UNIVERSITY OF CALIFORNIA, SAN DIEGO

Harnessing Microbial Metabolic Exchange for the Discovery of Biologically Active Molecules

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

in

Chemistry

by

Wei-Ting Liu

Committee in charge:

Professor Pieter C. Dorrestein, Chair
Professor Joshua S. Figueroa
Professor William H. Gerwick
Professor Susan S. Taylor
Professor Roy Wollman

2012

The dissertation of Wei-Ting Liu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego
2012

DEDICATION

This dissertation is dedicated to my family.

Without them, I will not be here.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

ACKNOWLEDGEMENTS

I would like to take this opportunity to acknowledge the contributions of those individuals without whom this thesis will not be possible.

Foremost, I wish to express my sincere gratitude to my research advisor and my committee chair, Prof. Pieter C. Dorrestein, for leading me into the mass spectrometry and natural products fields which disclose a new world for me. I also thank him for providing uncountable support to my life and research. Also, I want to thank all Dorrestein lab members for creating such an exciting lab atmosphere. It has been an absolute pleasure and an honor working with these folks.

Next I would like to thank all the professors who has been served in my committee, Prof. Gerwick, Prof. Taylor, Prof. Figueroa, Prof. Wollman, and Prof. Nakagawa for spending time leading me through my PhD tenure and provided valued comments.

I sincerely thank Prof. Kit Pogliano, Prof. Joe Pogliano, and Anne Lamsa for great collaboration on bioactivity characterization using fluorescence microscopy and Prof. Roger Linington and Weng Ruh Wong for BioMAP assay which adds significant science to the molecules we reported, Prof. Pavel Pevzner, Prof. Nuno Bandeira, and Hosein Mohimani for computational input, and NPAG (natural product affinity group) folks especially Bill Gerwick and Bradley Moore labs for numerous discussions and valuable comments. Without their support my research and scientific findings would have not progressed to a level suitable for a doctorate degree.

I honestly thanks all my friends especially Pi-Han Tsai, Li-Chih Hu, Eugene Lin, Miao-Ping Chien for endless support.

Chapter 1, in full, is a reprint of the material as it appears in a perspective article published in Nature Chemical Biology, 2012. Vanessa Phelan, Wei-Ting Liu, Kit Pogliano, and Pieter C. Dorrestein. The thesis author and V Phelan were the primary investigators and authors of this paper.

Chapter 2, in full, is a reprint of the material as it appears in PNAS, 2010. Wei-Ting Liu, Yu-Liang Yang, Yuquan Xu, Anne Lamsa, Nina M. Haste, Jane Y. Yang, Julio Ng, David Gonzalez, Craig D. Ellermeier, Paul D. Straight, Pavel A. Pevzner, Joe Pogliano, Victor Nizet, Kit Pogliano, and Pieter C. Dorrestein. The thesis author, YL Yang and Y Xu were the primary investigators and authors of this paper.

Chapter 3, in full, is a reprint of the material as it appears in JACS, 2011. Wei-Ting Liu, Roland Kersten, Yu-Liang Yang, Bradley Moore and Pieter C. Dorrestein. The thesis author was the primary investigator and author of this paper.

Chapter 4, in full, is the material that is being prepared for publication. Wei-Ting Liu, Anne Lamsa, Weng Ruh Wong, Paul D. Boudreau, Roland Kersten, Peng Yao, Wilna Moree, Brendan M. Duggan, Bradley S. Moore, William H. Gerwick, Roger G. Linington, Kit Pogliano, Pieter C. Dorrestein. The thesis author was the primary investigator and author of this paper.

VITA

EDUCATION

2012 Doctor of Philosophy, in Chemistry, University of California, San Diego
2009 Master of Science, in Chemistry, University of California, San Diego
2006 Bachelor of Science, in Biological Science,
National Chiao Tung University, Hsinchu, Taiwan

FUNDING & SELECTED AWARDS

2012 Teddy Traylor Award, University of California, San Diego, 2012.
2010, 2011 Study Abroad Grant SAS-98116-2-US-108, Ministry of Education, Taiwan.
2005, 2006 Undergraduate Research Grant, National Science Council, Taiwan.
2002-2006 Full Scholarship for Undergraduate Study, National Chiao Tung University.

TEACHING & CONSULTANT EXPERIENCE

May, 2012 Provide New Mass Spectrometry Approaches Training Sections at Hong Kong University of Science &Technology, Hong Kong
May, 2011 Provide Imaging Mass Spectrometry and Proteomics Training Sections at China Medical University, Taiwan
2007-2009 Chem143A (Organic Chemistry Laboratory), University of California, San Diego

PUBLICATIONS

REPRESENTATIVE PAPERS (*designates equal contribution)

1. **Liu WT**, Lamsa A, Wong WR, Boudreau PD, Kersten RD, Peng Y, Moree W, Duggan BM, Moore BS, Gerwick WH, Linington RG, Pogliano K, Pieter C. Dorrestein. Molecular networking and peptidogenomics guided genome mining revealed the rich metabolic output of Streptomyces roseosporus. ***Manuscript prepared for submission.***

2. Phelan VV*, **Liu WT***, Pogliano K, Dorrestein PC. Microbial Metabolic Exchange-the chemotype-phenotype link. ***Nat Chem Biol***. (Perspective). 8, 26-35 (2011)

3. **Liu WT**, Kersten RD, Yang YL, Moore BS, Dorrestein PC. Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycin in Streptomyces roseosporus. ***J Am Chem Soc***. 133, 18010-3 (2011).

4. **Liu WT***, Yang YL*, Xu Y*, Lamsa A, Haste NM, Yang JY, Ng J, Gonzalez D, Ellermeier CD, Straight PD, Pevzner PA, Pogliano J, Nizet V, Pogliano K, Dorrestein PC. Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of Bacillus subtilis. ***Proc Natl Acad Sci U S A.*** 107,16286-90 (2010).

Featured in "Imaging Reveals Key Metabolic Factors of Cannibalistic Bacteria", Science Daily.
Featured in "Cannibal bacteria could lead to new antibiotics", Ars Technica.
Selected by Faculty of 1000 Biology and rated as "Must Read".

5. **Liu WT***, Ng J*, Meluzzi D, Bandeira N, Gutierrez M, Simmons TL, Schultz AW, Linington RG, Moore BS, Gerwick WH, Pevzner PA, Dorrestein PC. Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. ***Anal Chem.*** 81, 4200-9 (2009).

CO-AUTHORED PAPERS

6. Boudreau PD, Byrum T, **Liu WT**, Dorrestein PC, Gerwick WH. Viequeamide A, a Cytotoxic Member of the Kulolide Superfamily of Cyclic Depsipeptides from a Marine Button Cyanobacterium. ***J Nat Prod.*** 75:1560-70 (2012).

7. Gonzalez DJ, Xu Y, Yang YL, Esquenazi E, **Liu WT**, Edlund A, Duong T, Du L, Molnár I, Gerwick WH, Jensen PR, Fischbach M, Liaw CC, Straight P, Nizet V, Dorrestein PC. Observing the invisible through imaging mass spectrometry, a window into the metabolic exchange patterns of microbes. ***J Proteomics.*** 75, 5069-76 (2012).

8. Lamsa A, **Liu WT**, Dorrestein PC, Pogliano K. The Bacillus subtilis cannibalism toxin SDP collapses the membrane potential and induces autolysis. ***Molecular Microbiology.*** 84, 486-500 (2012).

9. Mohimani H, **Liu WT**, Mylne JS, Poth AG, Colgrave ML, Tran D, Selsted ME, Dorrestein PC, Pevzner PA. Cycloquest: Identification of Cyclopeptides via database search of their mass spectra against genome database. ***J Proteome Res.*** 10, 4505-12 (2011).

10. Mohimani H, Yang YL, **Liu WT**, Hsieh PW, Dorrestein PC, Pevzner PA. Sequencing cyclic peptides by multistage mass spectrometry. ***Proteomics.*** 11, 3642-50 (2011).

11. Yang YL, Xu Y, Kersten RD, **Liu WT**, Meehan MJ, Moore BS, Bandeira N, Dorrestein PC. Connecting chemotypes and phenotypes of cultured marine microbial assemblages by imaging mass spectrometry. ***Angew Chem Int Ed Engl.*** 50, 5839-42 (2011).

Featured on the journal cover.
Featured in "Tiny Talk On a Barnacle's Back", Science Daily.

12. Grindberg RV, Ishoey T, Brinza D, Esquenazi E, Coates RC, **Liu WT**, Gerwick L, Dorrestein PC, Pevzner P, Lasken R, Gerwick WH. Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. ***PLoS One.*** 6, e18565 (2011).

13. Udwary DW, Gontang EA, Jones AC, Jones CS, Schultz AW, Winter JM, Yang JY, Beauchemin N, Capson TL, Clark BR, Esquenazi E, Eustáquio AS, Freel K, Gerwick L, Gerwick WH, Gonzalez D, **Liu WT**, Malloy KL, Maloney KN, Nett M, Nunnery JK, Penn K, Prieto-Davo A, Simmons TL, Weitz S, Wilson MC, Tisa LS, Dorrestein PC, Moore BS. Significant natural product biosynthetic potential of actinorhizal symbionts of the genus frankia, as revealed by comparative genomic and proteomic analyses. ***Appl Environ Microbiol.*** 77, 3617-25 (2011).

14. Mevers E, **Liu WT**, Engene N, Mohimani H, Byrum T, Pevzner PA, Dorrestein PC,

and Gerwick WH. Cytotoxic Veraguamides, Alkynyl Bromide Containing Cyclic Depsipeptides from the Marine Cyanobacterium cf. Oscillatoria margaritifera. ***J. Nat. Prod.*** 74, 928-36 (2011).

15. Mohimani H, **Liu WT**, Yang YL, Guadencio S, Fenical W, Dorrestein PC, and Pevzner PA. Multiplex de-novo sequencing of peptide antibiotics, ***J Comput Biol.*** 18, 1371-81 (2011).
16. Liu W, Tanasa B, Tyurina OV, Zhou TY, Gassmann R, **Liu WT**, Ohgi KA, Benner C, Garcia-Bassets I, Aggarwal AK, Desai A, Dorrestein PC, Glass CK, Rosenfeld MG. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. ***Nature.*** 466, 508-12 (2010).
17. Leão PN, Pereira AR, **Liu WT**, Ng J, Pevzner PA, Dorrestein PC, König GM, Vasconcelos VM, Gerwick WH. Synergistic allelochemicals from a freshwater cyanobacterium. ***Proc Natl Acad Sci U S A.*** 107, 11183-8 (2010).
18. He C, Nora GP, Schneider EL, Kerr ID, Hansell E, Hirata K, Gonzalez D, Sajid M, Boyd SE, Hruz P, Cobo ER, Le C, **Liu WT**, Eckmann L, Dorrestein PC, Houpt ER, Brinen LS, Craik CS, Roush WR, McKerrow J, Reed SL. A Novel Entamoeba Histolytica Cysteine Proteinase, EHCP4, is Key for Invasive Amebiasis and a Therapeutic Target. ***J Biol Chem.*** 285, 18516-27 (2010).
19. Watrous J, Burns K, **Liu WT**, Patel A, Hook V, Bafna V, Barry CE 3rd, Bark S, Dorrestein PC. Expansion of the mycobacterial "PUPylome". ***Mol Biosyst.*** 6, 376-85 (2010).
20. Hughes CC, Yang YL, **Liu WT**, Dorrestein PC, La Clair JJ, Fenical W. Marinopyrrole A target elucidation by acyl dye transfer. ***J Am Chem Soc.*** 131, 12094-6 (2009).
21. Ng J, Bandeira N, **Liu WT**, Ghassemian M, Simmons TL, Gerwick WH, Linington R, Dorrestein PC, Pevzner PA. Dereplication and de novo sequencing of nonribosomal peptides. ***Nat Methods.*** 6, 596-9 (2009).
    Featured in "New Drugs Faster from Natural Compounds", Science Daily.
22. Burns KE, **Liu WT**, Boshoff HI, Dorrestein PC, Barry CE 3rd. Proteasomal protein degradation in Mycobacteria is dependent upon a prokaryotic ubiquitin-like protein. ***J Biol Chem.*** 284, 3069-75 (2009).

CONFERENCE POSTERS

1. **Liu WT**, Bandeira N, Dorrestein PC. Harnessing Microbial Metabolic Exchange for the Discovery of Biologically Active Molecules. *Gordon Research Conference – Marine Natural Products.* **February 2012.** Ventura.
2. **Liu WT**, Kersten RD, Yang YL, Pevzner PA, Moore BS, Dorrestein PC. Mass spectrometry based structural characterization of peptidic natural products (PNPs). *52nd Annual meeting of America society of Pharmacognosy*. **July 2011.** San Diego.
3. **Liu WT**, Kersten RD, Yang YL, Pevzner PA, Moore BS, Dorrestein PC. Mass

spectrometry based structural characterization of peptidic natural products (PNPs). *Gordon Research Conference - High Throughput Chemistry and Chemical Biology.* **June 2011.** New London.

4. **Liu WT**, Mohimani H, Pevzner PA, Dorrestein PC. Mass spectrometry based structural characterization of peptidic natural products (PNPs). *ASMS Conference on Mass Spectrometry and Allied Topics.* **June 2011**. Denver.

5. **Liu WT**, Yang YL, Xu Y, Lamsa A, Yang JY, Ng J, Straight PD, Pevzner PA, Pogliano J, Pogliano K, Dorrestein PC. Metabolic profiling using imaging mass spectrometry identifies new cyclic natural products from *Bacillus subtilis*. *51st Annual meeting of America society of Pharmacognosy*. **July 2010**. Florida.

6. **Liu, WT**, Ng, J, Meluzzi, D, Bandeira, N, Gutierrez, M, Simmons, TL, Schultz AW, Linington, R G, Moore, BS, Gerwick, WH, Pevzner, PA, Dorrestein, PC. The Interpretation and Annotation of Tandem Mass Spectrometry Data Obtained from Non-ribosomal Derived Cyclic Peptides. *The 2008 BIO International Convention.* **June 2009**. San Diego.

ORAL PRESENTATIONS

1. Harnessing Microbial Metabolic Exchange for the Discovery of Biologically Active Molecules. *UCSD Pharmacology Research Discussion.* June 8, 2012.

2. Linking phenotypes, chemotypes with genotypes through imaging mass spectrometry and short sequence tagging (SST). *NPAG meeting*. San Diego. May 13, 2011.

2. Therapeutic lead discovery via identification of the *B. subtilis* cannibalism factors. *NPAG meeting*. San Diego. Aug 13, 2010.

ABSTRACT OF THE DISSERTATION

Harnessing Microbial Metabolic Exchange for the Discovery of Biologically Active Molecules

by

Wei-Ting Liu

Doctor of Philosophy in Chemistry

University of California, San Diego, 2012

Professor Pieter C. Dorrestein, Chair

Microbial metabolic exchange mediates microbial interactions and plays key roles in regulating biology and has shaped modern healthcare, agriculture and other commercial processes. In this thesis, cutting edge mass spectrometry techniques, new genome mining approaches, and innovative bioinformatics tools were coupled and developed into investigating microbial metabolic exchange and led to the identification and characterization of biological active molecules that may have therapeutic values.

The thesis begins by informing the reader current understanding of microbial metabolic exchange as well as the emergence of new technologies that are allowing us to eavesdrop on microbial conversations.

Chapter 2 describes the use of imaging mass spectrometry (IMS) to study a microbial intraspecies interaction which led to the identification and structural elucidation of two complex metabolites, sporulation killing factor (SKF) and sporulation delaying protein (SDP), that are active in a *Bacillus subtilis* cannibalism system. Further bioassays

investigate the cannibalistic activity of SDP and also reveal SDP inhibits *Staphylococci* pathogens with comparable IC50 to vancomycin.

Chapter 3 introduces a new genome-mining approach (peptidogenomics), and when combined with IMS leads the discovery of anti-infective agent arylomycin and its biosynthetic gene cluster in *S. roseosporus*.

Chapter 4 goes beyond studying one or two metabolites at a time, presents the systematic characterization of microbial metabolic profile by the combination of peptidogenomics and an innovative bioinformatics tool (MS/MS spectral molecular networking) that is capable of clustering molecules of similar structure classes which leads to efficient characterization of the "molecular universe" of *S. roseosporus*. One of the molecules, stenothricin, after further investigations suggest it appears to be unique in its mechanism of antibiotic activity.

Lastly, chapter 5 puts forth future directions envisioning how these tools and approaches described in the thesis can be further extended into investigating more complex microbial interactions that are of significant biological or clinical importance. Here, a proposal for investigating the molecular insight of fecal transplantation for the treatment of *Clostridium difficile*-associated diarrhea was presented. Similar approach could be undertaken to investigate other important human microbiota associated diseases/disorders such as Crohn's diseases, obesity, and diabetes.

Chapter I

Microbial Metabolic Exchange

the Chemotype-to-Phenotype Link

1.1 Abstract

The function of microbial interactions is to enable microorganisms to survive by establishing a homeostasis between microbial neighbors and local environments. A microorganism can respond to environmental stimuli using metabolic exchange—the transfer of molecular factors, including small molecules and proteins. Microbial interactions not only influence the survival of the microbes but also have roles in morphological and developmental processes of the organisms themselves and their neighbors. This, in turn, shapes the entire habitat of these organisms. Here we highlight our current understanding of metabolic exchange as well as the emergence of new technologies that are allowing us to eavesdrop on microbial conversations comprising dozens to hundreds of secreted metabolites that control the behavior, survival and differentiation of members of the community. The goal of the rapidly advancing field studying multifactorial metabolic exchange is to devise a microbial 'Rosetta stone' in order to understand the language by which microbial interactions are negotiated and, ultimately, to control the out- come of these conversations.

1.2 Overview

Microbial interactions (**Fig. 1.1**) exist in nearly every niche on this planet, ranging from the oral cavity, intestine and skin of humans, to the cocoons of wasps and down to grains of sand. When at equilibrium, many microorganisms coexist in stable mixed communities. When these communities are perturbed, our ecosystems can be considerably affected, resulting in catastrophic events that have an impact on our society, such as loss of food supplies, destruction of concrete buildings, deadly animal diseases and pandemics. Additionally, modern health care, agriculture and other commercial processes have been shaped by biologically active metabolites produced by fungi and bacteria[1-8]. For instance, the antibiotics penicillin and vancomycin facilitate the control of microbial infections, the immunosuppressant rapamycin allows routine organ transplantation and paclitaxel (Taxol) is a critical treatment for many cancers. Similarly, microbially produced molecules protect our food supplies from microbial and insect invasions; enhance growth of plants, poultry and cattle; and are also used in many consumer products, including soap, toothpaste and paints. Thus, when considering microbially produced metabolites, we often think in terms of how these metabolites influence our quality of life but frequently overlook their impact on complex microbial interactions and as initiators of multicellular behavior in microbial communities—the purposes for which these metabolites are primarily produced. For microbes themselves, microbial interactions provide access to nutrients and protection from external communities and allow adaptation to changing ecological niches.

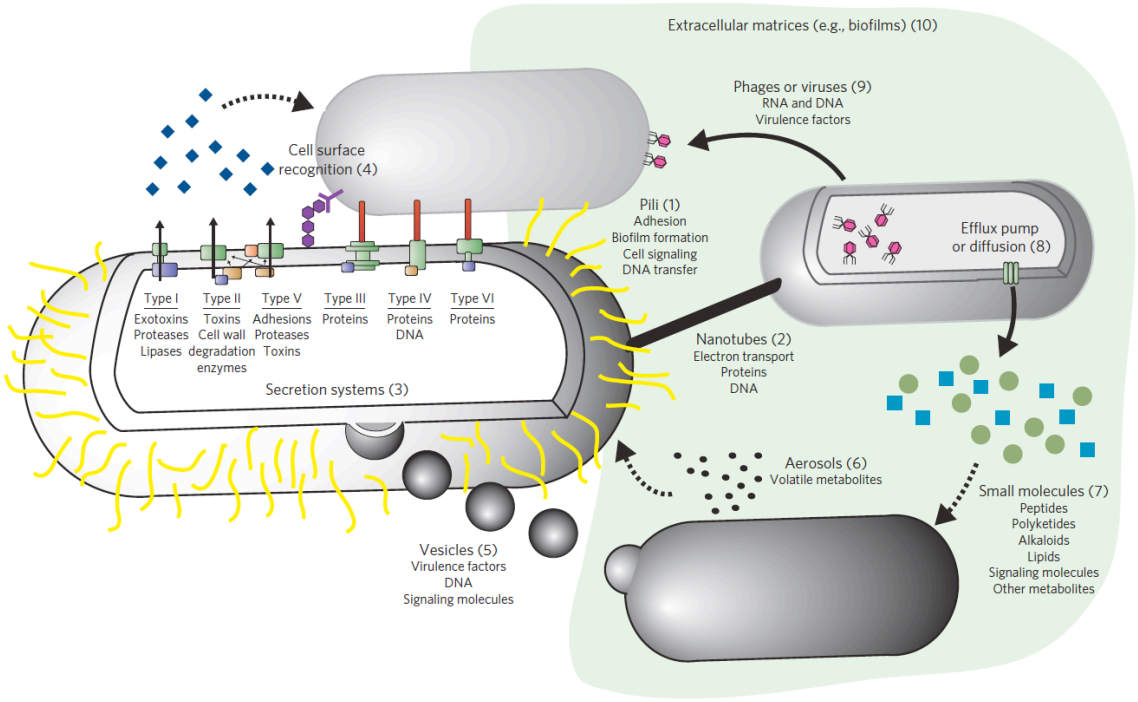


Figure 1.1: Microbial interactions. Microbial interactions may be parasitic, such that one organism benefits at the cost of another; mutualistic, such that both organisms benefit; or commensal, such that one organism benefits at no cost or benefit to the other. All of these interactions, regardless of the outcome, occur through a diverse set of mechanisms by which genetic and molecular information is transferred. The most widely studied mechanisms of microbial interaction, some of which remain controversial, are shown. These include pili (1)[69,70], nanotubes (2)[71–73], secretion systems (3)[74–77], cell surface recognition (4)[78,79], vesicles (5)[80–82], aerosols (6)[83–85], small molecules (7)[86–91] transported via efflux pumps or diffusion (8)[92], phages or viruses (9)[93–95] and biofilms (10)[96]. Each of these types of interaction plays a vital part in microbial metabolic exchange and provides the basis for microbial survival. Although some of these interactions are dependent on cell-to-cell contact, many do not occur through physical contact. Contact-independent metabolic exchange is advantageous because the signals are dispersed, enabling them to reach many neighboring cells and communities as opposed to only one cell at a time. The dispersion of metabolic exchange factors allows them to serve as nutrients or cues to neighboring microbes, thereby controlling the behavior of the larger microbial community and, in effect, leading to behavior as a multicellular entity.

Microbes dedicate enormous resources to microbial interactions. The percentages of microbial genomes that are dedicated to the production of secondary metabolites, a subset of metabolic exchange factors, have been defined by several studies as approximately 5–15%. Amazingly, however, the total number of open reading frames (ORFs) dedicated to microbial interactions has not been determined despite the importance of microbial interactions to the survival and fitness of the individual microbe and the larger microbial community as a whole[9-13]. To estimate the proportion of the bacterial proteome involved in microbial interactions, the genomes of *Staphylococcus aureus* subsp. *aureus* USA300_FPR3757, *Pseudomonas aeruginosa* str. PAO1 and *Bacillus subtilis* subsp. *subtilis* str. 168 were obtained from the Pathosystems Resource Integration Center (PATRIC) database[14] and were manually curated using the National Center for Biotechnology Information (NCBI) basic local alignment search tool (BLAST) in an attempt to assign a function to every predicted ORF.

According to our analysis, 17–42% of the predicted ORFs are dedicated to microbial interactions (**Fig. 1.2**). The ability to assign functions to several ORFs was limited by misannotations within the NCBI database, which includes annotations that do not have a basis in the biology of a given organ- ism. For instance, several ORFs of P. aeruginosa strains have been assigned to be involved in sporulation. As *P. aeruginosa* does not sporulate, these annotations are unfounded and were therefore corrected in our analysis. Additionally, after manual BLAST analysis, many ORFs did not seem to encode the conserved domains required for the previously assigned functions. This analysis, albeit rudimentary, indicates that a much larger portion of the proteomic capacity of these microbes is involved in establishing their interactome than is appreciated at present. By

and large, the interdependencies and diversity of interactions in the interactome are severely underappreciated and poorly understood.

In this Perspective, we discuss one of the major aspects of microbial interactions, microbial metabolic exchange, highlighting the chemical and functional diversity of the metabolites involved, the important roles they have in cellular differentiation within microbial colonies and in the maintenance of ecosystems, the emerging tools that allow us to study microbial interactions in a spatially organized and systematic approach, and how we anticipate this knowledge will shape research and biotechnological applications.

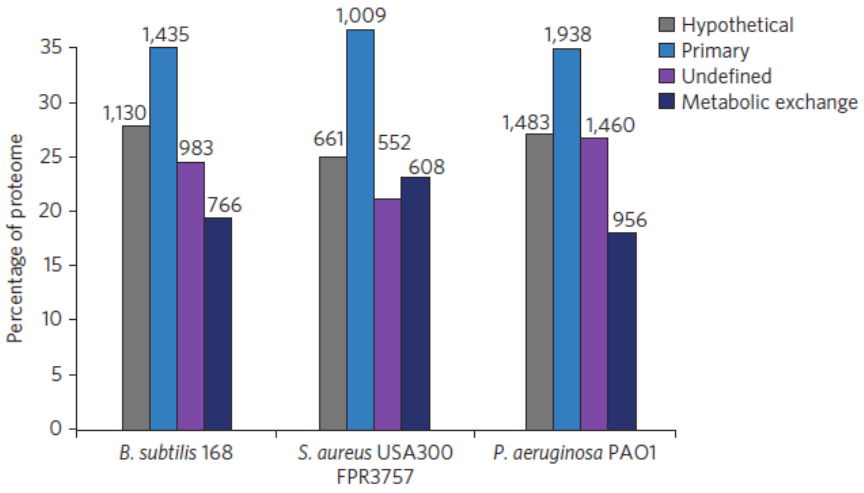


Figure 1.2: Percentages of the predicted ORFs used in microbial interactions. On the basis of BLAST analysis, the predicted ORFs of *S. aureus* subsp. *aureus* USA300 FPR3757, *P. aeruginosa* str. PAO1 and *B. subtilis* subsp. *subtilis* str. 168 were categorized by function into four groups: hypothetical or unassigned ORFs (gray), ORFs involved in primary metabolism (light blue), ORFs for which homologs exist but whose role in metabolic exchange is unclear (purple), and ORFs involved in microbial interactions and metabolic exchange (dark blue). The number above each group corresponds to the number of ORFs in that category. The roles of these ORFs were putatively assigned on the basis of BLAST analysis or inferred from clustering within the genome. For example, hypothetical or primary genes that clustered within gene clusters involved in the production of secondary metabolites were assigned to metabolic exchange, even though their roles in biosynthesis remain unknown. It should be noted that nutrition-based sensing and signaling were not included in this assessment.

1.3 Metabolic Exchange Factors

Quorum-sensing factors are the most widely recognized and most thoroughly studied metabolic exchange factors produced by microbial populations[15–18]. The secretion and detection of quorum-sensing factors is used to gauge cell density and to sense the presence of neighboring species by eavesdropping on the quorum-sensing factors they produce, resulting in the initiation of appropriate group behaviors. The structural diversity of quorum-sensing factors is broad (**Fig. 1.3**). Furthermore, depending on the molecules involved, quorum-sensing responses of microbial communities can vary substantially and may be affected by neighboring microbes and by abiotic factors such as pH, temperature and nutrient availability. Quorum sensing controls developmental processes such as cell differentiation and, in turn, influences many microbial systems of interest, including symbiotic interactions, virulence, competence, conjugation, antibiotic production, motility, sporulation and biofilm formation. Although quorum-sensing inhibition has been speculated to be a good therapeutic approach for targeting many pathogens, there are no molecules in this category with FDA approval at present.

Quorum-sensing responses are often studied as responses to independent molecules rather than in the context of a multifactorial metabolic exchange. For instance, quorum sensing in the human pathogen *S. aureus* involves the autoinducer system locus *agr* (accessory gene regulator), which encodes the quorum sensor autoinducing peptide. This quorum-sensing system is thought to regulate at least 23 secreted factors, including δ-toxin, α-hemolysin, and other toxins and proteases that affect interactions with neighboring microbes as well as with hosts[19]. Thus, in this case and, we suspect, in many others, quorum sensing is not responsible for only one specific phenotype but rather is the

first instruction in a negotiation involving multiple metabolic exchange factors.

Quorum-sensing factors are just one well-studied portion of an even larger collection of molecules that develops and maintains distinct cell populations within communities. Microbes produce a large repertoire of structurally varied metabolic exchange factors that is critical for establishing microbial communities composed of one or several species (**Fig. 1.4**). The largest force shaping microbial communities is nutrition. Mutualistic metabolic dependence (or syntrophism) has led to the evolution of sophisticated mechanisms for nutrient sensing and signaling, allowing mixed communities to thrive by efficiently using even marginal nutritional strategies[20,21]. Akin to quorum-sensing factors, many syntrophic molecules are directives to neighboring organisms. Recent reports highlighting the importance of syntrophic interactions have shown that diffusible metabolic exchange factors from one species can enable cultivation of another species. Specifically, it was shown that isolated marine bacteria would only grow in the presence of a shared 'helper' molecule produced by other organisms. The helper molecule was identified as a siderophore required for iron uptake[22,23]. Such helper molecules have also been observed in interactions between microbes on iron-rich media, and other molecules, such as the cell wall component *N*-acetylglucosamines (GlcNAcs), have been shown to regulate the production of these helper molecules[24,25]. Therefore, the proposal that multimolecular signal response systems exist and many secreted molecules have a specific story to tell, especially when viewed in combination, is an attractive one. What has been largely ignored in research is that multiple signals, loosely defined here as secreted molecules that influence community behavior, alter the multicellular and/or social behavior of microorganisms.

The boundary between metabolites involved in quorum sensing and those involved in other signaling roles often remains unclear. For example, in *B. subtilis*, two quorum-sensing peptides, competence pheromone (ComX) and surfactin, are responsible for paracrine signaling, in which some cells within the population produce a signal targeted to a specific subpopulation[26]. ComX triggers the production of surfactin, which in turn causes a subpopulation of *B. subtilis* cells within the colony to produce the extracellular matrix that holds the biofilm together. In addition to its role in intraspecies signaling, surfactin is an antibiotic and an interspecies signal that controls the production of aerial hyphae in *Streptomyces* species[27]. It is the most effective biological surfactant identified thus far, and this surfactant property is a trait that *B. subtilis* cells harness to move over solid surfaces, including those of plants[28]. Clearly, surfactin is a critical molecule for *B. subtilis*, but even in this case one suspects that we are only beginning to piece together the many key roles it has in the natural history of this species.

Overall, it is striking how little we know about the roles of metabolic exchange factors, including antibiotics, in natural communities. Although traditionally antibiotics were thought of as agents of microbial warfare, these molecules have also been proposed to have roles as quorum-sensing signals or to have other functions that help establish and stabilize microbial communities[29,30]. In the vicinity of the producer, an antibiotic concentration may be high; however, it remains unclear whether antibiotic concentrations in the natural microbial environment are sufficient to kill competing organisms[31]. For example, although high concentrations of the antibiotics tobramycin, ciprofloxacin and tetracycline lead to decreased growth or replication and, eventually, cell death in *P. aeruginosa*, subinhibitory concentrations of these antibiotics increase transcription of the

genes involved in biofilm formation, in effect serving as a signal. Other documented microbial processes affected by sublethal concentrations of antibiotics include motility, hypha formation and sporulation, and it is likely that social interactions and multicellular behavior are also affected[26–28].

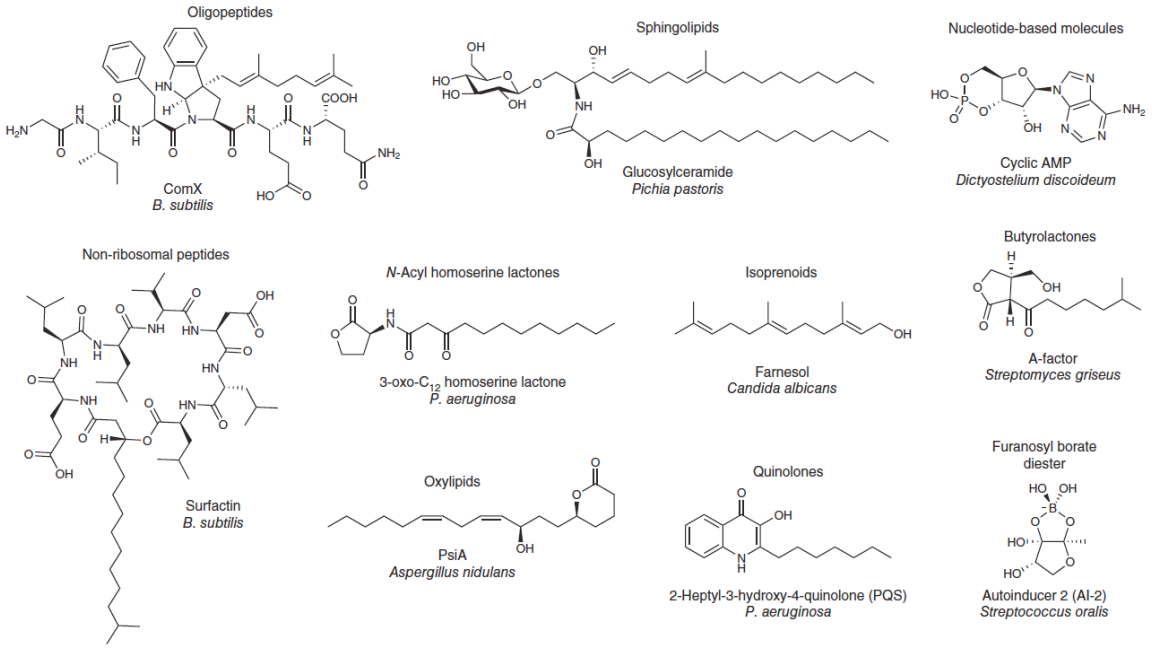


Figure 1.3: Chemical diversity of quorum-sensing molecules. The diversity of quorum-sensing molecules described in the literature is shown. The chemical scaffolds of quorum-sensing factors range in structural complexity from simple isoprenoids and cyclic nucleotides to quinolones to complex peptide scaffolds.

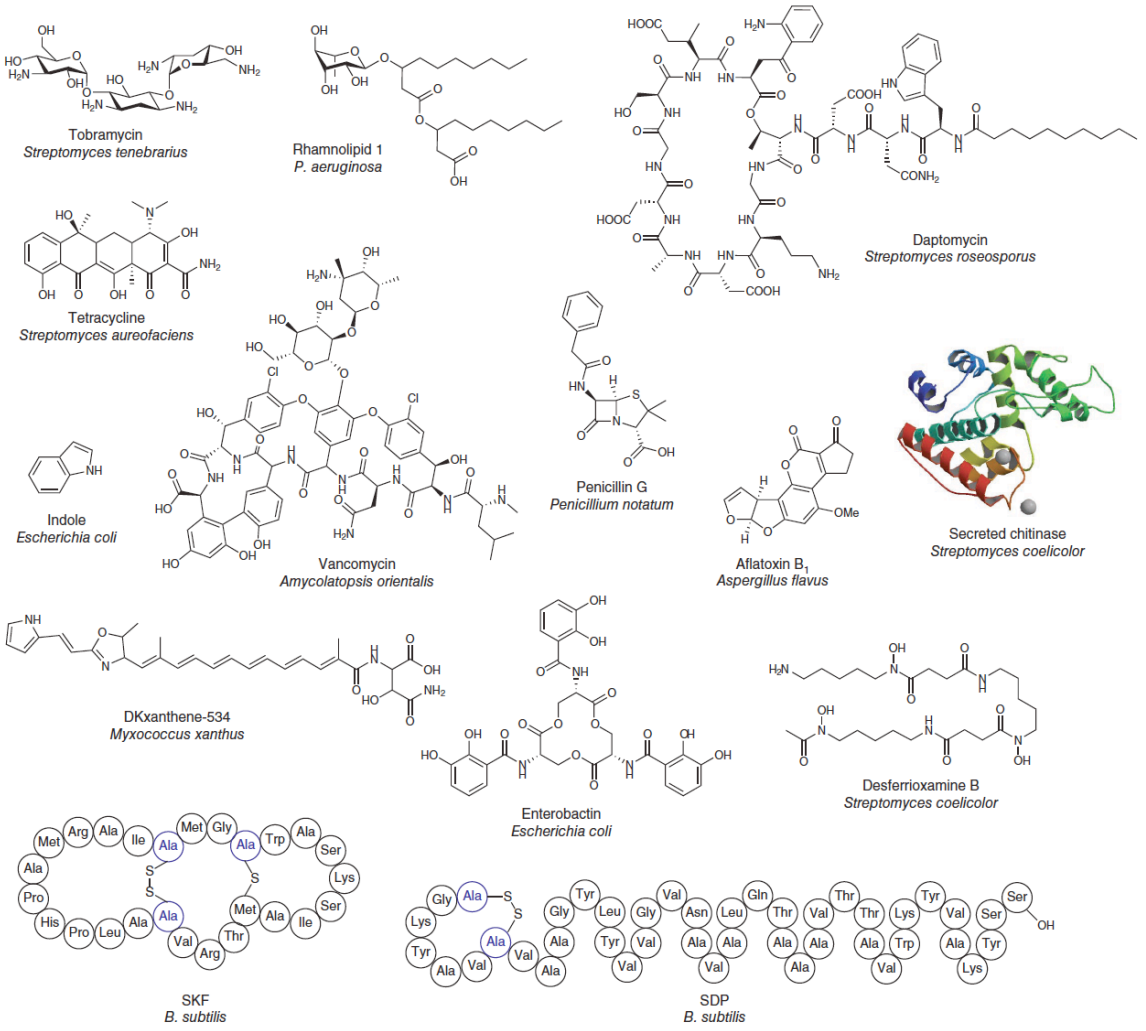


Figure 1.4: Chemical diversity of metabolic exchange factors. The metabolites involved in metabolic exchange have diverse structural scaffolds, ranging from small molecules and peptides to proteins (hydrolases, chitinases, protease and so on).

1.4 Metabolic Exchange in Multicellular Behavior

Understanding how multicellular behavior of microbial colonies is influenced by metabolic exchange factors is crucial for the development of a microbial Rosetta stone. This multicellular behavior is visible in the metabolic output of microbial colonies at both the colony level (**Fig. 1.5a**) and the cellular level (**Fig. 1.5b**), as seen in *B. subtilis*[26]. Although the distribution and localization of metabolic exchange factors are important, temporally regulated production of different classes of factor is essential for mediating cellular differentiation within a microbial colony. An array of secreted quorum-sensing factors and metabolites controls differentiation, but it remains unclear exactly how the combination of signals is processed and integrated to allow the elegant spatial organization of *B. subtilis* colonies to develop[32]. We speculate that, as we are beginning to see in *B. subtilis*, the primary function of metabolic exchange may be to spatially and temporally govern cell differentiation within the colony and to modulate this differentiation as necessitated by neighboring species and environmental conditions.

We anticipate that an understanding of the role of metabolic exchange in bacterial colony differentiation and the formation of mixed communities will lead to an improved ability to cure or prevent disease. This knowledge could potentially lead to an increased understanding of biofilm formation or to ways of decreasing the number of latent cells in infections such as tuberculosis[33]. Furthermore, we anticipate that through understanding and manipulating the multicellular behavior of microorganisms, we could improve the commercial production of microbial compounds either by increasing the number of cells actively producing an antibiotic or by inducing the expression of orphan genes that are involved in antibiotic biosynthesis but whose expression is not observed in pure culture[34].

Simply increasing the number of antibiotic-or biofuel-producing cells from 1 in 1,000 to 4 in 1,000 may increase the yield enough to make a product commercially viable. Thus, systematic approaches that allow the identification of conditions that induce metabolite production represent an important opportunity for biotechnology applications. To fully exploit this opportunity, we will have to understand the nature of the forces driving multicellularity, and in particular how multiple metabolic exchange factors converge on a cell and how this drives cellular differentiation. Ultimately, an appreciation of the roles of metabolic exchange in multicellular microbial communities in the context of biotechnological applications could be exploited for modern multibillion-dollar, biology-based economies (bioeconomies).

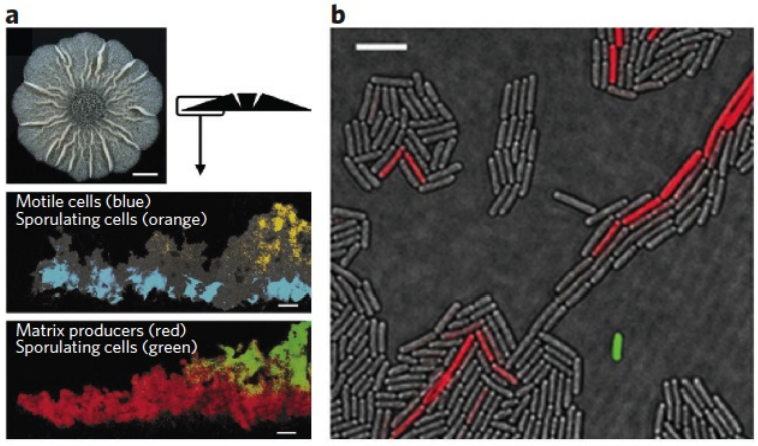


Figure 1.5: Cell differentiation of *Bacillus subtilis* at the colony and cellular levels. Monospecies bacterial communities are in fact multicellular communities with various subpopulations. (a) Using an overlay of fluorescence and transmitted light micrographs, distinct populations of a *B. subtilis* biofilm can be observed where motile cells, sporulating cells and matrix-producing cells are false-colored. Scale bars, 50 mm. (b) Using an overlay of fluorescence and transmitted light microscopy, distinct populations of surfactin-producing cells (expressing surfactin synthase subunit 1 (SrfAA)-YFP; artificially colored green) and matrix-producing cells (expressing YqxM-CFP; YqxM is a protein involved in anchoring cells together in *B. subtilis* biofilms; artificially colored red) can be observed. Although just 1 in 1,000–3,000 cells in a colony produces surfactin, this allows production of up to 1 g l−1 in liquid culture (4b)[97,98]. Scale bar, 3 mm. Figure adapted from ref. 26 with permission.

1.5 Ecological effects of metabolic exchange

In addition to being a potential driver of future bioeconomies, metabolic exchange drives ecology. One elegant early example of the critical part played by microbial metabolic exchange in marine ecology was described in the late 1980s (**Fig. 1.6a**)[35]. It was demonstrated that the small molecule istatin, produced by an *Alteromonas* sp. bacterial symbiont of brine shrimp (*Palaemon macrodactylus*), controls the growth of the pathogenic fungus *Lagenidium callinectes*, a recognized infective agent of many crustaceans. When embryos of the brine shrimp were treated with antibiotics to remove the *Alteromonas* strain, the fungal pathogen killed the embryos within 6 h. It is now clear that there is a rich diversity of microbial symbionts that use metabolic exchange to maintain the balance within the marine biome[36,37].

This rich diversity of metabolic exchange and symbiosis in the marine environment is mirrored in terrestrial ecosystems. A recently uncovered example is the microbial community associated with the life cycle of the leaf-cutting ant (**Fig. 1.6b**). Leaf-cutting ants, such as *Acromyrmex octospinosus*, maintain a mutualistic relationship with the fungus *Leucoagaricus gongylophorus* in which the ants provide the fungus with nutrition by supplying it with harvested leaf material, and in turn the fungus serves as a major food source for the ants[38,39]. Microbial symbionts of leaf-cutting ants, mainly *Pseudonocardia* spp. and *Streptomyces* spp., protect the fungal garden against the microbial pathogen of the genus *Escovopsis* by producing antifungal compounds such as actinomycins, valinomycin, antimycins, candicidin macrolides and the cyclodepsipeptide dentigerumycin, thereby protecting the ants' food supply and colony from destruction[40–43].

