preliminary Results

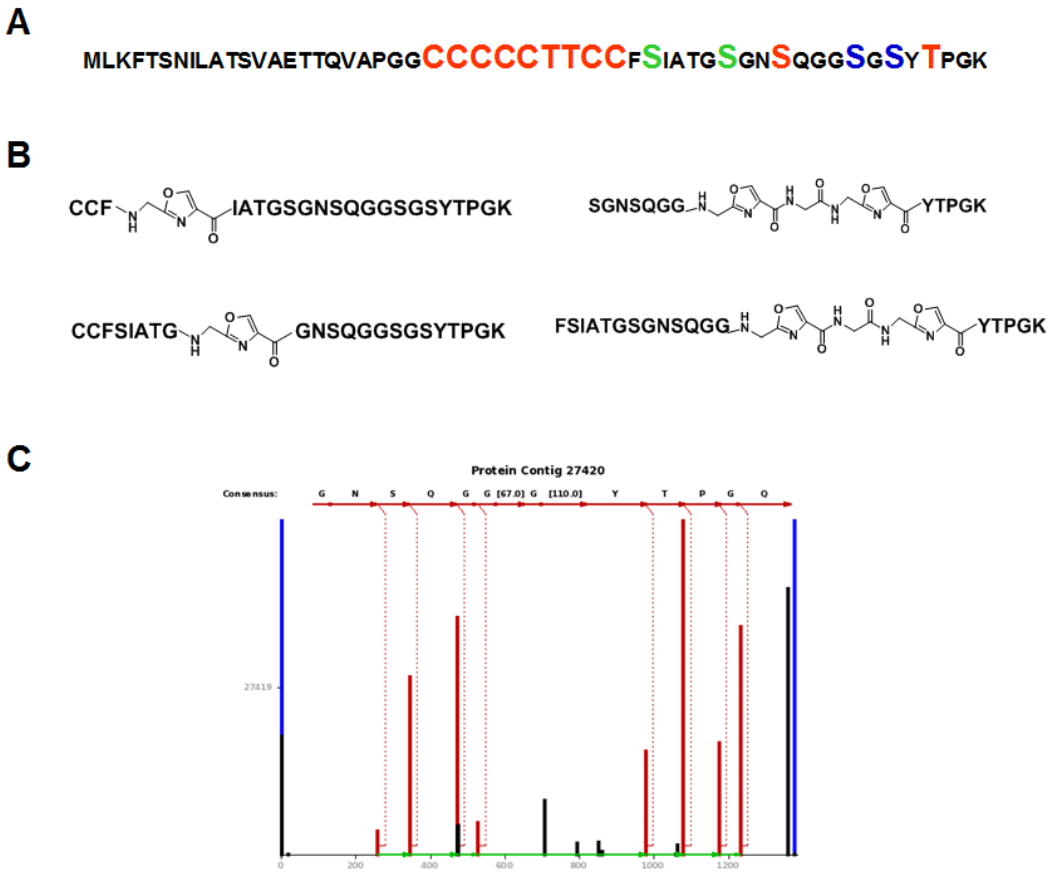

Figure 6.2 A. The primary sequence of the SagA prepropeptide. The C-terminal Ser, Thr, and Cys are predicted to undergo heterocyclization to form thiazole or methyl oxazole moieties. Red bold lettering depicts the amino acids that have not been identified as containing modifications by Spectral Networks. The green bold lettering depicts novel post-translationally modified peptides recovered by the proposed methodology. Blue bold lettering depict post-translationally modified peptides uncovered by Spectral Networks but have previously been published (12). **B.** Structures of oxazole containing peptides uncovered by the hypothesis driven identification of heterocyclized peptides proposed herein. **C.** Representative data output by Spectral Networks showing a Serine modification of minus 20 Da within (GGS-20) the SagA peptide.

Observations

The initial hypothesis of using a peptide known to contain the desired post-translational modifications as a point of propagation failed. Data sets of the varying prepared SagA peptides did not cluster/pair with the tetraoxazole peptide. One fundamental flaw with the experiment was the assumption that the SLS peptide is converted to contain oxazoles at the position synthesized. Chemical modifications synthetically installed into the SagA substrate (tetraoxazole) were based on point mutation studies that altered an active phenotype *in vitro*. The fact that a point mutation disrupts activity does not necessarily mean the residue contains a post-translational modification. It is possible that the single amino acid modified plays a structural role in the maturated toxin in regards to proper protein folding. These initial assumptions could be the reason why no peptides clustered with the tetraoxazole LCMS/MS data set.

Another unknown in the experiment is the manner in which the tetraoxazole peptides fragment once in the mass spectrometer. The tetraoxazole fragmentation patterns and potential rearrangements in the mass spectrometer could account for unique mass spectrum fingerprints. In result, Spectral Networks would not have means to match spectral partners as was observed. Future experiments could take a more systematic approach, as a peptide with one oxazole could be used as a template to propagate and search other similarity post-translationally modified peptides.

Peptides from the unmodified SLS 21mer peptide clustered with modified SagA peptides originating from biologically confirmed hemolytic assays or reactions that were quenched at different time points. The fact that an unmodified peptide, SLS

21mer, clustered with full length SagA modified substrate attest to the likelihood of success for this proposal. At the end two previously unreported post-translational modified peptides, S33 and S38 of the SagA peptide were identified. The two modifications at S33 and S38 have been shown by genetic studies to contribute to SLS activity (5). Further studies need to be performed to assess the reproducibility of these findings.