In addition to maintaining the external environment, microbial metabolic exchange plays an important part in human microecology. In the human body, it is estimated that mammalian cells are outnumbered by microbial cells ten to one. These microbes protect us from infection, degrade unused substrates, train the immune system and produce vitamins such as biotin or vitamin K[44–46]. The diversity of microbes and their metabolic exchange factors in human communities is staggering. For example, the human mouth provides a habitat for ~500 different species of naturally occurring and transient bacteria. Initial colonization is undertaken by several *Streptococcus* species that bind to salivary receptors and then act as a platform for additional bacteria to aggregate together. In an artificial dental plaque, initial colonization was dominated by *Streptococcus*, *Prevotella*, *Actinomyces* and *Veillonella* species[47]. Cells of the genera *Prevotella* and *Actinomyces* showed the most interspecies associations, and it is therefore proposed that the role of these genera is to establish and maintain biofilm complexity[47]. One example of the role of metabolic exchange factors in the human oral microbiome comes from the quorum-sensing molecules of the autoinducer 2 (AI-2) family, which are essential for mutualistic and abundant biofilm growth of two bacteria: *Actinomyces naeslundii* str. T14V and *Streptococcus oralis* str. 34[48]. In this case, direct contact between *A. naeslundii* and *S. oralis* through coaggregation and coadhesion leads to the upregulation of AI-2 by *S. oralis*. The locally high concentration of AI-2 subsequently allows the formation of a mutualistic mixed biofilm. However, this example is just a sentence in the conversation between the 500 species that compose a mature dental film[49]. To understand these communities, we must identify the molecules that dictate behavior and then understand their meaning and the hierarchy with which they are listened to and decoded.

It is clear that metabolic exchange plays a critical part in environmental and human microecology, but it is evident from sequencing of environmental DNA that these communities are even more complex than is illustrated by these examples[50]. Microbial metabolic exchange has a vital role in maintaining larger multicellular systems, and disruptions of the community may lead to the proliferation of pathogens. However, the system-wide effect of perturbing these interactions has not been established for any system. To truly examine the complexities of microbial metabolic exchange in ecological terms, new tools must be developed and applied.

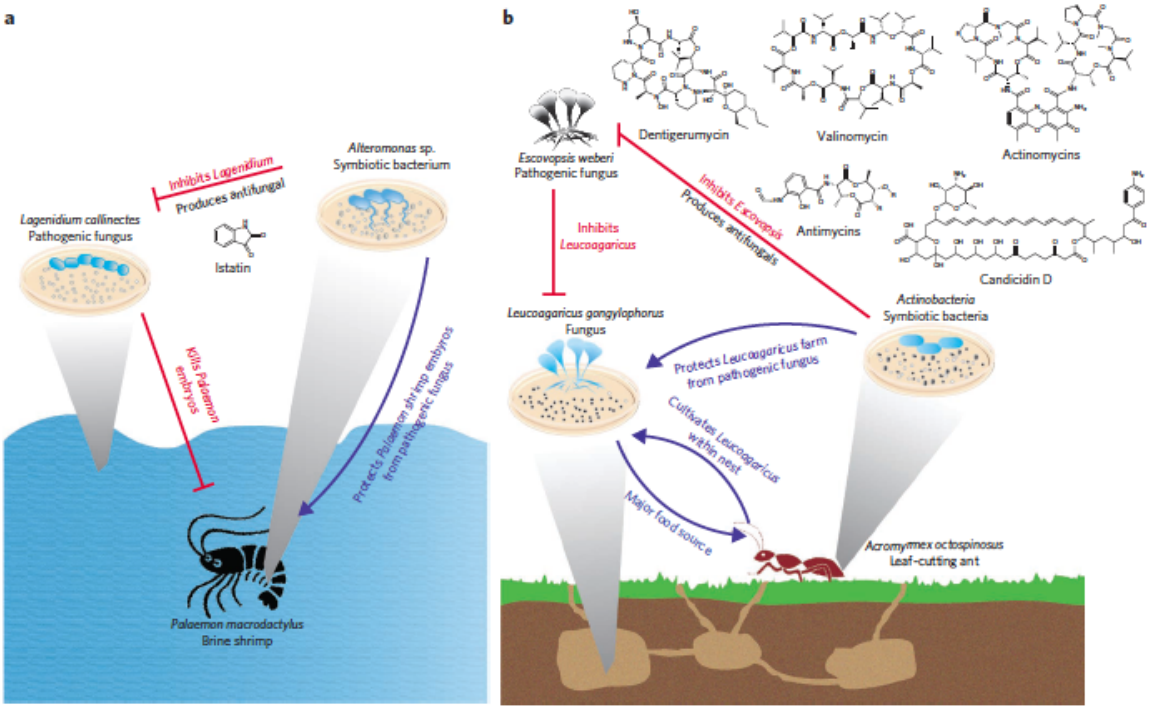


Figure 1.6: Ecological roles of microbial metabolic exchange. Microbial metabolic exchange has important roles in ecology and the survival of higher organisms. (a) Symbiotic bacteria of the brine shrimp produce the antifungal compound istatin, thereby protecting shrimp embryos from pathogenic fungi. (b) *Actinomyces* spp. symbionts of leaf-cutting ants produce metabolites that protect the fungus farmed by the ants from a pathogenic fungus.

1.6 Challenges in studying microbial metabolic exchange

The discovery of penicillin, the result of an interaction between a fungus and *S. aureus*, by Alexander Fleming ignited interest in studying single microbial metabolic exchange factors for medical applications. However, few studies have investigated the effects of the multitude of factors in multicellular communities. Daunting challenges remain, including the thorough identification of all microbes and metabolites involved in microbial metabolic exchange. Complex microbial communities, such as those found in the human intestine, dental plaque, the rhizosphere and biofouling communities, are estimated to contain several hundred to a few thousand different organisms. In the human intestine alone, as many as 1,000 different species of naturally occurring and transient microbes may participate in metabolic exchange[51]. On the basis of the available microbial genome sequences, it is reasonable to estimate that each of the microbes involved in these complex communities has the genetic capacity to produce approximately ten molecules capable of influencing the behavior of neighboring cell populations[52]. Therefore, it is likely that metabolic exchange in these communities involves thousands of molecules, posing a substantial challenge for understanding how they affect behavior. Furthermore, single molecules can participate in several aspects of the community, as exemplified by the roles of surfactin in motility and in intraspecies signaling and as an interspecies antibiotic[26]. It is possible that other metabolic exchange factors also have multiple roles in microbial interactions, but our ability to identify and verify these roles, or to identify alternative responsible factors, is hampered by the lack of tools available to study these complex interactions in context.

1.7 Emerging methods for studying metabolic exchange

The challenges of unraveling microbial interactions, combined with the ecological and medicinal importance of microbial metabolic exchange, have motivated the recent adaptation of a variety of powerful scientific techniques to the study of microbe-microbe and microbe-host interactions. Successful investigations of microbial metabolic exchange require traditional methods in microbiology, such as molecular biology, genetics, systems biology or biochemistry, to be combined with emerging and newly developed methods (**Table 1**), such as multiplexed fluorescence microscopy, NMR imaging, imaging mass spectrometry (IMS), microfluidics or approaches designed to investigate global genomic, proteomic, peptidomic or metabolic profiles. Ultimately, to truly understand the complexity of metabolic exchange in microbial communities, it is necessary to develop tools to identify the microbial players, understand their overall physical and metabolic states, characterize the metabolic outputs associated with those states and the surrounding environment and completely integrate results from diverse experiments.

Table 1.1: Investigational methods for microbial interactions and communities.

Method	Advantages	Disadvantages
IMS	Gathers mass information in both a chemical and a spatial manner	Each sample can be measured only at a single time point
	Correlates specific phenotypes with metabolites	Not all media are compatible with current tools
NMR imaging	Gathers mass information in both a spatial and a temporal manner	Requires specialized equipment
	Direct, non-invasive monitoring of metabolite concentrations	Deconvolution is complex
'Omics' approaches	Provides genetic, proteomic and metabolomic information for single species	Requires biochemical verification
	Correlates genotypes to chemotypes	Reliable, user-friendly databases are unavailable
Metagenomics	Generates a catalogue of microorganisms in a microbiome	Limited to identifying the microbes involved
	Allows comparison of microbiomes in healthy versus diseased individuals	Data analysis is complex
Microfluidics	Can precisely control the microenvironment	Requires expert knowledge for use
	Allows the confinement and quantification of cells	Device materials may be toxic to cells

The study of microbial interactions has led to the convergence of many traditional techniques, as well as the development of new methods. Highlighted are several approaches now being used to investigate microbial metabolic exchange and microbial communities.

The first step in studying microbial metabolic exchange is the identification of the members of a microbial community. This might include traditional methods such as biochemical assays, microscopy and 16S and 18S ribosomal RNA sequencing that are robust but low throughput, still high in cost and not feasible in all cases (not all organisms are amenable to facile DNA extractions). There is a need for the development of tools to efficiently and effectively identify all microbes in a community. To facilitate the quick identification of microbes, fluorescence *in situ* hybridization (FISH) methods are continuously being developed[53]. In FISH, specifically designed fluorescent nucleotide sequences are used as probes to label and identify targeted strains. This was recently accomplished for an artificial dental plaque in which 15 different phylotypes were simultaneously visualized and differentiated.

Another path toward understanding microbial interactions is to identify and characterize the metabolic exchange factors involved in a specific interaction. Typically, a single species is grown in liquid culture, and the resulting cell broth is chemically extracted by organic solvents, solid-phase resins or methods that allow identification of small polar metabolites. These crude extracts undergo bioactivity-guided fractionation against a panel of microbial or cell lines. Using analytical techniques such as HPLC, MS, NMR or X-ray crystallography, individual compounds are purified and structurally characterized. Complementing traditional approaches, recent advances in genome sequencing and protein structure prediction have greatly facilitated identification of biosynthetic genes and prediction of the biochemical activities of their encoded products. A number of tools are available for *in silico* identification of gene clusters involved in the production of potential metabolic exchange factors[54–56]. The most recent addition to these

tools is antiSMASH (antibiotics and secondary metabolite analysis shell), a web-based and stand-alone software pipeline for the identification, annotation and analysis of secondary metabolite gene clusters[57]. However, important challenges remain in this area. Specifically, although there are many public databases for the analysis of proteomic, RNA sequencing, lipidomic and metabolomic data, our ability to analyze the vast amounts of data generated by 'omic' tools and to integrate findings from diverse approaches is limited by a lack of centralization or maintenance and by unfriendly user interfaces.

The integration of biochemical and analytical techniques, as in IMS, also provides a new dimension to investigating the physical and metabolic states of interacting microbes[58]. MALDI-IMS, for example, has been adapted by our laboratory to agar-based microbial cultures, enabling the investigation of the chemical identity and spatial distribution of metabolic exchange factors that ionize by MS. The MALDI-IMS approaches in our laboratory (outlined in **Fig. 1.7**) require the sample to be covered with matrix and then dehydrated for analysis. We used MALDI-IMS to investigate *B. subtilis* cannibalism[59], as well as the bacterium's single-neighbor interactions with *Streptomyces coelicolor* and *S. aureus*[60,61] and its interactions within marine microbial assemblages[62]. Using MALDI-IMS, we were able to identify several metabolites produced by these organisms, including the cannibalistic factors sporulation delaying protein (Sdp) and sporulation killing factor (Skf) of *B. subtilis*. IMS allowed direct observation of these cannibalistic metabolites, which eased their isolation and subsequent structural analysis by MS and targeted NMR. Although MALDI-IMS has provided us with a new way to investigate microbial interactions at the chemical level, it is limited by the required

application of matrix and the necessary dehydration before sample analysis. The process of matrix application and dehydration kills the microbes and therefore restricts analysis to only a single colony at one point in time. In addition, in using our current MALDI-IMS methods, not all types of agar medium are compatible; therefore, each medium must be evaluated before an experiment is initiated. Furthermore, not all molecules ionize or are stable under the conditions in use at present[60]. To address these limitations of MALDI-IMS of microbial cultures, we are currently developing other MS-based techniques to complement MALDI-IMS to enable metabolic analysis of live microbial colonies.

Concurrent with the development of live MS analysis of microorganisms, NMR imaging is being developed as a tool to investigate live biofilms[63,64]. These tools have been used to interrogate the structure and the dynamic metabolic processes of *Shewanella oneidensis* str. MR-1 and *Streptococcus mutans* str. UA159 by integrating NMR with confocal laser scanning microscopy[63,64]. NMR provides direct, time-resolved, noninvasive monitoring of metabolite concentration, metabolic pathways and flux rates, while a confocal laser scanning microscope within the NMR magnet can monitor fluorescent tags to follow gene expression or an individual strain. The combination of these techniques provides an exciting opportunity to produce a full three-dimensional image of a live biofilm that includes NMR data showing the spatial distribution of small metabolites.

Recently, microfluidics has been presented as a tool that can be used to study metabolic exchange[65]. For a cell to differentiate into a specific cell type within a microbial colony, several signals may need to converge at specific concentrations. Microfluidics may provide us with a method to achieve those concentrations *in vitro*.

This ability will enable researchers to investigate the impact of gradients of multiple metabolic exchange factors on the behaviors of single cells or biofilm communities.

The methods mentioned above are a small subset of the advanced tools that have been developed to facilitate the study of microbial metabolic exchange. Although there have been many advances in the tools and methods available for investigating microbial interactions, research still suffers from the limitations of these tools, even when used in combination. We can visualize single cells with microscopy. We can visualize metabolites involved in metabolic exchange with IMS. However, the current tools available are not capable—or have not yet been demonstrated to be capable—of the resolution required to differentiate single cells and simultaneously detect the metabolites produced and their distribution in communities on a surface of choice. Ideally, we need to identify each microbe and visualize each individual cell and the metabolites it produces on materials such as mucosal epithelium, plant stems, flower petals or soil and to connect this information to genomic signatures. This will require substantial technological advances.

Although many current studies of microbial metabolic interactions are simplified to focus on just a single molecule or one interaction at a time, we envision a future in which multiplexed interactions can be studied in their natural context and the outcome of interspecies interactions can be controlled to benefit human health and agriculture. Thus, the next frontier is the development of tools that connect genotypes, chemotypes and phenotypes in complex environments and communities. To accomplish this, we need to continue to push the boundaries of what is possible, often with creative adaptations of existing instrumentation and the fusion of heretofore separate approaches and fields of

microbiology. Our laboratories use IMS, in addition to other MS and cell biology approaches, as a tool to connect chemotypes to phenotypes, providing insight into the largely invisible molecular sphere and an opportunity to observe the multiplexed nature of spatial systems microbiology[66]. However, to continue to push this field forward, it is critical to develop new methods that correlate the observed chemistry with changes in cell architecture, behavior and development to elucidate the function of newly discovered molecules. Connecting the chemistry to the genes responsible for metabolite biosynthesis and cellular responses and determining the developmental state of individual cells within those regions of the colonies that are producing specific metabolites will be critical for our understanding of microbial interactions in nature.

We already know that metabolic exchange factors serve as cues or signals in microbial communities; they are required for morphological and developmental processes, the survival of individual microbes and the fitness of the microbial community as a whole[67,68]. Typically, the language of microbes is studied one signal at a time, with entire subfields focusing on just a few molecules. In reality, there are many different classes of molecules necessary for the social behavior of microbes. Describing just one molecule at a time, typically out of context, is akin to us using one sentence instead of a full dialogue when we speak[49]. Fortunately, tools are emerging that enable us to listen in on microbial conversations consisting of several molecules, allowing microbial metabolic exchange to be documented in outstanding detail and in context. It is our opinion that these approaches provide the only way to develop a Rosetta stone for microbial communication as they will allow us to establish functional translations of the molecules that microbes secrete and the signaling networks that these molecules affect. In our quest

to generate a Rosetta stone, we should not underestimate the complexity of microbial interactions in the context of multicellular and community behavior. If every organism in a community of 500 microbes differentiates into 5 cell types, each producing 3 molecules that provide specific instructions to neighboring cells, there would be 7,500 molecules present and actively providing instruction in this community. Perhaps other disciplines, such as those concerning human population dynamics and interactions, can provide some critical insight. Typically, attention has been given to single factors and not to the multifactorial nature of metabolic exchange. To develop a Rosetta stone, we must understand what types of molecule are produced and how they act in concert, and we must continue to develop tools to study metabolic exchange.

There is little doubt that the next two decades will see the development of user-friendly tools to study the metabolic exchange of microbes both at the cellular level and in real time. Accomplishing this goal will require continued collaborative contributions from biologists, chemists, chemical engineers, informaticians, mathematicians and other investigators to design and build tools and methods for observing genetic and molecular changes spatially and temporally. It is these tools that will allow us to develop the Rosetta stone for microbial conversations, enabling us to understand, in molecular detail, how microbial communities are established and maintained in nature and in human hosts.

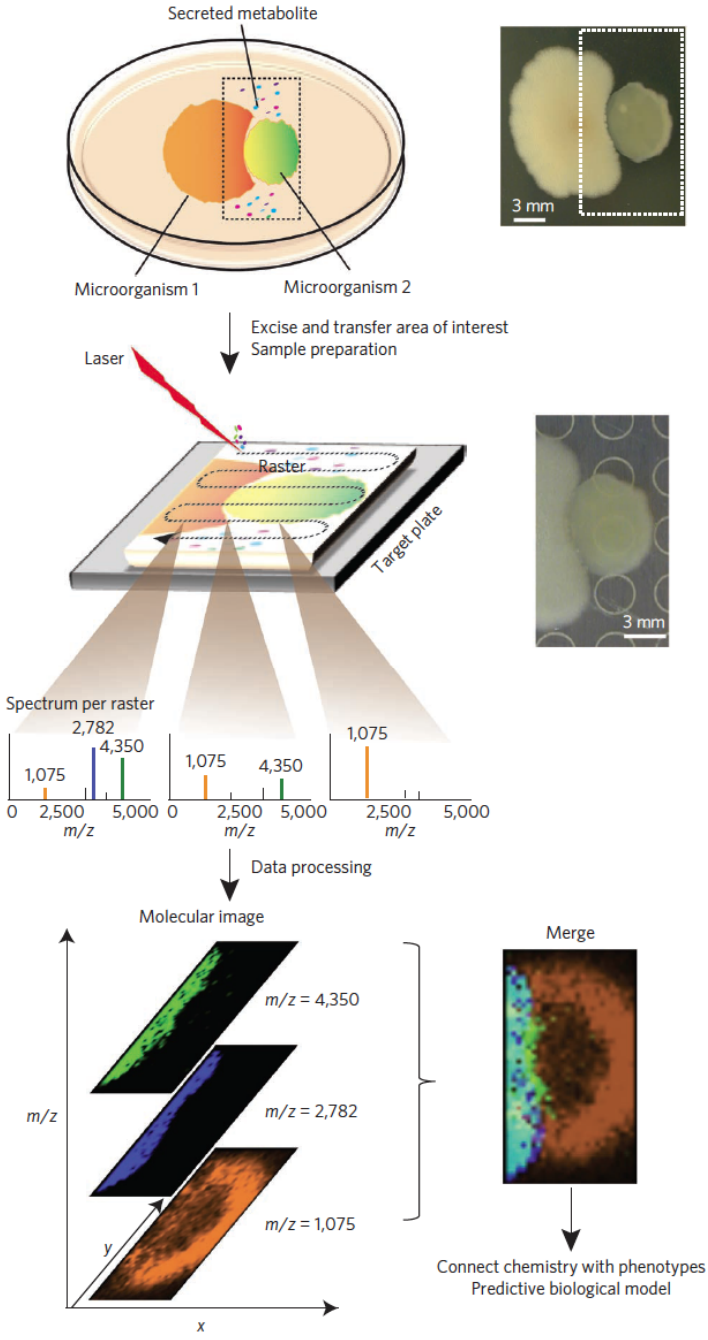


Figure 1.7: MALDI-IMS links chemistry to bacterial phenotypes. MALDI-IMS can be used to visualize the spatial distribution of metabolic exchange factors and to aid in their identification. To prepare the sample for MALDI-IMS, microbial colonies are cultured on an agar plate, excised, transferred to a MALDI target surface, covered with matrix, dried and subjected to rastering MALDI-MS. AMALDI-MS image is generated by directing a laser at different *x*,*y* positions of a sample in a predefined manner, creating a two-dimensional molecular profile of the molecules that are present in the top layer of the sample. Any ion observed in these spectra can be spatially visualized with a false-color that reflects the intensity of the MS signal[60].

References

1. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* **70**, 461–477 (2007).

2. Banat, I.M., Makkar, R.S. & Cameotra, S.S. Potential commercial applications of microbial surfactants. *Appl. Microbiol. Biotechnol.* **53**, 495–508 (2000).

3. Lugtenberg, B. & Kamilova, F. Plant-growth-promoting rhizobacteria. *Annu. Rev. Microbiol.* **63**, 541–556 (2009).

4. Hasan, F., Shah, A.A. & Hameed, A. Industrial applications of microbial lipases. *Enzyme Microb. Technol.* **39**, 235–251 (2006).

5. Fujinami, S. & Fujisawa, M. Industrial applications of alkaliphiles and their enzymes–past, present and future. *Environ. Technol.* **31**, 845–856 (2010).

6. Fusetani, N. Antifouling marine natural products. *Nat. Prod. Rep.* **28**, 400–410 (2011).

7. Dayan, F.E., Cantrell, C.L. & Duke, S.O. Natural products in crop protection. *Bioorg. Med. Chem.* **17**, 4022–4034 (2009).

8. Babalola, O.O. Beneficial bacteria of agricultural importance. *Biotechnol. Lett.* **32**, 1559–1570 (2010).

9. Omura, S. *et al.* Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci. USA* **98**, 12215–12220 (2001).

10. Ikeda, H. *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**, 526–531 (2003).

11. Bentley, S.D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147 (2002).

12. Ohnishi, Y. *et al.* Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **190**, 4050–4060 (2008).

13. Nett, M., Ikeda, H. & Moore, B.S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26**, 1362–1384 (2009).

14. Snyder, E.E. *et al.* PATRIC: the VBI PathoSystems Resource Integration Center. *Nucleic Acids Res.* **35**, D401–D406 (2007).

15. Lowery, C.A., Dickerson, T.J. & Janda, K.D. Interspecies and interkingdom communication mediated by bacterial quorum sensing. *Chem. Soc. Rev.* **37**, 1337–

1346 (2008).

16. Straight, P.D. & Kolter, R. Interspecies chemical communication in bacterial development. *Annu. Rev. Microbiol.* **63**, 99–118 (2009).

17. Ng, W.L. & Bassler, B.L. Bacterial quorum-sensing network architectures. *Annu. Rev. Genet.* **43**, 197–222 (2009).

18. Schuster, M. & Greenberg, E.P. A network of networks: quorum-sensing gene regulation in *Pseudomonas aeruginosa*. *Int. J. Med. Microbiol.* **296**, 73–81 (2006).

19. Antunes, L.C., Ferreira, R.B., Buckner, M.M. & Finlay, B.B. Quorum sensing in bacterial virulence. *Microbiology* **156**, 2271–2282 (2010).

20. Little, A.E., Robinson, C.J., Peterson, S.B., Raffa, K.F. & Handelsman, J. Rules of engagement: interspecies interactions that regulate microbial communities. *Annu. Rev. Microbiol.* **62**, 375–401 (2008).

21. Kaeberlein, T., Lewis, K. & Epstein, S.S. Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296**, 1127–1129 (2002).

22. D'Onofrio, A. *et al.* Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* **17**, 254–264 (2010).

23. Lewis, K., Epstein, S., D'Onofrio, A. & Ling, L.L. Uncultured microorganisms as a source of secondary metabolites. *J. Antibiot. (Tokyo)* **63**, 468–476 (2010).

24. Rigali, S. *et al.* Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep.* **9**, 670–675 (2008).

25. van Wezel, G.P. & McDowall, K.J. The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat. Prod. Rep.* **28**, 1311–1333 (2011).

26. López, D. & Kolter, R. Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. *FEMS Microbiol. Rev.* **34**, 134–149 (2010).

27. Straight, P.D., Willey, J.M. & Kolter, R. Interactions between *Streptomyces coelicolor* and *Bacillus subtilis*: Role of surfactants in raising aerial structures. *J. Bacteriol.* **188**, 4918–4925 (2006).

28. Angelini, T.E., Roper, M., Kolter, R., Weitz, D.A. & Brenner, M.P. *Bacillus subtilis* spreads by surfing on waves of surfactant. *Proc. Natl. Acad. Sci. USA* **106**, 18109–18113 (2009).

29. Davies, J. Everything depends on everything else. *Clin. Microbiol. Infect.* **15** (suppl. 1): 1–4 (2009).

30. Yim, G., Wang, H.H. & Davies, J. Antibiotics as signalling molecules. *Phil. Trans. R. Soc. Lond. B* **362**, 1195–1200 (2007).

31. Fajardo, A. & Martinez, J.L. Antibiotics as signals that trigger specific bacterial responses. *Curr. Opin. Microbiol.* **11**, 161–167 (2008).

32. Shapiro, J.A. Thinking about bacterial populations as multicellular organisms. *Annu. Rev. Microbiol.* **52**, 81–104 (1998).

33. Strauss, E. Grand challenge commentary: Exploiting single-cell variation for new antibiotics. *Nat. Chem. Biol.* **6**, 873–875 (2010).

34. Winter, J.M., Behnken, S. & Hertweck, C. Genomics-inspired discovery of natural products. *Curr. Opin. Chem. Biol.* **15**, 22–31 (2011).

35. Gil-Turnes, M.S., Hay, M.E. & Fenical, W. Symbiotic marine bacteria chemically defend crustacean embryos from a pathogenic fungus. *Science* **246**, 116–118 (1989).

36. Piel, J. Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **26**, 338–362 (2009).

37. Hentschel, U. *et al.* Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.* **68**, 4431–4440 (2002).

38. Currie, C.R., Scott, J.A., Summerbell, R.C. & Malloch, D. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* **398**, 701–704 (1999).

39. Currie, C.R. *et al.* Ancient tripartite coevolution in the attine ant-microbe symbiosis. *Science* **299**, 386–388 (2003).

40. Oh, D.C., Poulsen, M., Currie, C.R. & Clardy, J. Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. *Nat. Chem. Biol.* **5**, 391–393 (2009).

41. Schultz, T.R. & Brady, S.G. Major evolutionary transitions in ant agriculture. *Proc. Natl. Acad. Sci. USA* **105**, 5435–5440 (2008).

42. Haeder, S., Wirth, R., Herz, H. & Spiteller, D. Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis*. *Proc. Natl. Acad. Sci. USA* **106**, 4742–4746 (2009).

43. Schoenian, I. *et al.* Chemical basis of the synergism and antagonism in microbial communities in the nests of leaf-cutting ants. *Proc. Natl. Acad. Sci. USA* **108**, 1955–1960 (2011).

44. Guarner, F. & Malagelada, J.R. Gut flora in health and disease. *Lancet* **361**, 512–519 (2003).

45. Grice, E.A. & Segre, J.A. The skin microbiome. *Nat. Rev. Microbiol.* **9**, 244–253 (2011).

46. Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L. & Gordon, J.I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327–336 (2011).

47. Valm, A.M. *et al.* Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proc. Natl. Acad. Sci. USA* **108**, 4152–4157 (2011).

48. Rickard, A.H. *et al.* Autoinducer 2: a concentration-dependent signal for mutualistic bacterial biofilm growth. *Mol. Microbiol.* **60**, 1446–1456 (2006).

49. Schloss, P.D. & Handelsman, J. The last word: books as a statistical metaphor for microbial communities. *Annu. Rev. Microbiol.* **61**, 23–34 (2007).

50. Tringe, S.G. & Rubin, E.M. Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* **6**, 805–814 (2005).

51. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).

52. Baltz, R.H. Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol.* **8**, 557–563 (2008).

53. Moter, A. & Gobel, U.B. Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. *J. Microbiol. Methods* **41**, 85–112 (2000).

54. de Jong, A., van Heel, A.J., Kok, J. & Kuipers, O.P. BAGEL2: mining for bacteriocins in genomic data. *Nucleic Acids Res.* **38**, W647–W651 (2010).

55. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D.H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799–5808 (2005).

56. Ansari, M.Z., Yadav, G., Gokhale, R.S. & Mohanty, D. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* **32**, W405–W413 (2004).

57. Medema, M.H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal sequences. *Nucleic Acids Res.* **39**, W339–W346 (2011).

58. Watrous, J.D. & Dorrestein, P.C. Imaging mass spectrometry in microbiology. *Nat. Rev. Microbiol.* **9**, 683–694 (2011).

59. Liu, W.T. *et al.* Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **107**, 16286–16290 (2010).

60. Yang, Y.L., Xu, Y., Straight, P. & Dorrestein, P.C. Translating metabolic exchange with imaging mass spectrometry. *Nat. Chem. Biol.* **5**, 885–887 (2009).

61. Gonzalez, D. *et al.* Microbial competition between *Bacillus subtilis* and *Staphylococcus aureus* monitored by imaging mass spectrometry. *Microbiology* published online, doi:10.1099/mic.0.048736-0 (30 June 2011).

62. Yang, Y.L. *et al.* Connecting chemotypes and phenotypes of cultured marine microbial assemblages by imaging mass spectrometry. *Angew. Chem. Int. Edn. Engl.* **50**, 5839–5842 (2011).

63. McLean, J.S. *et al.* Investigations of structure and metabolism within *Shewanella oneidensis* MR-1 biofilms. *J. Microbiol. Methods* **74**, 47–56 (2008).

64. McLean, J.S., Ona, O.N. & Majors, P.D. Correlated biofilm imaging, transport and metabolism measurements via combined nuclear magnetic resonance and confocal microscopy. *ISME J.* **2**, 121–131 (2008).

65. Boedicker, J.Q., Vincent, M.E. & Ismagilov, R.F. Microfluidic confinement of single cells of bacteria in small volumes initiates high-density behavior of quorum sensing and growth and reveals its variability. *Angew. Chem. Int. Edn. Engl.* **48**, 5908–5911 (2009).

66. Kersten, R.D. *et al.* A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* **7**, 794–802 (2011).

67. De Sordi, L. & Muhlschlegel, F.A. Quorum sensing and fungal-bacterial interactions in *Candida albicans*: a communicative network regulating microbial coexistence and virulence. *FEMS Yeast Res.* **9**, 990–999 (2009).

68. Shank, E.A. & Kolter, R. New developments in microbial interspecies signaling. *Curr. Opin. Microbiol.* **12**, 205–214 (2009).

69. Craig, L., Pique, M.E. & Tainer, J.A. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2**, 363–378 (2004).

70. Proft, T. & Baker, E.N. Pili in Gram-negative and Gram-positive bacteria -structure, assembly and their role in disease. *Cell. Mol. Life Sci.* **66**, 613–635 (2009).

71. Dubey, G.P. & Ben-Yehuda, S. Intercellular nanotubes mediate bacterial communication. *Cell* **144**, 590–600 (2011).

72. Gorby, Y.A. *et al.* Electrically conductive bacterial nanowires produced by

Shewanella oneidensis strain MR-1 and other microorganisms. *Proc. Natl. Acad. Sci. USA* **103**, 11358–11363 (2006).

73. Reguera, G. *et al.* Extracellular electron transfer via microbial nanowires. *Nature* **435**, 1098–1101 (2005).

74. DiGiuseppe Champion, P.A. & Cox, J.S. Protein secretion systems in Mycobacteria. *Cell. Microbiol.* **9**, 1376–1384 (2007).

75. Hayes, C.S., Aoki, S.K. & Low, D.A. Bacterial contact-dependent delivery systems. *Annu. Rev. Genet.* **44**, 71–90 (2010).

76. Holland, I.B. The extraordinary diversity of bacterial protein secretion mechanisms. *Methods Mol. Biol.* **619**, 1–20 (2010).

77. Natale, P., Bruser, T. & Driessen, A.J. Sec-and Tat-mediated protein secretion across the bacterial cytoplasmic membrane–distinct translocases and mechanisms. *Biochim. Biophys. Acta* **1778**, 1735–1756 (2008).

78. Lebeer, S., Vanderleyden, J. & De Keersmaecker, S.C. Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. *Nat. Rev. Microbiol.* **8**, 171–184 (2010).

79. Ubbink, J. & Schar-Zammaretti, P. Probing bacterial interactions: integrated approaches combining atomic force microscopy, electron microscopy and biophysical techniques. *Micron* **36**, 293–320 (2005).

80. Haurat, M.F. *et al.* Selective sorting of cargo proteins into bacterial membrane vesicles. *J. Biol. Chem.* **286**, 1269–1276 (2011).

81. Mashburn, L.M. & Whiteley, M. Membrane vesicles traffic signals and facilitate group activities in a prokaryote. *Nature* **437**, 422–425 (2005).

82. Roze, L.V., Chanda, A. & Linz, J.E. Compartmentalization and molecular traffic in secondary metabolism: a new understanding of established cellular processes. *Fungal Genet. Biol.* **48**, 35–48 (2011).

83. Kai, M. *et al.* Bacterial volatiles and their action potential. *Appl. Microbiol. Biotechnol.* **81**, 1001–1012 (2009).

84. Minerdi, D., Bossi, S., Gullino, M.L. & Garibaldi, A. Volatile organic compounds: a potential direct long-distance mechanism for antagonistic action of *Fusarium oxysporum* strain MSA 35. *Environ. Microbiol.* **11**, 844–854 (2009).

85. Schulz, S. & Dickschat, J.S. Bacterial volatiles: the smell of small organisms. *Nat. Prod. Rep.* **24**, 814–842 (2007).

86. Borges-Walmsley, M.I. & Walmsley, A.R. cAMP signalling in pathogenic fungi: control of dimorphic switching and pathogenicity. *Trends Microbiol.* **8**, 133–141 (2000).

87. Boyer, M. & Wisniewski-Dye, F. Cell-cell signalling in bacteria: not simply a matter of quorum. *FEMS Microbiol. Ecol.* **70**, 1–19 (2009).

88. Pesavento, C. & Hengge, R. Bacterial nucleotide-based second messengers. *Curr. Opin. Microbiol.* **12**, 170–176 (2009).

89. Ratcliff, W.C. & Denison, R.F. Microbiology. Alternative actions for antibiotics. *Science* **332**, 547–548 (2011).

90. Singh, A. & Del Poeta, M. Lipid signalling in pathogenic fungi. *Cell. Microbiol.* **13**, 177–185 (2011).

91. Vendeville, A., Winzer, K., Heurlier, K., Tang, C.M. & Hardie, K.R. Making 'sense' of metabolism: autoinducer-2, LuxS and pathogenic bacteria. *Nat. Rev. Microbiol.* **3**, 383–396 (2005).

92. Putman, M., van Veen, H.W. & Konings, W.N. Molecular properties of bacterial multidrug transporters. *Microbiol. Mol. Biol. Rev.* **64**, 672–693 (2000).

93. Boyd, E.F. & Brussow, H. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol.* **10**, 521–529 (2002).

94. Coleman, D.C. *et al. Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of beta-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. *J. Gen. Microbiol.* **135**, 1679–1697 (1989).

95. Kuroda, M. *et al.* Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* **357**, 1225–1240 (2001).

96. Flemming, H.C. & Wingender, J. The biofilm matrix. *Nat. Rev. Microbiol.* **8**, 623–633 (2010).

97. Kearns, D.B. & Losick, R. Swarming motility in undomesticated *Bacillus subtilis*. *Mol. Microbiol.* **49**, 581–590 (2003).

98. López, D., Vlamakis, H., Losick, R. & Kolter, R. Cannibalism enhances biofilm development in *Bacillus subtilis*. *Mol. Microbiol.* **74**, 609–618 (2009).

Chapter 1, in full, is a reprint of the material as it appears in a perspective article published in Nature Chemical Biology, 2012. Vanessa Phelan, Wei-Ting Liu, Kit Pogliano, and Pieter C. Dorrestein. The thesis author and Vanessa Phelan were the primary investigators and authors of this paper.

Chapter II

Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*

2.1 Abstract

During bacterial cannibalism, a differentiated subpopulation harvests nutrients from their genetically identical siblings to allow continued growth in nutrient-limited conditions. Hypothesis-driven imaging mass spectrometry (IMS) was used to identify metabolites active in a *B. subtilis* cannibalism system in which sporulating cells lyse non-sporulating siblings. Two candidate molecules with sequences matching the products of *skfA* and *sdpC*, genes for the proposed cannibalistic factors sporulation killing factor (SKF) and the sporulation delaying protein (SDP), respectively, were identified and the structures of the final products elucidated. SKF is a cyclic 26-amino acid peptide that is post-translationally modified with one disulfide and one cysteine thioether bridged to the α-position of a methionine, a post-translational modification not previously described in biology. SDP is a 42-residue peptide with one disulfide bridge. In spot test assays on solid medium, overproduced SKF and SDP enact a cannibalistic killing effect with SDP having higher potency. However, only purified SDP affected *B. subtilis* cells in liquid media in fluorescence microscopy and growth assays. Specifically, SDP treatment delayed growth in a concentration-dependent manner, caused increases in cell permeability and ultimately cell lysis accompanied by the production of membrane tubules and spheres. Similarly, SDP but not SKF was able to inhibit the growth of the pathogens *Staphylococcus aureus* and *Staphylococcus epidermidis* with comparable IC_{50}'s to vancomycin. This investigation, with the identification of SKF and SDP structures, highlights the strength of imaging mass spectrometry in investigations of metabolic exchange of microbial colonies and also demonstrates IMS as a promising approach to discover novel biologically active molecules.

2.2 Introduction

Metabolic exchange describes the process of exchanging signals or nutrients between cells or populations and is a common feature of all living systems. Bacteria produce a wide array of signaling molecules to control metabolic as well as morphological and developmental changes in either an interspecies or intraspecies manner[1]. *Bacillus subtilis,* for example, has a complex life cycle and thrives in diverse living conditions ranging from soil, contaminated wounds and the intestinal tract[2-4]. To accommodate this, *B. subtilis* dedicates ~10% of its genome to the production of specific molecules involved in intra- and interspecies metabolic exchange[5]. Two of these molecules are SDP and SKF that, based on genetic experiments, are proposed to lyse a subpopulation of *B. subtilis* cells to provide nutrients for the remaining cells, a process referred to as bacterial cannibalism[6-10]. This behavior is dependent on Spo0A, a master transcriptional regulator that also controls biofilm formation and sporulation[6-13]. We set out to characterize these cannibalistic compounds in order to establish their roles in the *B. subtilis* life cycle and to understand their structure and biosynthesis. Additionally, previous reports suggested that the *skf* and *sdp* gene clusters preferentially target non-*B. subtilis* cells, suggesting the cannibalistic factors might represent promising new antibiotic leads[14-15].

Prior to this study, SDP had been partially purified[6] while SKF had not been identified or structurally characterized, although *B. subtilis* is the model organism for Gram-positive bacteria and many laboratories had investigated its metabolic output. The difficulties of identifying these molecules could arise from the fact that the cannibalistic effect was only observed on solid media, but not in liquid media[6]. Therefore, we decided

to use imaging mass spectrometry (IMS) to visualize the process by growing the domesticated strain PY79 directly on the matrix-assisted laser desorption ionization (MALDI) target plate[16-18]. With this approach, we were able to purify and determine the structure of mature SKF and SDP. In solid medium, both molecules act as cannibalistic killing factors, however, only SDP inhibited the growth of *B. subtilis* and the gram positive pathogens *S. aureus* and *S. epidermidis* in liquid medium. This investigation demonstrates that IMS is an effective tool to identify cell-to-cell interaction signals and provides an approach to the discovery of bioactive molecules.

2.3 Results and Discussion

2.3.1 Identification of SKF and SDP via imaging mass spectrometry

The *B. subtilis* cannibalistic phenotype only presents itself on solid media[6], yet there are very few tools that can spatially characterize metabolic output on solid surfaces. We therefore used thin-layer agar IMS, which is capable of capturing the spatial distribution of metabolites in growing colonies and can be used to understand microbial metabolic exchange, to visualize the cannibalistic factors SKF and SDP[18]. In this IMS experiment, a growing culture of *B. subtilis* PY79, used in the original report on cannibalism[6], was cultured adjacent to a Δ*spo0A* strain (KP648) and its metabolic output was profiled by IMS (**Fig. 2.1a**). We anticipated that a zone of clearing on the Δ*spo0A* strain would be observed and that IMS could be used to identify candidate signals involved in the lysis of Δ*spo0A* cells. Indeed we observed a decreased growth phenotype as well as a glassy appearance in the region of Δ*spo0A* adjacent to the co-cultured PY79 cells (**Fig. 2.1a**, **Fig. 2.2**).

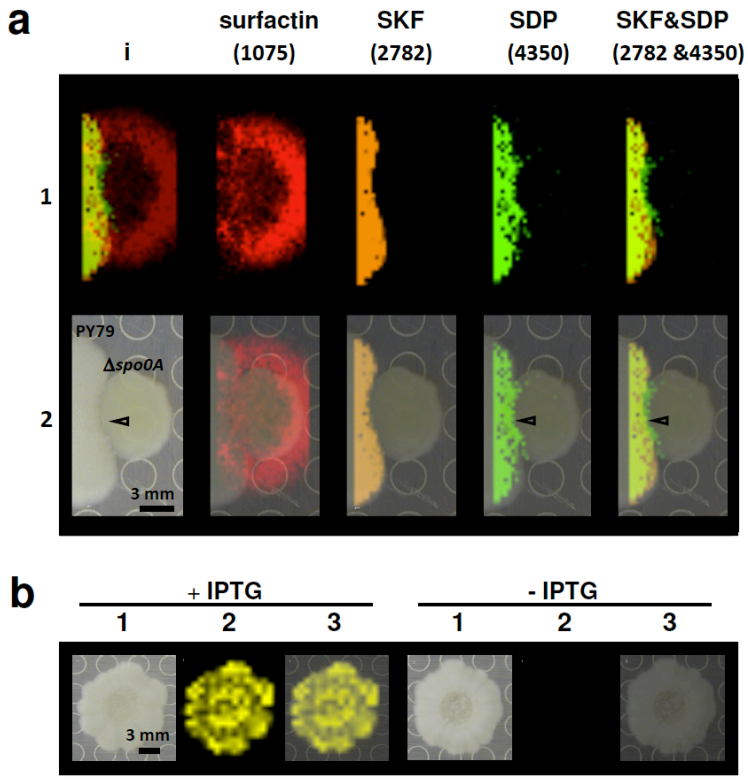


Figure 2.1: Imaging mass spectrometry of the intraspecies metabolic exchange. (**a**) Imaging mass spectrometry of PY79 and Δ*spo0A* (KP648) co-culture. 1) Ion distributions observed for surfactin ($[M+K]^+$), SKF ($[M+H]^+$), and SDP ($[M+K]^+$). 1i is a superposition of all ions. 2) Superposition of the photograph and IMS data. The arrows point to the glassy region of Δ*spo0A* and SDP overlap. (**b**) Imaging mass spectrometry of EG208 with or without IPTG. 1) photographs of the colonies; 2) imaging MS pictures of ion 2782; 3) superposition of 1 and 2.

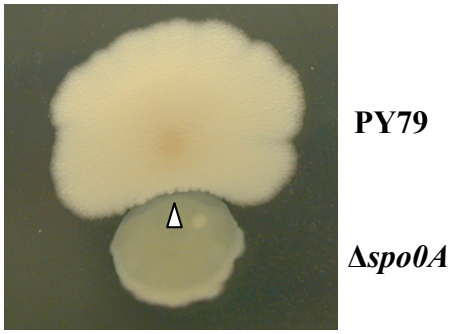


Figure 2.2: PY79 inhibits Δ*spo0A* (KP648) strain. The arrows (Δ) indicate the glassy region in the Δ*spo0A* colony.

Several ions were observed in the IMS data. First, KP648 (PY79, Δ*spo0A*) produced the non-ribosomal peptide synthetase derived lipopeptide antibiotic surfactin. This is unexpected because PY79 has a frameshift mutation in the gene encoding for the phosphopantetheinyl transferase protein, Sfp[19], and we confirmed by DNA sequencing that this mutation is present in the Δ*spo0A* strain. This indicates that the surfactin non-ribosomal peptide synthetases are activated via a lower efficiency phosphopantetheinylation pathway that is upregulated in the absence of Spo0A. However, judging from their relative mass spectrometry intensities, the signal of surfactin at 96 hours was at least 10 fold less than the amount produced by the non-domesticated *B. subtilis* strain 3610 (**Fig. 2.3**), which has a functional *sfp* gene. Second, some signals produced by PY79 appeared to define a boundary between the PY79 and Δ*spo0A* colonies, including the ions at *m/z* 2782 and 4350 (**Fig. 2.1a**). Ion 2782 was associated with the border between the two cultures and ion 4350 was found in the regions where the Δ*spo0A* strain was glassy and displayed reduced growth. Thus the 4350 *m/z* ion was the favored candidate that caused the majority of the cannibalistic killing effect, because of the overlap with the zone of reduced growth on Δ*spo0A*, while the 2782 ion stopped at the interface of the two colonies.

After an *n*-butanol extraction and a desalting step, the two uncharacterized ions were measured at *m/z* 2782 $[M+H]^+$ and 4312 $[M+H]^+$ suggesting that ion *m/z* 4350 is the potassium adduct form $[M+K]^+$ (**Fig. 2.4**). These two ions were then subjected to fragmentation by MALDI time-of-flight tandem mass spectrometry (TOF/TOF). The ion at *m/z* 4312 $[M+H]^+$ gave a long and unambiguous sequence tag corresponding to VAAGYLYVVGVNAVALQTAAAVTTAVW and matched residues from Val148 to

Trp174 of SdpC (**Fig. 2.5a**, **Fig. 2.6**, and **Table 2.1**) while the TOF/TOF fragmentation of the 2782 ion gave a candidate sequence tag, LPHPA (**Fig. 2.5b**). This sequence tag matched a sequence within SkfA, the proposed precursor for the mature form of SKF[6]. However, the sequence tag was insufficient to positively identify SKF because none of the ion masses could be directly matched with the linear sequence of SkfA and many ions in the spectrum remained unexplained. To confirm the identity of this ion, strain EG208 with an IPTG inducible promoter in front of the *skf* gene cluster was subjected to IMS. If the ion with *m/z* of 2782 is SKF, it is expected only to be present when IPTG is added. Indeed, in the absence of IPTG, the ion at *m/z* 2782 is absent as judged by IMS (**Fig. 2.1b**). In addition, the production of this ion was abolished when the *skfA* gene was inactivated, solidifying that this 2782 ion is indeed the mature form of SKF (**Fig. 2.3**, EG165). Similarly, the 4350 signal was not observed in an *sdpC* deletion strain, in agreement with the identification of this ion as SDP (**Fig. 2.3**, EH273).

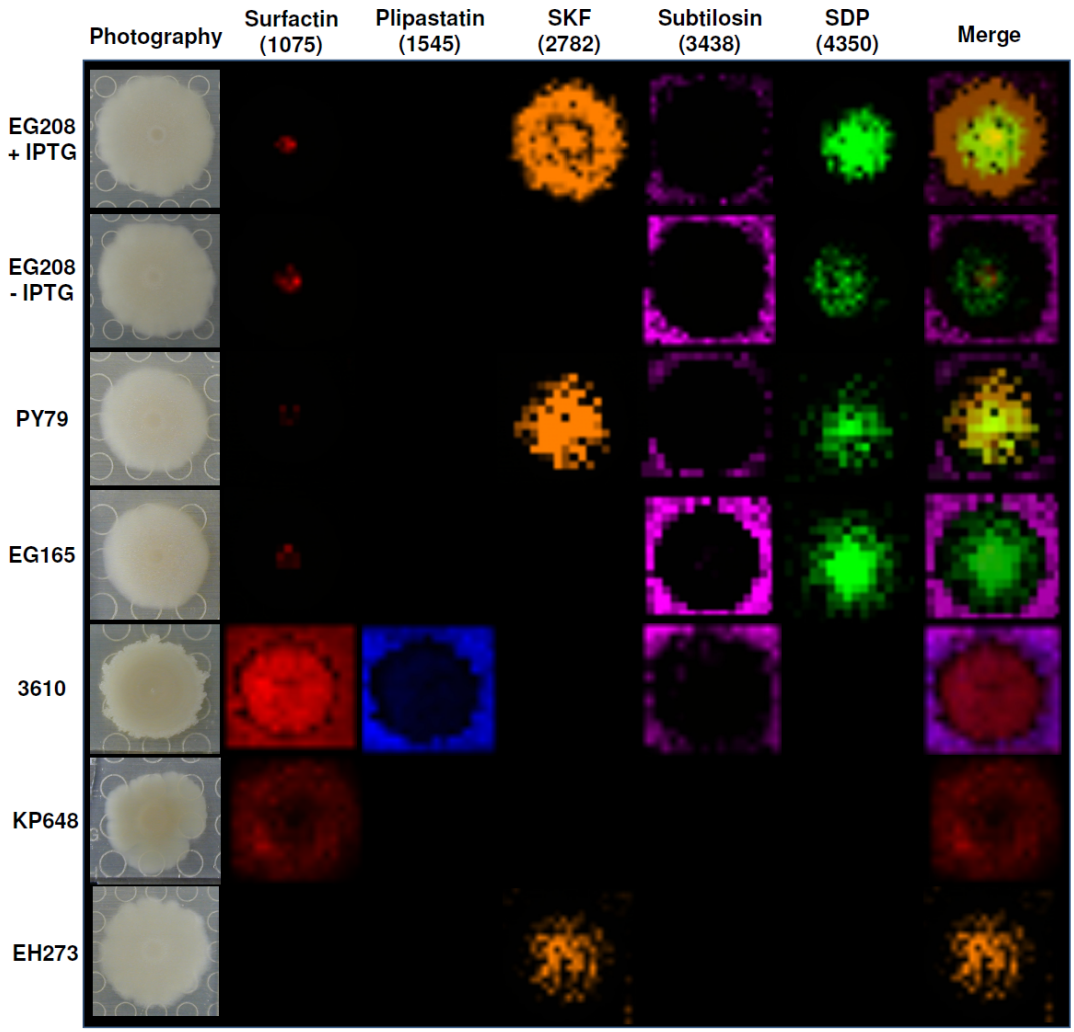


Figure 2.3: Metabolic profile of the strains used in this study. Strains were inoculated on an ISP2 agar plate and allowed to grow for 96 hours at room temperature and subjected to thin-layer agar IMS. The number labeled below each metabolite is the representative ion observed (*m/z*). *m/z* 1045, 1545, 3438 and 4350 are potassium adduct form $[M+K]^+$. For brief genotype description, 3610 is a wild type strain; PY79 is a laboratory domesticated wild type; EG208 contains the *skf* gene cluster under control of an IPTG inducible promoter; EG165, KP648, EH273 are *skfA, spo0A, sdpABC* deletion strains, respectively. All four mutants are constructed under PY79 background. The ion intensity was reflected by the intensity of colors. Each column of ions was displayed using same intensity scale optimized per each metabolite. The scale range in normalized relative ion intensity for each ion was specified in FlexImaging 2.0 (Bruker) as follows: surfactin 10%-100%; plipastatin 10-80%; SKF 30-60%; subtilosin 20-60%; SDP 1-30%.

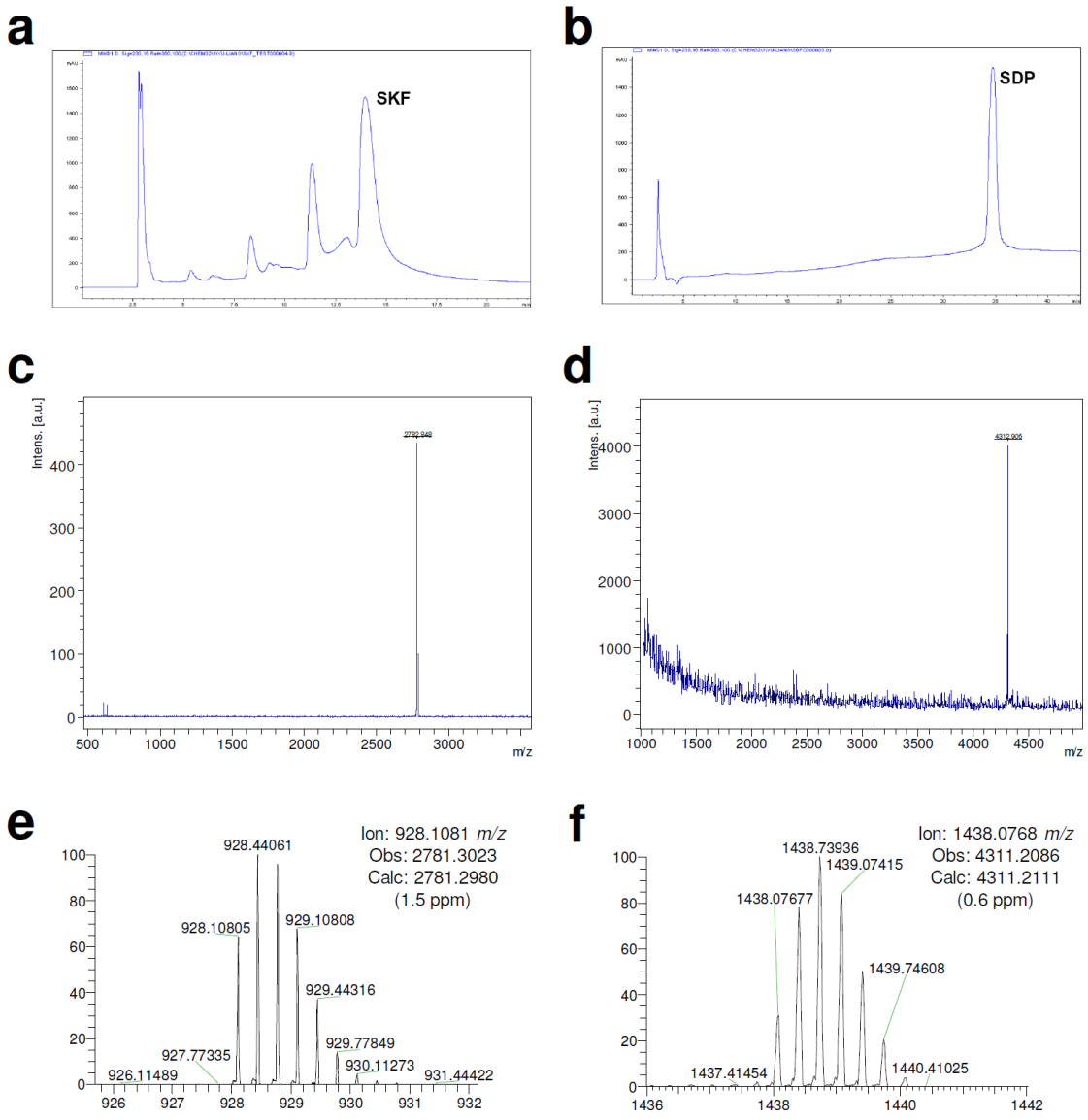


Figure 2.4: SKF and SDP purification. a, b) Crude SKF and SDP fractions obtained from an initial separations using Sephadex LH-20 column were further purified by HPLC (solvent system and gradients were specified in materials and methods section). UV 230 nm was used to detect SKF and SDP. SKF and SDP were eluted at 14, and 34.5 min, respectively. **c)** MALDI-TOF MS spectrum of purified SKF. **d)** MALDI-TOF MS spectrum of purified SDP. **e)** FT-ICR MS spectrum of purified SKF. **f)** FT-ICR MS spectrum of purified SDP.

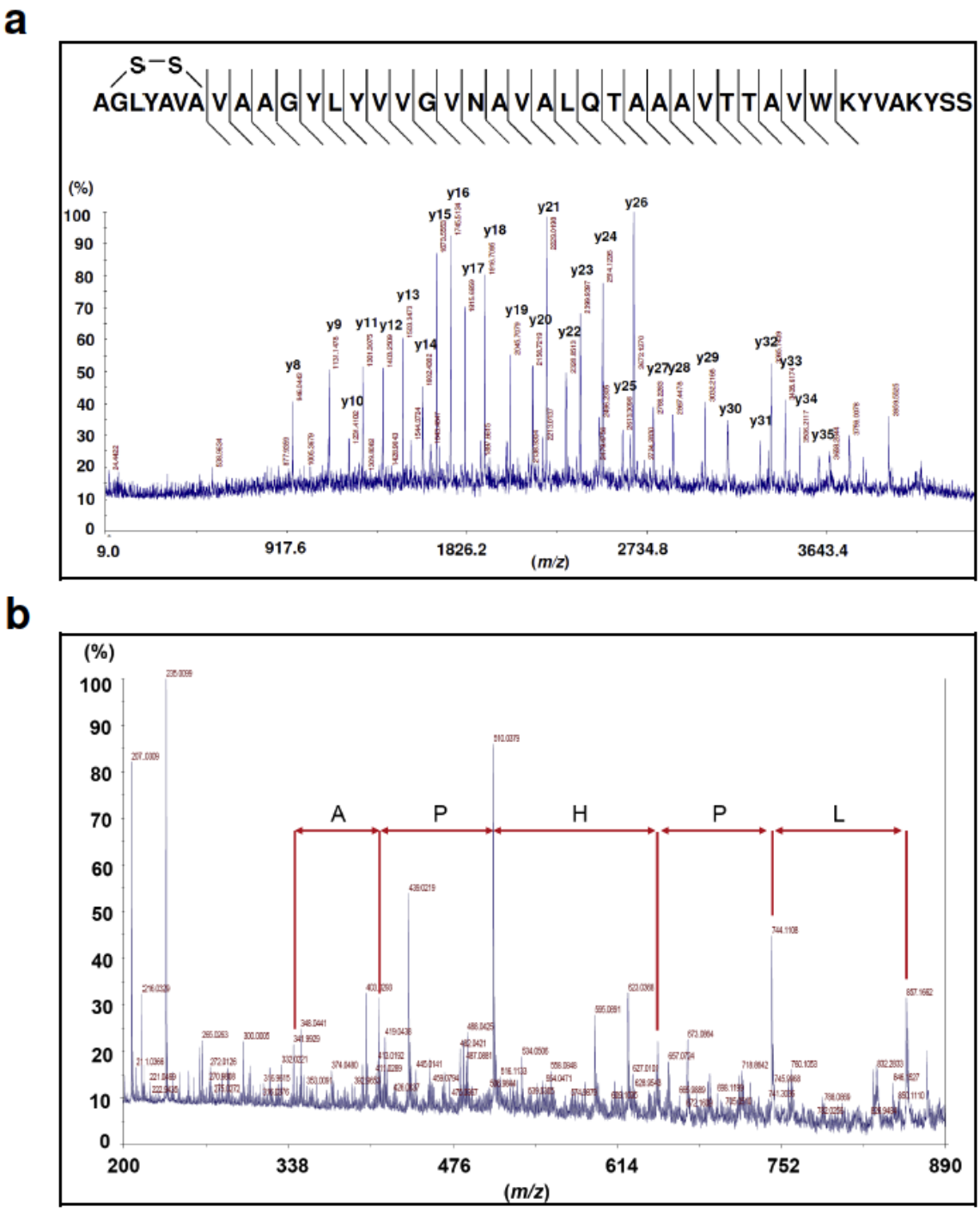


Figure 2.5: Intact cell MALDI TOF/TOF spectrum. a) Intact cell MALDI TOF/TOF spectrum of ion at *m/z* 4312. The corresponding ion table is showed in Table 2.2. **b)** Intact cell MALDI TOF/TOF spectrum of ion at *m/z* 2782. The observed sequence tag that matched to SkfA is shown.

a)

```
atgaaaagtaaattacttaggctattgattgtttccatggtaacgatattggtttttca
ttagtaggactctctaaggagtcaagtacatctgctaaagaaaaccatacattttctgga
gaagattactttagaggacttttatttggacaaggggaagttggtaaattaatttcaaac
gatttggaccctaaactcgtaaaagaggcaaatagtacagaaggtaaaaagttagtaaat
gatgtagtcaaatttataaaaaaagatcaaccacaatatatggatgaattgaaacaatcg
attgacagcaaagaccctaaaaaactcattgaaaatatgaccaaagcagaccaacttatc
caaaaatatgctaagaaaaatgaaaacgtaaaatactcttctaataaagttactccatct
tgtgggctttatgccgtctgtgtagcagctggatatttatatgttgtgggcgttaacgca
gttgcattacaaacggctgccgcagtaacaactgcagtgtggaaatacgttgccaaatat
tcctcttcagcttctaataattctgatttagaagcggctgctgcaaaaaccctaaaattg
attcatcaataa
```

b)

```
        10         20         30         40         50         60
MKSKLLRLLI VSMVTILVFS LVGLSKESST SAKENHTFSG EDYFRGLLFG QGEVGKLISN
        70         80         90        100        110        120
DLDPKLVKEA NSTEGKKLVN DVVKFIKKDQ PQYMDELKQS IDSKDPKKLI ENMTKADQLI
       130        140        150        160        170        180
QKYAKKNENV KYSSNKVTPS CGLYAVCVAA GYLYVVGVNA VALQTAAAVT TAVWKYVAKY
       190        200
SSSASNNSDL EAAAAKTLKL IHQ
```

Figure 2.6: a) *sdpC* gene sequence. b) SdpC protein sequence. Mature SDP is highlighted in yellow. A disulfide bridge exists between C141 and C147.

Table 2.1: Annotation of ion *m/z* 4312.6 intact cell MALDI TOF/TOF MS spectra.

Seq.	Ion	Calc. mass	Obs. Ion	Obs. Mass	Error (Da)	Seq.	Ion	Calc. mass	Obs. Ion	Obs. Mass	Error (Da)
C*	y42	4311.21				**A**	y21	2228.18	2229.02	2228.01	0.17
G	y41	4210.22				**L**	y20	2157.15	2158.72	2157.71	0.57
L	y40	4153.20				**Q**	y19	2044.06	2045.71	2044.70	0.64
Y	y39	4040.11				**T**	y18	1916.00	1916.71	1915.70	0.30
A	y38	3877.05				**A**	y17	1814.96	1815.69	1814.68	0.28
V	y37	3806.01				**A**	y16	1743.92	1745.51	1744.51	0.59
C*	y36	3706.94				**A**	y15	1672.88	1673.56	1672.55	0.33
V	y35	3603.93	3605.58	3604.57	0.64	**V**	y14	1601.85	1602.44	1601.43	0.41
A	y34	3504.87	3506.21	3505.20	0.34	**T**	y13	1502.78	1503.35	1502.34	0.44
A	y33	3433.83	3435.62	3434.61	0.78	**T**	y12	1401.73	1403.25	1402.24	0.51
G	y32	3362.79	3365.74	3364.74	1.95	**A**	y11	1300.68	1302.31	1301.30	0.62
Y	y31	3305.77	3307.02	3306.01	0.24	**V**	y10	1229.64	1231.41	1230.40	0.76
L	y30	3142.71	3143.85	3142.84	0.14	**W**	y9	1130.58	1131.15	1130.14	-0.44
Y	y29	3029.62	3032.22	3031.21	1.59	**K**	y8	944.50	946.04	945.034	0.54
V	y28	2866.56	2867.45	2866.44	0.12	**V**	y7	816.40			
V	y27	2767.49	2768.23	2767.22	0.27	**Y**	y6	653.34			
G	y26	2668.42	2669.21	2668.20	0.22	**A**	y5	554.27			
V	y25	2611.40	2613.31	2612.30	0.90	**K**	y4	483.23			
N	y24	2512.33	2514.12	2513.12	0.78	**Y**	y3	355.14			
A	y23	2398.29	2399.94	2398.93	0.64	**S**	y2	192.07			
V	y22	2327.25	2328.85	2327.84	0.59	**S**	y1	105.04			

2.3.2 Isolation and structural elucidation of SDP and SKF

With the masses of SKF and SDP in hand it became possible to use a mass spectrometry-guided isolation for both molecules. Unlike the antibiotics surfactin and subtilosin, SDP and SKF did not readily diffuse into the solid media as judged by IMS (**Fig. 2.1**, **Fig. 2.3**). This implied that they are hydrophobic in nature, and thus we adapted our purification protocol accordingly (see materials and methods section). The isolation and subsequent structural analysis using tandem mass spectrometry and NMR enabled us to determine that SDP is a 42 amino acid peptide with a disulfide crosslink and that SKF is a 26 amino acid disulfide-containing cyclic peptide with a thioether crosslink of a cysteine to the α-carbon of a methionine (**Fig. 2.7**). This ribosomally encoded peptide is unusual in terms of structure but is consistent with the transport, and biosynthetic enzymes found on the *skf* gene cluster as described in the supporting information[6]. A full description of the data and methods that led us to the determination of these structures is provided in the supplementary information.

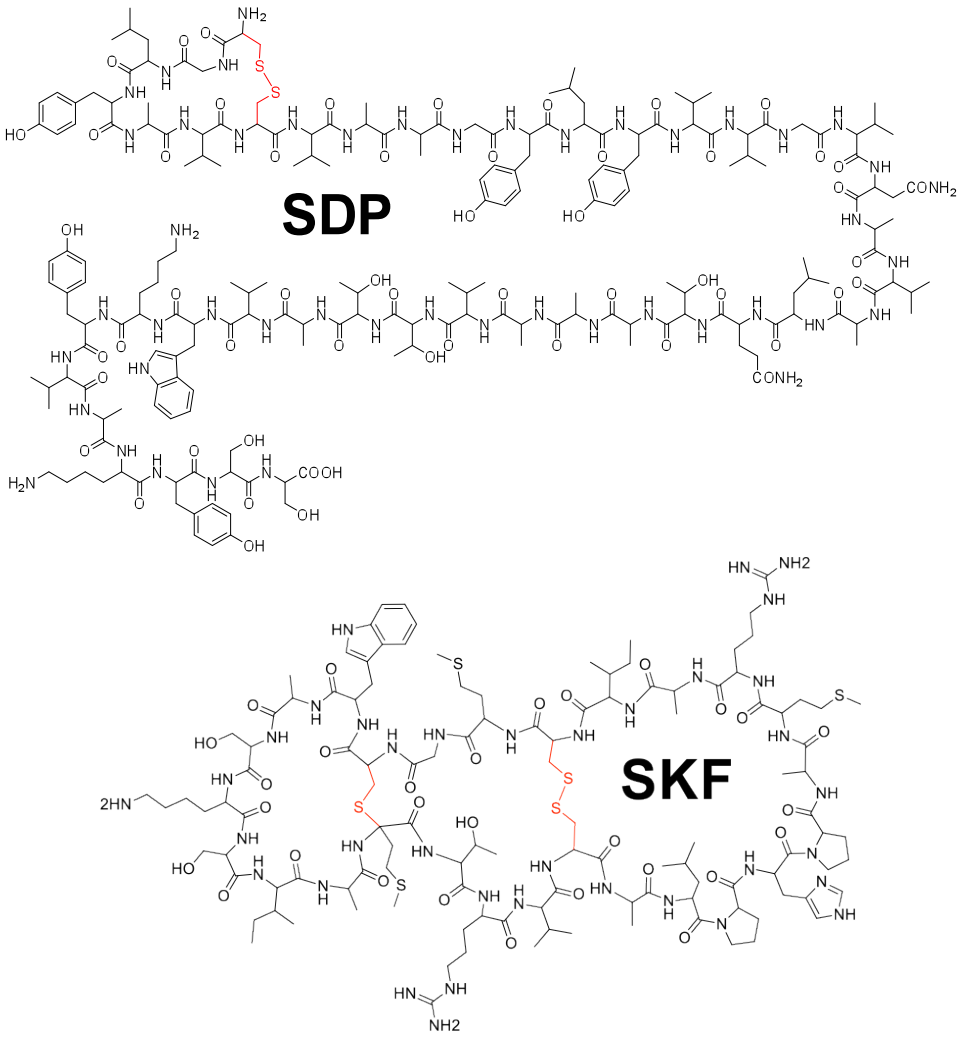


Figure 2.7: The structures of SKF and SDP

2.3.3 Biological effects of SDP and SKF

With the availability of microgram quantities of purified SKF and SDP, the biological effects of these compounds on *B. subtilis* growth in liquid and solid cultures, and the effects on cell structure were evaluated. First, we added the purified compounds to liquid cultures of the undomesticated wild-type strain 3610, its domesticated laboratory descendent PY79, and a PY79 strain containing the Δ*spo0A* mutation (KP648). 3610 was included to verify the relevance of our findings to an undomesticated strain of *B. subtilis*.

In rich media (ISP2 or LB media), 20 mg/mL purified SDP significantly and rapidly inhibited growth of 3610, PY79, and Δ*spo0A,* while 20 mg/mL of SKF, surprisingly, had little observable effect on growth (**Fig. 2.8**). PY79 was much less affected by the addition of SDP than the other strains, presumably because PY79 produces SKF and SDP during growth and likely expresses the resistance genes, whereas we were unable to detect the compounds in the Δ*spo0A* mutant under any conditions tested and 3610 produced low levels of the compounds only at late times on LB and DSM (**Fig. 2.3**, **Fig 2.9**).

To determine the growth inhibitory activity of SDP in more detail we investigated the effects of different concentrations of SDP on growth of the Δ*spo0A* strain KP648. A concentration-dependent growth effect was observed (**Fig. 2.10a**). Upon addition of 5 μg/ml to 20 μg/ml SDP, growth was halted but was able to resume after several hours of continued incubation. The recovery represents survival of a subpopulation of SDP-resistant cells that are able to resume growth after a lag period (see below). The concentration of SDP significantly affected the degree to which growth was inhibited and the length of the growth lag. To evaluate how rapidly SDP inhibited growth, SDP was added to an exponentially growing culture. The addition of 20 mg/mL SDP at different time points rapidly caused growth to level off with very little decrease in optical density (**Fig. 2.10b**). These results demonstrate that purified SDP has rapid effects on growth of *B. subtilis* cells, but that SKF does not at the highest concentration we tested (20 μg/ml).

We next performed fluorescence microscopy on living cells of 3610, PY79 and the Δ*spo0A* derivatives of these strains (ALB1035 and KP648, respectively) following treatment with 20 mg/mL SDP for various times in liquid culture. Cells were stained with FM 4-64, a fluorescent membrane stain that inserts into the outer leaflet of the bilayer, as

well as Sytox Green and DAPI, two DNA stains that do not efficiently cross the bilayer unless the cells are permeabilized (**Fig. 2.10c**). The first effect was noted approximately 60 min after the addition of SDP to 3610, when we observed that the cells often showed partial, asymmetric septa and localized bright staining of membranes (arrows), suggesting the presence of deformations in the cytoplasmic membrane, particularly at division sites. Approximately 5% of the cells at this time also showed increased permeability to Sytox Green and DAPI, and a subset of these cells showed large gaps in the membrane staining (double arrowheads), suggesting that these cells lack an intact cytoplasmic membrane. By 120 min, more cells showed increased permeability to Sytox Green, DAPI and discontinuous membranes, and we observed many spherical and tubular membrane projections (arrowheads). By 120 min, ~33% of all morphologically intact 3610 cells stained with Sytox Green indicating they are permeabilized and 5% of the cells had protruding tubules (**Table 2.2** and **Table 2.3**). By 300 min, very few intact cells remained in the Δ*spo0A* culture, whereas the PY79 culture still contained dividing cells, confirming the increased sensitivity of the Δ*spo0A* mutant to SDP. Time-lapse microscopy revealed that the membrane tubules were formed and released in a matter of seconds (**Fig. 2.10d**).

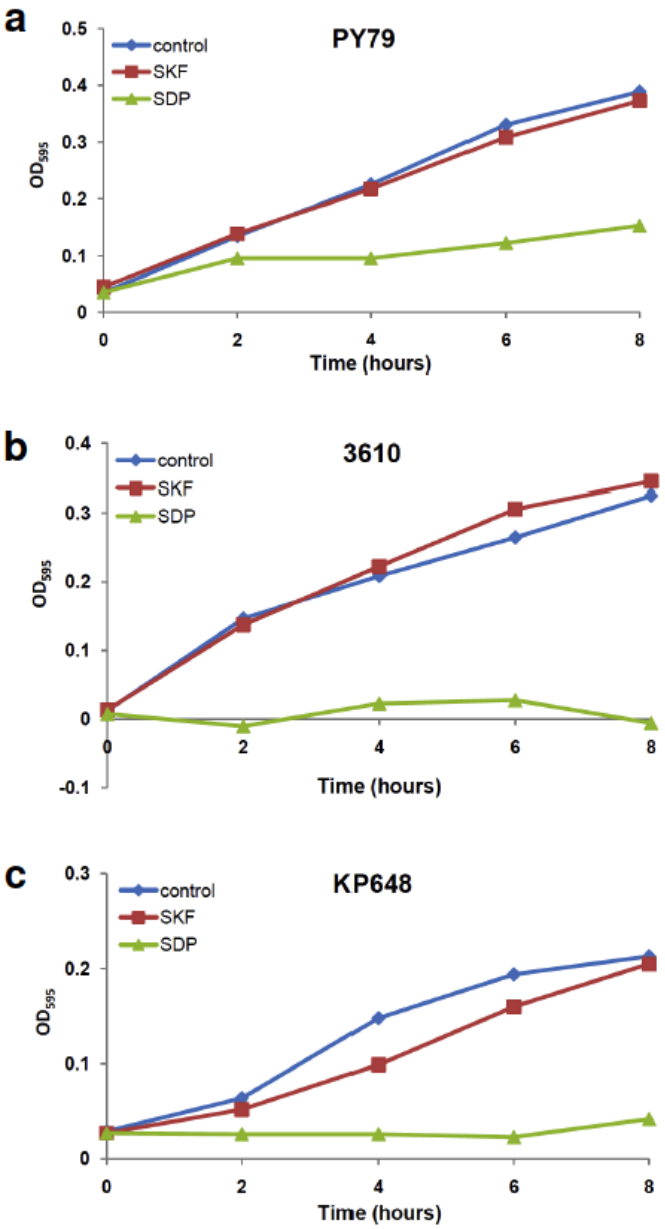


Figure 2.8: The effects of SDP and SKF on growth of *B. subtilis* strains in ISP2 media. The overnight culture was diluted into ISP2 media to OD_{595} 0.03, and SKF or SDP were added into 100 µL of diluted cultures to a final concentration of 20 µg/mL. The plate was shaken at 37 ^{o}C, 120 rpm. The OD_{595} was measured at each time point.

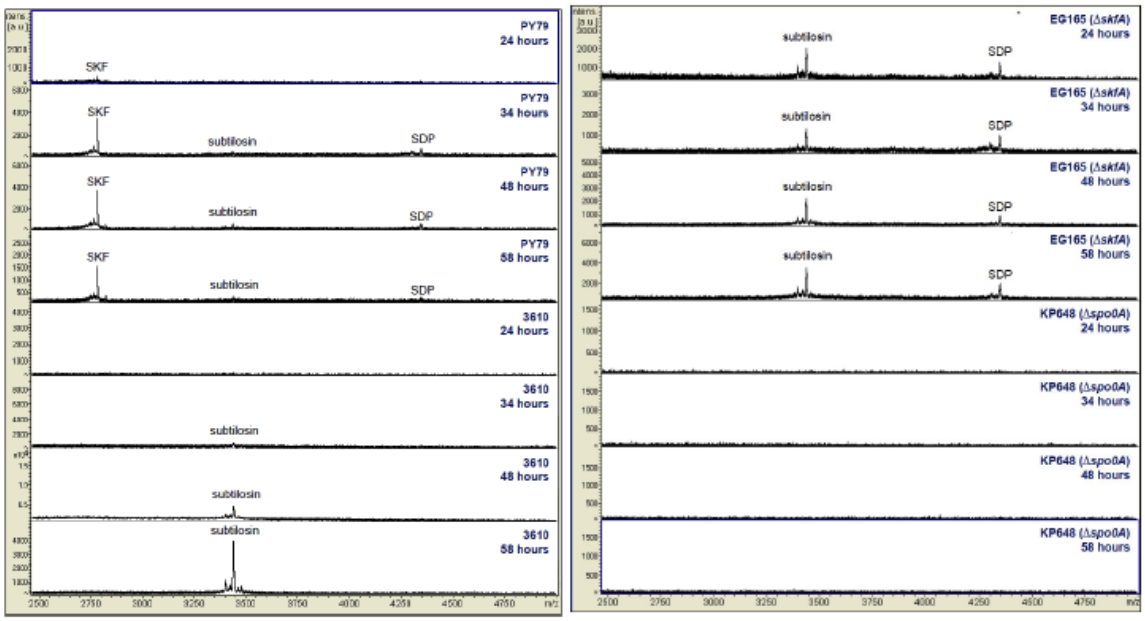


Figure 2.9: Time course of *B. subtilis* wildtype and mutant strains collected by intact cell MALDI-TOF MS[37,38]. Each strain was allowed to grow on ISP2 solid agar for 24, 34, 48 and 58 hours at 28 ºC. Each spectrum shown is an average of 200 single spectra. EG165 and KP648 are *skfA* or s*po0A* gene deletion strains, respectively. The metabolic outputs for 96 hours cultures are shown in Fig. 2.3.

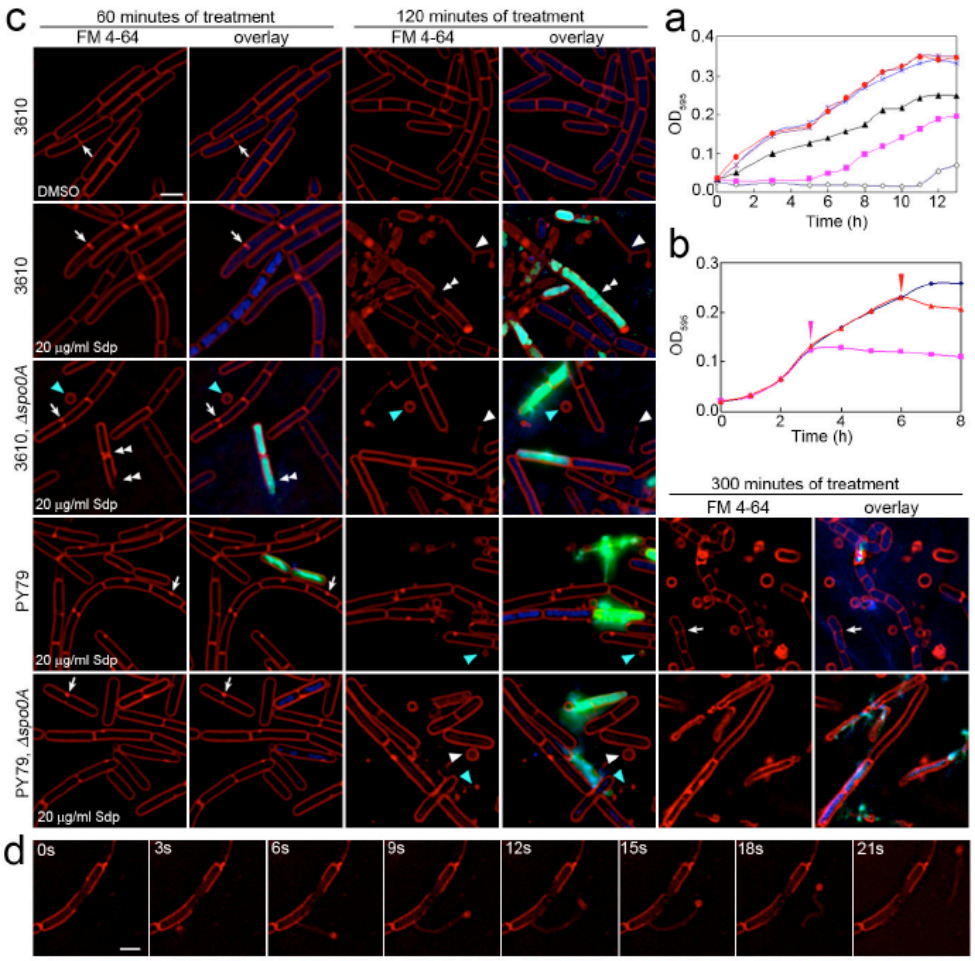


Figure 2.10: Biological effects of SDP on *B. subtilis*. (**a**) Growth curves of KP648 (D*spo0A*) in ISP2 media with various concentration of SDP. SDP was added at 0 h. –◇– is 20 mg/mL; –■– is 10 mg/mL; –▲– is 5 mg/mL; –✳– is 2 mg/mL; –✳– is 0.2 mg/mL; –●– is the DMSO control. (**b**) Growth curves of KP648 (D*spo0A*) in ISP2 media with 20 mg/mL of SDP. SDP was added at 3 h (–■–) and 6 h (–▲–). –◆– is the DMSO control. (**c**) Fluorescence micrograph of growing cells of 3610, PY79, ALB1035 (3610, D*spo0A*), and KP648 (PY79, D*spo0A*) treated with DMSO or 20 mg/mL SDP for the time indicated. The red stain is FM 4-64, a fluorescent membrane stain; the blue and green stains are DAPI and Sytox Green, two DNA stains that are membrane impermeable. Sytox Green is the least permeable and provides the greatest increase in fluorescence in permeabilized cells. The white arrows point to dividing cells. In the DMSO control the arrow points to a normal division, while in the other images the septa are asymmetric; the double arrowheads point to large gaps in membrane staining; the light blue triangles point to membrane spheres while the white triangles point to tubular membranes. After 300 minutes of treatment of PY79 with SDP, surviving cells are smaller, dividing and impermeable to Sytox. (**d**) Timelapse microscopy collected at 110 minutes after SDP treatment. Still images from 0 ~21 seconds are shown.

Table 2.2: Sytox Green cell permeability over time of SDP treatment.

Average Percent Permeabilized Cells[1] ±SD[2] (# of cells scored in each experiment)				
Timepoint	**3610[3]**	**ALB1035[3]**	**PY79**	**KP648**
30 min DMSO	1.9% (313)	0% (111)	0.6% (353)	0.3% (292)
60 min DMSO	1.1% (379)	0.7% (153)	0.8%±0.3% (186,313,318,260)	1.1%±0.1% (284,437)
90 min DMSO	2.7% (331)	0.3% (337)	0.5%±0.5% (247,320,326,435)	0.6%±0.4% (326,350)
120 min DMSO	0.3% (290)	0% (265)	0.5%±0.5 (166,329,365,321)	1.2%±1.6% (225,300)
30 min SDP	10.3% (339)	1.3% (156)	0.6% (174)	2.4% (252)
60 min SDP	4.7% (342)	7.5% (308)	4.2%±2.9% (224,250,313,215)	3.5%±1.5% (309,367)
90 min SDP	28.0% (336)	16.3% (1295)	7.5%±2.2% (271,375,290,304)	13.6%±15.1% (274,305)
120 min SDP	32.8% (344)	34.3% (507)	13.7%±1.5% (134,248,203,307)	19.8%±5.8% (301,255)

[1]The percentage of permeabilized cells for individual experiments were calculated and then averaged to generate the average percent permeabilized cells. Cell debris was not scored unless it clearly was derived from a single cell. Membrane spheres and tubules were not counted as cells.

[2]The standard deviation was calculated based on the percent permeabilized cells in repeated experiments.

[3]Experiments on these strains were performed once, so no standard deviation is available.

Table 2.3: Membrane staining irregularities in strain 3610 after 120 minutes of SDP treatment.

Cell type	**% cells (# scored)**
Intact cells[1]	64% (314)
Irregular membrane	29% (140)
Clear gaps in membrane	7% (36)
Tubules	5% (27)[2]
Spheres	5% (26)[2]

[1]Includes permeabilized cells that had no obvious membrane deformation.

[2]Membrane tubules and spheres are subcellular particles, so the scoring indicates the frequency with which such structures were observed relative to the total number of cells scored.

The domesticated strain PY79 and its Δ*spo0A* derivative (KP648) responded more slowly to SDP, with only ~4% of cells showing increased Sytox Green permeability at 60 min after treatment and with major changes in cell morphology first observed 90 min after treatment. At 120 min, 13.7% of all morphologically intact PY79 cells stained with Sytox and increased to 19.8% in the PY79, Δ*spo0A* strain (**Table 2.2**). These results confirm our initial hypothesis that the 4350 *m/z* ion, due to the overlap with the decreased growth phenotype of Δ*spo0A* in the co-culture of Δ*spo0A* with PY79, is the major cannibalistic factor. Further, the data indicate that SDP does not rapidly lyse all the Spo0A-OFF cells and that a subpopulation of Δ*spo0A* cells remains viable even after several hours of treatment, suggesting an additional degree of multicellular behavior in a population of genetically identical *B. subtilis* cells.

2.3.4 Dual nature of SKF- and SDP-mediated killing of sister cells

Having failed to detect any biological affect of purified SKF using the above liquid culture assays, we wanted to evaluate the ability of SKF and SDP to work independently on solid media. To assess this, we set out to determine if purified SDP or SKF inhibited the growth of Δ*spo0A* on solid media. Spotting 2 μg of SDP resulted in a large zone of decreased growth and 2 μg of SKF resulted in a smaller zone in which the lawn appeared somewhat thinner, while the control 10% DMSO did not have an effect on the growth (**Figure 11a**). This indicates that both SDP and SKF reduce the growth of *B. subtilis* on solid media, although purified SDP has a much stronger effect.

We next sought to determine if SDP and SKF produced independent killing effects of similar magnitude when overproduced on solid medium. To do so, we used a

spot assay in which PY79, Δ*spo0A,* Δ*skf*, Δ*sdp* and SKF or SDP overexpressing strains are spotted on *B. subtilis* lawns. When PY79 was spotted on the Δ*spo0A* lawn, a large zone of clearing was observed (**Fig. 11b**), as previously reported[6]. This phenomenon is mostly dependent on SDP since the inhibitory effect was still observed with a Δ*skf* strain, but was abolished when *sdpABC* genes were deleted or not induced (**Fig. 11b**). We next overexpressed SKF in an *sdpABC* deletion background to eliminate the effect of SDP (P_{hyspac}*skf,* Δ*sdpABCIR*), and observed a killing effect towards lawns of PY79, Δ*skf* and Δ*spo0A* (**Fig. 11c**). These results indicate that although most of the killing effect of PY79 on a Δ*spo0A* lawn is mediated by SDP rather than SKF, over-expression of the *skf* operon still mediates a killing activity. Thus, either SKF or SDP can independently mediate cannibalism on solid culture medium, but SDP is much more potent than SKF.