B. Functional Characterization of ClosE, a Putative Protease in the CLS Gene Cluster

Introduction

The genus Clostridium consists of relatively large, gram-positive, rod-shaped bacteria in the phylum firmicutes. The species *Clostridium Botulinum* are ancient organisms that dwell in virtually all anaerobic habitats of nature where organic nutrients are available, including soils, aquatic sediments and the intestinal tracts of animals (13). *C. Botlinum* is infamous for producing the most potent neurotoxins known to man, the botulinum neurotoxins (BoNTs) (14), which have been used as biological warfare agents (15). As pathogens of food-borne botulism, they owe their virulence almost entirely to elaboration of their differential toxins (16). Recently a novel, the post-translationally modified β-hemolytic/cytolytic toxin was discovered in *C. Botulinum*, indentified as clostridiolysin S (CLS). The CLS toxin is of particular interest given it resembles the key virulence determinant Streptolysin S (SLS) in Group A Streptococcus (GAS) in gene and structure organization. Murine models of infection showed SLS deficient GAS strains loss efficiency to necrotize soft tissue and thus hinder proliferation of the bacteria upon a host.

In theory, as was determined in GAS, disruption of the CLS gene locus could help combat *C. Botulinum* from becoming systemic, at which point stopping the release of its dangerous, highly potent BoNTs. We hypothesize that a molecular understanding of the CLS gene cluster can lead to alternative therapeutic strategies targeting compounds that inhibit any functional gene that contributes to the maturation

of CLS. Therefore we decided to further investigate the CLS locus found in *C. botulinum*.

To gain insight into the mechanism of CLS biosynthesis, we initiated a bioinformatics-based search of bacteriocin gene clusters that resembled the Clos locus. There were striking similarities between the Clos operon and the Microcin B17 bacteriocin operon found in *Escherichia coli*. Microcin B17 (MccB17) is a 3.1 kDa peptidyl antibiotic produced by strains of *E. coli* that harbor a plasmid containing the microcin B17 operon (17). This operon contains seven genes, designated *mcbA-G,* that are solely responsible for toxin production. Mature microcin B17 is derived from a 69 amino acid pre-propeptide encoded by the *mcbA* gene. This peptide is post-translationally modified by the gene products of *mcbBCD*, which convert 14 of the 69 residues into four oxazoles and four thizaoles to generate promicrocin B17 (18). The N-terminal 26 residues are then proteolytically removed by an unknown mechanism to yield mature microcin B17. The toxin is finally exported from the cell by an ATP-binding cassette (ABC)-type transporter encoded by the products within the gene locus (19).

In vitro studies have shed insight into the functional roles of ClosA, a peptide substrate that is post-translationally modified, and ClosB-D characterized as modifying enzymes which act upon ClosA. Of particular interest within this proposal are the genes of unknown function, *closE* and *closF* given there uncharacterized role in the maturation of CLS. BLAST analysis shows ClosE is similar to CaaX-domain proteases, while ClosF has no homology to any known protein. The likelihood that ClosE is a protease is reinforced by the fact that several clusters within the widely

distributed family of toxins contain a gene encoding a CaaX-type protease similar in sequence to ClosE (5). It is interesting to speculate that ClosE is the protease responsible for N-terminal cleavage of the CLS toxin precursor. This proposal will attempt to functionally characterize ClosE, guided by the hypothesis that the protein is a protease. Two approaches will be taken (Figure 6.3): (i) Co-lysis of an affinity tagged MBP-ClosA and non-affinity tagged ClosE, followed by protein purification of MBP-ClosA and assessment of proteolysis by an SDS-PAGE gel shift relative to the control (BL21 *E. coli* transformed with shuttle alone). (ii) Co-transformation of MBP-closA and closE into BL21 *E. coli*, followed by protein purification of MBP-ClosA and assessment of proteolysis by an SDS-PAGE gel shift relative to the control (BL21 *E. coli* transformed with shuttle alone). Based on BLAST analysis we hypothesize ClosE will act upon ClosA by cleaving the N-terminal leader peptide. If protease activity is observed on the ClosA peptide assessed by SDS-PAGE gel, the cleavage sight will be annotated by mass spectral means.