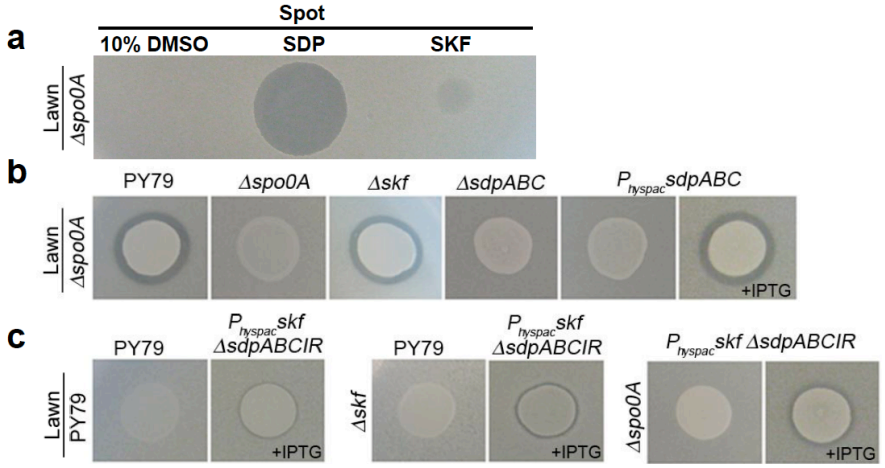


Figure 2.11: Spot assays to compare the effect of exogenously supplied and endogenously produced SDP and SKF. Lawns of the indicated strains were prepared in top agar. After solidification the lawns were spotted with either (A) purified SDP, SKF and DMSO or (B, C) the indicated strains that overexpress or lack SDP and SKF.

2.3.5 SDP but not SKF has antibacterial activities against pathogens

The above results indicate that both SKF and SDP mediate cannibalistic effects. We therefore set out to determine if these two molecules would also display activity against human pathogens by screening purified SKF and SDP for inhibitory activity against a panel of pathogenic microbes (including *B. subtilis* strains PY79 and its Δ*spo0A* derivative for comparison). We employed a growth assay that measures differences in cell density compared to untreated controls. This screen revealed that SKF had no effect on growth, while SDP decreased cell density of members of the Gram-positive genus *Staphylococcus* to a greater extent than *B. subtilis,* but it did not affect the tested Gram-negative pathogens *Pseudomonas aeruginosa* or *Klebsiella pneumoniae* (**Fig.12**). SDP exhibited potent inhibitory activity against two *S. aureus* variants, the Newman strain used extensively in laboratory studies of *S. aureus* virulence, as well as clinical isolate of methicillin-resistant *Staphylococcus aureus* (MRSA) sequence type ST59. The IC_{50} against these *S. aureus* strains were 210 and 110 nM, respectively, slightly more active than the leading contemporary pharmacological agent for treatment of MRSA infection, vancomycin (IC_{50} 360, 270 nM, respectively). SDP also inhibited growth of *S. epidermidis*, an opportunistic pathogen associated with nosocomial infections of catheters and the urinary tract and invasive infections in human premature neonates (IC_{50} 990 nM). SDP has a relatively simple chemical structure, and might therefore provide an antibiotic candidate for future development of derivatives with smaller size and optimized activity against MRSA and closely related species.

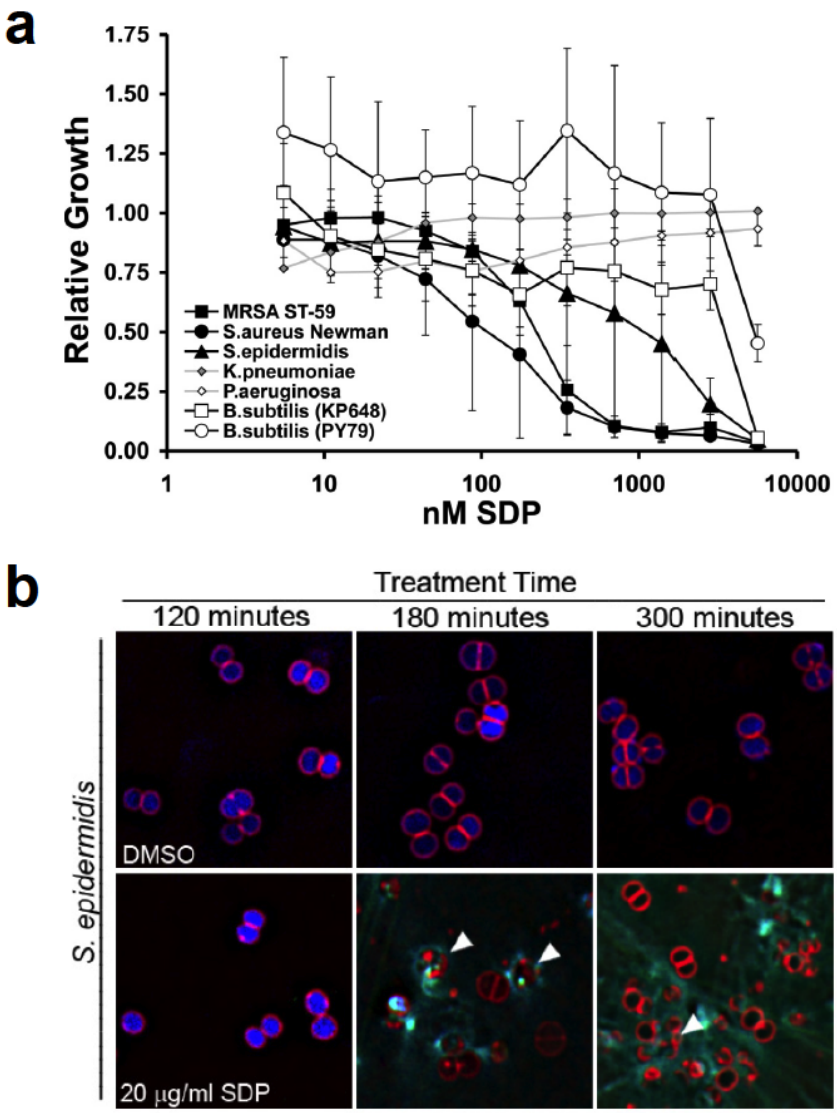


Figure 2.12: Biological activity of SDP on clinically relevant human pathogens. (a) SDP inhibition curves for pathogenic microbes. Relative growth of *P. aeruginosa*, *K. pneumonia*, *B. subtilis* strains KP648 (Δ*spo0A)* and PY79, *S. aureus* Newman strain, a clinical isolate of methicillin resistant *S. aureus*, and *S. epidermidis* with the presence of increasing concentrations of SDP is shown in the curve. (b) Fluorescence micrograph showing growing cells of *S. epidermidis* treated with DMSO or 20 μg/mL SDP in DMSO for the time indicated and stained with FM 4-64 (red), DAPI (blue) and Sytox (green) as in Fig 2.10.

2.4 Summary

In summary, this paper highlights the need for the development of new technologies to study and discover biologically active molecules, a cornerstone of chemical biology as well as therapeutic discovery[20]. Thin-layer agar MALDI-IMS of co-cultures of bacterial colonies enabled the discovery, isolation and structural elucidation of two biologically active factors SKF and SDP, one of which was cyclized and uniquely post-translationally modified with a thioether linkage to the α-position of a methionine. MALDI-IMS can also be used to predict the function of metabolites. In this study we observed an overlap of SDP extending from the co-cultured strain PY79 with the region of Δ*spo0A* that was inhibited, enabling us to formulate a hypothesis that SDP, but not SKF, was the main cannibalistic factor. Indeed when we tested the biological activities of the purified compounds, only SDP showed inhibitory activity in liquid cultures. On the other hand both SDP and SKF mediated inhibition on solid cultures when over-expressed in whole cells grown on solid medium or when the purified compounds were spotted onto solid media. However, SKF was much less potent than SDP and our data indicate that SDP is the primary toxin that mediates cannibalism. Finally, screening both SKF and SDP for their anti-bacterial activities allowed us to demonstrate that SDP, but not SKF, inhibit growth of *Staphylococci* to a greater extent than it inhibits *B. subtilis* growth. This suggests that SDP also participates in defensive or predatory behavior directed at other species[14-15], as well as cannibalism[6]. The fact that SDP inhibited clinically relevant pathogens also demonstrates a novel application of studying metabolic exchange by imaging mass spectrometry in the discovery of biologically active molecules and

therefore represents that IMS may be used as a tool in the efforts to discover new classes of therapeutic agents.

2.5. Supporting information

2.5.1 Structural elucidation of SDP and SKF

Purified SDP was subjected to high resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) to obtain a monoisotopic mass *m/z* 4311.209 $[M+H]^+$ (**Fig. 2.4**), and this mass matched Cys141-Ser182 of SdpC-2.016 Da, indicating the likely presence of a disulfide crosslink. The fragmentation by collision-induced dissociation (CID) and FT-ICR MS analysis confirmed the sequence with a disulfide crosslink localized to residues Cys141 and Cys147 (**Fig. 2.7**, **Fig. 2.6**, **Table 2.4**). The N-terminal boundary of SDP is in agreement with the N-terminal Edman amino acid sequencing previously performed on SDP[6].

Purified SKF was also analyzed by high-resolution mass spectrometry and found to have a mass of 2781.302 Da (**Fig. 2.4**). This mass could not be readily matched to the C- or N-terminal portions of the 55 amino acid protein precursor SkfA[6]. When purified SKF was subjected to high-resolution tandem mass spectrometry, the sequence tag WASKSI was obtained (**Fig. 2.13a**, **Fig. 2.14**). To define the number of amino acids that were involved in the mature SKF metabolite, a ^{15}N feeding experiment was performed. The feeding experiment with 98% pure [^{15}N]ammonium chloride resulted in a 36 Da increase in mass, indicating that SKF contained 36 nitrogen atoms which matched to the number of nitrogens in the last 26 amino acids of the SkfA sequence, CMGCWASKSIAMTRVCALPHPAMRAI (**Fig. 2.13b**). The calculated mass of this peptide is 2803.340, and therefore, the mature form of SkfA is 22.038 Da less than that expected from the intact peptide sequence. The absence of observable *y* ions provided an indication that SKF may be cyclic[21]. We postulated that SKF was cyclic accounting for

18.011 of the 22.038 Da mass differences. The remaining 4.027 Da difference between the parent SkfA peptide and the cyclized precursor could be explained by two crosslinks, possibly a disulfide and a thioether.

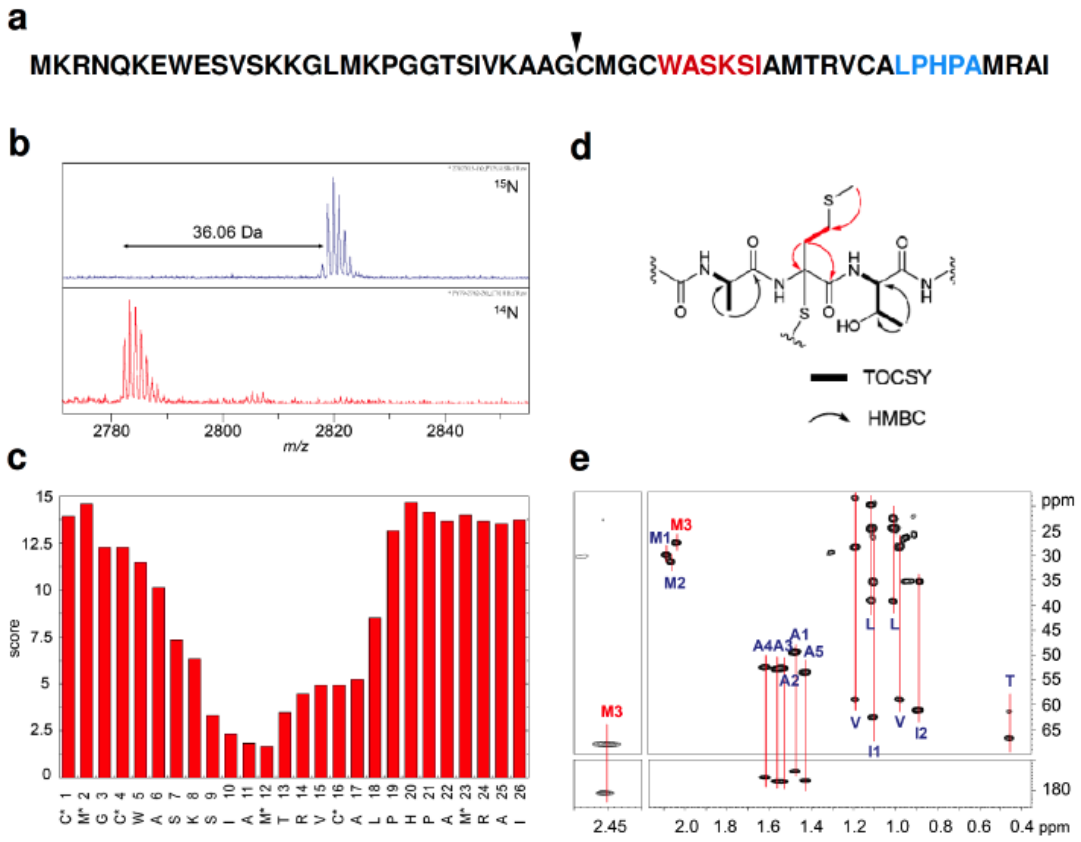


Figure 2.13: The structural characterization of SKF. (**a**) The SkfA sequence. ▼ indicates a protease cleavage site. The sequence tags observed by tandem mass spectrometry generated by FT-ICR MS/MS (red) and MALDI TOF/TOF (blue) are highlighted. (**b**) The MALDI-TOF analysis of SKF from cells grown in ^{15}N-media. (**c**) Comparative dereplication of deuterated-dethiolated SKF. (**d**) A schematic representation of the ^{1}H-^{13}C long range correlations and ^{1}H-^{1}H TOCSY correlations of the modified methionine and neighbor residues in SKF. (**e**) The ^{1}H-^{13}C HMBC spectrum of SKF showing the important ^{13}C-chemical shift supporting the α-connection to the methionine. The full TOCSY and HMBC spectra and tables of the observed chemical shifts are provided in the supporting information.

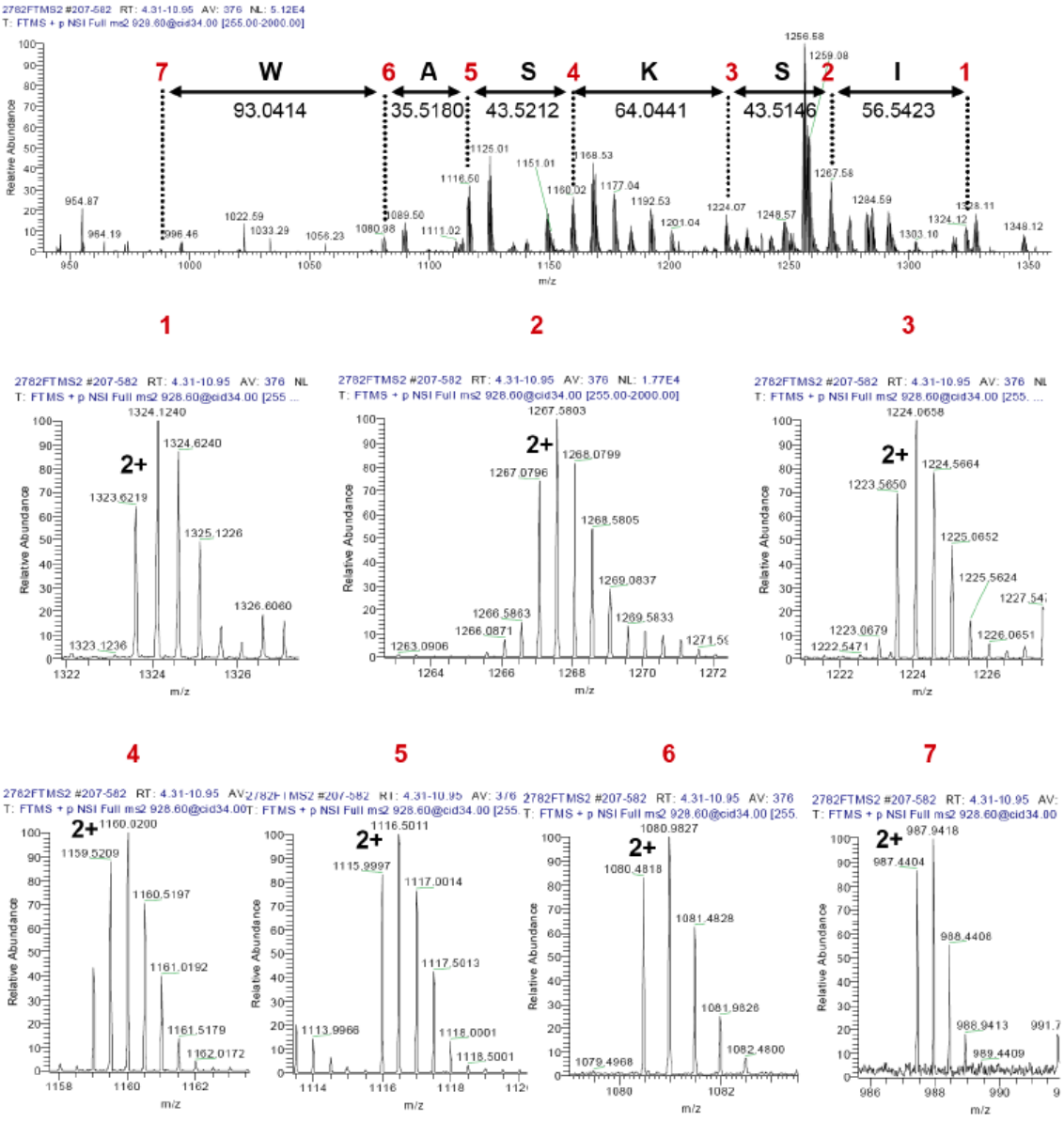


Figure 2.14: FT MS/MS spectrum of ion *m/z* 928.60 (2+ charge state). The observed mass difference and isotope profiles of each ion (2+ 3 charge state) are show.

Table 2.4: Annotations of SDP FT MS/MS spectrum. Errors are in ppm.

Seq.	Ion	Calc. mass	Obs. ion	Obs. mass	Error	Ion	Calc. mass	Obs. ion	Obs. mass	Error
C*	b1	103.0092				y42	4311.2111			
G	b2	160.0307				y41	4210.2176			
L	b3	273.1147				y40	4153.1961			
Y	b4	436.1780				y39	4040.1121			
A	b5	507.2152				y38	3877.0487			
V	b6	606.2836				y37	3806.0116			
C*	b7	707.2771	708.2770	707.2697	10.42	y36	3706.9432			
V	b8	806.3455	807.3433	806.3360	11.76	y35	3603.9340	1802.9338	3603.8530	22.47
A	b9	877.3826	878.3798	877.3725	11.50	y34	3504.8656	1753.4027	3504.7908	21.34
A	b10	948.4197	949.4153	948.4080	12.34	y33	3433.8285	1717.8869	3433.7592	20.17
G	b11	1005.4412	1006.4351	1005.4278	13.29	y32	3362.7914	1682.3684	3362.7222	20.57
Y	b12	1168.5045	1169.4938	1168.4865	15.40	y31	3305.7699	1653.8592	3305.7038	20.00
L	b13	1281.5886	1282.5748	1281.5675	16.42	y30	3142.7066	1572.3325	3142.6504	17.87
Y	b14	1444.6519	1445.6347	1444.6274	16.95	y29	3029.6226	1515.7906	3029.5666	18.45
V	b15	1543.7203	1544.7019	1543.6946	16.64	y28	2866.5592			
V	b16	1642.7887	1643.7614	1642.7541	21.06	y27	2767.4908	1384.7287	2767.4428	17.33
G	b17	1699.8102	1700.7851	1699.7778	19.04	y26	2668.4224	1335.1968	2668.3790	16.25
V	b18	1798.8786	1799.8453	1798.8380	22.55	y25	2611.4009	1306.6889	2611.3632	14.44
N	b19	1912.9215	1913.8818	1912.8745	24.57	y24	2512.3325	1257.1545	2512.2944	15.16
A	b20	1983.9586				y23	2398.2896	1200.1330	2398.2514	15.91
V	b21	2083.0270				y22	2327.2525	1164.6165	2327.2184	14.63
A	b22	2154.0642				y21	2228.1841	1115.0821	2228.1496	15.46
L	b23	2267.1482				y20	2157.1470	1079.5659	2157.1172	13.78
Q	b24	2395.2068				y19	2044.0629	1023.0255	2044.0364	12.95

Table 2.4: Continued. Annotations of SDP FT MS/MS spectrum. Errors are in ppm.

T	b25	2496.2545	y18	1916.0043	958.9989	1915.9832	11.01
A	b26	2567.2916	y17	1814.9567	1815.9249	1814.9176	21.50
A	b27	2638.3287	y16	1743.9195	1744.8932	1743.8859	19.28
A	b28	2709.3658	y15	1672.8824	1673.8595	1672.8522	18.06
V	b29	2808.4342	y14	1601.8453	1602.8207	1601.8134	19.91
T	b30	2909.4819	y13	1502.7769	1503.7552	1502.7479	19.29
T	b31	3010.5296	y12	1401.7292	1402.7101	1401.7028	18.84
A	b32	3081.5667	y11	1300.6816	1301.6685	1300.6612	15.63
V	b33	3180.6351	y10	1229.6444	1230.6321	1229.6248	15.96
W	b34	3366.7144	y9	1130.5760	1131.5661	1130.5588	15.22
K	b35	3494.8094	y8	944.4967	945.4916	944.4843	13.13
V	b36	3657.8727	y7	816.4018	817.3993	816.3920	11.93
Y	b37	3756.9411	y6	653.3384			
A	b38	3827.9782	y5	554.2700			
K	b39	3956.0732	y4	483.2329			
Y	b40	4119.1365	y3	355.1380			
S	b41	4206.1685	y2	192.0746			
S	b42	4293.2006	y1	105.0426			

To confirm that SKF was indeed cyclic, all thiols were replaced with a proton using a reductive dethiolation reaction composed of $NiCl_2$ and $NaBH_4$ (**Fig. 2.15**)[22]. Subjecting SKF to this reaction resulted in an ion with a mass of 2551.444 Da. Tandem mass spectrometry by collision induced dissociation (CID) confirmed the sequence to the 26 amino acids on the C-terminal end of SkfA as well as a cyclic head-to-tail linkage between isoleucine and cysteine. The calculated mass of this molecule is 2551.450. To define the connection of the thioether linkage, the reductive dethiolation was repeated but with deuterated solvents and $NiCl_2/NaBD_4$ and resulted in a species with mass of 2558.491, 7.041 Da larger compared to the product resulted from $NiCl_2/NaBH_4$ reaction, suggesting 7 thiol connections. Six deuteriums were introduced from the replacement of three methionine and three cysteine side chains with deuterons while the remaining deuterium was introduced at the site of the thioether linkage (**Fig. 2.16a**). Therefore, it became possible to map out the position of the thioether linkage by locating this extra deuterium that gave rise to the 1 Da mass shift. To map the thioether linkage, multiple stages of tandem mass spectrometry were obtained on the deuterated dethiolated SKF. To find the position of the additional deuterium, a mass list with manually deconvoluted fragment ions was analyzed by an algorithm NRP-comparative dereplication[23] against the theoretical structure of SKF with a deuterium labeled on each desulfurized position (**Fig. 2.16a**). NRP-dereplication matched the MS fragments with the structure and Met12 had the lowest score indicated the strongest correlation of the 1 Da increase between the observed ions in the tandem mass spectrometry experiment when compared to the theoretically predicted ions of the deuterated dethiolated SKF template (**Fig. 2.13c** and

Fig. 2.16b). Guided by the NPR-dereplication result, manual annotation was performed, again verifying the extra deuterium on Met12 (**Fig. 2.16*c*** and **Table 2.5**, **Table 2.6**).

The position of the disulfide bond was determined by reduction, iodoacetamide alkylation and tandem mass spectrometry. Reduction and alkylation of SKF resulted in a mass increase of 116.061 Da, in agreement with two free thiols (**Fig. 2.15**). Via tandem mass spectrometry, the alkylated residues were found to be Cys1 and Cys16, positioning the disulfide between these two cysteines. Even though the tandem mass spectrometry of the desulfurized SKF indicated that the thioether linkage is connected to the methionine, it did not provide regiochemical information to which carbon of Met12 that Cys4 was connected. To determine the regiospecificity of the tetrahedral linkage, we resorted to nuclear magnetic resonance (NMR) spectroscopy (**Fig. 2.17** and **Table 2.7**). To determine the relevant proton signals, the NMR signal that corresponded to the methionine involved in the thioether cross-link needed to be identified. To find the modified methionine with an absent proton, a ^{1}H-^{1}H total correlation spectroscopy (TOCSY) was obtained first. From the TOCSY, one set of methionine correlations lacking an α-proton was observed, and suggested that the linkage of Cys4 to Met12 is via the α-carbon of the latter residue (**Fig. 2.17b** and **c**). An α-connection would result in a ^{13}C-chemical shift at this α-carbon of about 70 ppm while a β-connection would result in ^{13}C-chemical shift of 40-50 ppm[22]. The ^{13}C-chemical shift information was obtained indirectly by heteronuclear multiple bond correlation (HMBC) (**Fig. 2.13d** and **e**, **Fig. 2.17d**), because of the small quantities of pure SKF available. The same methionine that was missing the proton in the TOCSY was scrutinized in the HMBC spectrum. In the HMBC, in agreement with the findings by TOCSY, there was no evidence of an α-proton

in Met12; however, the β-proton possessed long range correlations between the β-protons and two quaternary carbons, located at δ 67.9 and 180.7 (**Fig. 2.13d** and **e**), and thus consistent with a thioether bridge connecting to the α-carbon of Met12. A post-translational modification of a cysteine to the α-carbon of methionine has not been previously reported[24].

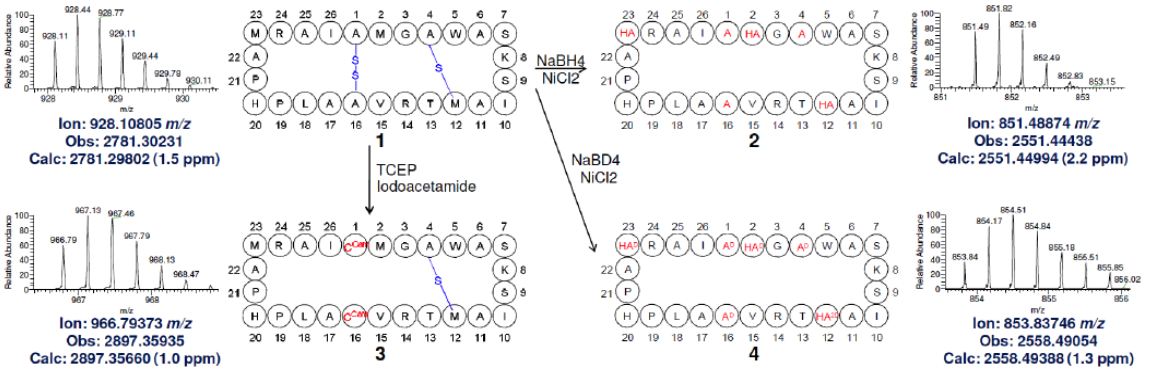


Figure 2.15: Chemical derivatization of SKF. TCEP reduction followed by iodoacetamide alkylation yielded a species with 116 Da addition (compound 2), corresponding to two carboxyamidomethylated cysteines (C^{Cam}) with the retaining of the thioether linkage. $NaBH_4$ and $NaBD_4$ reduction yielded dethiolated SKF derivatives with cysteine reduced to alanine, and methionine to homoalanine (HA) (compound 3,4). Deuteriums were labeled on the reductive cleavage positions under deuterated condition (compound 4). Zoom in MS1 spectrum of 3+ charge state species analyzed by FT-ICR MS as well as observed (Obs) and calculated (Calc) masses were showed beside each derivative.

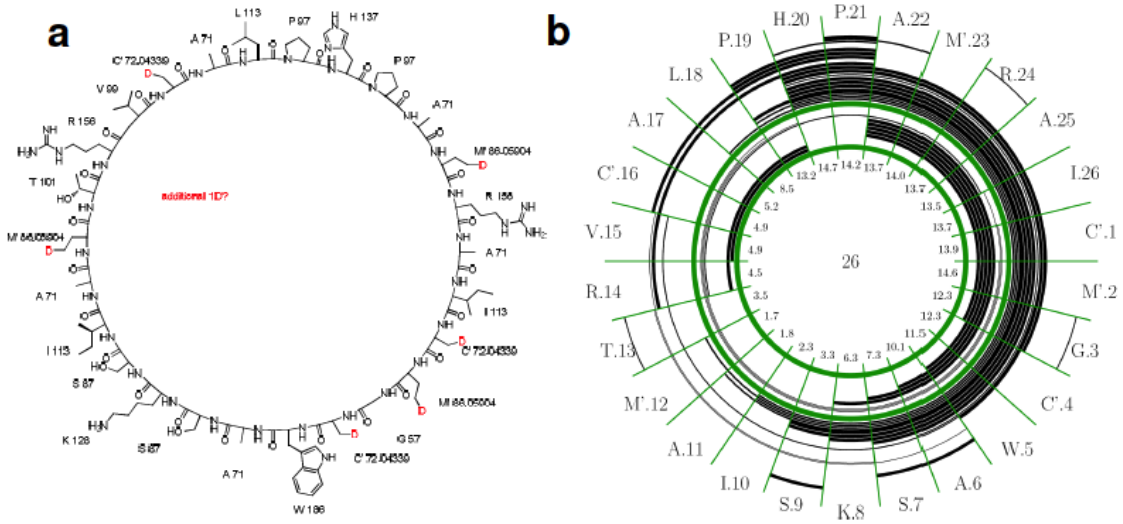


Figure 2.16: Comparative dereplication of SKF. a) SKF template used for comparative dereplication. Under $NaBD_4/NiCl_2$ dethiolation reaction, seven deuteriums were obtained. Six deuteriums can be predicted to be introduced from the replacement of three methionine and three cysteine side chains (red color labeled), whereas the remaining deuterium is introduced at the site of the thioether linkage. This template and the dethiolated SKF MS/MS spectrum were subjected to NPR-dereplication algorithm to compute for correlation. The least correlated residue indicates the location of the extra deuterium which allows to infer the location of thioether linkage. **b)** Comparative dereplication suggested thioether linkage on Met^{12}. Each semicircle represents an annotated peak in the MS spectrum. Peaks with multiple annotations split their count equally among the repeats. Inner numbers are the count/score of the supporting peaks for the conservation of the given amino acid. Amino acid codes have an extra index number to disambiguate repeated amino acids by their position. $Methionine^{12}$ has the lowest score indicating it bears the extra deuterium.

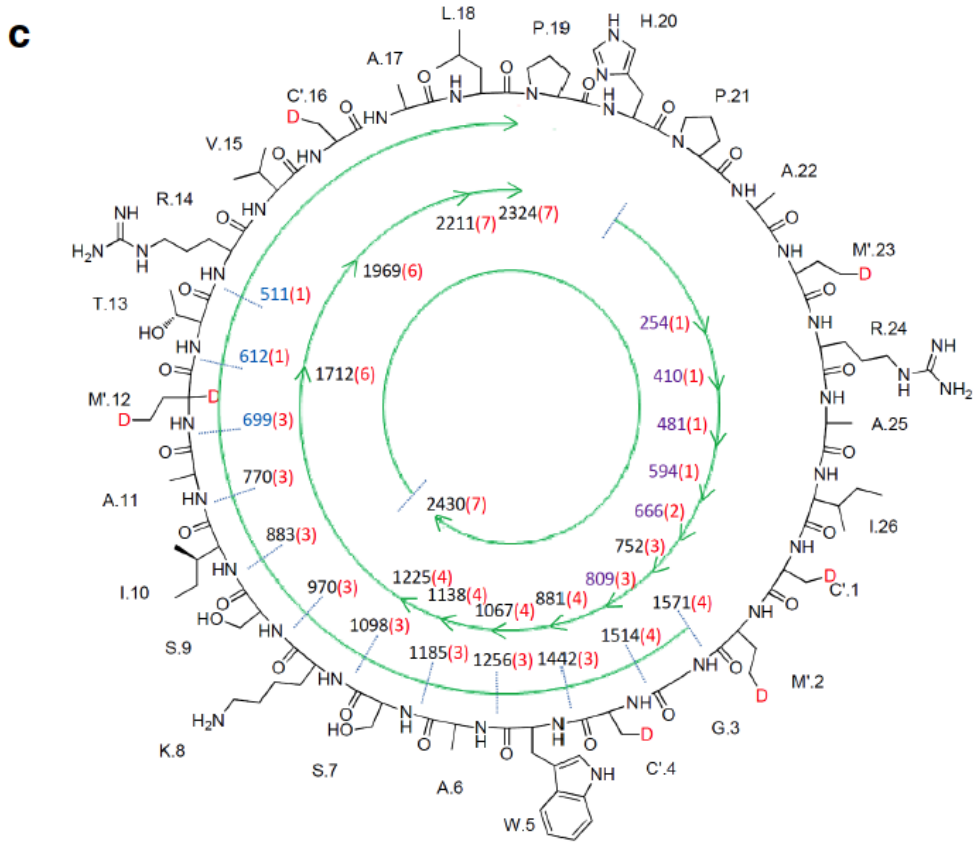


Figure 2.16: Comparative dereplication of SKF. Continued. c) Summary of fragments observed by CID fragmentation of dueterated dethiolated SKF. Incorporated deuteriums were labeled on structure with D (in red). Each number labeled inner the chemical structure is the mass of observed fragment with semi-circle representing the sequence. Associated number in parenthesis represents the number of deuterium within certain fragment. Each residue was labeled with one letter amino acid code as well as corresponding masses showed outside the chemical structure. Ions observed in MS^2 were showed in black color. Further supporting ions observed from MS^3 spectrum of ion *m/z* 883 were shown in blue and ions observed from MS^3 spectrum of ion *m/z* 881 were shown in purple. The corresponding ion tables are showed in **Tables 2.5** and **2.6**.

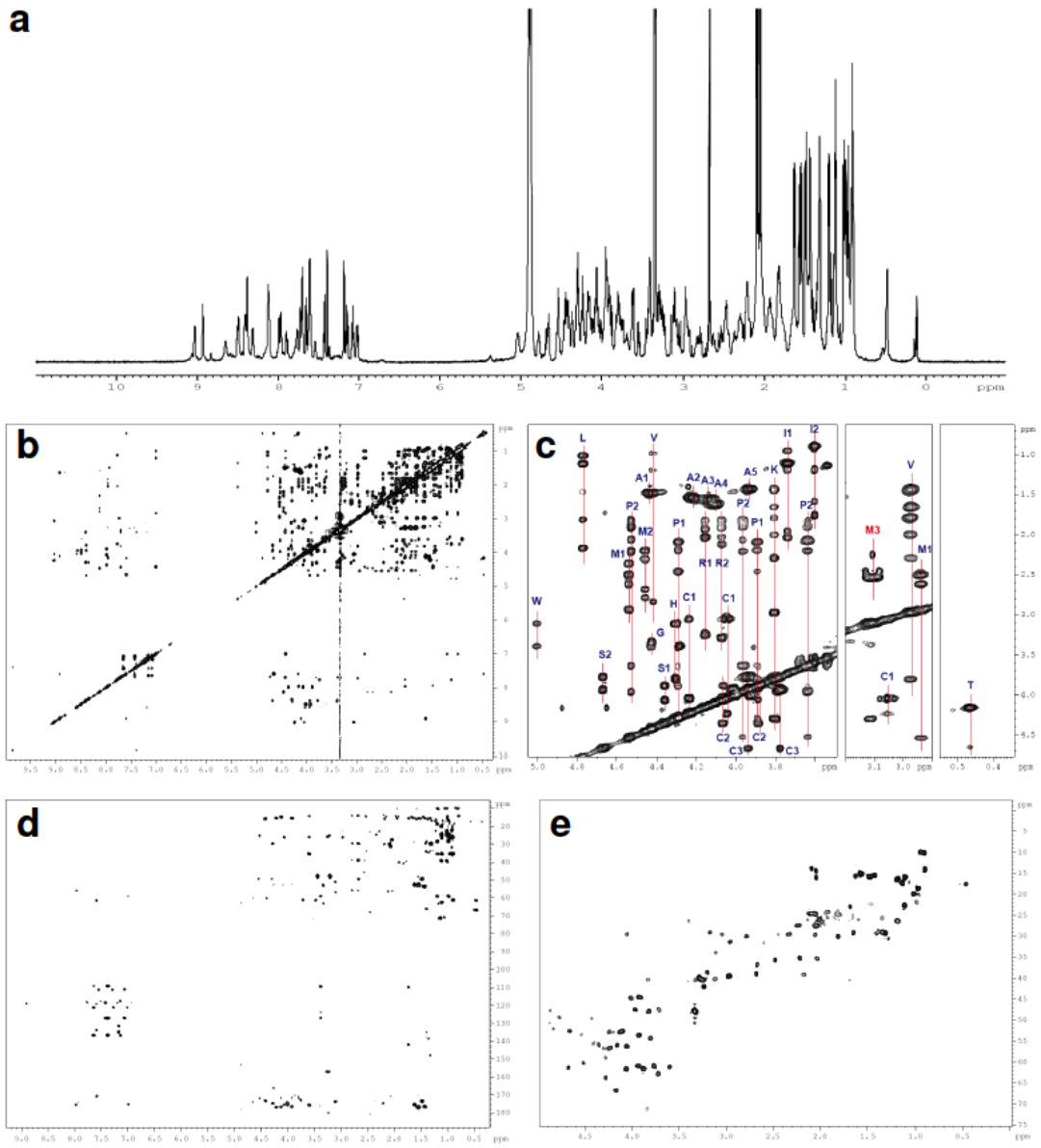


Figure 2.17: NMR spectra of SKF. a) ^{1}H NMR spectrum of SKF. The spectrum was observed in CD_3OD, 600 MHz. The detailed annotations were listed in **Table 2.7. b, c**) ^{1}H-^{1}H TOCSY spectra and annotations of SKF. The spectrum was observed in CD_3OD, 600 MHz, with mixing time = 90 ms. subfigure b is a full spectrum whereas subfigure c is a zoom in spectrum as well as annotations. **d**) ^{1}H-^{13}C HMBC spectrum of SKF. The spectrum was observed in CD_3OD, 600 MHz, with $^{2,3}J_{1H/13C}$ = 7 Hz. Full ^{1}H-^{13}C HMBC spectrum is shown. The annotations for critical signals supporting modified methionine are displayed in **Fig 2.7e**. **e**) ^{1}H-^{13}C HSQC spectrum of SKF. The spectrum was observed in CD_3OD, 600 MHz, with $^{1}J_{1H/13C}$ = 145 Hz. ^{1}H-^{13}C HSQC spectrum was collected to assist in HMBC spectrum annotation.

Table 2.5: Annotations of critical ions observed in dueterated dethiolated SKF MS/MS spectrum analyzed by FT-ICR MS.

Calc. Mass 1[a]	D number[b]	Calc. Mass 2[c]	Obs Mass	Error (ppm)	Annotation[d]
749.4547	3	752.4736	752.4733	-0.3	PAM'RAIC'M'
767.4654	3	770.4842	770.4802	-5.2	AM'TRVC'AL
877.5133	4	881.5384	881.5395	1.2	PAM'RAIC'M'GC'
880.5495	3	883.5683	883.5648	-3.9	IAM'TRVC'AL
967.5815	3	970.6003	970.6006	0.3	SIAM'TRVC'AL
1063.5926	4	1067.6177	1067.6171	-0.6	PAM'RAIC'M'GC'W
1095.6765	3	1098.6953	1098.6955	0.2	KSIAM'TRVC'AL
1134.6298	4	1138.6548	1138.6530	-1.6	PAM'RAIC'M'GC'WA
1182.7085	3	1185.7273	1185.7267	-0.5	SKSIAM'TRVC'AL
1221.6618	4	1225.6869	1225.6763	-8.6	PAM'RAIC'M'GC'WAS
1253.7456	3	1256.7644	1256.7656	1.0	ASKSIAM'TRVC'AL
1439.8249	3	1442.8438	1442.8402	-2.4	WASKSIAM'TRVC'AL
1510.8621	4	1514.8871	1514.8793	-5.2	C'WASKSIAM'TRVC'AL
1567.8835	4	1571.9086	1571.8966	-7.7	GC'WASKSIAM'TRVC'AL
1705.9627	6	1712.0004	1711.9916	-5.1	PAM'RAIC'M'GC'WASKSIAM'
1963.1115	6	1969.1492	1969.1343	-7.5	PAM'RAIC'M'GC'WASKSIAM'TR
2204.2542	7	2211.2981	2211.3032	2.3	PAM'RAIC'M'GC'WASKSIAM'TRVC'A
2317.3384	7	2324.3822	2324.3689	-5.7	PAM'RAIC'M'GC'WASKSIAM'TRVC'AL
2423.3550	7	2430.3989	2430.3984	-0.2	SIAM'TRVC'ALPHPAMRAIC'M'GC'WAS

a. Theoretical mass of fragment ions resulted from $NaBH_4$ reduction (dethiolated SKF).
b. D number represents for the number of deuterium labeled within certain fragments.
c. Theoretical mass of fragment ions resulted from $NaBD_4$ reduction (dueterated dethiolated SKF).
d. Residue C and M are marked due to the reason that these two residues are derivatized after reaction.

Table 2.6: Annotations of critical ions observed in additional fragmentation (MS3) of dueterated dethiolated SKF analyzed by IT-MS.

Calc. Mass 1[a]	D number[b]	Calc. Mass 2[c]	Obs Mass	Error (Da)	Annotation[d]
510.33	1	511.33	511.26	-0.07	RVC'AL
611.38	1	612.38	612.35	-0.03	TRVC'AL
696.43	3	699.45	699.44	0.00	M'TRVC'AL
806.48	3	809.50	809.44	-0.05	PAM'RAIC'M'G
664.40	2	666.41	666.44	0.03	PAM'RAIC'
593.36	1	594.37	594.44	0.07	PAM'RAI
480.28	1	481.29	481.35	0.07	PAM'RA
409.24	1	410.25	410.35	0.10	PAM'R
253.14	1	254.15	253.99	-0.16	PAM’

a. Theoretical mass of fragment ions resulted from $NaBH_4$ reduction (dethiolated SKF).
b. D number represents for the number of deuterium labeled within certain fragments.
c. Theoretical mass of fragment ions resulted from $NaBD_4$ reduction (dueterated dethiolated SKF).
d. Residue C and M are marked due to the reason that these two residues are derivatized after reaction.

Table 2.7: ^{1}H NMR data of SKF.