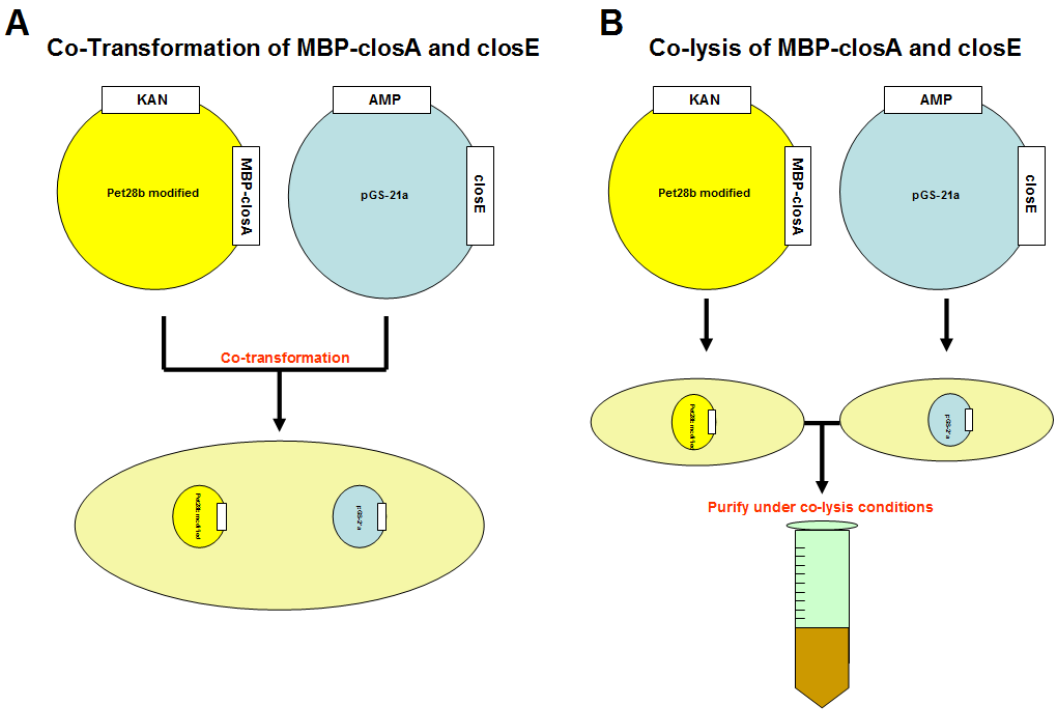

Figure 6.3 A. Schematic of overall strategy for the co-transformation of *closE* and MBP-*closA*. **B.** Schematic for the method in which ClosE and MBP-ClosA will be co-lysed. The goal of this experimental scheme is to determine if ClosE acts upon ClosA in a proteolyic manner. Proteolysis will be assessed by SDS-PAGE gel and mass spectrometry.

Method

Approaches to assess the functionality of ClosE. (i) Co-lysis of MBP-ClosA and ClosE: The *closA* gene was subcloned into a MBP modified pet28b vector using the restriction enzymes BamHI/NotI. A Puc vector harvesting *closE* was subcloned into pGS-21a using NdeI/BamHI NdeI digestion of the vector pGS-21a. Colony PCR and sequencing of pGS-21a-closE and MBP modified pet28b-closA were positive. Each verified plasmid DNA were transformed separately into BL21 *E. coli* then grown, induced, and pellets separately purified under co-lysis conditions. Briefly, the pellets were individually resuspended in lysis buffer then incubated together 1.5 hrs after adding lysozyme and sonicating. Affinity purification was performed to recover the yielded MBP-ClosA protein. Qualitative assessment of proteolysis was performed via SDS-page gel. The data obtained from the SDS-page gel was further analyzed for assessment of proteolysis by in-gel digestion and LCMS/MS analysis. (ii) *Co-Transformation of MBP-closA and closE into BL21 E. coli*: Transformation of MBP-closA and closE was performed into BL21 *E. coli*. Transformations into BL21 were performed using different ratios of closA to closE and plated on AMP100/Kan50 plates. Fifteen colonies from each plate were picked and assessed by a small induction test. Thereafter, affinity purification was performed in attempt to recover the MBP-ClosA protein and assessment of proteolytic cleavage was performed via SDS-PAGE gel and LCMS/MS analysis.

Preliminary results

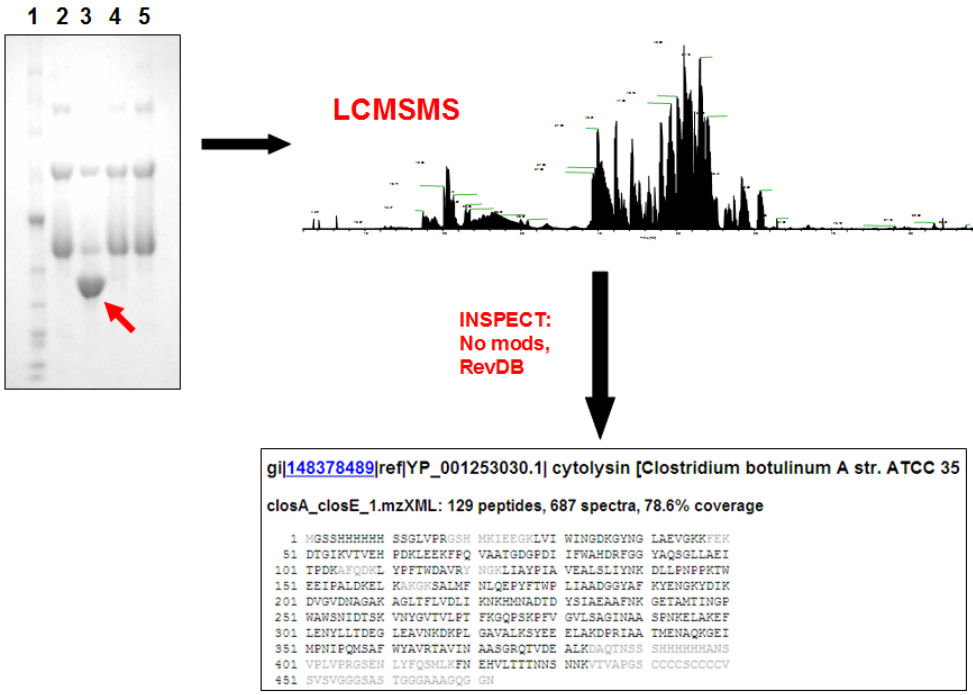

Figure 6.4 SDS-PAGE analysis from the co-transfomation is shown. Lane 1 SDS-PAGE protein ladder standard; Lane 2 MBP-ClosA alone; Lane 3 MBP-ClosA truncated purified from the co-transformation experiment; Lane 4 MBP-ClosA isolated from the co-transformation with the pGS-21a vector alone; Lane 5 MBP-ClosA alone purified from previous protein purification batches. Following the flow chart, the major band in Lane 3 was excised (indicated by red arrow) and processed for LCMS/MS analysis. The resulting LCMS/Ms data was then processed by INSPECT analysis in order to annotate peptide coverage.

Observations

Proteolysis of ClosA was observed by SDS-PAGE when co-lysed and co-transformed with ClosE (co-lysis data not shown), which strengthens the hypothesis of ClosE functioning as a protease (Figure 6.4). Because proteolysis of ClosA was observed in both experimental strategies, the annotation of the cut sight by LCMS/MS was pursued. An in-gel digestion of the SDS-PAGE gel protein band was examined, which led to ambiguous data given parcel annotation of the truncated MBP-ClosA fusion protein was indentified. Additionally, only a parcel annotation of the MBP fusion tag was achieved.

In hindsight, top-down mass spectrometry would have been the better experiment to perform on the truncated MBP-ClosA protein since LCMS/MS could have resulted in peptide losses resulting from chemical properties gained or lost hindering ionization (e.g. non-ionizable peptide). Using a top-down mass spectrometry approach could elevate these issues, as an intact mass of the resulting truncated MBP-ClosA protein would be obtained at high mass accuracy using the FT-ICRMS. The obtained mass of the truncated MBP-ClosA would then be compared to the control MBP-ClosA protein that showed an accurate mass on SDS-PAGE gel analysis.

Another experimental difficulty to consider is ClosE could potentially recognize a sequence in the MBP fusion tag and in result is cleaved as seen by the absence of peptides in the LCMS/MS analysis (Figure 6.4). To overcome this potential hurdle another type of fusion tag can be used (e.g. His tag or GST) to affinity-purify ClosA once cleaved. Further experiments need to be performed to assess the reproducibility of the experimental findings.

C. An interaction between skin microbiome organisms: *Staphylococcus epidermidis* and *Staphylococcus aureus*

Introduction

The skin microbiome represents a mega-community of microorganisms that interact with one another affecting human health. The flora on the skin is dynamic and little is known about the initiation of these changes (20). Two model organisms reported to be found on the human skin are *Staphylococcus epidermidis* and *Staphylococcus aureus* (21, 22). Both organisms have been found in human microbiome cataloging projects (23, 24). *S. epidermidis* is an opportunistic pathogen that is found in the digestive tract, epithelial wounds, extremities of the human body and surgical instrumentation (25, 26). Because *S. epidermidis* is commonly found on the skin of healthy humans, the bacteria is reported to be commensal resident (27). Recent reports have shown *S. epidermidis* has developed a mutualistic relationship with humans as it has adaptive strategies to cooperate with components of the innate immune system via the production of genetically encoded molecules that enhance evading pathogen killing (28). Similarly, *S. aureus* is found on human skin, digestive tracts, nares, livestock and surgical instrumentation; making up a minor component of soil (29, 30). At minimum, 30% of the world population is colonized with *S. aureus* although some studies suggest that a larger percentage is a carrier of the bacteria. *S. aureus* is the world leading disease causing bacterial pathogen that has developed significant resistance against a variety of antibiotics (31). Although *S. aureus*

colonizes a large segment of the world inhabitants, it is only infectious to a subset of this population.

S. epidermidis has been shown to enhance pathogen killing by innate immunity cooperation and *S. aureus* only colonizes a portion of the human population. Therefore, we hypothesize that *S. epidermidis,* once in its niche, is a major player in governing *S. aureus* colonization on a host. As a means to retain an established niche upon a host, we predict *S. epidermidis* can respond to *S. aureus* by secreting an array of molecules to control developmental phenotypes and thereby alter its ability to proliferate. Gaining insight into the molecular means by which *S. epidermidis* interacts with other skin dwelling organisms, can shed insight into the how a homeostatic environment is maintained among the thousands of microorganism comprising the skin microbiota. Additionally if the compounds show clinically relevant bioactivity against *S. aureus*, they can be used as therapeutics.

To explore our hypothesis we decided to use the emerging tool of imaging mass spectrometry (IMS) to monitor the co-culturing experiment of *S .epidermidis* and *S. aureus* when grown on agar media. IMS allows for the visualization of the molecules involved in chemical communication between the two bacteria, which will give insight on how *S. epidermidis* controls *S. aureus* growth if the proposed hypothesis is validated. Additionally, fluorescence microscopy will be used to monitor any morphological changes at the single cell level caused by the co-culturing experiment.