	NH	α	β	γ	others
A1	7.93	4.43	1.48		
A2	7.98	4.22	1.54		
A3	8.11	4.13	1.56		
A4	8.91	4.1	1.62		
A5	8.40	3.93	1.43		
C1	9.02	4.23	3.05, 4.04		
C2	7.98	4.35	3.88, 4.06		
C3	8.38	4.67	3.78, 3.93		
G	8.39	3.39, 4.28			
H	8.03	5.00	3.10, 3.40		8.58 (NH), 7.18
I1	7.60	3.74	2.04	1.11 (CH_3), 1.19, 1.96	1.01 (d)
I2	8.47	3.6	1.76	0.89 (CH_3), 1.19, 1.59	0.91 (d)
K	7.73	4.28	2.29	1.79, 1.99	1.43 (d), 1.65 (d), 2.97 (e)
L	7.60	4.77	1.46, 2.17	1.81	1.01 (d), 1.12 (d)
M1	7.52	4.54	2.36, 2.51	2.62, 2.94	2.09 (SCH_3)
M2	8.13	4.46	2.20, 2.31	2.69, 2.79	2.07 (SCH_3)
M3	ND	-	2.25, 2.46	2.54, 3.10	2.04 (SCH_3)
P1	-	4.29	2.09, 2.46	2.09, 2.19	3.65 (d), 3.90 (d)
P2	-	4.53	1.91, 2.21	1.81, 2.07	3.64 (d), 3.97 (d)
R1	7.96	4.15	2.03	1.83, 1.94	7.44 (NH), 3.25 (d)
R2	7.98	4.07	2.10	1.82, 1.92	7.53 (NH), 3.28 (d)
S1	7.99	4.36	3.88, 4.07		
S2	8.38	4.67	3.77, 3.93		
T	7.59	4.65	4.16	0.47	
V	7.77	4.42	2.84	0.98, 1.20	
W	8.34	4.30	3.11, 3.80		9.85 (NH), 7.39 (s), 7.66 (d), 7.07 (t), 7.15 (t), 7.43 (d)

2.5.2 The functional annotation of the *skf* biosynthetic operon.

Bsed on the structure of SKF, it is now possible to propose the role of each gene on the *skf* operon involved in SKF biosynthesis. The *skf* operon was previously predicted to contain a stop codon due to sequencing errors in the original genome that resulted in an incorrect open reading frame assignment for *skfC* and *skfD*[6]. Thus, *skfC* and *skfD* are in fact a single open reading frame, and therefore we have omitted the designation *skfD*. Therefore the *skf* operon has seven genes, *skfABCEFGH* (**Fig. 2.18**). SkfA is a prepropeptide that is post-translationally modified to the mature SKF[6]. SkfB belongs to the radical SAM superfamily that includes genes such as *albA*, *lipA*, *bioB*, which are involved in the biosynthesis of C-S linkages in subtilosin, lipoic acid, and biotin, respectively (**Fig. 2.19a**)[22, 25-27]. We propose that SkfB is responsible for the C-S linkage of Cys4 to the α-carbon of Met12 similar to the proposed reaction catalyzed by AlbA on the subtilosin pathway[22, 28]. SkfC belongs to the CaaX family of proteases (**Fig. 2.19b**) which, in eukaryotes, are responsible for the proteolysis of C-terminal prenylated cysteines[29]. Since this is the only protease candidate on the *skf* gene cluster, we propose it is involved in the cyclization reaction, which would represent a new function for a CaaX family member of proteins. SkfE is homologous to the cytoplasmic ATPase domain of ABC transporters, while SkfF is a polytopic membrane protein; both are predicted to be involved in the export and immunity of SKF[6]. SkfG showed homology to several HEAT-repeat containing proteins, although the role of SkfG remains unclear. SkfH is a thioredoxin-oxidoreductase like protein and may be involved in the generation of the disulfide bond analogous to the proposed function of the oxidoreductase on the sublancin 168 biosynthetic pathway (**Fig. 2.19c**)[30].

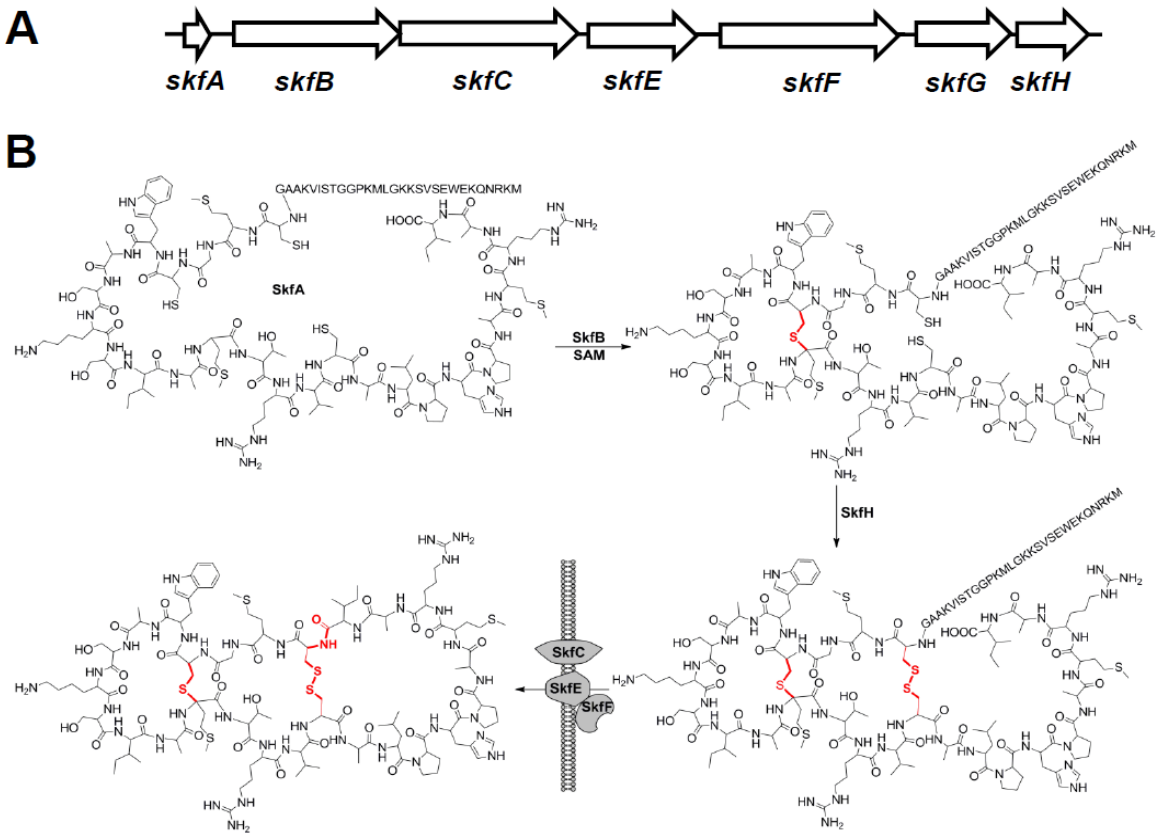


Figure 2.18: The functional annotation of the *skf* operon. (**a**) The *skf* operon. (**b**) The proposed biosynthetic pathway of SKF.

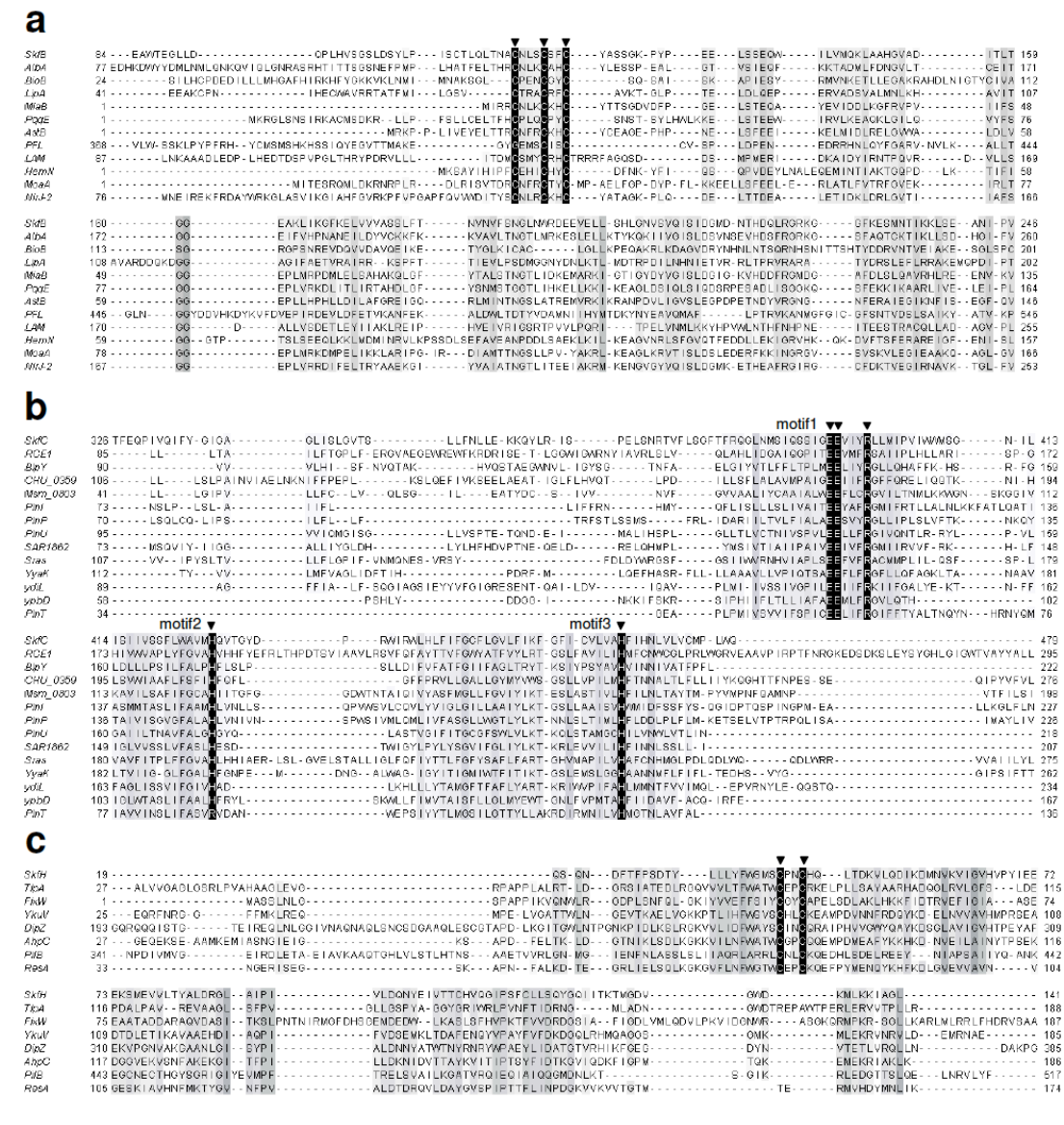


Figure 2.19: Multiple sequence alignment of SkfB, SkfC, and SkfH. a) Multiple sequence alignment of SkfB. SkfB belongs to radical SAM superfamily, along with some well-known proteins such as LAM, PFL, BioB, LipA. All members of this family show a conserved motif of CXXXCXXC (marked by ▼) which act to coordinate the iron in [4Fe-4S] cluster. **b**) Multiple sequence alignment of SkfC. Three conserved motifs are showed. The first motif contains two adjacent glutamic acid usually follows by a highly conserved arginine spaced by three resides. The first Glu in motif1 as well as the two histidines are thought to involve in zinc binding whereas the second Glu is responsible for catalytic activity. **c**) Multiple sequence alignment of SkfH. SkfH homologs to thioredoxin showed a characteristic CXXC motif. Only the highly conserved regions are displayed due to the length variation of each protein. Alignment was done by kalign (22) using default settings. Only the highly conserved regions are displayed due to the length variation of each protein.

2.6 Materials and Methods

2.6.1 Strains used in this study

Bacillus	Genotype	Background	Source – Ref.
3610	Prototroph		S. Branda[31]
PY79	Prototrophic derivative of *B. subtilis* 168		K. Pogliano[32]
KP648	Δ*spo0A::erm*	PY79	K. Pogliano[33,34]
ALB1035	Δ*spo0A::erm*	3610	This study
EG208	$P_{spac\text{-}hy}$-*skfABCDEFGH(kan)*	PY79	R. Losick[6]
EG165	Δ*skfA::spc*	PY79	R. Losick[6]
EH273	*sdpABC::kan*	PY79	R. Losick[7]
ALB1046	Δ*skf::cat*	PY79	This study
ALB1085	$P_{spac\text{-}hy}$-*skfABCDEFGH(kan)*, *ΔsdpABIC::tet*	PY79	This study
CDE1160	*amyE::*$P_{spac\text{-}hy}$-*sdpC(spec)*, *sdpC::tet*, $P_{spac\text{-}hy}$-*sdpAB(cat)*	PY79	This study

2.6.2 Thin-layer agar MALDI-IMS

Sample preparation for thin-layer agar MALDI-IMS experiments was performed using a modified method adjusted from reference 17. 0.2-1 μL of bacterial overnight cultures grown in LB media were spotted on 100 O.D. x 25 mm Petri dishes (Fisherbrand) containing ISP2 solid agar. For testing individual cultures in isolation, strains were spotted at the center of Petri dishes; for the PY79 and Δ*spo0A* co-culturing experiment, both strains were spotted on the same plate with a distance between spots of 0.75 cm. The Petri dishes were sealed with parafilm and incubated for 2-4 days as indicated. After culturing, a rectangular section containing the colonies was transferred to a MALDI target plate. It is critical to avoid any air bubbles because it will cause the agar to peel off during the IMS process. After taking a photograph, a 1:1 mixture of α-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid was sprinkled on top of the culture using a 20 μm sieve method[18, 35], and was dried in a 37 ^{o}C oven for three hours. The detailed instrument parameters for collecting image data were described in reference 17. Briefly, the sample was subjected to Bruker microflex MALDI-TOF MS for imaging MS acquisition and the data was analyzed using the FlexImaging 2.0 software.

2.6.3 Purification of SKF and SDP

PY79 was cultured on approximately 1000 IPS2 agar plates and cultured at 28 ^{o}C for 2 days. The bacteria were scraped from agar plates and then re-suspended in milli-Q water. Equal amount of *n*-BuOH was used to extract SKF and SDP from aqueous layer. The crude *n*-BuOH extract was lyophilized, re-dissolved and then eluted via Sephadex LH-20 column using isopropanol/MeOH = 1:9. Each fraction was analyzed by MALDI

MS and the fractions containing signals *m/z* 2782 or *m/z* 4312 were collected into two separate tubes. The two tubes that contained crude SKF or SDP, respectively, were purified subsequently by HPLC (C-18, 250 x 4.6 mm) running a gradient from 25% solvent A to 70% solvent A in 60 minutes with flow rate 1mL/min. Solvent A is isopropanol/MeCN = 7:3 containing 0.1% TFA; solvent B is 0.1% TFA (aq). SKF and SDP were eluted at 14, and 34.5 min, respectively (Fig. S3). Purified SKF and SDP were lyophilized and stored at -80ºC before using for bioassay and structural elucidation. The yield for SKF and SDP was ~ 0.5 μg each per plate.

2.6.4 SKF derivatization

Dethiolated SKF was prepared by dissolving 1 μg of SKF with 1.5 μg $NaBH_4$/$NaBD_4$ and 1.5 μg $NiCl_2$ in 6.25 μL of 60% MeOH/MeOD. This reaction was incubated at 50 ^oC, and an additional 1.5 μg of $NaBH_4$/$NaBD_4$ and $NiCl_2$ were added into the reaction 5 and 10 minutes after initiation of the reaction to ensure complete conversion of SKF into dethiolated SKF. The mixture was then centrifuged for 1 min at 14500 rpm to remove the insoluble particles and then purified by HPLC using an Agilent Eclipse XDB-C18 column running MeCN gradients or by C18 ZipTip (Millipore) following the manufacturer's protocol prior to MS analysis.

For disulfide bond reduction, 1 μg SKF was dissolved in 40 mM ammonium bicarbonate buffer, pH 8.0, containing 10% MeCN. TCEP was added to reach a final concentration of 20 mM and incubated at 85 ^oC for 1hr. To prevent the free thiols from reforming disulfide bond, iodoacetamide was used to cap the cysteine thio group. To accomplish this, 50 mM Tris buffer, pH 7.4 was added to TCEP treated SKF solution to a

final concentration of 1 mM to bring up the pH to neutral (checked by pH paper). 5 µg of iodoacetamide powder was directly added into the reaction mixture and allowed to react at RT for 5 min followed by quenching with an equal volume of 10% formic acid.

2.6.5 General MS procedure for the characterization of SKF

For the MS data acquisition, each compound was dissolved in spray solvent 50:50 MeOH/H_2O containing 1% formic acid, and underwent nanoelectrospray ionization on a biversa nanomate (Advion Biosystems, Ithaca, NY) using a back pressure of 0.3-0.5 p.s.i. and the spray voltage of 1.4 -1.5 kV. MS spectra were acquired on a 6.42 T Finnigan LTQ-FTICR MS or a Finnigan LTQ-MS (Thermo-Electron Corporation, San Jose, CA) running Tune Plus software version 1.0 and Xcalibur software version 1.4 SR1. The instrument was first autotuned on the *m*/*z* value of the ion to be fragmented. Then, the ions were isolated by the linear ion trap and fragmented by collision induced dissociation (CID) (isolation window: 3-10 *m/z*; collision energy: 30).

2.6.6 NMR measurement of SKF

400 µg SKF was dissolved in 40 µL of CD_3OD for NMR data acquisition. NMR spectra were recorded on Bruker Avance III 600 MHz NMR with 1.7 mm Micro-CryoProbe at 300 K, with standard pulse sequences provided by Bruker. 2D TOCSY spectra were recorded with mixing times of 90 ms. 2D ^{1}H-^{13}C HMBC spectra were recorded with ^{2}J or $^{3}J_{H\text{-}C}$ coupling constants at 7 Hz, 2D ^{1}H-^{13}C HSQC spectra were recorded with $^{1}J_{H\text{-}C}$ coupling constants at 145 Hz.

2.6.7 Effect of SKF and SDP on *B subtilis* cell growth curve

The effect of SKF and SDP on *B. subtilis* cell growth was performed using 96 well microtiter plates. A 2 mL overnight culture in LB media was centrifuged at 6000g for 10 minutes and supernatants discarded. The cell pellets were resuspended using 2 mL of ISP2 media. OD_{595} of the resuspended cells were measured (ELx808 Ultra Microplate Reader, Bio-TEK Instruments), and the final OD_{595} was adjusted to 0.03 with ISP2 media. 100 µL diluted culture with indicated working concentrations of SKF or SDP were aliquoted into each well. The plate was shaken at 37^{o}C, 120 rpm. OD_{595} were measured and recorded at each time point. To evaluate the effect of SDP on an exponentially growing culture, SDP was added at 3 or 6 hours to a final concentration of 20 µg/mL, and the OD_{595} was measured at each time point**.**

2.6.8 Fluorescence microscopy

The effects of SDP on individual *B. subtilis* cells were investigated in 15 µL cultures prepared in the following manner. Cultures were grown in LB media to an OD_{600} of 0.3, centrifuged, resuspended in 1/10 the volume and 14.25 µL of concentrated cells were added to 1.7 mL microcentrifuge tubes. At t = 0, 0.75 µL of 10% DMSO or 400 µg/mL SDP (in 10% DMSO) was added to different aliquots of cells. The tubes were capped and incubated at 37°C in a roller. Samples were collected for imaging every 30 minutes. 2 µL of cells were added to 0.5 µL of a stain mix containing 30 µg/mL FM 4-64, 2.5 µM Sytox Green and 1 µg/mL DAPI prepared in 1X T-base. Cells were immobilized with Poly-L-Lysine and imaged on an Applied Precision Spectris Microscope[36]. Images were deconvoluted and the medial focal planes shown. Time-lapse imaging showing the

formation of membrane tubules and projections was performed on these slides, collecting images of cell membranes every 3 seconds for 1 minute.

Quantification of the amount of cellular lysis was performed by determining the percent cells showing clearly discontinuous membranes and increased permeability to Sytox Green relative to the number of intact cells. This data showed that the 3610 strain, both with and without the Δ*spo0A* mutation (ALB1035) showed the most rapid onset of cell lysis (first evident at t = 60 minutes) and the highest frequency of lysis. The strains in PY79 background were affected more slowly and in a lower percent of cells. We did not score later time points in this manner, because the extensive cell lysis made it impossible to determine the percent cell lysis, since it was unclear how many cells produced the debris.

2.6.9 Spot assay

Lawns were created by mixing 50 µl exponentially growing cells (OD_{600} = 0.4-0.6) with 3.5 ml 0.35% LB agar and pouring the mixture onto LB plates. When indicated, IPTG was added to a final concentration of 1 mM. After the top agar solidified, 5 µl of exponentially growing cells or purified compounds were spotted on top of the lawn and allowed to dry. The plates were then incubated overnight at 30°C.

2.6.10 Screen of antibacterial activities against pathogens

The activities of SDP and SKF were tested in a microtiter based screen for growth inhibitory activity against a variety of Gram-positive and Gram–negative bacterial species. Then, the IC_{50} of these compounds was assessed against a smaller set of

representative organisms. For this assessment, SDP was tested against three organisms whose growth it inhibited, methicillin-resistant *Staphylococcus aureus* (MRSA) sequence type 59 (ST59)*, S. aureus* Newman, *Staphylococcus epidermidis* (ATCC35984), and two that were unaffected, *Klebsiella pneumoniae* (ATCC700603) and *Pseudomonas aeruginosa* (ATCC 10145). SKF had no effect in the microtiter assay for any species and the MIC assay was performed with *S. aureus* MRSA ST59*, K. pneumoniae* (ATCC700603 and ATCC 35657), *Burkholderia cepacia* (ATCC 17765) and *Escherichia coli* (ATCC 25922).

IC_{50} assays were performed by a microbroth dilution assay. The overnight culture of the tested strain was diluted 1:200 in cation-adjusted Mueller-Hinton broth (MHB, Hardy Diagnostics, Santa Maria, CA) and grown with shaking at 37°C to mid-logarithmic phase after which they were centrifuged and pellets were resuspended in phosphate-buffered saline to an OD_{600} of 0.4 – 0.5. Prior to the addition of this pre-culture, 96-well polystyrene test plates (Costar® #3288, Corning, NY) containing duplicate samples of serially diluted test compounds, SDP or SKF, and appropriate antibiotic controls were prepared in CAMHB. Bacteria were added to the test plate to a final concentration of $5x10^5$ CFU/ml in a volume of 80 ml/well. The control antibiotics included vancomycin (Hospira, Lake Forest, IL, USA) for *Staphylococcal* strains, ciprofloxacin (Fluka, Sigma-Adlrich) for *Pseudomonas aeruginosa* and sulfamethoxazole, trimethoprim (SMX-TMP Sicor ™ Irvine, CA) for *K. pneumoniae* and *E. coli* strains. Following the addition of bacteria, test plates were incubated at 37°C in a shaking incubator for 20 - 22h. Finally the plates were assessed for bacterial growth by the presence of turbidity at OD_{600}. The absorbance at each tested

concentration were normalized to the negative control (absence of test compound) to determine relative growth at a given concentration of SDP or SKF. The IC_{50} values of SDP were determined from this data.

References

1. Shank, E.A. & Kolter, R. New developments in microbial interspecies signaling. *Curr Opin Microbiol.* **12**, 205–214 (2009).

2. Tam, N.K.M. *et al*. The intestinal life cycle of *Bacillus subtilis* and close relatives. *J Bacteriol.* **188**, 2692–2700 (2006).

3. Aguilar, C., Vlamakis, H., Losick, R. & Kolter, R. Thinking about *Bacillus subtilis* as a multicellular organism. *Curr Opin Microbiol.* **10**, 638–643 (2007).

4. Earl, A.M., Losick, R. & Kolter, R. Ecology and genomics of *Bacillus Subtilis*. *Trends Microbiol.* **16**, 269–275 (2008).

5. Kunst, F. *et al*. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256 (1997).

6. Gonzalez-Pastor, J.E., Hobbs, E.C. & Losick, R. Cannibalism by sporulating bacteria. *Science.* **301**, 510–513 (2003).

7. Ellermeier, C.D., Hobbs, E.C., Gonzalez-Pastor, J.E. & Losick, R. A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. *Cell.* **124**, 549–559 (2006).

8. Engelberg-Kulka, H., Amitai, S., Kolodkin-Gal, I. & Hazan, R. Bacterial programmed cell death and multicellular behavior in bacteria. *PLoS Genetics.* **2**, 1518–1526 (2006).

9. Claverys, J.P. & Havarstein, L.S. Cannibalism and fratricide: Mechanisms and raisons d'être. *Nat Rev Microbiol.* **5**, 219–229 (2007).

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

11. Burbulys, D., Trach, K.A. & Hoch, J.A. Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell.* **64**, 545–552 (1991).

12. Fawcett, P., Eichenberger, P., Losick, R. & Youngman, P. The transcriptional profile of early to middle sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci* USA **97**, 8063–8068 (2000).

13. Molle, V, et al. The Spo0A regulon of *Bacillus subtilis. Mol Microbiol.* **50**,1683–1701 (2003).

14. Lin, D., Qu, L.J., Gu, H., & Chen, Z.A. 3.1-kb genomic fragment of *Bacillus subtilis* encodes the protein inhibiting growth of *Xanthomonas oryzae* pv. *oryzae*. *J Appl Microbiol.* **91**, 1044–1050 (2001).

15. Nandy, S.K., Bapat, P.M. & Venkatesh, K.V. Sporulating bacteria prefers predation to cannibalism in mixed cultures. *FEBS Letters.* **581**, 151–156 (2007).

16. Cornett, D.S., Reyzer, M.L., Chaurand, P. & Caprioli, R.M. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat Methods.* **4**, 828–833 (2007).

17. Seeley, E.H. & Caprioli, R.M. Molecular imaging of proteins in tissues by mass spectrometry. *Proc Natl Acad Sci* USA **105**, 18126–18131 (2008).

18. Yang, Y.L., Xu, Y., Straight, P. & Dorrestein, P.C. Translating metabolic exchange with imaging mass spectrometry. *Nat Chem Biol.* **5**, 885–887 (2009).

19. Zeigler, D.R., *et al*. The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *J Bacteriol.* **190**, 6983–6995 (2008).

20. Li, W.H.J. & Vederas, J.C. Drug discovery and natural products: end of an era or an endless frontier? *Science.* **325**, 161–165 (2009).

21. Ngoka, L.C.M. & Gross, M.L. Multistep tandem mass spectrometry for sequencing cyclic peptides in an ion-trap mass spectrometer. *J Am Soc Mass Spectrom.* **10**, 732–746 (1999).

22. Kawulka, K.E., *et al*. Structure of subtilosin A, a cyclic antimicrobial peptide from Bacillus subtilis with unusual sulfur to alpha-carbon crosslinks: Formation and reduction of alpha-thio-alpha-amino acid derivatives. *Biochemistry.* **43**, 3385–3395 (2004).

23. Ng, J., *et al*. Dereplication and *de novo* sequencing of nonribosomal peptides. *Nat Methods.* **6**, 596–599 (2009).

24. Oman, T.J. & van der Donk, W.A. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol.* **6**, 9–1(2010).

25. Miller, J.R., *et al*. Escherichia coli LipA is a lipoyl synthase: In Vitro biosynthesis of lipoylated pyruvate dehydrogenase complex from octanoyl-acyl carrier protein. *Biochemistry.* **39**, 15166–15178 (2000).

26. Sofia, H.J., Chen, G., Hetzler, B.G., Reyes-Spindola, J.F. & Miller, N.E. 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. *Nucl Acids Res.* **29**,1097–1106 (2001).

27. Berkovitch, F., Nicolet, Y., Wan, J.T., Jarrett, J.T. & Drennan, C.L. Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. Science. **303**, 76–79 (2004).

28. McIntosh, J.A., Donia, M.S., Schmidt, E.W. Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. *Nat Prod Rep.* **26**, 537–559 (2009).

29. Pei, J. & Grishin, N.V. Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends Biochem Sci.* **26**, 275–277 (2001).

30. Dorenbos, R., et al. Thiol-disulfide oxidoreductases are essential for the production of the lantibiotic sublancin 168. *J Biol Chem.* **277**, 16682–16688 (2002).

31. Branda, S.S., Gonzalez-Pastor, J.E., Ben-Yehuda, S., Losick, R. & Kolter, R. Fruiting body formation by *Bacillus subtilis*. *Proc Natl Acad Sci* USA **98**, 11621–11626 (2001).

32. Youngman, P., Perkins, J.B. & Losick, R. Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in Bacillus subtilis or expression of the transposon-borne erm gene. *Plasmid.* **12**, 1–9 (1984).

33. Pogliano, J., Sharp, M. & Pogliano, K. Chromosome partitioning during establishment of cellular asymmetry in *Bacillus subtilis*. *J Bact.* **184**, 1743–1749 (2002).

34. Ireton, K., Rudner, D.Z., Siranosian, K.J. & Grossman, A.D. Integration of multiple developmental signals in Bacillus subtilis through the Spo0A transcription factor. *Genes Dev.* **7**, 283–294 (1993).

35. Puolitaival, S. M., Burnum, K.E., Cornett, D.S. & Caprioli, R.M. Solvent-free matrix dry-coating for MALDI imaging of phospholipids. *J Am Soc Mass Spectrom.* **19**, 882–886 (2008)

36. Becker, E.C. & Pogliano, K. Cell-specific SpoIIIE assembly and DNA translocation polarity are dictated by chromosome orientation. *Mol Microbiol.* **66**, 1066–1079 (2007).

37. Leenders, F., Stein, T.H., Kablitz, B., Franke, P. & Vater, J. Rapid Typing of Bacillus subtilis Strains by their Secondary Metabolites Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Intact Cells *Rapid Commun Mass Spectrom.* **13**, 943–949 (1999).

38. Erhard, M., von Döhren, H. & Jungblut, P. Rapid typing and elucidation of new secondary metabolites of intact cyanobacteria using MALDI-TOF mass spectrometry. *Nat Biotechnol* **15**, 906–909 (1997).

39. Lassmann, T. & Sonnhammer, E.L.L. Kalign, Kalignvu and Mumsa: web servers for multiple sequence alignment. *Nucl Acids Res.* **34**, 596–599 (2006).

A continued study featured mode of action of SDP has recently been published in Molecular Biology, 2012. Lamsa A, Liu WT, Dorrestein PC, Pogliano K. The thesis author was the collaborative author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in PNAS, 2010. Wei-Ting Liu, Yu-Liang Yang, Yuquan Xu, Anne Lamsa, Nina M. Haste, Jane Y. Yang, Julio Ng, David Gonzalez, Craig D. Ellermeier, Paul D. Straight, Pavel A. Pevzner, Joe Pogliano, Victor Nizet, Kit Pogliano, and Pieter C. Dorrestein. The thesis author, YL Yang and Y Xu were the primary investigators and authors of this paper.

Chapter III

Imaging Mass Spectrometry and Genome Mining via Short Sequence Tagging Identified the Anti-Infective Agent Arylomycin in *Streptomyces roseosporus*

3.1 Abstract

Here, we described the discovery of anti-infective agent arylomycin and its biosynthetic gene cluster in an industrial daptomycin producing strain *Streptomyces roseosporus.* This was accomplished via the use of MALDI imaging mass spectrometry (IMS) along with peptidogenomic approach in which we have expanded to short sequence tagging (SST) described herein. Using IMS we have observed that prior to the production of daptomycin, a cluster of ions (**1**-**3**) were produced by *S. roseosporus* and correlated well with the decreased staphylococcal cell growth. Further adopted SST peptidogenomics approach, which relies on the generation of sequence tags from tandem mass spectrometric data and query against genomes to identify the biosynthetic genes, we were able to identify these three molecules (**1**-**3**) to arylomycins, a class of broad-spectrum antibiotics that targets type I signal peptidase. The gene cluster responsible for arylomycin production in *S. roseosporus* was then identified. The identification of arylomycins and their biosynthetic gene cluster from intensely studied microorganism highlights the strength of IMS and MS guided genome mining approaches in effectively bridging the gap between phenotypes, chemotypes and genotypes.

3.2 Introduction

Natural products that are made by non-ribosomal peptide synthetases (NRPS) have an unrivaled track record as anti-infective agents in the clinic[1,2]. Penicillin, vancomycin, and daptomycin are examples of antibiotics that are NRPS-derived[3-6] (**Fig. 3.1**). With the emergence of antibiotic-resistant microbes, there is a great interest in molecules that target drug resistant microbes[7,8]. However, the last broad-spectrum antibiotic introduced on the market was over 50 years ago.

Our laboratory has been interested in the development of mass spectrometric methodologies that interconnects phenotypes, chemotypes, and genotypes. A part of the motivation for these tools is not only to discover new biology but also apply these tools to the discovery of antimicrobials. Here we report the use of imaging mass spectrometry in combination with a short sequence tagging (SST)-based genome mining approach that connects phenotypes and chemotypes with genotypes. We applied this approach to the discovery of the arylomycins (**1-3, Fig. 3.1**) and their biosynthetic pathway in *Streptomyces roseosporus*.

To connect phenotypes with chemotypes, our laboratory has recently developed methods to investigate microbial metabolic interactions via imaging mass spectrometry (IMS)[9-11]. One of the goals for the development of IMS approaches to detect metabolic exchange is to enable the discovery of new therapeutic leads. Herein the pathogens *S. aureus*[12] and *S. epidermidis*[13] were co-cultured with *Streptomyces roseosporus* NRRL 15998, whose genome has been sequenced[13]. This actinomycete produces daptomycin, an antibiotic used in the clinic to treat gram-positive bacterial infections[4,6,14-17].

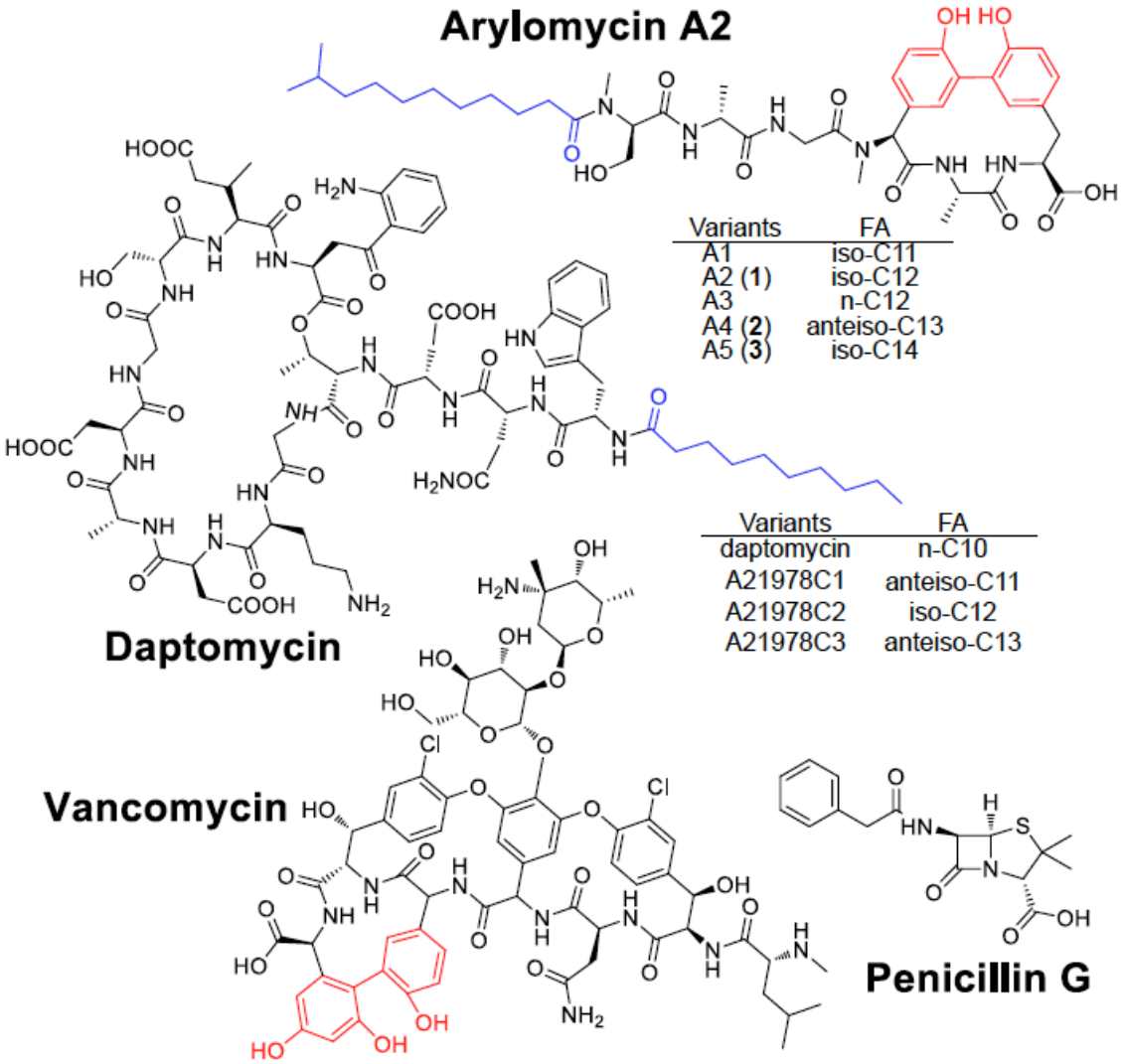


Figure 3.1: Structures of NRPS-derived compounds. Arylomycins and daptomycin have various components with different fatty acid (FA) (labeled in blue) chain lengths. Bis-aryl bridges are highlighted in red.

3.3 Results and Discussion

To demonstrate that IMS can be used to observe the molecules responsible for the inhibition of pathogens, we prepared lawns of *S. epidermidis* and *S. aureus* and then spotted *S. roseosporus* in the center (**Fig. 3.2**). After 36 hours incubation, inhibition zones were observed as expected in both staphylococcal lawns. Surprisingly, even though we determined that the IMS methodology can detect as little as 10 pmole of daptomycin, ions corresponding to daptomycin were not observed. Instead, a cluster of ions at *m/z* 863, 877, and 891, referred to as compounds **1**-**3** in this paper, were observed to localize at the zone of inhibition area. The absence of daptomycin-related signals in the zone-of-inhibition experiment suggested that *S. roseosporus* produced additional antibiotics.

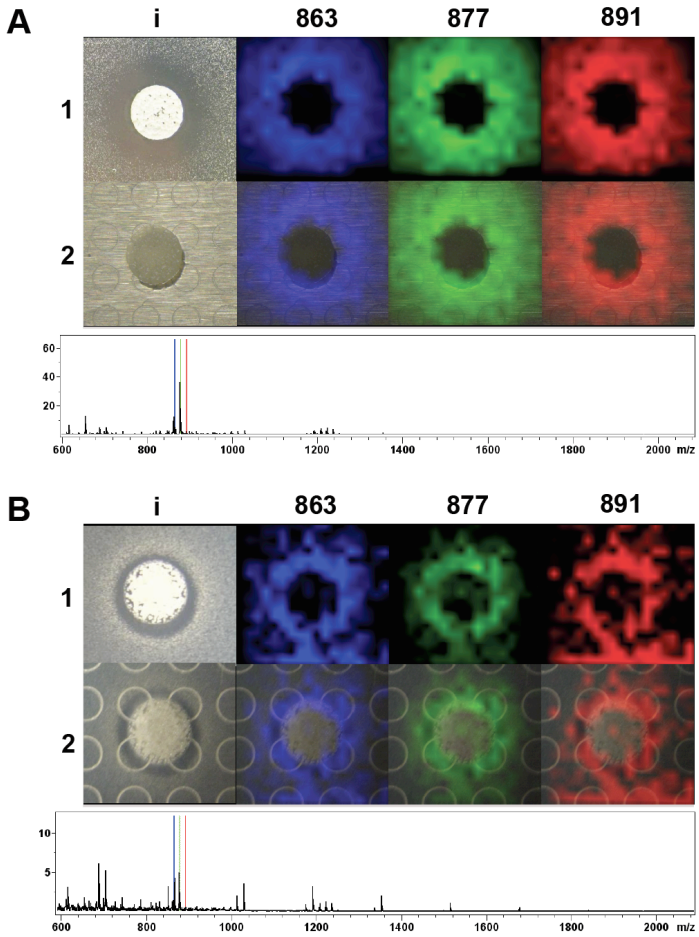


Figure 3.2: IMS of *S. roseosporus* co-cultured with *S. epidermidis* or *S. aureus*. (A) IMS of *S. roseosporus* spotted on top of a *S. epidermidis* lawn. (B) IMS of *S. roseosporus* spotted on top of a *S. aureus* lawn. (1) ion distribution of compound **1-3** (863, 877, 891) observed in IMS. 1i is a photograph showing *S. roseosporus* inhibit *Staphylococci* growth. (2) Superimposition of the photograph with IMS data on top of MALDI target plate. Average mass spectrum of each IMS experiment was shown below IMS images with signals correlated to compound **1-3** labeled with corresponding color as displayed in images.

A time course experiment of methanol extracts of *S. roseosporus* starter cultures revealed that compounds **1-3** were observed at 36 hours (**Fig. 3.3**), in agreement with the incubation time in the zone-of-inhibition experiment described above. Not until 48 hours, the production of signals at *m/z* 1634.72, 1648.74, 1662.75, which correspond to daptomycin variants (A21978C1-3, **Fig. 3.1**) were observed. That daptomycin is not

produced until 48 hours is consistent with the absence of daptomycin variants signals in the IMS data. MS-guided purification revealed that the molecules at *m/z* 863, 877, and 891 have monoisotopic masses of 825.439 (**1**), 839.455 (**2**), and 853.471 (**3**) Da, suggesting that the ion cluster observed in IMS exists as the potassium adduct. Compound **2** was purified and shown to exhibit antibiotic activity against *S. epidermidis* with similar efficacy to daptomycin but milder activity towards *S. aureus*, in agreement with the smaller zone of clearing for the *S. aureus* observed in **Fig. 3.2, Fig. 3.4)**.

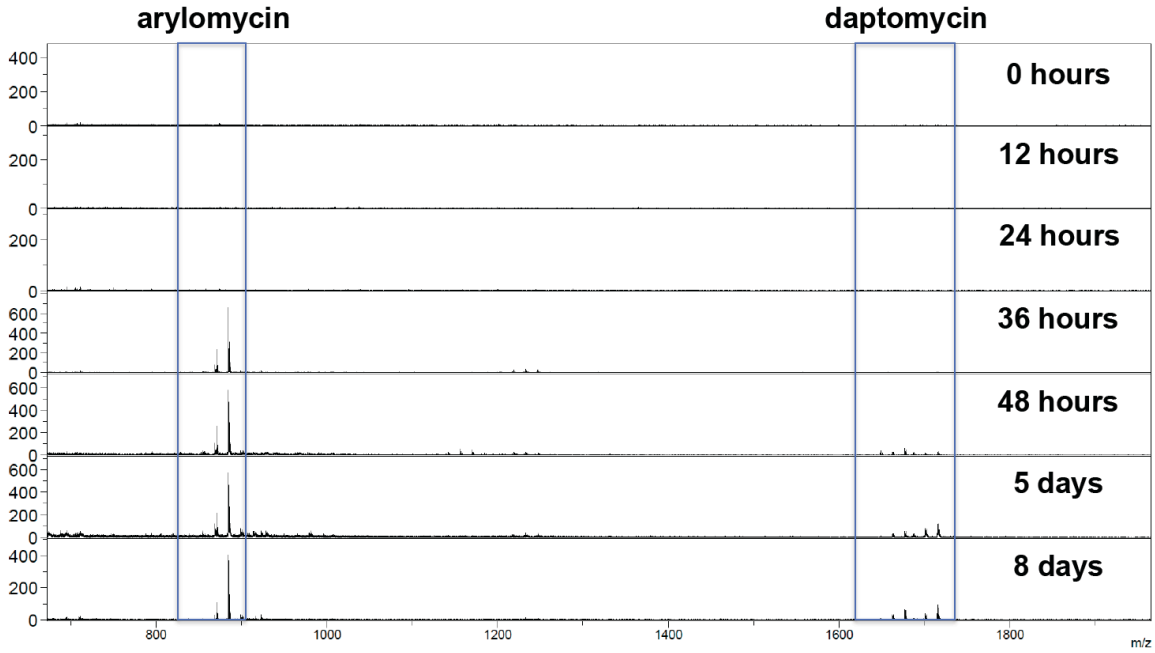


Figure 3.3: Time course for the production of arylomycin and daptomycin variants by *Streptomyces roseosporus.*

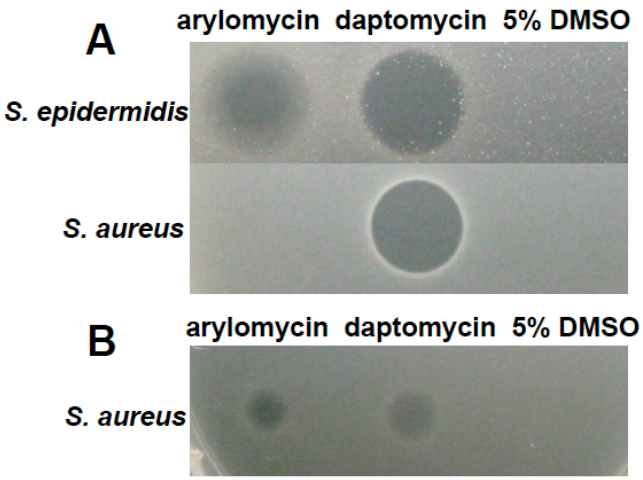


Figure 3.4: Spot assay showing arylomycin inhibit *S. epidermidis* and *S. aureus*. (A) 3µl of purified arylomycin (0.2 µg/ml), daptomycin (0.2 µg/ml), and DMSO (5%) or (B) 10µl of purified arylomycin (0.2 µg/ml), daptomycin (0.02 µg/ml), and DMSO (5%) were spotted on *staphylococcal* (indicated on the left) lawns.