The experimental set up will entail the co-culturing of a *S. epidermidis* colony adjacent to a *S. aureus* colony. As a control only one particular *S. epidermidis* isolate

will be used in the study. Variability will be introduced by the use of twenty-one differentially sequenced *S. aureus* isolates, which will be individually co-cultured with *S. epidermidis*. Because a niche suggests the cell density of the establishment bacteria is plentiful, experiments will be performed were the density of *S. epidermidis* is three times the amount of *S. aureus* and vice versa. As a pilot experiment, initially, the system will be examined when the cell density are equal.

The competitive outcome between two bacteria is most likely determined by the type and amounts of molecular species a given participant is able to produce. Therefore, before initiating the co-culturing experiments we set out to determine a formula to categorize the twenty one *S. aureus* isolates based on their metabolic output. To this end, intact MALDI-TOF was used to monitor the number of molecular species each of the twenty one *S. aureus* isolate produced within a 400-4000 Da range. Moreover, after categorizing the isolates the hemolytic behavior of each isolate was also examined. The data showed a correlation between the number of molecular species and hemolytic robustness (Figure 6.5). Using metabolic output as a factor in conjunction with hemolysis, the twenty one *S. aureus* isolates were categorized from high to low based on the number of molecules made within the monitored mass spectral window. From the varying categorized *S. aureus* isolates, three isolates were picked to interact with *S. epidermidis* at a high, medium and low molecule producing range within the established categorization.

Methods

Imaging mass spectrometry. The ISP2 media was prepared with yeast extract 2 g, malt extract 4 g, dextrose 2 g, and agar 10 g added to 500 mL of deionized water and autoclaved. 10 mL of autoclaved agar was poured into sterile Petri dishes under flame. Colony growth was initiated by streaking 3uL of overnight *S. aureus and S. epidermidis* at equal cell densities on the prepared thin layer agar ISP2 plates. The bacterial colonies were allowed to grow for 48 hrs at 30°C before transferring the co-culturing experiment to a Bruker MSP 96 MALDI anchor plate. After 48 hrs the co-culturing experiment grown on the ISP2 media was excised and transferred to a MALDI anchor plate. Thereafter, the target plate was uniformly covered with Sigma universal matrix (α-cyano-4-hydroxycinnamic acid and 2, 5-dihydroxybenzoic acid) by the use of a 50 μm sieve. Once that sample was completely covered with matrix, it was baked in a 37 °C oven for 3-4 hours until it was deemed dried at which point it was subjected to imaging mass spectrometry.

Fluorescence microscopy. *S. epidermidis* cells were harvested from a colony that was either adjacent to that of *S. aureus* or grown in isolation. Cells were resuspended in T-base, stained with a final concentration of 30 μg/ml FM 4-64, which stains the membranes red, and 2.5 uM Sytox Green (Invitrogen), which brightly stains only cells with membrane damage. The cells were immobilized with poly-L-Lysine and visualized with an Applied Precision Spectris microscope.

Preliminary Results

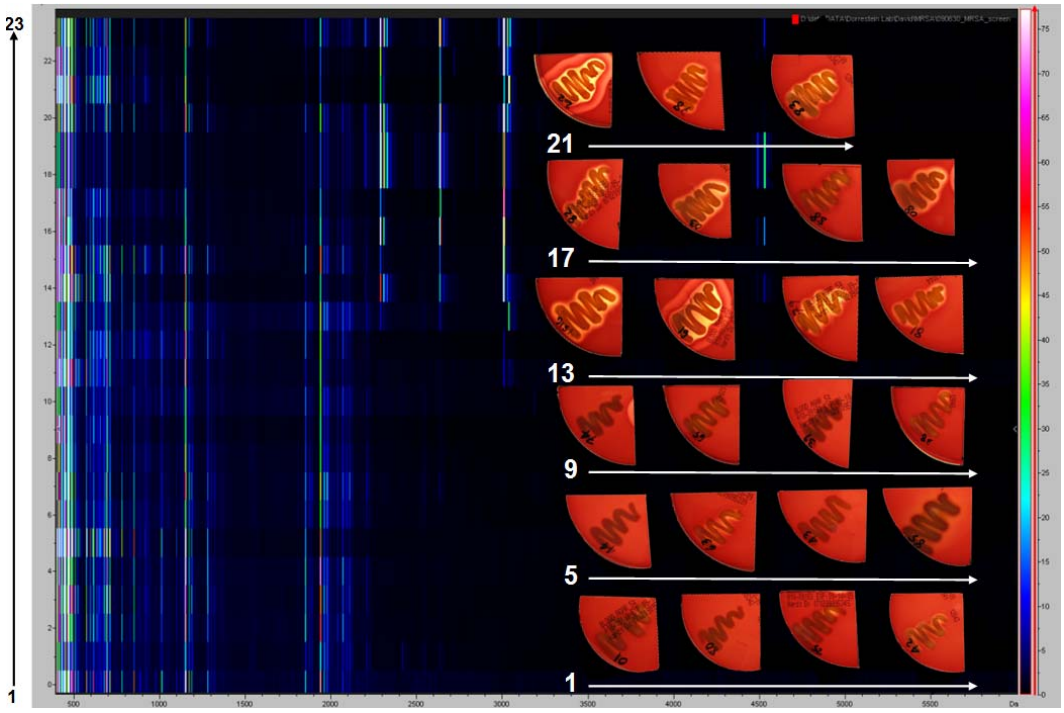

Figure 6.5 Gel stack view of the molecular ions produced by twenty three different S. aureus isolates. The gel stack view is a graph that displays a series of mass spectrometry spectra simultaneously. Intact cell MALDI-TOF analysis was obtained for the examined isolates and was organized in increasing number of molecular ions observed in the > 2000 Da range. A correlation was observed between the number of detectable ions and the amount of β-hemolysis observed on blood agar plates.