To link to genotypes, our laboratory has recently developed a peptidogenomic mining approach to the discovery of peptidyl natural products[18]. The approach itself relies on the generation of peptide sequence tags from tandem mass spectrometric data to query genomes and to identify the biosynthetic genes. In turn, in an iterative fashion, the biosynthetic gene cluster supports the identification of a peptide as either a ribosomal or non-ribosomal product and facilitates the prediction of a (partial) structure. For ribosomally-encoded peptides, a 5-6 consecutive amino acid residue sequence tag is often needed to successfully match to its precursor gene because of the larger proteomic search space. In this report, we show that for NRPS-derived peptides, this approach could be expanded to short sequence tagging (SST) with only one or two amino acid residues to identify the candidate biosynthetic gene clusters as we suggested would be possible[18]. SST can be employed to carry out genome mining with molecules that are NRPS-derived. This is possible because the search tags can be more specific due to additional non-proteinogenic amino acids and the much smaller query space because of the small number (often <10) of NRPS gene clusters within a microbial genome. This scenario is similar to matching a peptide to a small database in a proteomics experiment where it also becomes possible to match to the correct peptide with minimal fragmentation data while much more fragmentation information is needed when a large database is used. Therefore, even with a very short sequence tag we can still narrow down to the candidate biosynthetic gene cluster. The identification of candidate gene clusters, in turn, aids in the structural characterization of the molecule.

As proof-of-principle for SST, we first demonstrate how this works with daptomycin. The ions at *m/z* 1634.73, 1648.74, and 1662.75, corresponding to

daptomycin variants, were subjected to tandem MS using collision-induced dissociation and resulted in fragment masses at m/z 1051.43, 1166.46, 1280.50 which suggested a sequence tag of Asp-Asn (**Fig. 3.5**). Such a tag provides a minimal search unit that can be searched against all predicted NRPS biosynthetic pathways found on the *S. roseosporus* genome. All three tandem MS datasets of ions at *m/z* 1634.73, 1648.74, and 1662.75 resulted in an identical sequence tags (**Fig. 3.5**). The combination of NP.searcher and NRPS predictor[19,20], two programs designed to identify the amino acid specificity of non-ribosomal peptide synthetases were utilized to predict all possible NRPS gene clusters and their amino acid codes in the *S. roseosporus* NRRL 15998 genome. Seven gene clusters that display NRPS features were found by NP.searcher on the *S. roseosporus* genome. Matching the sequence tag obtained from the tandem MS data of daptomycin variants against the NRPS predictor and NP.searcher predicted amino acids identified the daptomycin gene cluster. Therefore the proof-of-principle experiment with daptomycin variants demonstrated that the correct gene cluster could be identified from the genome through the SST approach.

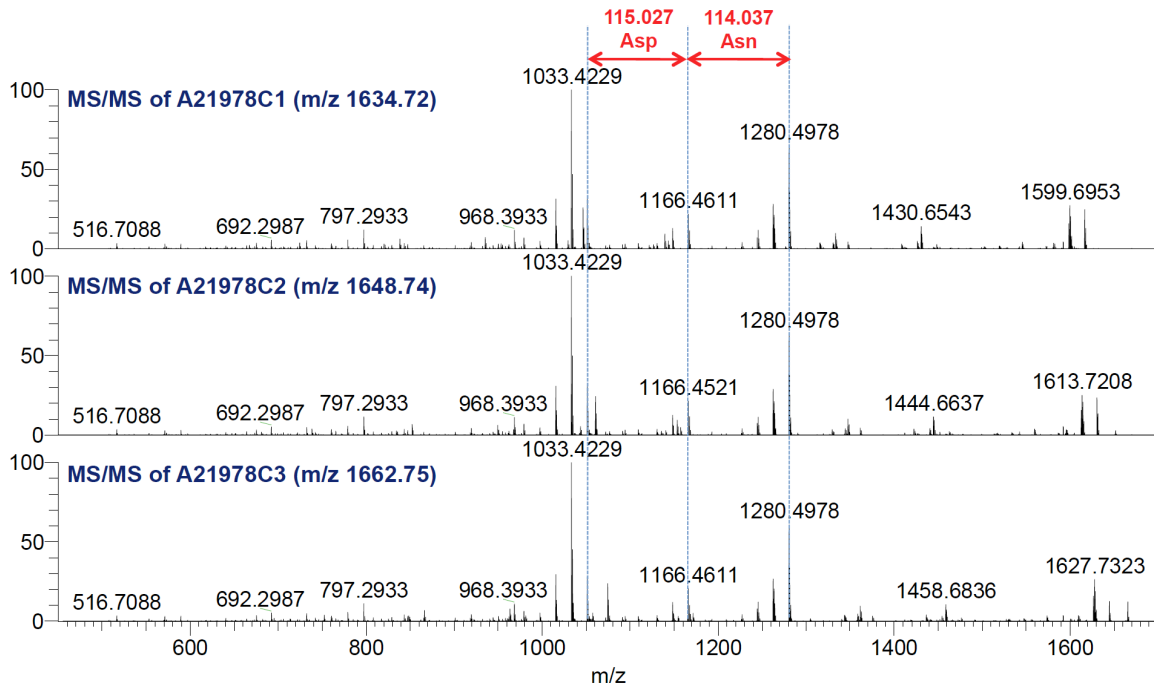
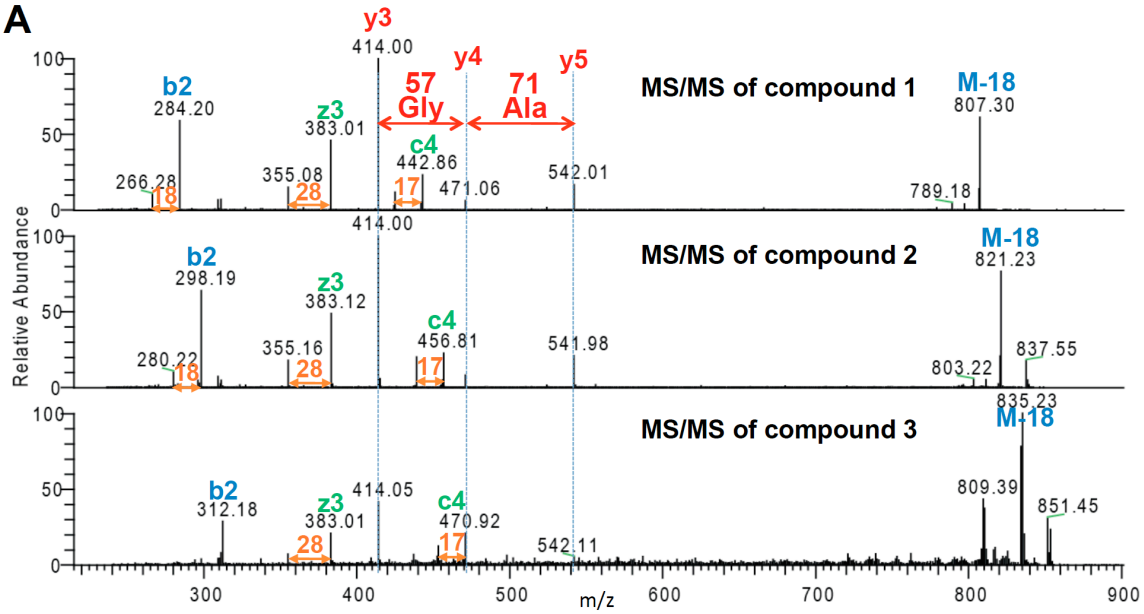


Figure 3.5: Alignment of FT MS/MS of daptomycin A21978C1-C3, the proof-of-principle experiment for SST, revealing the sequence tag Asp-Asn.

Next we set out to identify the series of ions at *m/z* 825.439, 839.455 and 853.471 (**1-3**). These molecules are separated by 14.013 Da consistent with CH_2 mass shifts. Possible explanations to account for this 14 Da difference may arise from different length of fatty acid chain, amino acid substitution, or methylation. Each scenario is commonly found in NRPS biosynthetic pathways. Therefore SST was employed to match these molecules to one of the remaining six NRPS gene clusters. To *de novo* sequence peptides, it is often challenging to separate the ions belonging to y-ion series from the b-ion series. In this experiment however, all three ions were first subjected to low-resolution tandem MS, and the spectra were aligned (**Fig. 3.6A**). This revealed a series of ions that displayed 14 Da mass shifts (shifting ions) and a series of ions that did not display the mass shifts (non-shifting ions). We were able to retrieve a sequence tag from the non-

shifting ions which displayed mass differences of 57 and 71 Da, suggesting glycine and alanine, respectively. There was only one predicted NRPS out of the remaining six NRPS gene clusters that were identified on the *S. roseosporus* genome that contained this sequence tag. That gene cluster is predicted to encode six amino acids, Ser, Ala/Gly, Gly, Hpg, Ala/Gly and Tyr. Although high-resolution MS spectra could provide more unambiguous sequence tag, we show that SST works with high-resolution as well as low-resolution MS data. We consulted the NORINE database that contains greater than 1000 NRPS-derived molecules and enables users to input specific residue(s) to search for molecules that have specified structural units[21]. Searching the NORINE database for the Ser, Ala/Gly, Gly, Hpg, Ala/Gly and Tyr tag resulted in one group of candidate molecules, the arylomycins[22,23].

A



B

	chemical formula	calc. mass	obs. Mass	error (Da)
b1	$C_{12}H_{22}O^+$	183.1743		
b2	$C_{16}H_{30}NO_3^+$	284.2220	284.20	-0.02
b3	$C_{19}H_{35}N_2O_4^+$	355.2591		
b4	$C_{21}H_{38}N_3O_5^+$	412.2806		
y3	$C_{21}H_{24}N_3O_6^+$	414.1660	414.00	-0.17
y4	$C_{23}H_{27}N_4O_7^+$	471.1874	471.06	-0.13
y5	$C_{26}H_{32}N_5O_8^+$	542.2245	542.01	-0.21
y6	$C_{30}H_{39}N_6O_{10}^+$	643.2722		
C4	$C_{22}H_{43}N_4O_5^+$	443.3228	442.86	-0.46
z3	$C_{20}H_{19}N_2O_6^+$	383.1238	383.01	-0.11

C

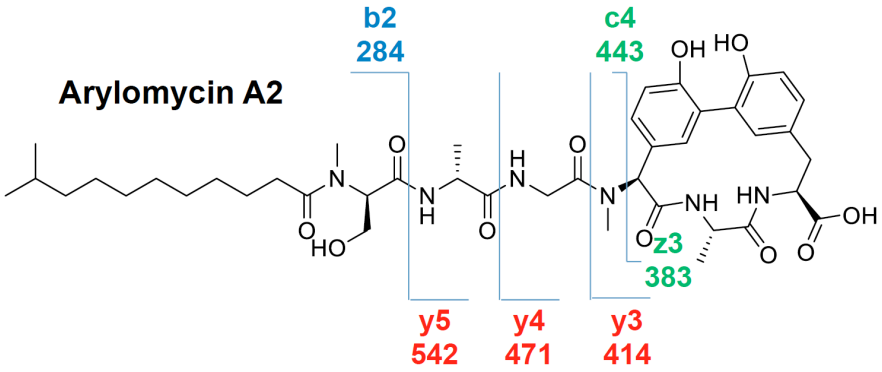


Figure 3.6: Correlate compound 1-3 to arylomycins. (A) Alignment of IT MS/MS of compounds **1-3** revealed sequence tag Gly-Ala. (B) Annotated ion table corresponds to the IT MS/MS of compound **1**. (C) Ion map showing the fragmentation pattern of compound **1** correlates to arylomycin A2.

Having a candidate molecule in hand, the intact masses of arylomycins were compared and the fragmentation data was re-inspected (**Fig. 3.6b and c**). The intact masses of the observed ions matched to the calculated masses of arylomycins within 0.5-2 ppm[22]. The analysis of the fragmentation of compound **1** revealed that the observed b, y and c and z ions were within 1 ppm, in agreement with arylomycin A2 (**Fig. 3.7**).

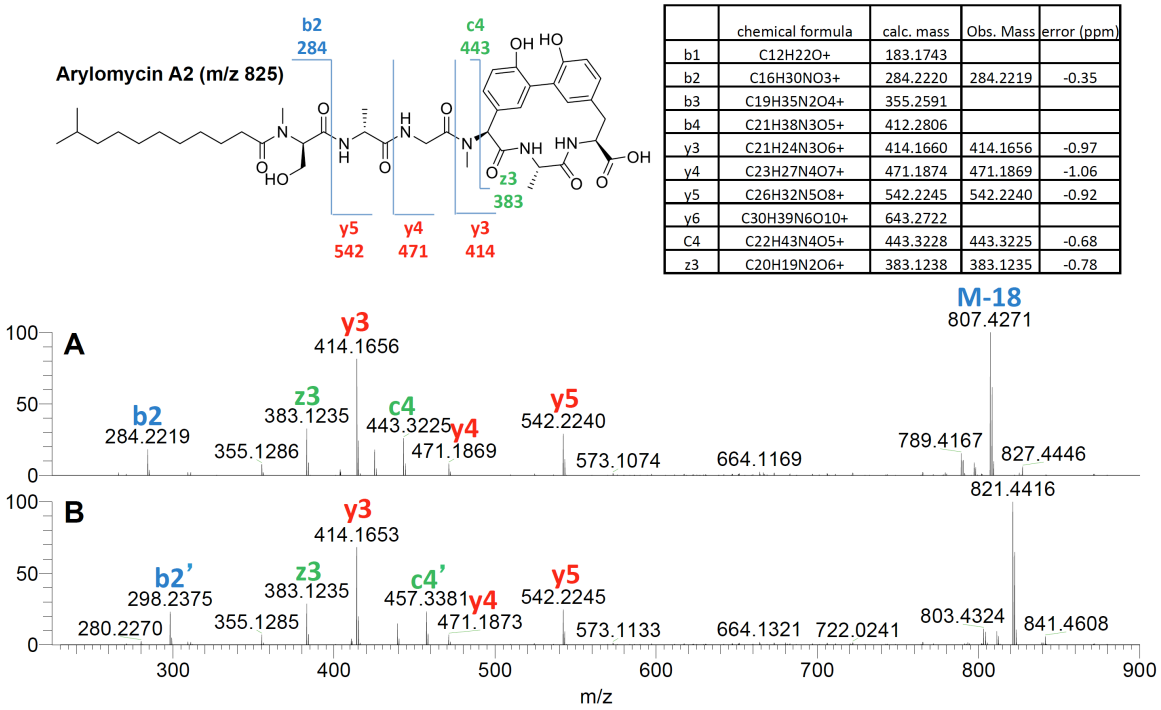


Figure 3.7: Annotated FT MS/MS spectra of compound 1 and 2. (A, B) FT MS/MS of compound **1** (arylomycin A2) and compound **2** (arylomycin A4), respectively. Ion b2 and c4 showed 14 Da shift due to the different length in fatty acid portion.

Arylomycins are an exciting set of biologically active molecules. Arylomycins are a class of broad-spectrum antibiotics that targets type I signal peptidases (SPase)[24-28]. SPase is responsible for the cleavage of signal peptides from secreted protein and is an attractive antimicrobial target because it is highly conserved among bacteria, located on the extracellular surface of the cytoplasmic membrane, and is essential for bacterial

viability. Therefore, it has been suggested that arylomycins as promising therapeutic leads[29,30]. This promise has led to a 2010 start-up company, RQx Pharmaceuticals, that is aiming to develop arylomycin and its analogs into the clinic. Arylomycins were first discovered from *Streptomyces* sp. Tü 6075 isolated in the tropical rain forest at Cape Coast, Ghana[22,23]. Independently, using assays screened for novel signal peptidase I inhibitors, scientists at Eli Lilly & Company reported a similar group of compounds that shares the same skeleton to arylomycins but with a glycosylation on the hydroxyphenylglycine residue[28]. Although natural resistance has been reported[31], arylomycins and its glycosylated congeners are effective against gram-negative bacteria, such as *Helicobacter pylori*, *Yersinia pestis*, and gram-positive bacteria *Streptococcus pneumonia*, *Streptococcus pyrogens, Staphylococcus epidermidis* and *Staphylococcus haemolyticus* with MICs of 4–16 μg/ml[22,23,28,31-34].

To verify that *S. roseosporus* NRRL 15998 produces arylomycins, the candidate arylomycin gene cluster was annotated (**Scheme 3.1, Supporting information section, Table 3.1**) and analyzed for biosynthetic consistency with the proposed arylomycin product. As the candidate gene cluster in the NRRL 15998 strain contained a frameshift and sequencing gaps, we based our analysis on the complete, almost identical gene cluster sequence of *S. roseosporus* strain NRRL 11379, which also produces the same set of molecules (**Fig. 3.8**). Arylomycins contains an N-acyl chain and six amino acids, serine (Ser), alanine (Ala), glycine (Gly), hydroxyphenyl glycine (Hpg), alanine (Ala) and tryrosine (Tyr). Ser1 and Hpg4 are *N*-methylated, Ser1 and Ala2 are in D-configuration and finally Tyr6 and Hpg4 are cross-linked by a biaryl carbon-carbon linkage reminiscent of vancomycin (**Fig. 1**). The gene cluster comprises 10 genes and is

consistent with the arylomycins core structure. The assembly-line NRPS contains 6 modules on 3 genes (*aryABD*) where all A domains have the predicted substrate specificity of the observed amino acids, Ser, Ala/Gly, Gly, Hpg, Ala/Gly and Tyr respectively[35,36]. The loading module has a C domain that clades with starter C domains based on a phylogenetic analysis (**Fig. 3.9**). This C domain incorporates the *N*-acyl group into arylomycins as starter C domains are known to catalyze initial *N*-acylation in NRP biosynthesis[37]. The two *N*-methyl groups at positions 1 (Ser) and 4 (Hpg) in arylomycins are in agreement with the methyltransferase domains in the corresponding NRPS modules. Furthermore, the two D-amino acid residues at positions 1 (D-Ser) and 2 (D-Ala) are consistent with the epimerization domains in corresponding NRPS modules, too. Finally the gene cluster contains a cytochrome P450 enzyme with 49% similarity to the vancomycin OxyC protein that is predicted to form the bis-aryl carbon linkages in vancomycin[38] (**Fig. 3.10**). Thus, AryC is likely to be responsible for the biaryl bond formation in arylomycins. It should be noted that we did not observe the nitrosated congeners that corresponding to the arylomycin B series described in previous reports[22,23] in *S. roseosporus,* which is in agreement with the absence of nitrosating enzymes in the gene cluster[39]. Therefore, based on the annotation of the MS data as well as the annotation of the gene cluster, the data suggest that SST enabled the discovery of the promising anti-infective agent arylomycins from *S. roseosporus.*

Scheme 3.1: Arylomycin biosynthetic gene cluster and proposed biosynthetic pathway.

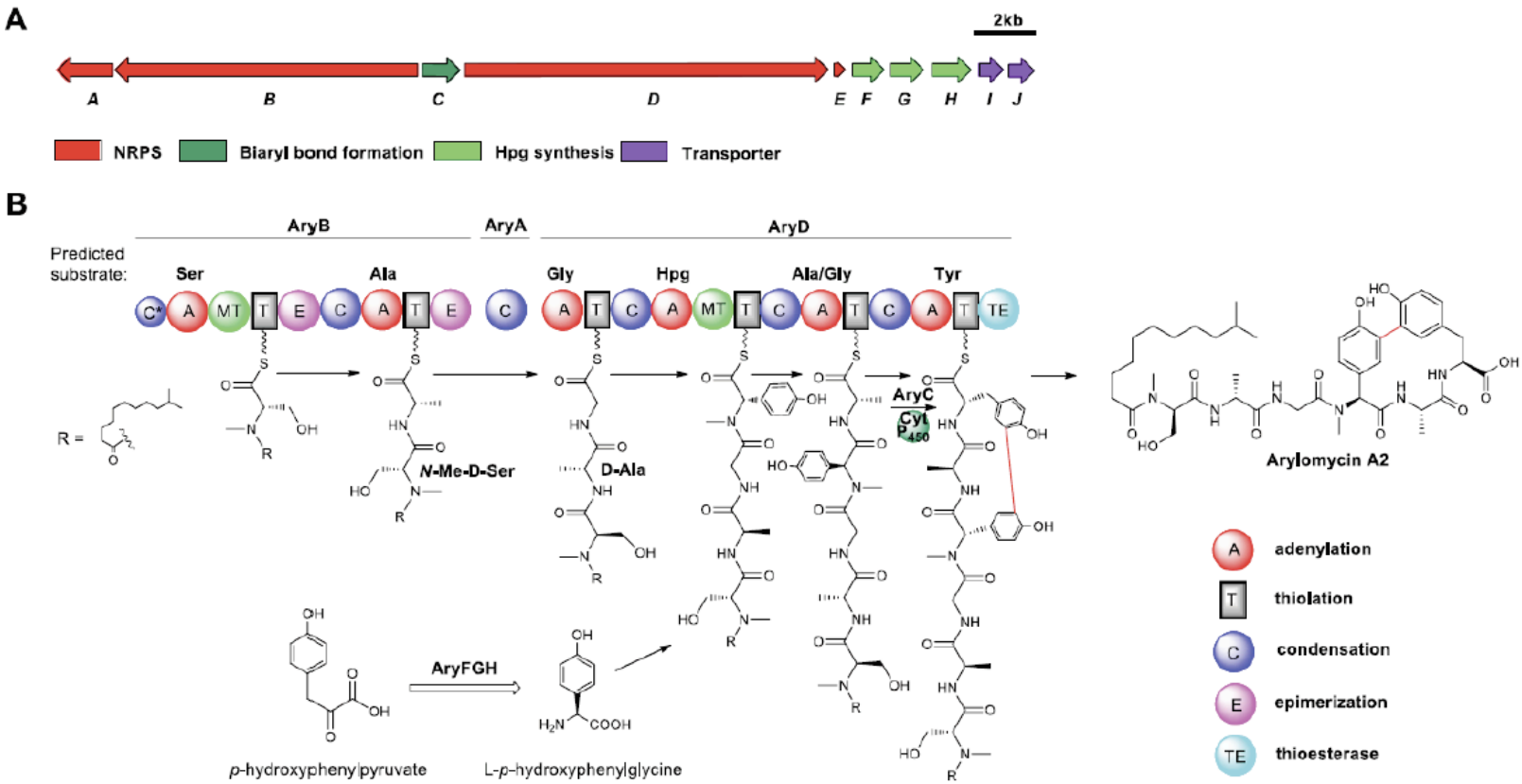


Table 3.1: Genes in the arylomycin biosynthetic cluster from *S. roseosporus* NRRL 11379 and predicted functions based on sequence homology and protein domain analysis.

Gene	Size [aa]	Predicted function	Protein homolog*	Accession number	Protein similarity/ identity [%/%]
aryA	584	NRPS2 (C)	amino acid adenylation domain protein [*Streptomyces violaceusniger* Tu 4113]	ZP_07611402.1	64/51
aryB	3485	NRPS1 (C^S-A-T-MT-E-C-A-T-E)	amino acid adenylation [*Streptomyces ghanaensis* ATCC 14672]	ZP_06575791.1	61/48
aryC	389	Cytochrome P450	cytochrome P450 [*Streptomyces steffisburgensis*]	CAJ42333.1	61/46
aryD	4297	NRPS3 (A-T-C-A-MT-T-C-A-T-C-A-T-TE)	putative pristinamycin I peptide synthase 3 and 4 [*Streptomyces pristinaespiralis*]	CBH31051.1	60/459
aryE	73	MbtH domain-containing protein	MbtH domain-containing protein [*Frankia* sp. EAN1pec]	YP_001510193.1	79/66
aryF	354	4-hydroxyphenylpyruvate dioxygenase	4-hydroxyphenylpyruvate dioxygenase [*Streptosporangium roseum* DSM 43021]	YP_003342414.1	69/57
aryG	371	FMN-dependent α-hydroxy acid dehydrogenase	FMN-dependent alpha-hydroxy acid dehydrogenase [*Streptomyces bingchenggensis* BCW-1]	ADI04246.1	72/61
aryH	4 9	Aminotransferase	aminotransferase class I and II [*Micromonospora* sp. ATCC 39149]	ZP_04604097.1	76/66
aryI	290	ABC transporter	ATP/GTP binding protein NosF [*Streptomyces albus* J1074]	ZP_04704494.1	70/62
aryJ	207	ABC transporter	hypothetical protein SalbJ_21220 [*Streptomyces albus* J1074]	ZP_04704495.1	71/56

* pBLAST analysis of manually annotated ORFs, *S. roseosporus* NRRL 15998 excluded (best homologs for all proteins)

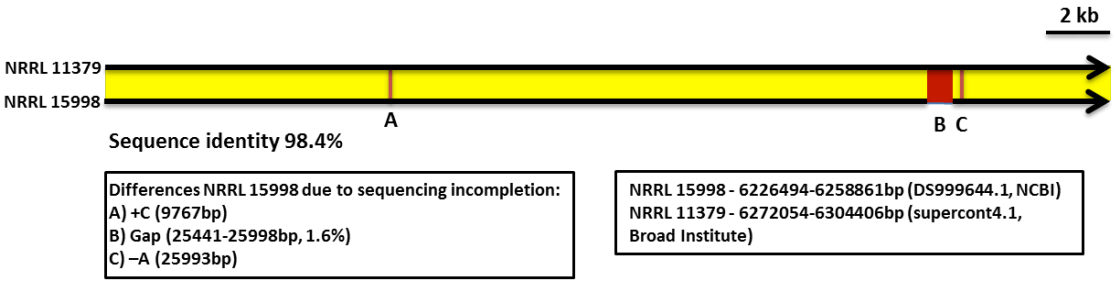


Figure 3.8: Sequence comparison of arylomycin gene cluster region in *S. roseosporus* NRRL 11379 and NRRL 15998. Sequences were compared by nblast-alignment. The total sequence identity was 98.4% due to a frameshift and missing sequencing information in NRRL 15998. The sequence identity in non-gap regions was 100%. Thus, the gene cluster annotation (Table S1) was done on the complete sequence of the NRRL 11379 arylomycin gene cluster region.

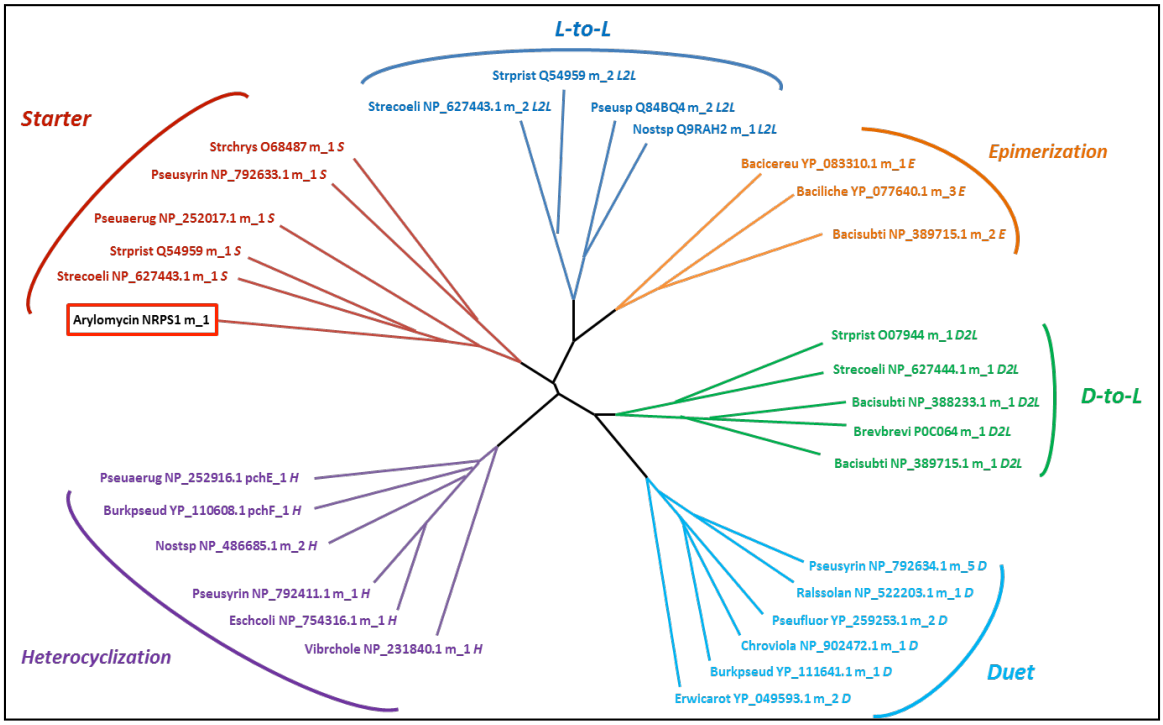


Figure 3.9: Phylogenetic analysis of arylomycin NRPS1m_1 C domain by ClustalW2 analysis. Arylomycin NRPS1 m_1C-domain clusters closest to starter C domains.

```
OxyC            MGHDIDQVAPLLREPANFQLRTNCDPHEDNFGLRAHGPLVRIVGESSTQLGRDFVWQAHG 60
AryC            ----------------------------MYLGGRRG---TEAVGESREPG----VWEVFR 25
                                             :* *     .. ****        **:..

OxyC            YEVVRRILGDHEHFTTRPQFTQSKSGAHVEAQFVGQISTYDPPEHTRLRKMLTPEFTVRR 120
AryC            YDEAVQVLGDHRTFSSDMNHFIPEEQRQLARAARGNFVGIDPPDHTQLRGLVSQAFSPRV 85
                *: . ::****. *::  :.  .:.  ::     *::   ***:**:** :::  *: *

OxyC            IRRMEPAIQSLIDDRLDLLEAEG---PSADLQGLFADPVGAHALCELLGIPRDDQREFVR 177
AryC            TAALEPRIGRLAEQLLDDIVAERGDKASCDLVGEFAGPLSAIVIAELFGIPESDHTMIAE 145
                   :** *  * :: ** : **    .*.** * **.*:.* .:.**:***..*:  :..

OxyC            RIRRNADLSRGLKARAADSAAFN----------RYLDNLLARQRADPDDGLLGMIVRDHG 227
AryC            WAKALLGSRPAGELSIADEAAMQNTADLVRRAGEYLVHHITERRARPQDDLTSRLATTEV 205
                  :   .   . :   **.**::          .** : ::.:** *:*.* . :.  .

OxyC            DN--VTDEELKGLCTALILGGVETVAGMIGFGVLALLDNPGQIELLFESPEKAERVVNEL 285
AryC            DGKRLDDEEIVGVIGMFLIAGYLPASVLTANTVMALDEHPAALAEVRSDPALLPGAIEEV 265
                *.  : ***: *:   :::.*  ..: : .  *:** ::*. :  : ..*     .::*:

OxyC            VRYLSPVQAPNPRLAIKDVVIDGQLIKAGDYVLCSILMANRDEALTPDPDVLDANRAAVS 345
AryC            LRWRPPLVR-DQRLTTRDADLGGRTVPAGSMVCVWLASAHRDPFRFENPDLFDIHRNAGR 324
                :*: .*:   : **: :*. :.*: : **. *   :  *:**     :**::* :* *

OxyC            DVGFGHGIHYCVGAALARSMLRMAYQTLWRRFPGLRLAVPIEEVKYRSAFVDCPDQVPVT 405
AryC            HLAFGKGIHYCLGAPLARLEARIAVETLLRRFERIEIPRDESVEFHESIGVLGPVRLPTT 384
                .:.**:*****:**.***   *:* :** ***  :.:.   .   :.*  *  * ::*.*

OxyC            W---- 406
AryC            LFARR 389
```

Figure 3.10: CLUSTALW2 alignment of Cytochrome P450 AryC and biaryl C-C coupling OxyC from the vancomycin biosynthetic pathway in *Amycolatopsis orientalis*. AryC and OxyC share the signature sequence FGHG*X*H*X*CLG (underlined, red – active site cysteine) of cytochrome P450 enzymes and are 49% similar and 30% identical based on blastp alignment.

To finally confirm the production of the arylomycins lipopeptides from the daptomycin-producing *S. roseosporus* strains, a larger scale fermentation, extraction, and purification using reversed-phase HPLC was undertaken. Compounds **1** and **2** were subjected to NMR analysis. The resulting ^{1}H-NMR spectra revealed the same peptide core as arylomycin by comparing with the NMR spectra described in the original arylomycin report[22,33] (**Fig 3.11**). The ^{1}H-NMR spectra of **1** and **2** matched to arylomycin A2 and A4, respectively[33].

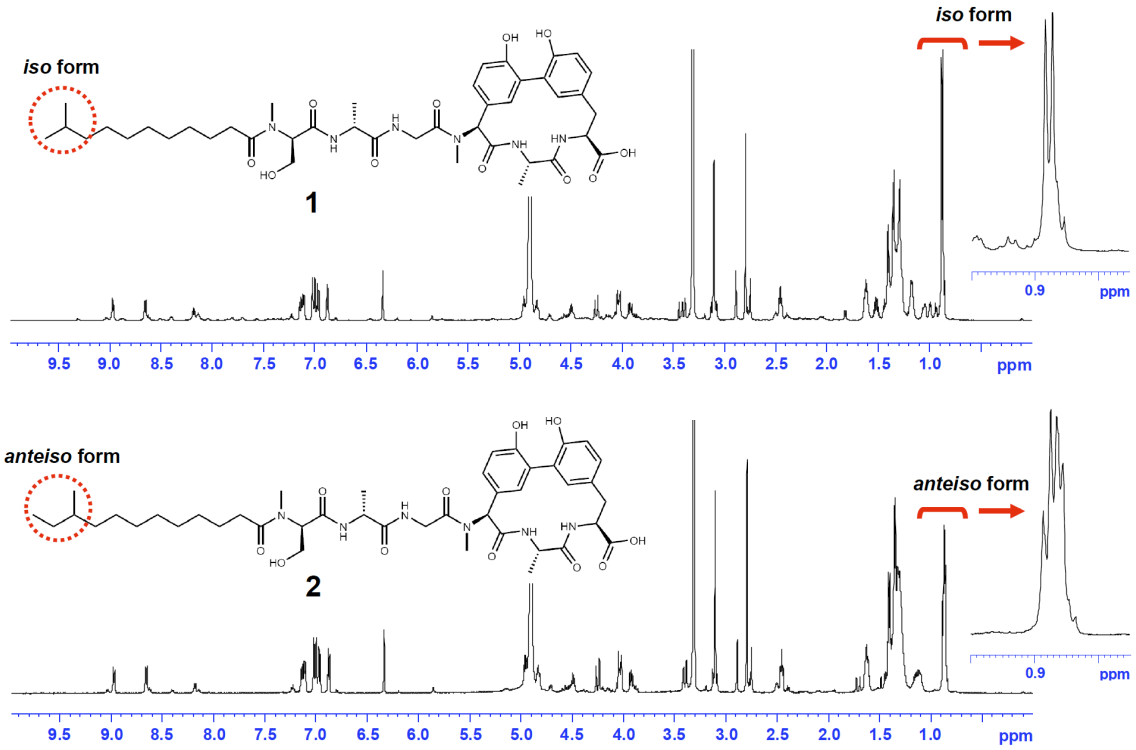


Figure 3.11: ^{1}H NMR spectra (600 MHz, CD_3OD) to confirm the discovery of arylomycins from *S. roseosporus.* NMR spectra are identical to arylomycin A2 (compound **1**) and A4 (compound **2**)[4,]

3.4 Summary

The identification of arylomycins and their biosynthetic gene cluster from intensely studied microorganism of commercial importance highlights the strength of MS-guided genome mining approaches and IMS effectively bridging the gap between phenotypes, chemotypes and genotypes. The SST approach described herein enables matching of molecules identified through imaging mass spectrometry to NRPS biosynthetic machinery using only a minimal sequence tag. We anticipate that SST will also prove capable of identifying the biosynthetic machinery for molecules that contain non-standard amino acids, which are often incorporated in NRPs. According to the NCBI genome database, there are now ~1700 fully sequenced bacterial genomes as assessed in July 2011 opposed to ~1000 in October 2009[40] which represents a more than 70% increase in less than 2 years (and there are ~5000 bacterial genome sequencing projects in progress). Since the full repertoire of genome sequence continues to expand at a rapid pace, there is a need to increase our effectiveness in genome mining to identify new natural products; genome mining of one molecule at a time will not be able to keep up the pace by which genomes are sequenced. Current high-throughput approaches (e.g. metabolomic or proteomic) do not efficiently identify natural products and therefore there is a need to develop genome mining approaches that enables us to rapidly connect chemotypes to genotypes and ultimately phenotypes. Imaging mass spectrometry and/or peptidogenomics, especially with its expansion to smaller peptides and SSTs, are approaches that expedite genome mining for amino acid containing natural products and their connection to phenotypes. Furthermore our findings will enable the discovery of the

arylomycin biosynthetic gene cluster thereby enabling future bioengineering to produce novel arylomycin analogs.

References

1. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the last 25 years. *J Nat Prod.* **70**, 461-477 (2007).

2. Li, J.W. & Vederas, J.C. Drug discovery and natural products: end of an era or an endless frontier? *Science.* **325**, 161-165 (2009).

3. Aharonowitz, Y., Cohen, G. & Martin, J.F. Penicillin and Cephalosporin Biosynthetic Genes - Structure, Organization, Regulation, and Evolution. *Annu Rev Microbiol.* **46**, 461-495 (1992).

4. Miao, V. et al. Daptomycin biosynthesis in Streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology.* **151**, 1507-1523 (2005).

5. van Wageningen, A.M.A. et al. Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol.* **5**, 155-162 (1998).

6. Robbel, L. & Marahiel, M.A. Daptomycin, a Bacterial Lipopeptide Synthesized by a Nonribosomal Machinery. *J Biol Chem.* **285**, 27501-27508 (2010).

7. Payne, D.J., Gwynn, M.N., Holmes, D.J. & Pompliano, D.L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov.* **6**, 29-40 (2007).

8. Fischbach, M.A. & Walsh, C.T. Antibiotics for Emerging Pathogens. *Science.* **325**, 1089-1093 (2009).

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

10. Liu, W.T. et al. Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of Bacillus subtilis. *P Natl Acad Sci USA* **107**, 16286-16290 (2010).

11. Yang, Y.L. et al. Connecting chemotypes and phenotypes of cultured marine microbial assemblages by imaging mass spectrometry. *Angewandte Chemie.* **50**, 5839-5842 (2011).

12. Chambers, H.F. & Deleo, F.R. Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nature reviews. Microbiology.* **7**, 629-641 (2009).

13. Otto, M. Staphylococcus epidermidis--the 'accidental' pathogen. *Nature reviews. Microbiology.* **7**, 555-567 (2009).

14. Gu, J.Q. et al. Structural characterization of daptomycin analogues A21978C1-3(d-Asn11) produced by a recombinant Streptomyces roseosporus strain. *J Nat Prod.* **70**, 233-240 (2007).

15. Baltz, R.H. Genomics and the ancient origins of the daptomycin biosynthetic gene cluster. *The Journal of antibiotics.* **63**, 506-511 (2010).

16. Nguyen, K.T. et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc Natl Acad Sci U S A* **103**, 17462-17467 (2006).

17. Baltz, R.H., Miao, V. & Wrigley, S.K. Natural products to drugs: daptomycin and related lipopeptide antibiotics. *Natural product reports.* **22**, 717-741 (2005).

18. Kersten, R.D. et al. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol.* **7**, 794-802 (2011).

19. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D.H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799-5808 (2005).

20. Li, M.H.T., Ung, P.M.U., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D.H. Automated genome mining for natural products. *Bmc Bioinformatics.* **10** (2009).

21. Caboche, S. et al. NORINE: a database of nonribosomal peptides. *Nucleic Acids Res.* **36**, D326-331 (2008).

22. Holtzel, A. et al. Arylomycins A and B, new biaryl-bridged lipopeptide antibiotics produced by Streptomyces sp. Tu 6075. II. Structure elucidation. *The Journal of antibiotics.* **55**, 571-577 (2002).

23. Schimana, J. et al. Arylomycins A and B, new biaryl-bridged lipopeptide antibiotics produced by Streptomyces sp. Tu 6075. I. Taxonomy, fermentation, isolation and biological activities. *The Journal of antibiotics.* **55**, 565-570 (2002).

24. Paetzel, M., Goodall, J.J., Kania, M., Dalbey, R.E. & Page, M.G. Crystallographic and biophysical analysis of a bacterial signal peptidase in complex with a lipopeptide-based inhibitor. *J Biol Chem.* **279**, 30781-30790 (2004).

25. Bockstael, K. et al. Evaluation of the type I signal peptidase as antibacterial target for biofilm-associated infections of Staphylococcus epidermidis. *Microbiol-Sgm.* **155**, 3719-3729 (2009).

26. Powers, M.E. et al. Type I Signal Peptidase and Protein Secretion in Staphylococcus epidermidis. *J Bacteriol.* **193**, 340-348 (2011).

27. Luo, C., Roussel, P., Dreier, J., Page, M.G. & Paetzel, M. Crystallographic analysis of bacterial signal peptidase in ternary complex with arylomycin A2 and a beta-sultam inhibitor. *Biochemistry.* **48**, 8976-8984 (2009).

28. Kulanthaivel, P. et al. Novel lipoglycopeptides as inhibitors of bacterial signal peptidase I. *J Biol Chem.* **279**, 36250-36258 (2004).

29. Clardy, J., Fischbach, M.A. & Walsh, C.T. New antibiotics from bacterial natural products. *Nature biotechnology.* **24**, 1541-1550 (2006).

30. Butler, M.S. & Buss, A.D. Natural products--the future scaffolds for novel antibiotics? *Biochemical pharmacology* **71**, 919-929 (2006).

31. Smith, P.A., Roberts, T.C. & Romesberg, F.E. Broad-Spectrum Antibiotic Activity of the Arylomycin Natural Products Is Masked by Natural Target Mutations. *Chem Biol.* **17**, 1223-1231 (2010).

32. Smith, P.A., Powers, M.E., Roberts, T.C. & Romesberg, F.E. In Vitro Activities of Arylomycin Natural-Product Antibiotics against Staphylococcus epidermidis and Other Coagulase-Negative Staphylococci. *Antimicrob Agents Ch.* **55**, 1130-1134 (2011).

33. Roberts, T.C., Smith, P.A., Cirz, R.T. & Romesberg, F.E. Structural and initial biological analysis of synthetic arylomycin A2. *Journal of the American Chemical Society.* **129**, 15830-15838 (2007).

34. Dufour, J., Neuville, L. & Zhu, J. Intramolecular Suzuki-Miyaura reaction for the total synthesis of signal peptidase inhibitors, arylomycins A(2) and B(2). *Chemistry.* **16**, 10523-10534 (2010).

35. Challis, G.L., Ravel, J. & Townsend, C.A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol.* **7**, 211-224 (2000).

36. Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol.* **6**, 493-505 (1999).

37. Imker, H.J., Krahn, D., Clerc, J., Kaiser, M. & Walsh, C.T. N-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem Biol.* **17**, 1077-1083 (2010).

38. Pylypenko, O., Vitali, F., Zerbe, K., Robinson, J.A. & Schlichting, I. Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. *J Biol Chem.* **278**, 46727-46733 (2003).

39. Kersten, R.D. & Dorrestein, P.C. Metalloenzymes: Natural product nitrosation. *Nat Chem Biol.* **6**, 636-637 (2010).

40. Walsh, C.T. & Fischbach, M.A. Natural products version 2.0: connecting genes to molecules. *Journal of the American Chemical Society.* **132**, 2469-2493 (2010).

Chapter 3, in full, is a reprint of the material as it appears in JACS, 2011. Wei-Ting Liu, Roland Kersten, Yu-Liang Yang, Bradley Moore and Pieter C. Dorrestein. The thesis author was the primary investigator and author of this paper.

Chapter IV

Revealing the molecular universe of *Streptomyces roseosporus* through MS/MS networking guided genome mining

4.1 Introduction

Since the discovery of penicillin in 1928, scientists have been mining fermentation products from microorganisms in the search for molecules to combat infectious diseases[1-3]. Most (75%) of the anti-infectives that save countless of lives and enormously improve quality of life originate from microbes found in Nature[3]. Meanwhile, due to the emergence of multidrug resistance, such as reports from the CDC that deaths due to *S. aureus* infections have overtaken deaths by HIV[4, 5], it is increasingly recognized that the search for new antimicrobials much continue with renewed vigor[6-9].

Our laboratories have been interested in the development of mass spectrometric and genome mining methodologies with the goals of improving our functional understanding of these molecules and applying these methodologies to therapeutic discovery from natural sources[10-15]. Thus, as opposed to studying one molecule at a time, we aim at a global visualization of molecules from a single microbe, which we define as the 'molecular universe' of that organism. Here, the molecular universe of *Streptomyces roseosporus* is visualized through MS/MS molecular networking[14], in combination with peptidogenomics[12, 13] (**Fig. 4.1**). MS/MS networking is a methodology where molecules that are characterized by mass spectrometry are subjected to fragmentation and are visualized as nodes (circles) and the relatedness of each node is defined by an edge (lines). The thickness of the edge defines the degree of similarity of the MS/MS spectra. The molecules that are described in this paper are summarized in **Fig. S4.1**.

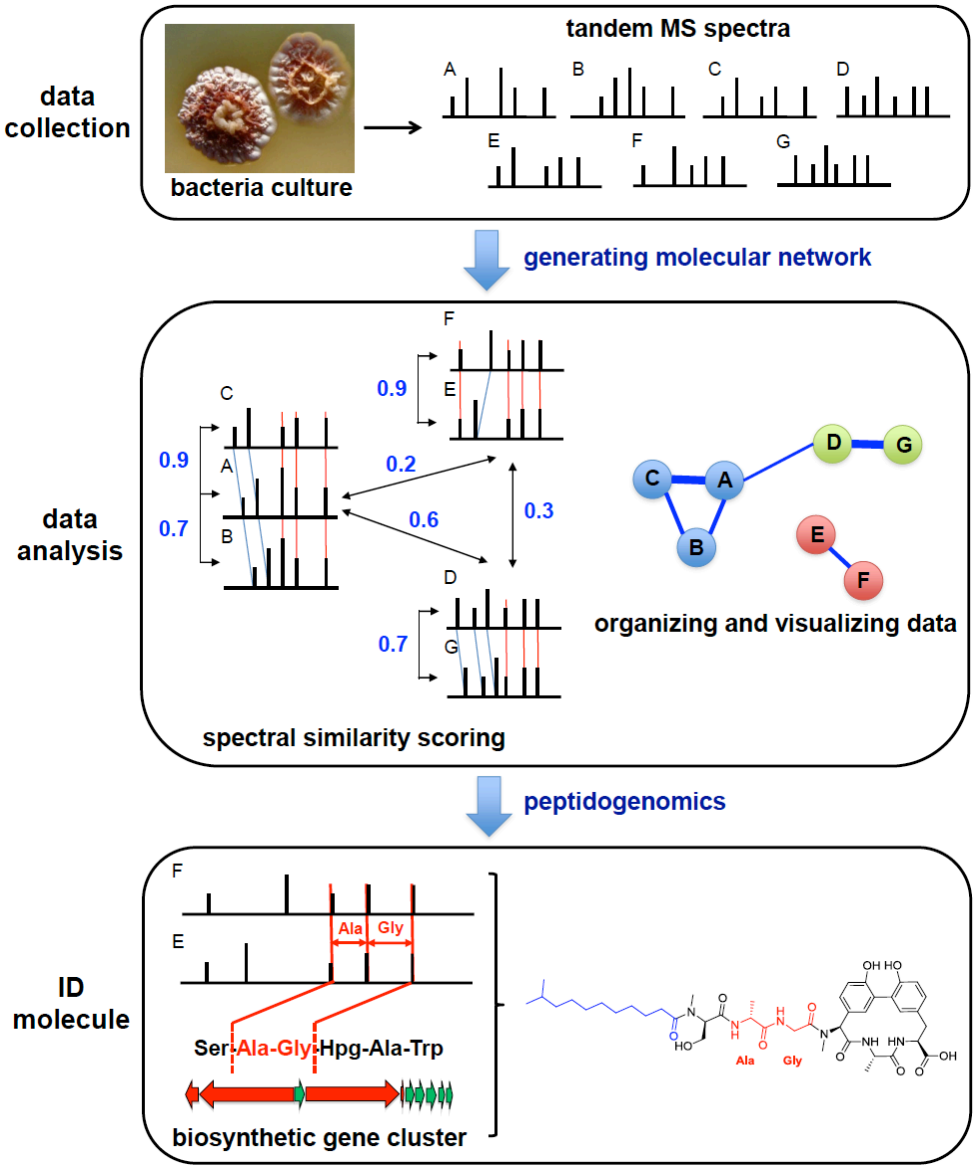


Figure 4.1: General workflow to characterize microbial metabolic profile. Bacteria were cultured and the tandem MS spectra of metabolites were collected. Spectra were aligned with the cosine scores calculating similarity with '1' indicating identical spectra and '0' no similarity at all. A molecular network was thus generated with each node representing a unique spectrum and the thickness of edge representing the cosine score. The spectra correlated to interesting nodes/clusters were subjected to sequence tagging and to a query against genome databases to identify biosynthetic gene clusters. By this process, the structure of molecules were predicted or identified with the information from both MS spectra and genomic information.

4.2 Results and Discussion

To visualize the molecular universe of *S. roseosporus*, extracts from a single petri dish of this microorganism were subjected to direct infusion and LC-MS/MS in a data-dependent manner, as commonly used for untargeted metabolomics analysis[16, 17]. To reduce the complexity of analysis, MS/MS spectra with parent ion masses within 0.3 Da and related MS/MS spectral patterns were merged, subjected to MS/MS networking[14] and displayed in Cytoscape[18, 19] (**Fig. 4.2)**. There are 837 nodes that made up the molecular universe of *S. roseosporus* (**Fig. 4.2**). In this current work, we focused on a few constellations that are composed of cluster of MS/MS spectra, that upon inspection of the MS/MS data had peptidic signatures (**Fig. 4.2 and S4.2**). We first looked for daptomycin and observed 15 nodes that included all four variants along with the natural isotopic forms[20-22] (**Fig. 4.2**, **Fig. S4.2**).

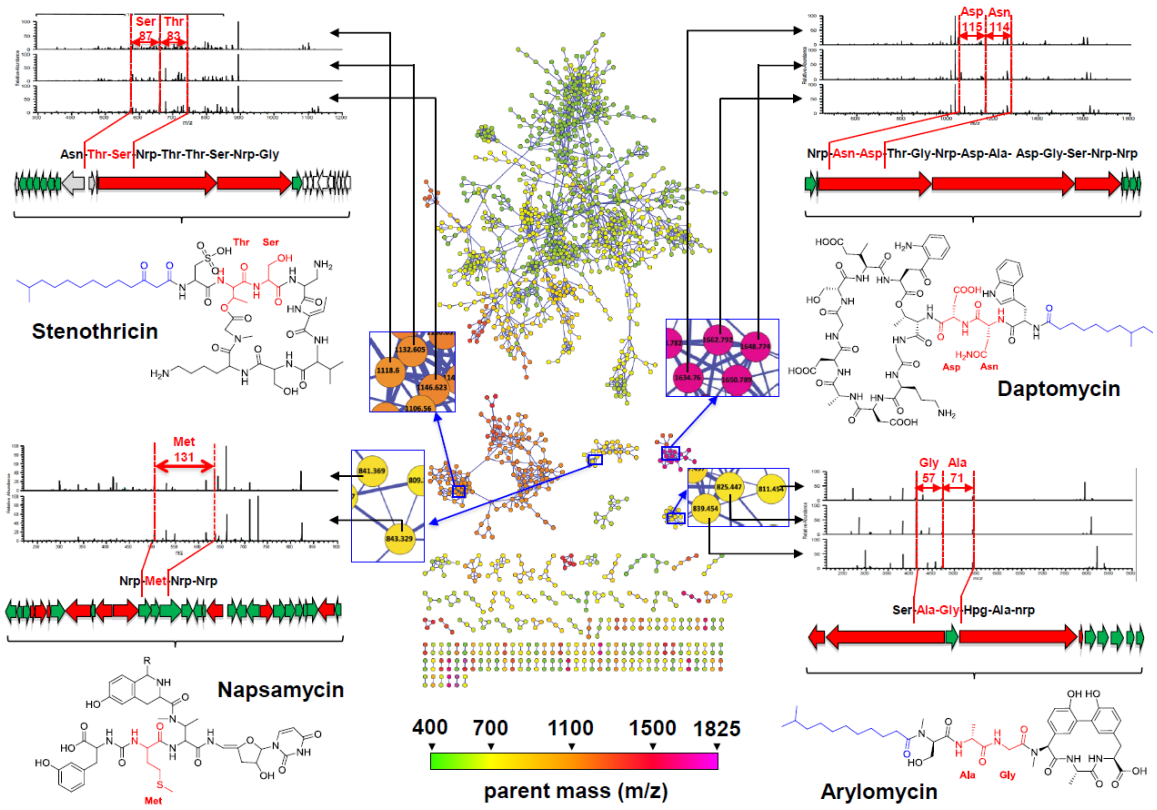


Figure 4.2: Mapping the molecular universe through MS/MS networking and peptidogenomics to reveal the rich metabolic potential of *Streptomyces roseosporus*. Datasets contained MS/MS spectra collected from *S. roseosporus* butanol extract were subjected to molecular networking analysis and revealed several groups of highly clustered ions. Subsequently, analysis of the clusters using SST peptidogenomics identified four classes of NRPs products: daptomycin, arylomycin, stenothricin, and napsamycin.

Unexpectedly, we also observed nodes in the daptomycin constellation that possess much lower molecular weights (*m/z* 1280, 1263) but clearly clustered with known daptomycin variants (**Fig. S4.2**). By comparing MS fragmentation patterns, the structures of these lower molecular weight analogs were determined as molecules that are structurally releated to daptomycin but missing the *N*-terminal lipid chain and tryptophan (**Fig. S4.3**).

We then looked for arylomycin, a molecular family we recently discovered to be

produced by *S. roseosporus*[13]. All arylomycin variants formed their own MS/MS constellation which was composed of 12 nodes (**Fig. S4.2**). In addition, we observed arylomycin Na and K adducts that were located within a branch of the largest constellation (**Fig. S4.4, S4.5**). Within this cluster there were ions that were 161 Da smaller than the parent molecular family. Inspection of the MS/MS data revealed that the -161 Da component is at the *C*-terminal end of the arylomycin family and lacks the biaryl linkage as well as the *C*-terminal tryptophan residue previously seen in an *aryC* knock out strain[23]. It is intriguing that even in wildtype *S. roseosporus*, one can capture such biosynthetic intermediates.

Within the *S. roseosporus* molecular universe, there are two other constellations that have peptidic signatures: one possesses 24 nodes, while the other has 141 nodes. Kaysser et al. suggested *S. roseosporus* has the biosynthetic potential to produce napsamycins by genome analysis, but whether *S. roseosporus* produces napsamycins was not documented[24]. The napsamycins are a group of peptididyl nucleoside antibiotics, and herein, we observed the napsamycin family formed a constellation that is composed of 24 nodes (**Fig. 4.2**, **Fig. S4.2**). The identity of these molecules as napsamycins was confirmed by comparing the MS/MS spectra with those reported in literature[24] (**Fig. S4.6**). Additionally, several higher molecular weight ions were found to cluster within the napsamycin family, and thus comprise as yet unknown napsamycin variants.

The constellation with 141 nodes possessed molecular ions ranging from *m/z* 1090 to 1473 and included the following series of fragment masses; *m/z* 1090, 1104, 1118, 1132, 1146, 1160, 1174, 1188 (**Fig. S4.2**), suggesting a high likelihood that they were non ribosomal peptide synthetase (NRPS) derived. Therefore, we aligned the MS/MS

spectra of these ions to reveal potential sequence tags, and matched these to the amino acid building blocks that are predicted to be loaded by adenylation domains of the NRPS gene clusters found by antiSMASH[25-27] (**Fig. 4.2**). Of the seven NRPS gene clusters found in the *S. roseosporus* genome, four of these are short NRPS gene clusters containing less than 4 modules, and thus are less likely to be responsible for forming a peptide of mass near 1100 Da. Two of the remaining three larger NRPS gene clusters were assigned as the daptomycin and arylomycin gene cluster[13, 22]. MS/MS spectra revealed a potential sequence tag 151-83-87-86-83 (**Fig. S4.7A**). Because the 83 Da mass could arise from dehydrobutyrine (derived from threonine), the sequence tag could be translated into X-Thr-Ser-X-Thr which partially matched the putative sequence of the third gene cluster (Asp-Thr-Ser-Dab-Thr-Thr-Ser-Met-Gly) predicted by the Stachelhaus A-domain specificity rules[28]. Searching all of the molecular information we had in the Antimarine© database resulted in a candidate molecule. The match was to stenothricin[29-31], which was discovered in 1974 from a soil *Streptomyces* sp. obtained from Zahedan, Iran[30]. Four variants have been reported, three of them were included in Antimarin[29, 31]. Correspondingly, four of the 141 nodes in the stenothricin constellation have *m/z* values of 1118.61, 1132.61, 1146.62, 1160.64 that are within 2 ppm of the calculated mass for these four variants. MS/MS fragmentation of the molecules correlated well with the literature reported stenothricin structures (**supporting info**, **Fig. S4.7, Fig. S4.8**).

In addition to the MS/MS data, the gene cluster with 21 predicted genes also agrees with the stenothricin structure (**Fig. 4.3, Table S4.1**). Stenothricin contains an *N*-acyl chain and nine amino acid residues; cysteic acid, threonine, serine, 2,3-diaminopropionic acid, dehydrobutyrine, valine, serine, lysine, and glycine. The *C*-

terminal glycine is *N*-methylated, and forms a macrocycle involving Thr2. Lastly, Thr5 is dehydrated to a dehydrobutyrine. The NRPS assembly-line contains nine NRPS modules on three genes (*stenPST*). The putative amino acid stereochemical configuration can be inferred by *C*-domain phylogenetic clustering[32] (**Fig. S4.9**). The *N*-methyl group at the *C*-terminal glycine is in agreement with the methyltransferase domains in the corresponding NRPS modules. Stenothricins contain two unusal amino acids, CysA and Dpr. The genes to make these residues are also present in the gene cluster (**Fig. S4.10-4.11**). Thus, both the MS/MS and the gene cluster matched to the structure of stenothricin. Finally, NMR data also supported the identification of this group of molecules as being in the stenothricin family of molecules (**Supporting info**, **Fig. S4.12-4.18, Table S4.2**). In addition, the large number of nodes in the stenothricin constellation revealed a substantial number of analogs with different sized lipid side chains, amino acid substitutions, hydrolysis products, and nodes that were 162 or 324 Da larger which are mass shifts often correlated with glycosylation[27, 33].

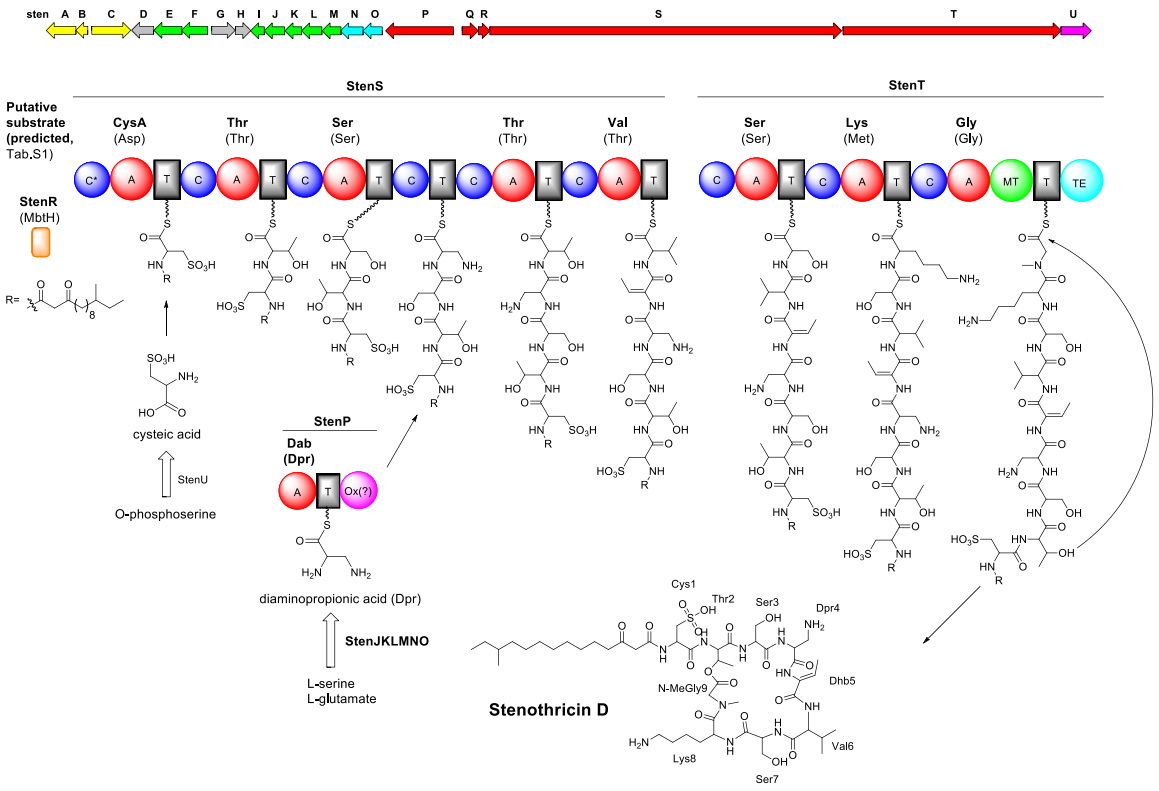


Figure 4.3: Proposed biosynthetic pathway of stenothricin

Lipopeptides often exhibit promising anti-microbial activities[34]. Significant examples are daptomycin and polymyxin that are well known anti-infective agents[35, 36]. With the limit availability of purified stenothricin single form, we subjected stenothricin variants mixture to BioMAP, a screening platform that tests for anti-microbial activity against 15 bacterial strains to generate a antimicrobial profile[37]. Contrary to published results[30], stenothricin is active against both Gram negative and Gram positive bacteria including *Bacillus subtilis* as well as a panel of clinically important pathogenic bacteria such as *Staphylococcus epidermis*, *Enterococcus faecium*, *Listeria ivanovii*, *Enterobacter aerogenes*, and *Acinetobacter baumannii* with IC_{50} values ranging from 1 to 27 µM. These are emerging pathogens which cause severe pneumonia, urinary tract infections,

and septicemia[38-40] (Figure S19). The antimicrobial profile of stenothricin was compared to 72 established profiles of known drugs covering 12 major activity classes. The results indicated that stenothricin has a distinct activity profile from these 72 known antibiotics, suggesting that it may exhibit a different mode of action compared to the major classes of antibiotics that include cell wall, protein, RNA, and nucleic acid synthesis inhibitors, DNA-damaging agents, and additional targets[37] (**Fig. 4.4A**).

The unique antibiotic signature of stenothricin in BioMAP suggests that it may possess a novel mode of action. To gain further provide insight into stenothricin's mode of action, we employed cytological profiling[41, 42]. The IC_{50} of stenothricin to *B. subtilis* was determined in BioMAP to be 4.56 μM. With the availability of singly purified stenothricin form D, which have the same activity as the mixture of stenothricin variants used in the BioMAP screen (**Fig. S4.20**), the biological effects on *B. subtilis* cell growth and architecture were evaluated. Treatment with 18 μM stenothricin resulted in cells with membranes that appear to have detached from their peptidoglycan cell wall. This was observed at the interface between two cells around the region of the septa. With stenothricin treatment, the septa can be seen as an unstained gap between the cells and membrane staining throughout the cell, a pattern not observed in the controls (**Fig. 4.4F**). At 1.8 μM, cells show no obvious phenotype and only a slight decrease in growth rate (**Fig. 4.4C,D**). Comparing stenothricin treated cells to those treated with daptomycin, there was no similarity in the resulting phenotype (**Fig. 4.4G**). Cells treated with 4.4 μM of stenothricin have a different appearance than those at 18 and 8.8 μM, with large membrane blobs or internal vesicles at their septa, similar but not identical to those seen with vancomycin or detergent treatments (**Fig. 4.4I,J,K**). In summary, none of the

antibiotics that have been tested in our laboratories thus far shows a phenotype similar to that given by stenothricin in either the BioMAP and cytological profiling assays (some representative comparisons are shown in **Fig. 4.4L**)[43], and thus, it appears that stenothricin is unique in its mechanism of antibiotic activity.

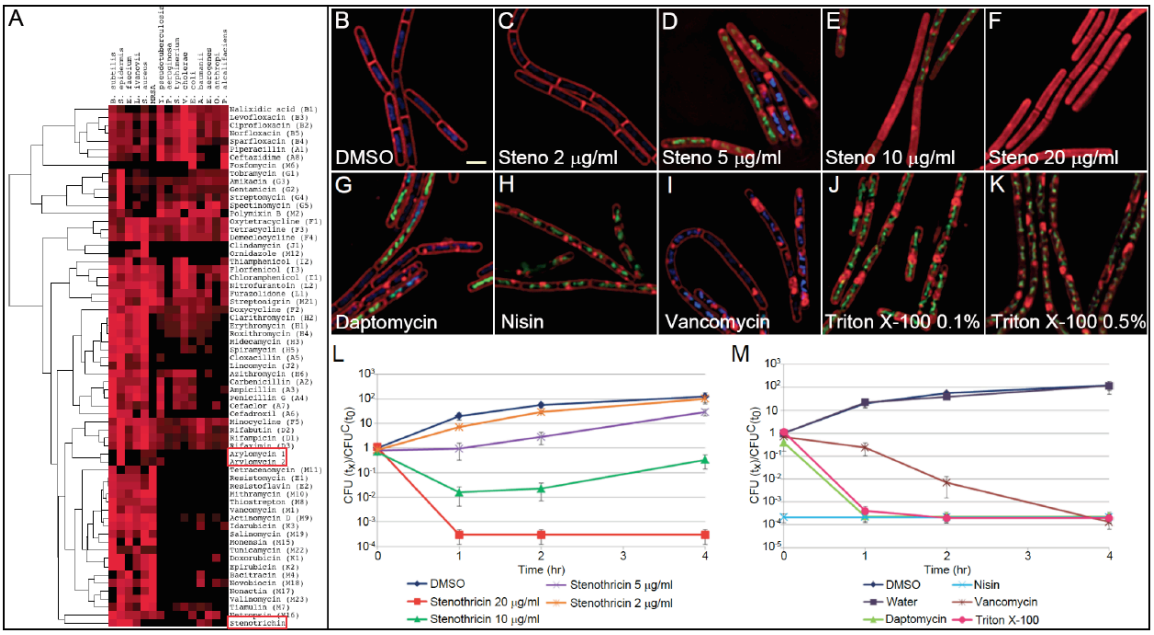


Figure 4.4: Bioactivities of stenothricin and other antimicrobials. (**A**) BioMAP profiling compared the antimicrobial activities of stenothricin and arylomycin with 72 known antibiotics covering 12 major classes. Antimicrobial potencies are represented by a red-black color scheme with a gradient from inactive (black) to most potent (red). Antibiotics are alphanumerically labeled as A to L according to structural class. Code M is assigned to antibiotics that are single derivatives without other family members as described in Wong et al. (**B-K)** Fluorescence micrographs revealed the effect on cell architecture of *B. subtilis* when treated with (**B**) 0.5% DMSO (**C**) 2 μg/ml (1.8 μM) stenothricin D (**D**) 5 μg/ml (4.4 μM) stenothricin D (**E**) 10 μg/ml (8.8 μM) stenothricin D (**F**) 20 μg/ml (18 μM) stenothricin 1132 (**G**) 10 μg/ml (6.2 μM) daptomycin (**H**) 6 μg/ml (1.8 μM) nisin (**I**) 1.25 μg/ml (1 μM) vancomycin (**J**) 0.1% Triton X-100 (**K**) 0.5% Triton X-100 for 20 minutes. Cells are stained with FM 4-64 (red), DAPI (blue), and SYTOX green. SYTOX green is membrane impermeable and only stains cells with compromised membranes. Scale bar represents 2 μm. (**L**) Effects of different amounts of stenothricin 1132 on PY79 cell viability. (**M**) Effects of control treatments on PY79 viability. Cell viability is shown as the ratio of colony-forming units (cfu) at the indicated time and treatment to the cfu at t_0 for the control (cfu^C). Error bars show the standard error of ≥3 experiments.

The above results show that *S. roseosporus* produced a large panel of NRPS-derived antibiotics, including previously unreported analogs in known series of compounds involving truncations, glycosylations, hydrolysis products, and biosynthetic intermediates and/or shunt products and that all can be visualized using MS/MS networking (**Fig. S4.3**). Specifically, MS/MS networking highlights that *S. roseosporus* produces at least four NRPS-derived molecules with different modes of antibiotic action. Daptomycin acts by disruption of membrane potential[44], arylomycin as a type II signal peptidase inhibitor[13, 45], napsamycin as a bacterial translocase I inhibitor[24], and stenothricin displays an antimicrobial activity different from any known antibiotic tested in this study. Overall, this study highlights the importance of developing new approaches that are able to visualize the 'molecular universe' of an organism and thus more deeply interrogate the biosynthetic capacities of microorganisms.

4.3 Supporting Information

4.3.1 MS/MS fragmentation pattern correlates well with stenothricin structure

MS/MS fragmentation of the molecules we observed were used to compare with the literature reported stenothricin structure. However, several of the major fragments of these ions do not readily match the stenothricin structure based on conventional cyclic peptide fragmentation pathway so we set out to understand this discrepancy[10, 46]. Stenothricin contains a lactone as a result of macrocyclization of an internal threonine attacking on the C-terminus, a reaction catalyzed by the thioesterase domains. Because of the lactone, it was likely that there was a rearrangements which would end up with the same end products as a Mclafferty type of rearrangement opening up the lactone[10, 47, 48] (**Fig. S4.7**, **S4.8**). With the arrangement mechanism taken into account, the major fragment ions could be explained and the MS/MS data is in full agreement with these molecules being stenothricin variants.

4.3.2 Biosynthesis of CysA and Dpr in stenothricin

Stenothricin contain two unusal amino acid, CysA and Dpr. CysA is found as the first residue at the N-terminal. Although having been described as building blocks in a few natural products such as corticiamide A, and oriamide isolated from marine sponges, the enzymology behind the incorporation of this building block into NRPS type molecules has not been described[49, 50]. StenU shares homology to cysteate synthate, a PLP (pyridoxal 5'-phosphate) dependent enzyme that converts phosphoserine to cysteic acid through a β-elimination of phosphate and then the β-addition of sulfite to produce cysteate found on the coenzyme M biosynthesis pathway in *Methanosarcina*

acetivorans[51]. Therefore, we proposed StenU catalyzes the biosynthesis of cysteic acid from phosphoserine as subsequently loaded onto the thiolation domain. The substrate specificity 10 aa code (DATKMGHVGK) of the first A domain in StenS was predicted to encode for an asparatic acid. Asparatic acid shares structural similiary to cysteic acid. It is therefore not unreasonable that based on the position and structural similarity that the cysteic acid is loaded by this first A domain in the assembly line. Although the A domain is predicted to load Asp, the A domain is very selective for cysteic acid as none of the 141 nodes in the stenothricin constellation come from Asp. The identification of the stenothricin biosynthetic gene cluster would facilitate further characterization the chemistry behind the incorporation of cysteic acid in NRPS.

The other unusual residue Dpr is found in the tuberactinomycin family of NRPS derived antibiotics that are essential components for the treatment of multidrug-resistant tuberculosis[52]. The biosynthetic pathways of some of the members within the tuberactinomycin family have been characterized. In the case of capreomycin and viomycin, the biosynthetic pathways contained enzymes that share 54%, 56% and 42%, 46% homology with StenN and StenO[52, 53]. The functions of these homologous enzymes have been elucidated as 2,3-diaminopropionate synthase and orithine cyclodeaminase. Therefore, we expected StenN carries out the biosynthesis of Dpr, but also requires a nucleophile generated by StenO. This nucleophile was generated while catalyzing the conversion of L-ornithine to L-proline (**Fig. S4.12**). A set of enzymes to make L-ornithine/L-arginine StenEFIJKLM are also closely clustered within the stenothricin biosynthetic gene cluster (**Fig. 4.3**, **Fig. S4.13**, **Table S4.1**). However, the arginine biosynthetic genes are not complete in this cluster which supports the hypothesis that the

product is ornithine for Dpr biosynthesis. The missing arginine biosynthetic genes are somewhere else in the genome. It is interesting to speculate that Arg biosynthesis might be translationally coupled to stenothricin production as the Arginine biosynthesis regulator ArgR is in the gene cluster. In terms of stenothricin biosynthesis, it is predicted that StenP selects Dpr which is loaded in trans onto module 4 of StenS. The A domain code of StenP is identical to NpsP5 or PacP from the napsamycin or pacidamycin gene cluster which loads diaminobutyric acid (Dab)[24, 54]. Although the high abundant stenothricin analogs we observed incorporated a Dpr residue as judged by MS/MS fragmentation pattern, benefiting by the ultrahigh sensitivity of MS, we were able to pull out some of the analogs that have tha Dpr residue instead of Dab residue. This led us to postulate StenP can load both Dpr and Dab substrates.

Although we did not determine the stereochemistry of stenothricin by Marfey's analysis as we do not have enough materials, the amino acid configuration can be inferred by phylogenetic clustering described above. Toward this end, the rest of C domains in stenothricin assembly lines were subjected to phylogenetic analysis, and were clustered with either a L to L ($^{L-}C_L$), or Dual E/C domain. Basesd on this analysis, we anticipated that most of the residues in stenothricin is in D-configuration with only Val 6, Lys 8, and Gly 9 in L-configuration (**Fig. S4.9**).

4.3.3 Structural verification of stenothricin with NMR spectroscopy

Structure determination based on the NMR spectra compared favorably to the structure proposed by biosynthetic analysis and mass spectrometry fragmentation pattern analysis (**Fig. S4.14-4.20**, **Table S4.2**). Comparison of the twelve spin systems revealed

by the TOCSY spectrum to HSQC and HMBC spectra suggested residues consistent with the proposed structure of stenothricin, namely cysteic acid, threonine, two serines, 2,3-diaminopropinoic acid, dehydrobutyrine, valine, lysine, and glycine. Efforts to determine the sequence of residues by NMR analysis, however, were frustrated by two difficulties. First, the loss of exchangeable amine, amide, alcohol, and β-keto amide enol protons to the solvent prevented these species from showing strong 2D-NMR correlations. In addition the presence of a pair of equilibrating *N*-methyl amide rotomers of *N*-methyl-glycine, which because of the constrained rotation of the cyclic structure cause rotomer peaks for the adjacent functional groups as well, specifically the methyl group of threonine, and the α-carbon of lysine.

Though not complete, some sequence information could be derived from the HMBC spectrum. A strong HMBC correlation between the α-carbon of valine and the carbonyl of dehydrobutyrine confirmed the connection of these residues by an amide bond. In turn the vinyl methyl of dehydrobutyrine showed HMBC correlations to the carbonyl carbon of the 2,3-diaminopropinoic acid residue, confirming an amide bond linked these residues. Unfortunately, correlations were not seen continuing on from either the 2,3-diaminopropinoic acid, or the valine. Separately, the β-carbon of threonine showed an HMBC correlation with the carbonyl carbon of glycine, which helped to confirm that the structure was cyclized onto the threonine alcohol. The *N*-methyl amide protons of glycine showed and HMBC correlation with the carbonyl carbon of lysine, which continuing from the lysine residue, showed an HMBC correlation to from the α-carbon to the adjacent carbonyl carbon within one of the serine residues. Other HMBC correlations that might have confirmed the connectivity of all of the residues, were not

detected above the baseline. As such the NMR spectra alone are insufficient to determine the sequence of residues in stenothricin, but strongly support the structural assignment arrived at by the orthogonal methods of MS/MS fragmentation analysis, and biosynthetic gene cluster comparison.

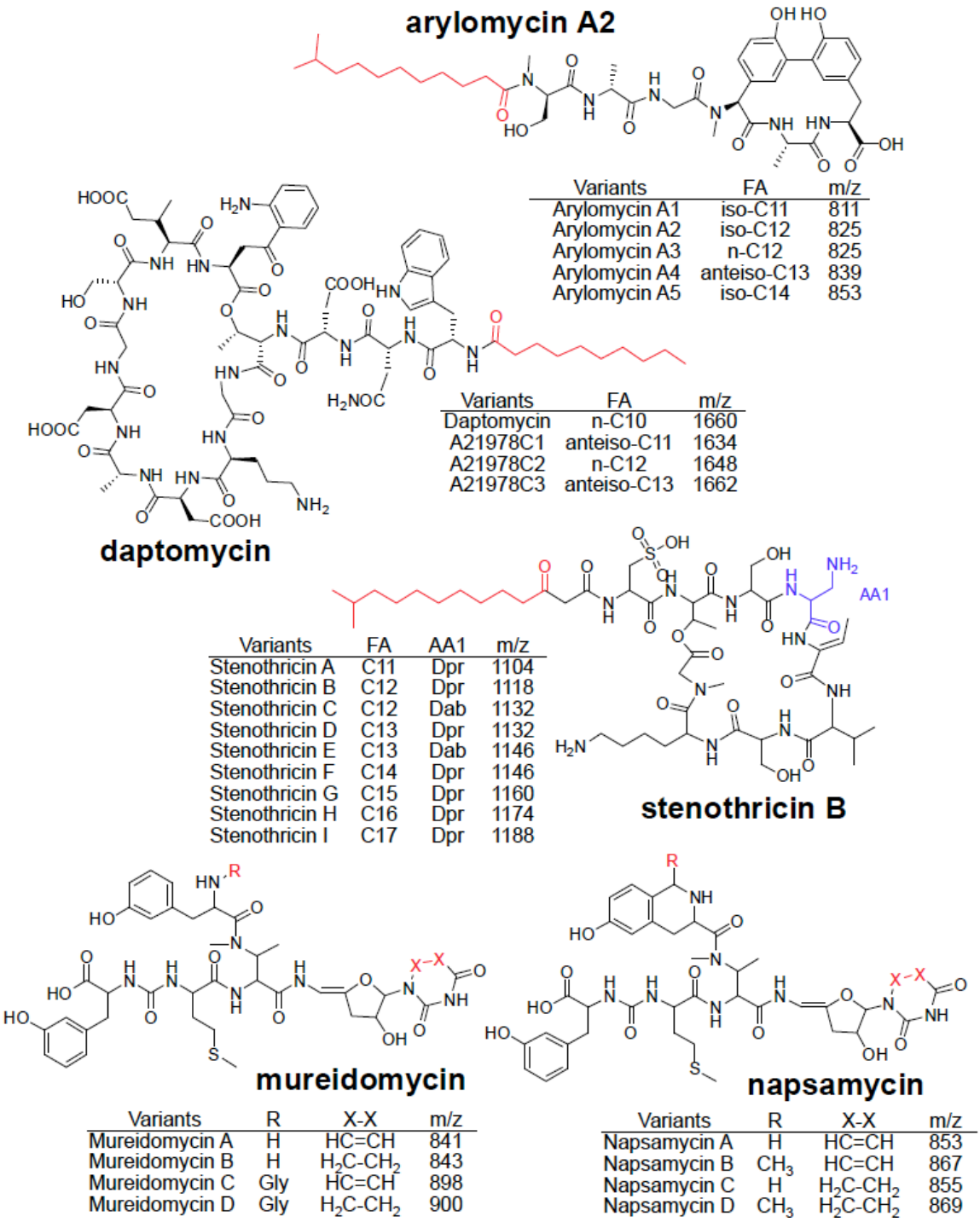


Figure S4.1: Molecules produced by *S. roseosporus* that are described in this paper.

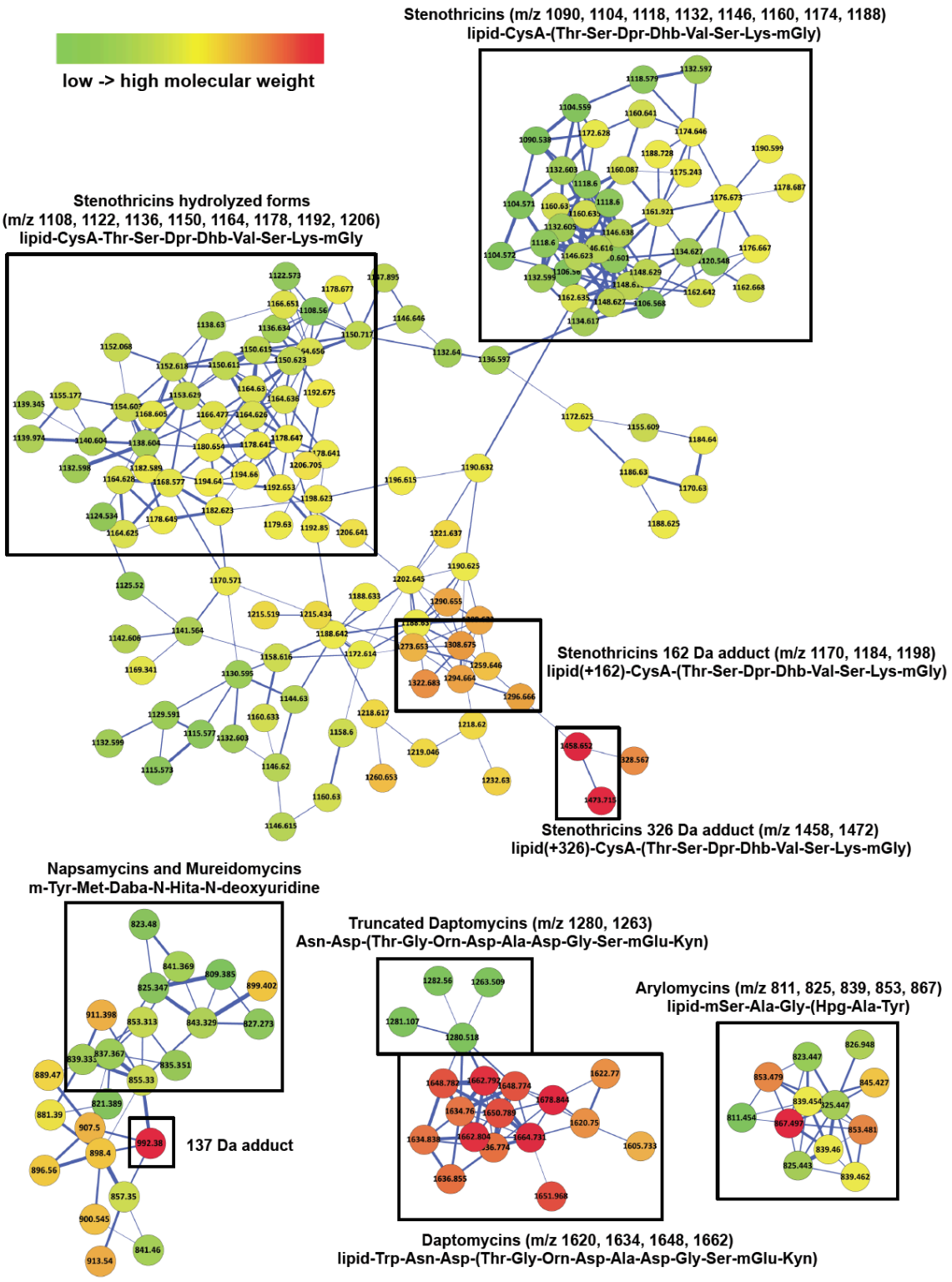


Figure S4.2: Zoom in of the molecular universe on the major classes of molecules that are described in this work.

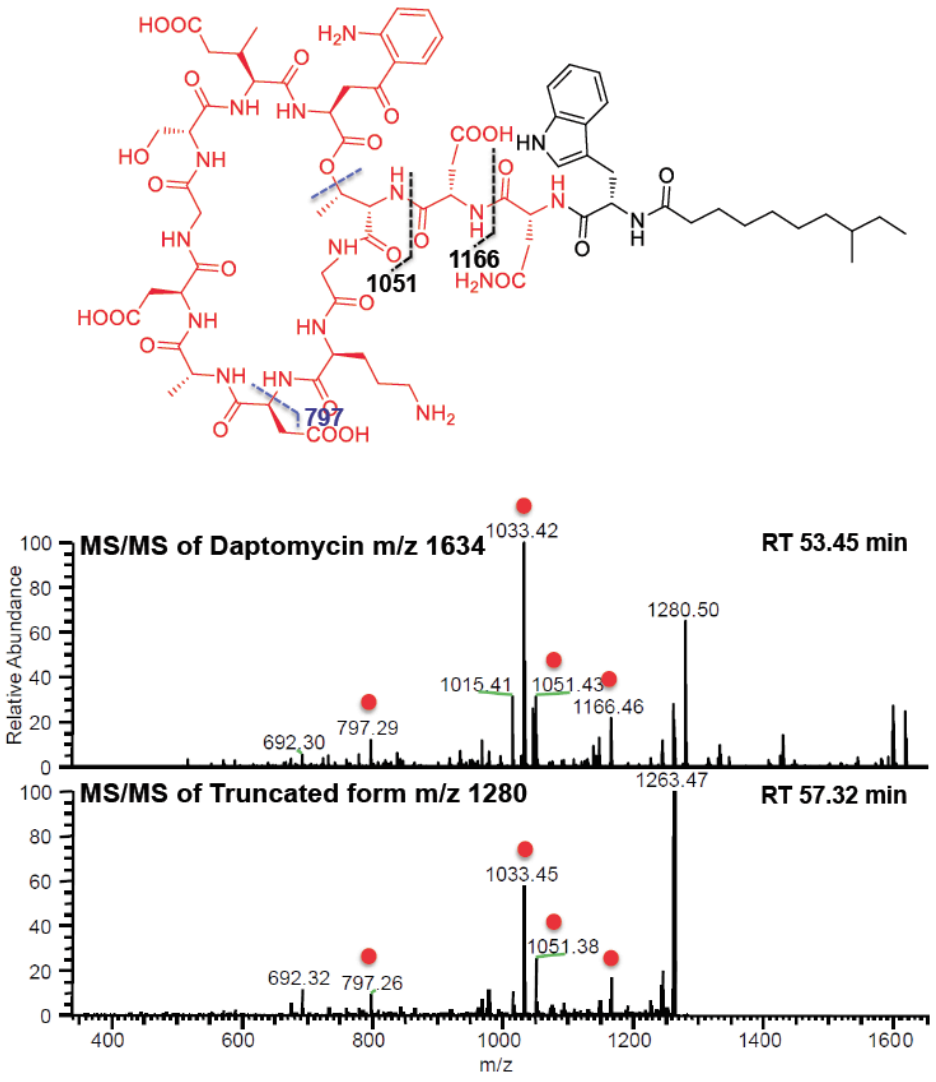


Figure S4.3: MS/MS fragmentation suggested ion at m/z 1280 as daptomycin truncated form that misses N-terminal lipid chain and tryptophan.