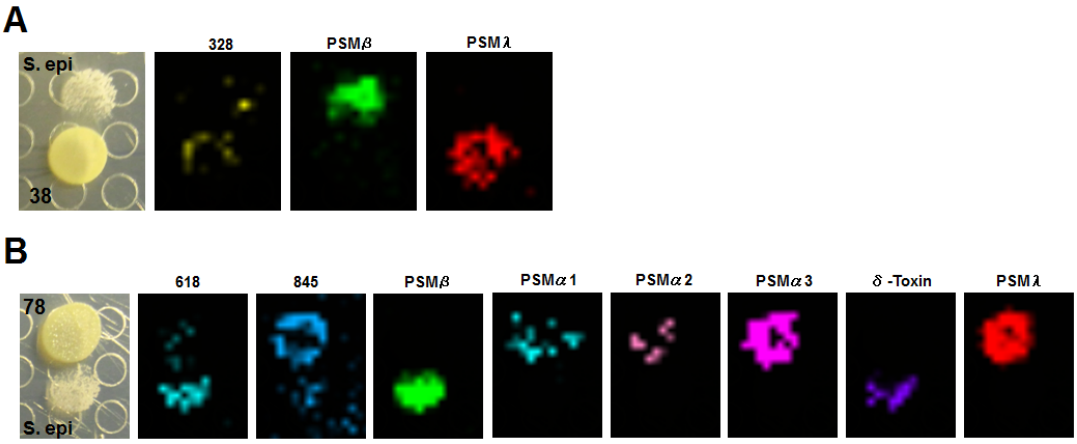

Figure 6.6 Interaction between *S. aureus* and *S. epidermidis* grown on ISP-s media and examined by IMS. **A.** Isolate MRSA-38 shows the ability to alter the growth and morphology of *S. epidermidis.* IMS of the interaction shows a limited number of ions present in the mass spectral window tested. **B.** Isolate MRSA-78 showed the ability to alter the growth and morphology of *S. epidermidis* when co-cultured. IMS of the interaction showed several molecular species present.

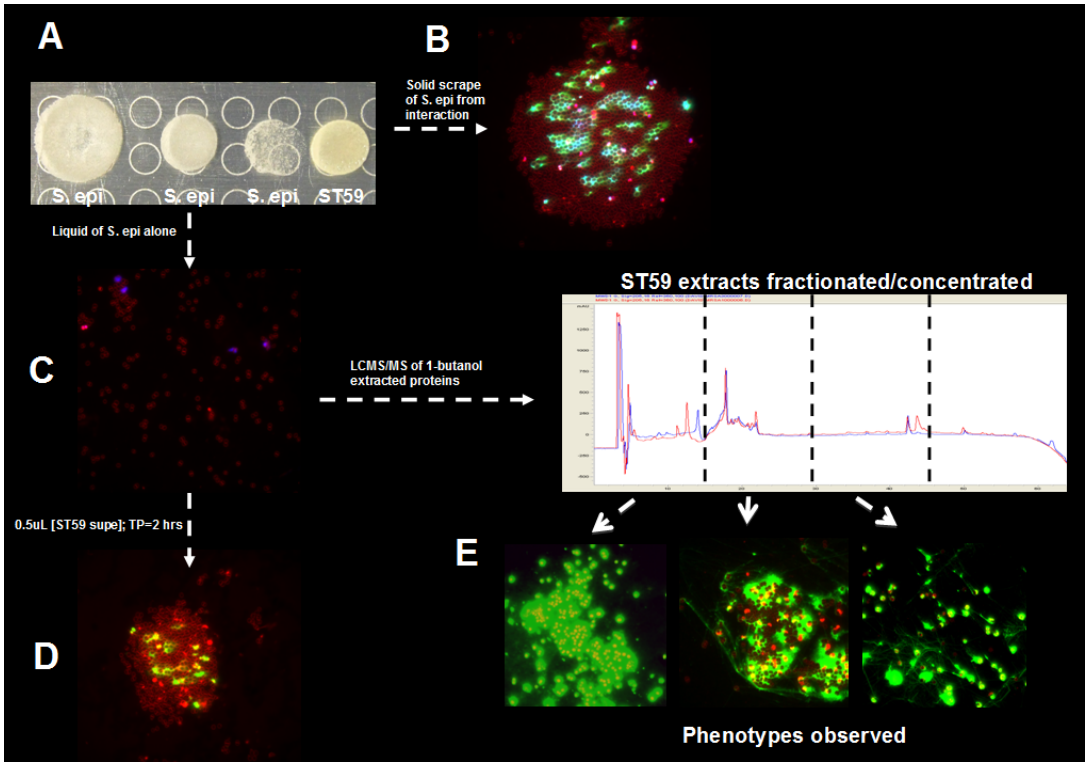

Figure 6.7 Fluorescence microscopy of *S. epidermidis* cells obtained from a co-culturing experiment with *S. aureus*. **A.** Co-cultures experiments with the bacteria were set up on ISP-2 media. **B.** Solid scraping of *S. epidermidis* obtained from the bacterial interface was examined by fluorescence microscopy **C.** Fluorescence microscopy obtained from solid scraping of *S. epidermidis* grown separate from the S. aureus. **D.** Fluorescence microscopy of *S. epidermidis* spiked with concentrated *S. aureus* supernatant. **E.** Differential phenotypes observed by fluorescence microscopy when *S. aureus* supernatants were fractionated by HPLC and incubated with *S. epidermidis*.

Observations

No obvious morphological differences were observed between the interactions using the differentially categorized *S. aureus* strains and *S. epidermidis* when examined by IMS and fluorescence microscopy in the pilot experiments. Additionally, the *S. epidermidis* strain tested, at an equal cell density, did not have the ability to clear nor alter the phenotype of *S. aureus*. In contrast, all three *S. aureus* isolates tested eradicated *S. epidermidis* at the bacterial interface and also seemed to alter its growth morphology (Figure 6.6). Based on the altered morphology and clearing observed on *S. epidermidis*, *S. aureus* has a competitive advantage in niche establishment at the parameters tested. Besides clearing, the interaction showed the morphology of *S. epidermidis* changed when grown adjacent to *S. aureus* as the cells seemed to "cluster" into microcolonies. The observation of microcolonies was very interesting and therefore we decided to further pursue the morphological phenomenon.

To reconstitute the formation of clustered microcolonies, supernatants obtained from *S. aureus* overnight growths were harvested and incubated with *S. epidermidis* cultures. Fluorescence microscopy showed secreted factors (supernatant) from *S. aureus* altered the phenotype of *S. epidermidis* (Figure 6.7), consistent with assessments made from observing the bacterial colonies when co-cultured. Preliminary studies performed on HPLC fractions of *S. aureus* supernatant also showed minimal clustering, but not to the extent of the tested supernantant or of the original co-culturing experiment (Figure 6.7). Further studies need to be performed to identify if one molecular factor from *S. aureus* is responsible for the phenotype alteration or if several molecular factors work in concert. Natural product isolation

strategies of *S. aureus* supernatants could be pursued to purify compounds identified by IMS and test if the compounds are active against S. *epidermidis*.

The rationale behind the alteration of the *S. epidermidis* colony is unclear. It could be hypothesized that *S. epidermidis* use the "clustering" event as a defensive mechanism or as a pathway to death. Co-culturing studies have shown *P. auregonisa* alters the colony morphology of *S. aureus* leading to small colony variants (SCVs), isolates which gain resistant to antibiotics and are associated with Cystic Fibrosis (32). Because an understanding of the molecular interplay between microbiome residents can lead to health related issues and insight, further studies need to be performed on the *S. epidermidis* colony variants in order to determine the biological response observed.

Because *S. aureus* makes an array of differential secreted molecular species beyond the mass spectral window monitored (400-4000 Da), it is unclear if the altered phenotype of the *S. epidermidis* colony is a result of the molecular factors corresponding to the ions monitored with IMS. Finally, further studies need to be performed using differential cell densities as proposed in order to fully test the hypothesis that *S. epidermidis* when established in a niche can affect *S. aureus* colonization upon a host.

Acknowledgements

Performed experiments: Pablo Reyna, Nuno Bandeira, Mary Hensler, David Gonzalez. Designed experiments: David Gonzalez, Pieter Dorrestein, Mary Hensler, Victor Nizet, Tinya Fleming, Kit Pogliano. Written by David Gonzalez.

Reference

1. Datta V, et al. (2005) Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Mol Microbiol 56:681–695.

2. Nizet V, et al. (2000) Genetic locus for streptolysin S production by group A streptococcus. Infect Immun 68:4245–4254.

3. Humar D, et al. (2002) Streptolysin S and necrotising infections produced by group G streptococcus. Lancet 359:124–129.

4. Nizet V, (2002) Streptococcal beta-hemolysins: Genetics and role in disease pathogenesis. Trends Microbiol 10:575–580.

5. Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V., and Dixon, J. E. (2008) *Proc. Natl. Acad. Sci. U.S.A.* 105, 5879–5884.

6. Yorgey P, Davagnino J, Kolter R (1993) The maturation pathway of microcin B17, a peptide inhibitor of DNA gyrase. Mol Microbiol 9:897–905.

7. Taketo A, Taketo Y. Effect of streptococcal extracellular nuclease on the carrier activity of RNA for streptolysin S. Z Naturforsch C. 1983 Jan-Feb;38(1-2):107-11.

8. Reid GE, McLuckey SA.'Top down' protein characterization via tandem mass spectrometry. J Mass Spectrom. 2002 Jul;37(7):663-75. Review.

9. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal Chem. 2005 Jul 15;77(14):4626-39.

10. Gonzalez DJ, Lee SW, Hensler ME, Markley AL, Dahesh S, Mitchell DA, Bandeira N, Nizet V, Dixon JE, Dorrestein PC. Clostridiolysin S, a post-translationally modified biotoxin from Clostridium botulinum. J Biol Chem. 2010 Sep 3;285(36):28220-8. Epub 2010 Jun 25.

11. Bandeira N. Protein identification by spectral networks analysis. Methods Mol Biol. 2011;694:151-68.

12. Mitchell DA, Lee SW, Pence MA, Markley AL, Limm JD, Nizet V, Dixon JE. Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. J Biol Chem. 2009 May 8;284(19):13004-12. Epub 2009 Mar 13.

13. Peck MW, Stringer SC, Carter AT. Clostridium botulinum in the post-genomic era. Food Microbiol. 2011 Apr;28(2):183-91. Epub 2010 Mar 17.

14. Horowitz BZ. Type E botulism. Clin Toxicol (Phila). 2010 Nov;48(9):880-95. Review.

15. Li B, Peet NP, Butler MM, Burnett JC, Moir DT, Bowlin TL.Small molecule inhibitors as countermeasures for botulinum neurotoxin intoxication.Molecules. 2010 Dec 30;16(1):202-20.

16. Garcia-Rodriguez C, Geren IN, Lou J, Conrad F, Forsyth C, Wen W, Chakraborti S, Zao H, Manzanarez G, Smith TJ, Brown J, Tepp WH, Liu N, Wijesuriya S, Tomic MT, Johnson EA, Smith LA, Marks JD.Neutralizing human monoclonal antibodies binding multiple serotypes of botulinum neurotoxin. Protein Eng Des Sel. 2011 Mar;24(3):321-31. Epub 2010 Dec 13.

17. Davagnino J, et al. (1986) The DNA replication inhibitor microcin B17 is a forty-three-amino-acid protein containing sixty percent glycine. Proteins 1:230–238.

18. Yorgey P, et al. (1994) Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. Proc Natl Acad Sci USA 91:4519–4523.

19. Frank DN, Feazel LM, Bessesen MT, Price CS, Janoff EN, Pace NR. The human nasal microbiota and Staphylococcus aureus carriage. PLoS One. 2010 May 17;5(5):e10598.

20. Beloborodova NV, Khodakova AS, Bairamov IT, Olenin AY. Microbial origin of phenylcarboxylic acids in the human body. Biochemistry (Mosc). 2009 Dec;74(12):1350-5.

21. Whitfeld M, Gunasingam N, Leow LJ, Shirato K, Preda V. Staphylococcus epidermidis: a possible role in the pustules of rosacea. J Am Acad Dermatol. 2011 Jan;64(1):49-52. Epub 2010 Oct 12.

22. Peker E, Kirimi E, Tuncer O, Ceylan A, Cagan E, Dogan M. Necrotizing fasciitis caused by Staphylococcus epidermidis in a neonate with extremely low birthweight.J Dermatol. 2010 Jul;37(7):671-3.

23. Grice EA, Segre JA. The skin microbiome. Nat Rev Microbiol. 2011 Apr;9(4):244-53.

24. Karlsson FH, Nookaew I, Petranovic D, Nielsen J. Prospects for systems biology and modeling of the gut microbiome. Trends Biotechnol. 2011 Mar 8.

25. Pennisi E. Human genome 10th anniversary. Digging deep into the microbiome. Science. 2011 Feb 25;331(6020):1008-9.

26. Ibrahem S, Salmenlinna S, Virolainen A, Kerttula AM, Lyytikäinen O, Jägerroos H, Broas M, Vuopio-Varkila J. Carriage of methicillin-resistant Staphylococci and their SCCmec types in a long-term-care facility. J Clin Microbiol. 2009 Jan;47(1):32-7. Epub 2008 Oct 29.

27. Busscher JF, van Duijkeren E, Sloet van Oldruitenborgh-Oosterbaan MM. The prevalence of methicillin-resistant staphylococci in healthy horses in the Netherlands. Vet Microbiol. 2006 Mar 10;113(1-2):131-6. Epub 2005 Nov 21.

28. Cogen AL, Yamasaki K, Muto J, Sanchez KM, Crotty Alexander L, Tanios J, Lai Y, Kim JE, Nizet V, Gallo RL. Staphylococcus epidermidis antimicrobial delta-toxin (phenol-soluble modulin-gamma) cooperates with host antimicrobial peptides to kill group A Streptococcus. PLoS One. 2010 Jan 5;5(1):e8557.

29. Cogen AL, Yamasaki K, Sanchez KM, Dorschner RA, Lai Y, MacLeod DT, Torpey JW, Otto M, Nizet V, Kim JE, Gallo RL. Selective antimicrobial action is provided by phenol-soluble modulins derived from Staphylococcus epidermidis, a normal resident of the skin. J Invest Dermatol. 2010 Jan;130(1):192-200.

30. Otto M. Basis of virulence in community-associated methicillin-resistant Staphylococcus aureus. Annu Rev Microbiol. 2010 Oct 13;64:143-62.

31. Deleo FR, Otto M, Kreiswirth BN, Chambers HF. Community-associated meticillin-resistant Staphylococcus aureus. Lancet. 2010 May 1;375(9725):1557-68.