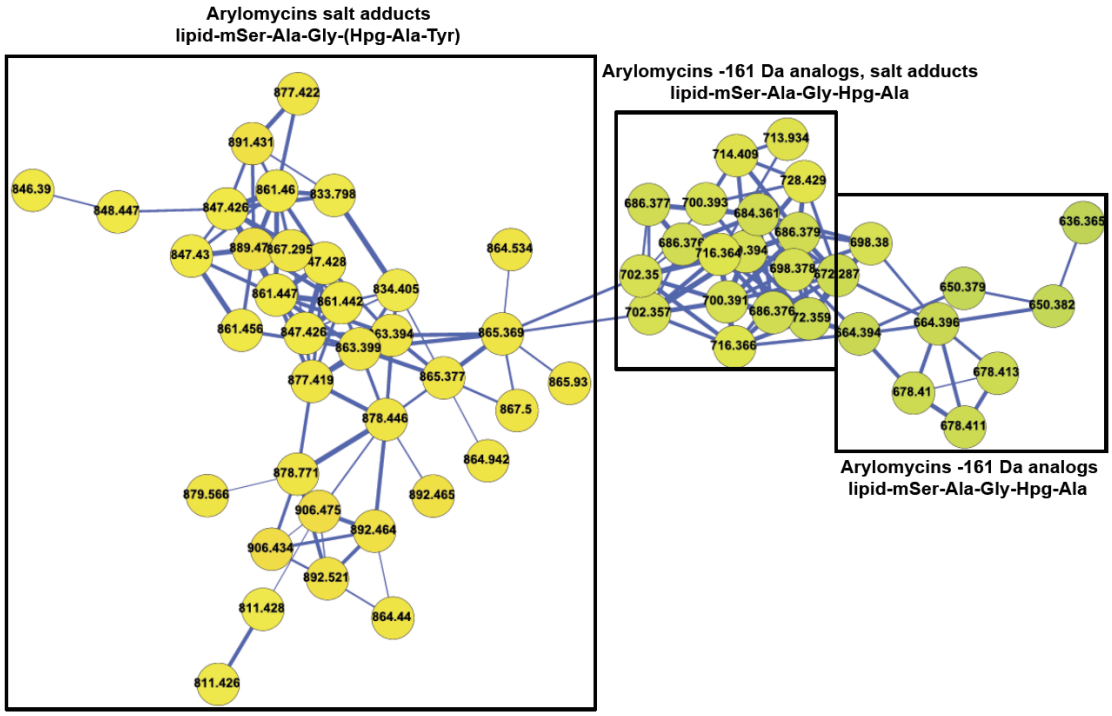
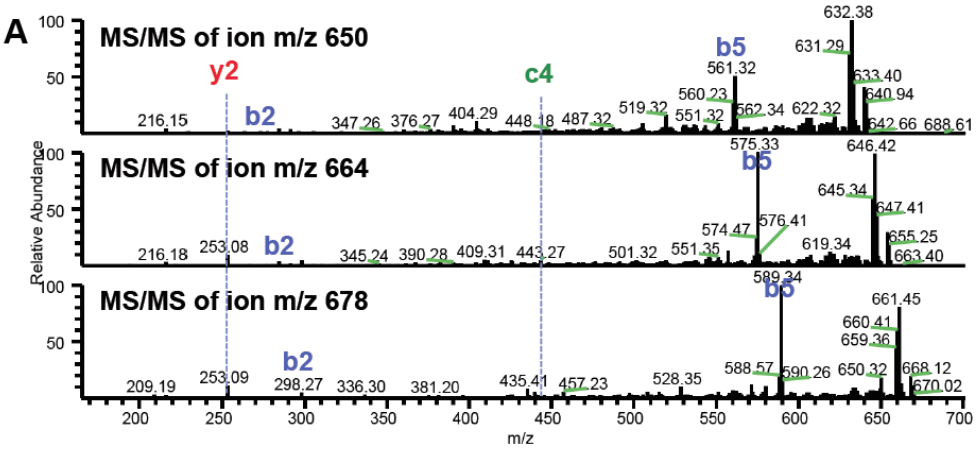


Figure S4.4: Zoom in of the molecular universe on the truncated arylomycin analogs.

A



B

	Chemical formula	calc. mass	obs. Mass	error (Da)
b2	$C_{16}H_{30}NO_3^+$	284.2220	284.31	0.09
b3	$C_{19}H_{35}N_2O_4^+$	355.2591		
b4	$C_{21}H_{38}N_3O_5^+$	412.2806		
b5	$C_{30}H_{47}N_4O_7^+$	575.3419	575.33	-0.01
y2	$C_{12}H_{17}N_2O_4^+$	253.1183	253.08	-0.04
y3	$C_{14}H_{20}N_3O_5^+$	310.1397		
y4	$C_{17}H_{25}N_4O_6^+$	381.1769		
y5	$C_{21}H_{32}N_5O_8^+$	482.2245		
c4	$C_{22}H_{43}N_4O_5^+$	443.3228	443.27	-0.05

C

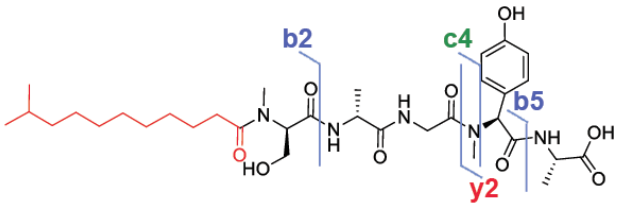


Figure S4.5: MS/MS of arylomycin truncated forms. (A) Alignment of IT MS/MS of ion at m/z 650, 664, 678. (B) Annotated ion table corresponds to the IT MS/MS of ion 664. (C) Ion map showing the fragmentation pattern of ion 664.

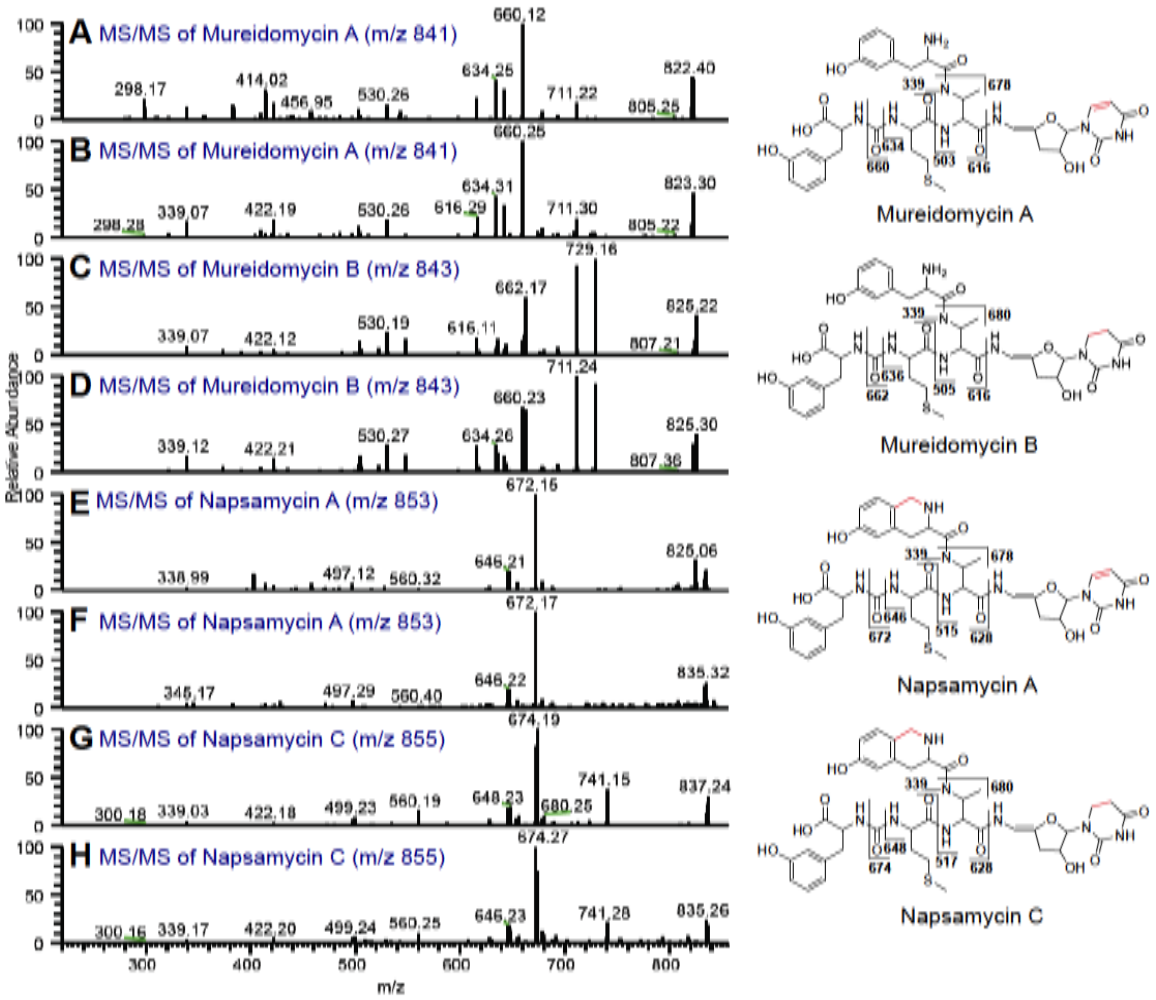


Figure S4.6: Alignment of MS/MS spectra of mureidomycin and napsamycin isolated from *S. roseosporus* and *S. sp.* DSM5940. A, **C**, **E**, **G** are MS/MS spectra observed from *S. roseosporus*. **B**, **D**, **F**, **H** are MS/MS spectra observed from S. s*p.* DSM5940. Fragmentation schemes are showed aside.

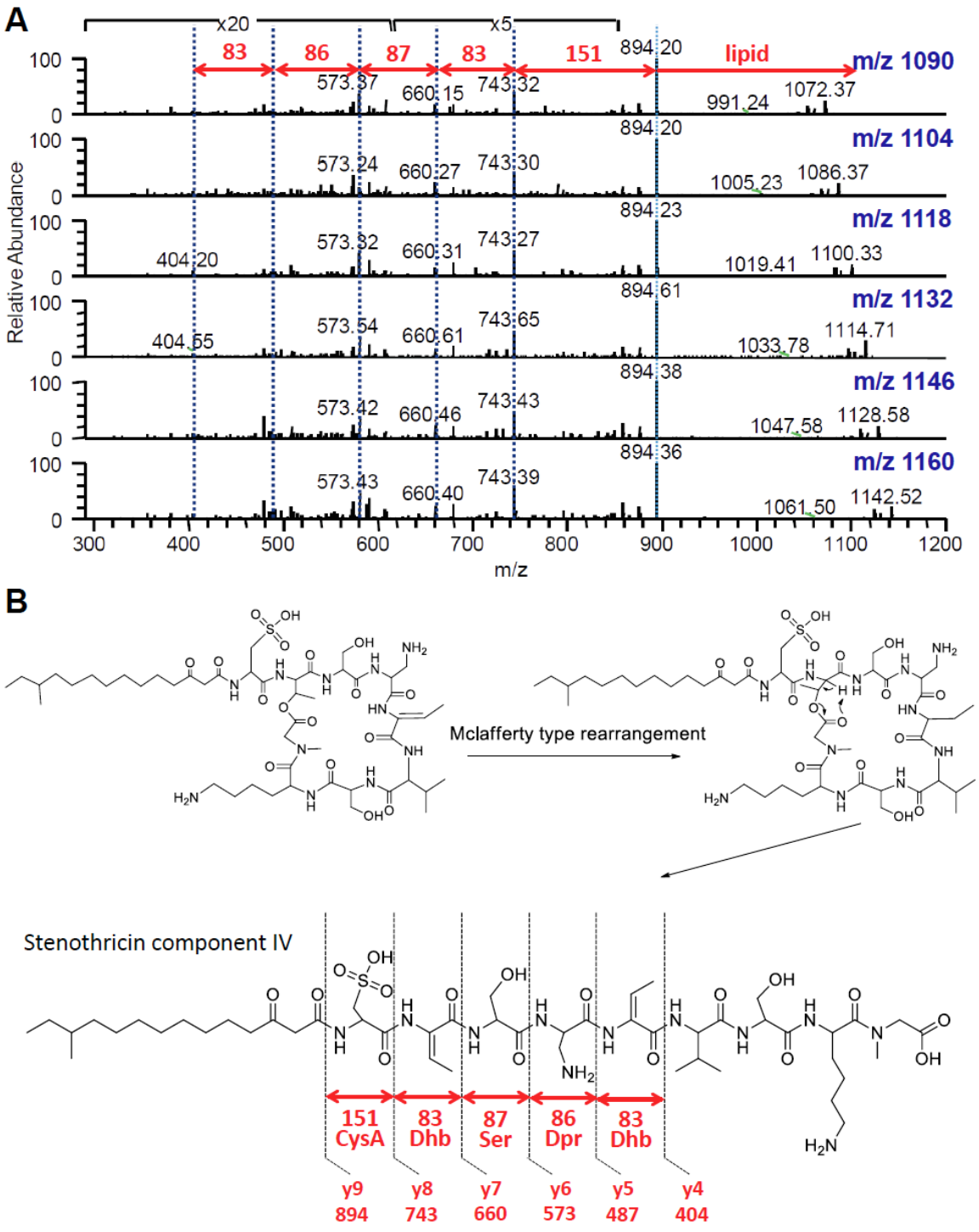
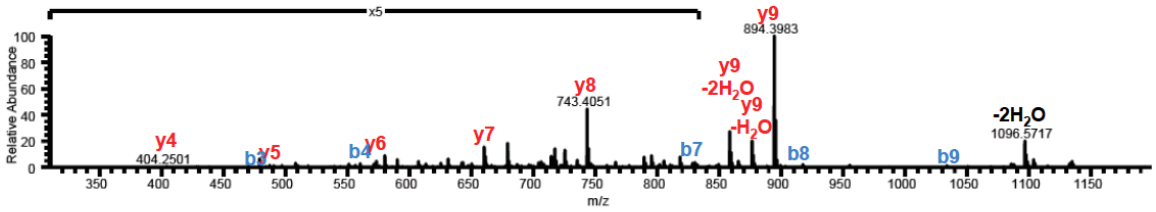


Figure S4.7: Peptidogenomic analysis of stenothricin.

A



B

	chemical formula	calc. Mw	obs. Mw	error (ppm)		chemical formula	calc. Mw	obs. Mw	error (ppm)
b1	$C_{15}H_{27}O_2^+$	239.3731	N/A		y1	$C_3H_8NO_2^+$	90.059	N/A	
b2	$C_{18}H_{32}NO_6S^+$	390.1945			y2	$C_9H_{20}N_3O3^+$	218.1499	N/A	
b3	$C_{22}H_{37}N_2O_7S^+$	473.2316	473.232	0.8	y3	$C_{12}H_{25}N_4O_5^+$	305.1819	N/A	
b4	$C_{25}H_{42}N_3O_9S^+$	560.2636	560.264	0.7	y4	$C_{17}H_{34}N_5O_6^+$	404.2504	404.250	-1.0
b5	$C_{28}H_{48}N_5O_{10}S^+$	646.3116			y5	$C_{21}H_{39}N_6O_7^+$	487.2875	487.287	-1.0
b6	$C_{32}H_{53}N_6O_{11}S^+$	739.3488			y6	$C_{24}H_{45}N_8O_8^+$	573.3355	573.337	2.6
b7	$C_{37}H_{62}N_7O_{12}S^+$	828.4172	828.420	3.4	y7	$C_{27}H_{50}N_9O_{10}^+$	660.3675	660.368	0.8
b8	$C_{40}H_{67}N_8O_{14}S^+$	915.4492	915.448	-1.4	y8	$C_{31}H_{55}N_{10}O_{11}^+$	743.4046	743.405	0.5
b9	$C_{46}H_{79}N_{10}O_{15}S^+$	1043.5442	1043.544	-0.2	y9	$C_{34}H_{60}N_{11}O_{15}S^+$	894.3986	894.399	0.4

C

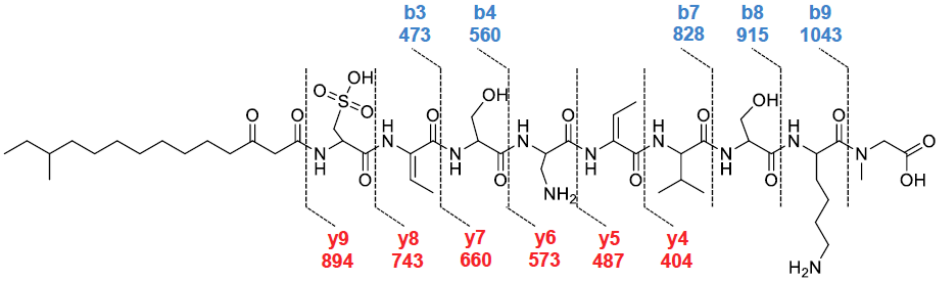


Figure S4.8: Annotated FT MS/MS spectra of stenothricin D. (A) FT MS/MS of stenothricin D and corresponding (B) ion table and (C) ion map.

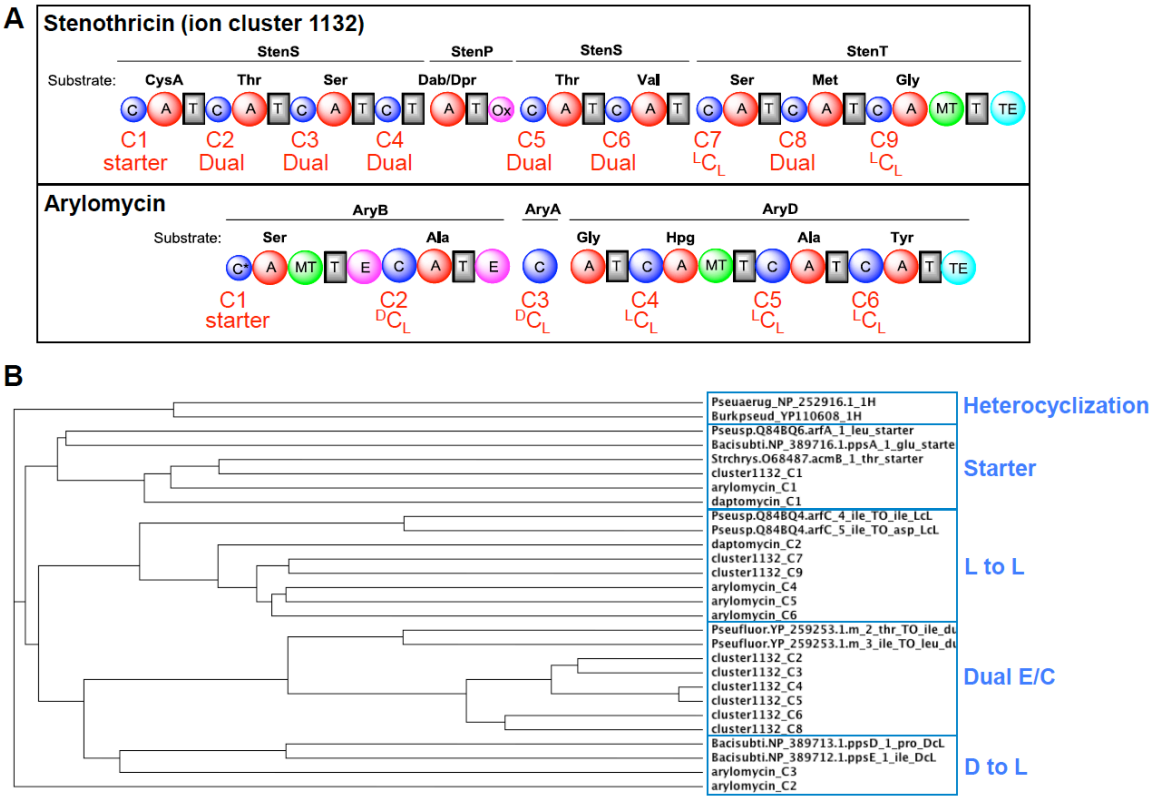


Figure S4.9: Phylogenetic analysis of stenothricin condensation (C) domains. (A) Summary of arylomycin and stenothricin NRPS modules organization with the predicted C domain functionality as judged by phylogenetic clustering with standard C domain subtypes. (B) Phylogenetic dendrogram suggests the first arylomycin and stenothricin C domain both cluster to starter C domains. The other C domains in arylomycin or stenothricin assembly lines are either clustered with L to L ($^{L}C_{L}$), D to L ($^{D}C_{L}$), or Dual E/C domains. An L to L C domain catalyzes the peptide bond between two L-amino acids, a D to L domain links an L-amino acid to a growing peptide ending with a D-amino acid, whereas Dual E/C domains catalyze both epimerization and condensation.

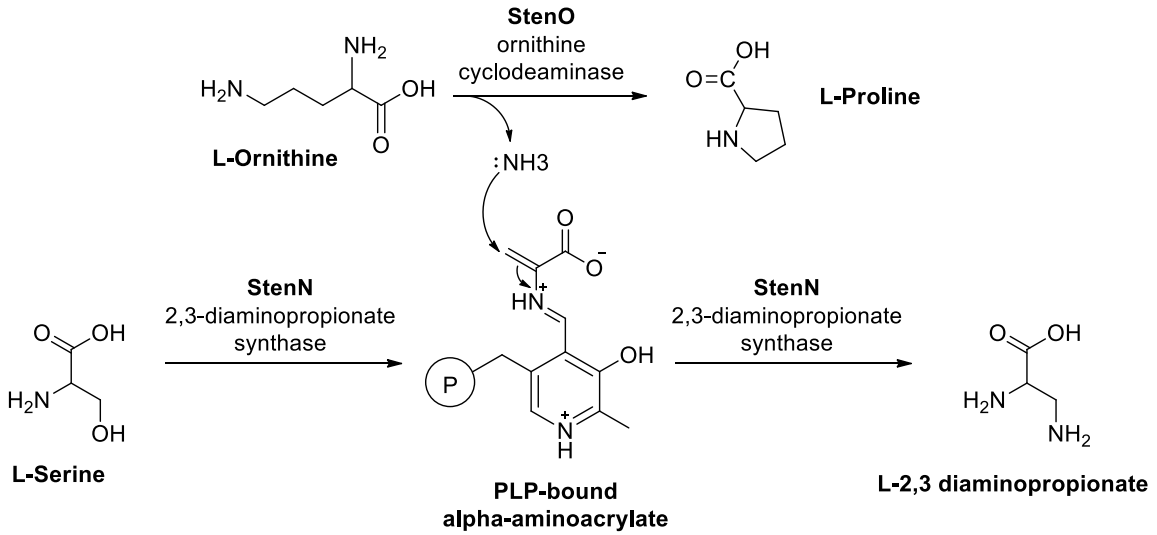


Figure S4.10: Proposed pathways of StenNO catalyze the formation of diaminopropionate from Orinithine and Serine.

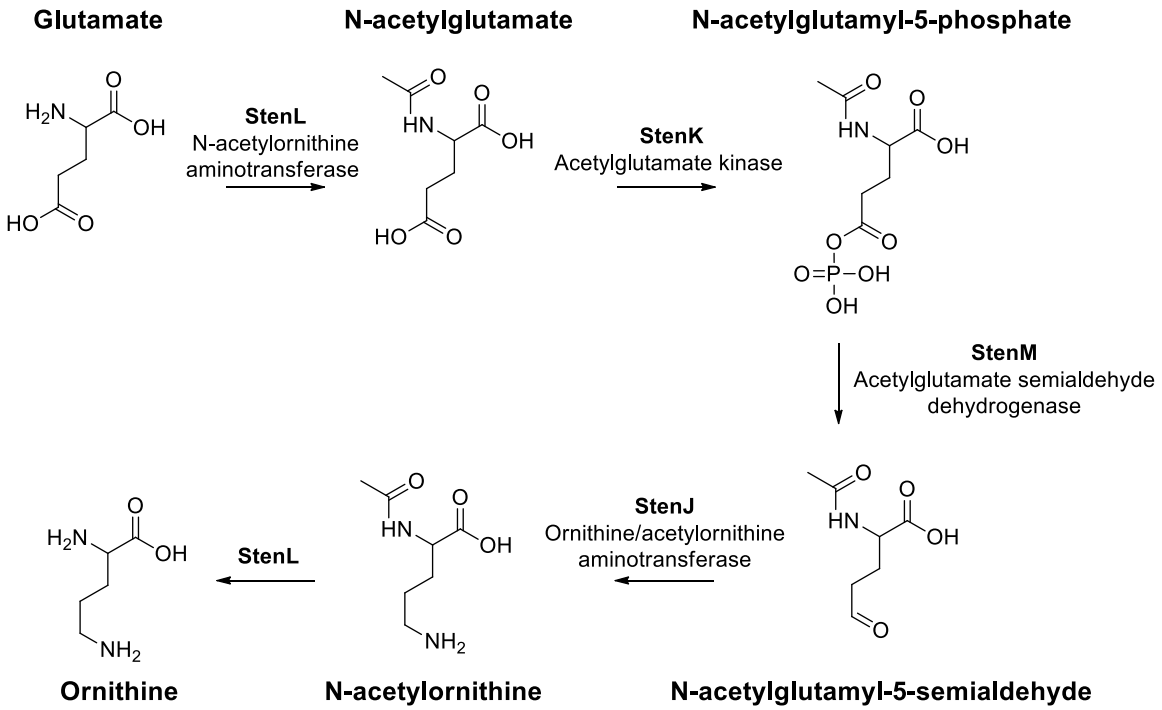


Figure S4.11: StenJKLM is responsible for Ornithine synthesis (proposed).

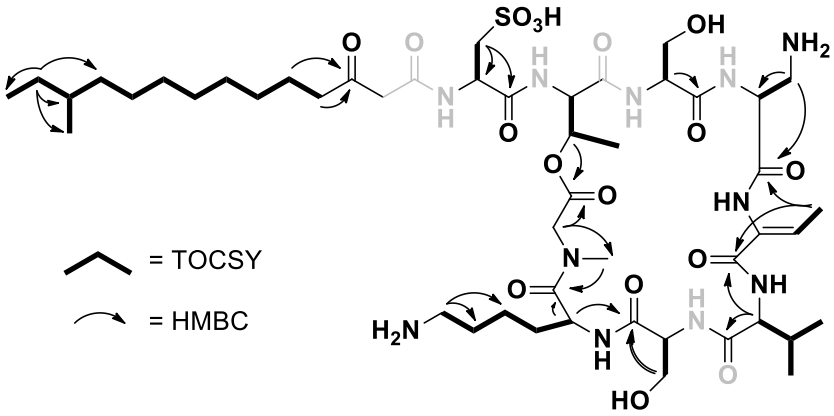


Figure S4.12: Key 2D-NMR correlations in determining connectivity of residues supporting the assignment that this is indeed stenothricin.

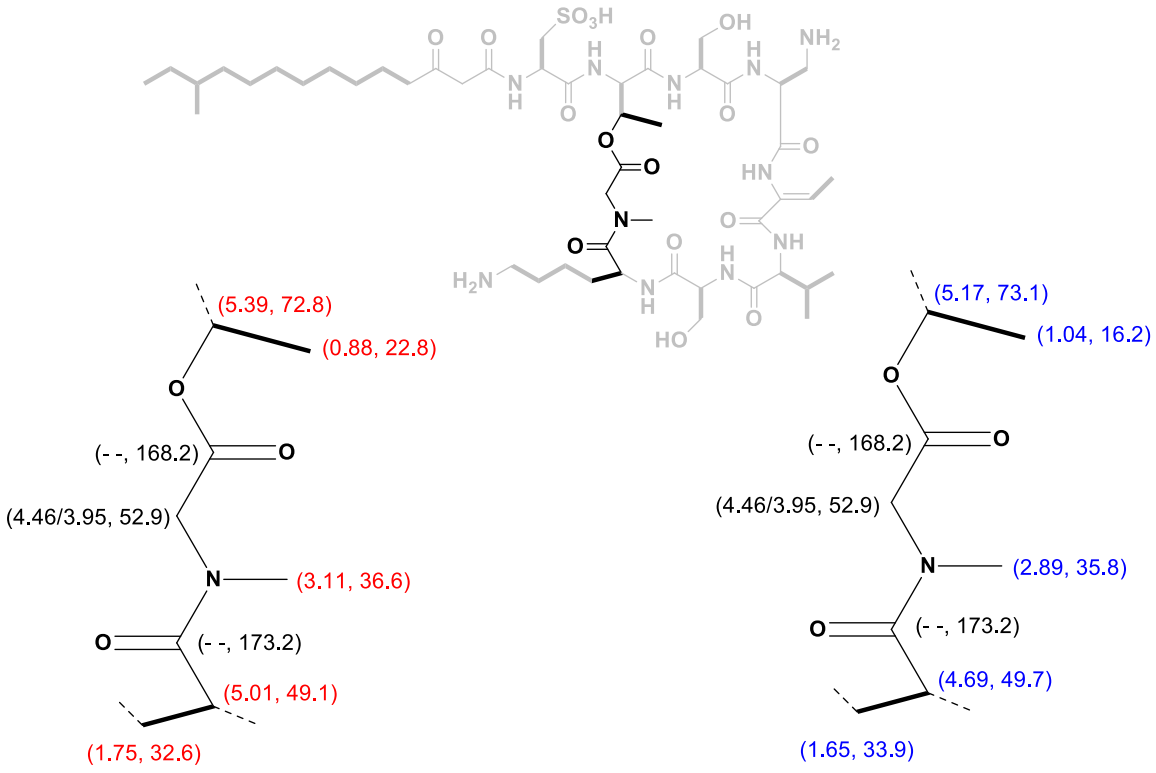


Figure S4.13: *N*-methyl amide associated rotomer shifts in stenothricin (δ_H, δ_C).

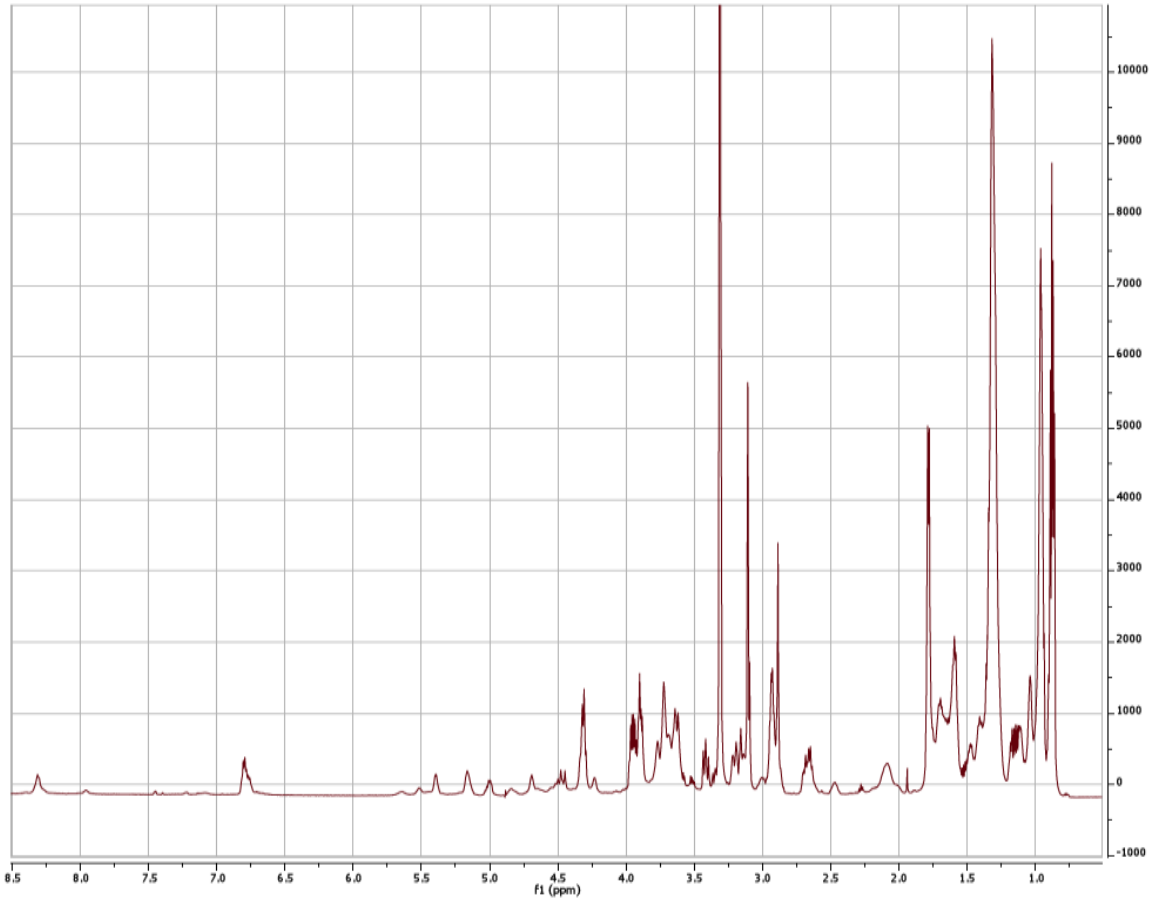


Figure S4.14: ^{1}H NMR (600 MHz, CD_3OD) Spectrum of Stenothricin.

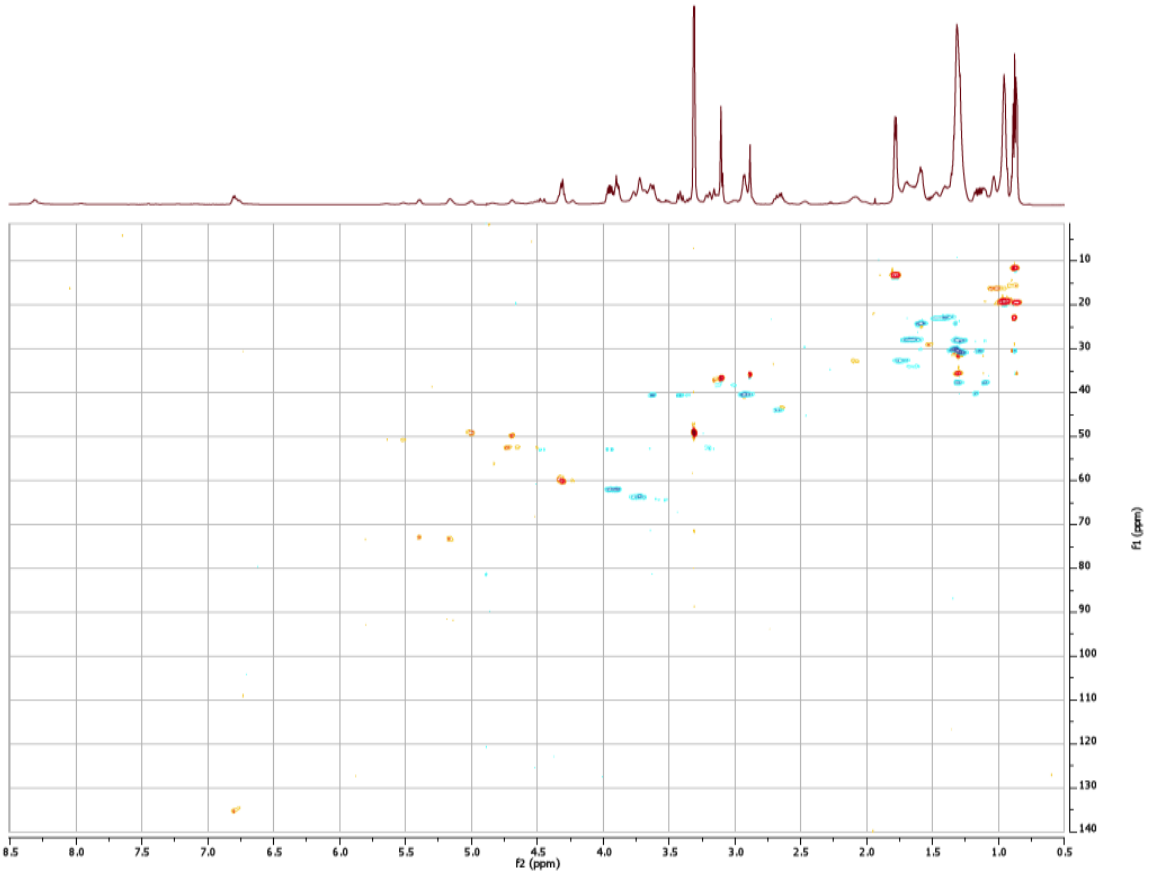


Figure S4.15: HSQC (600 MHz, CD_3OD) Spectrum of Stenothricin.

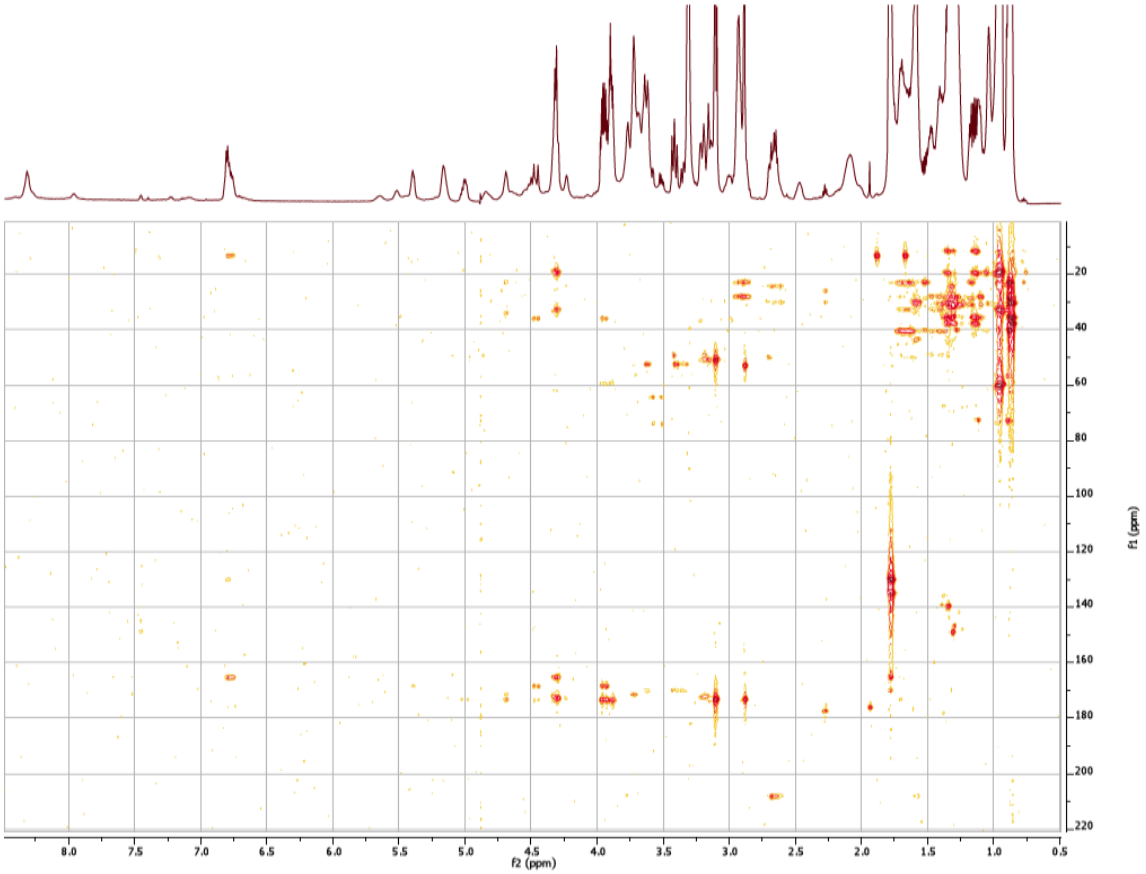


Figure S4.16: HMBC (600 MHz, CD_3OD) Spectrum of Stenothricin.

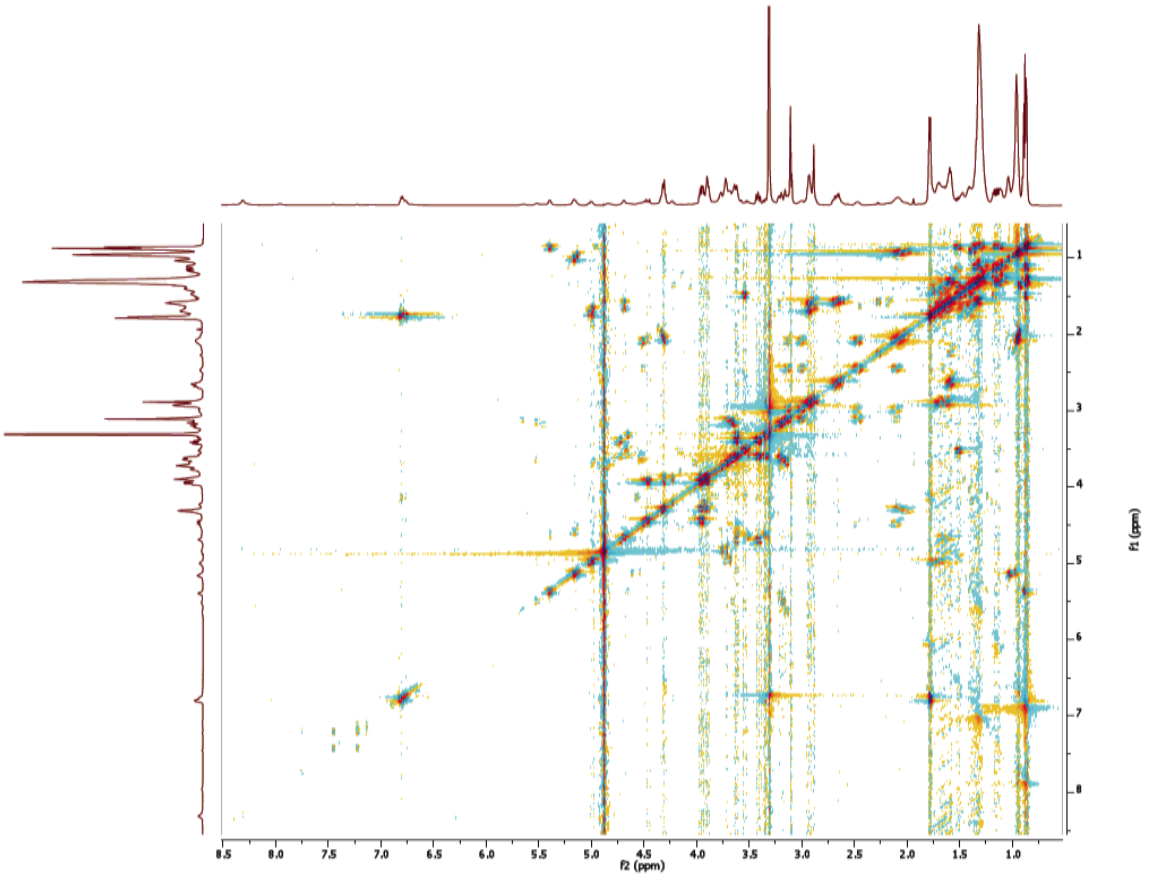


Figure S4.17: COSY (600 MHz, CD_3OD) Spectrum of Stenothricin.

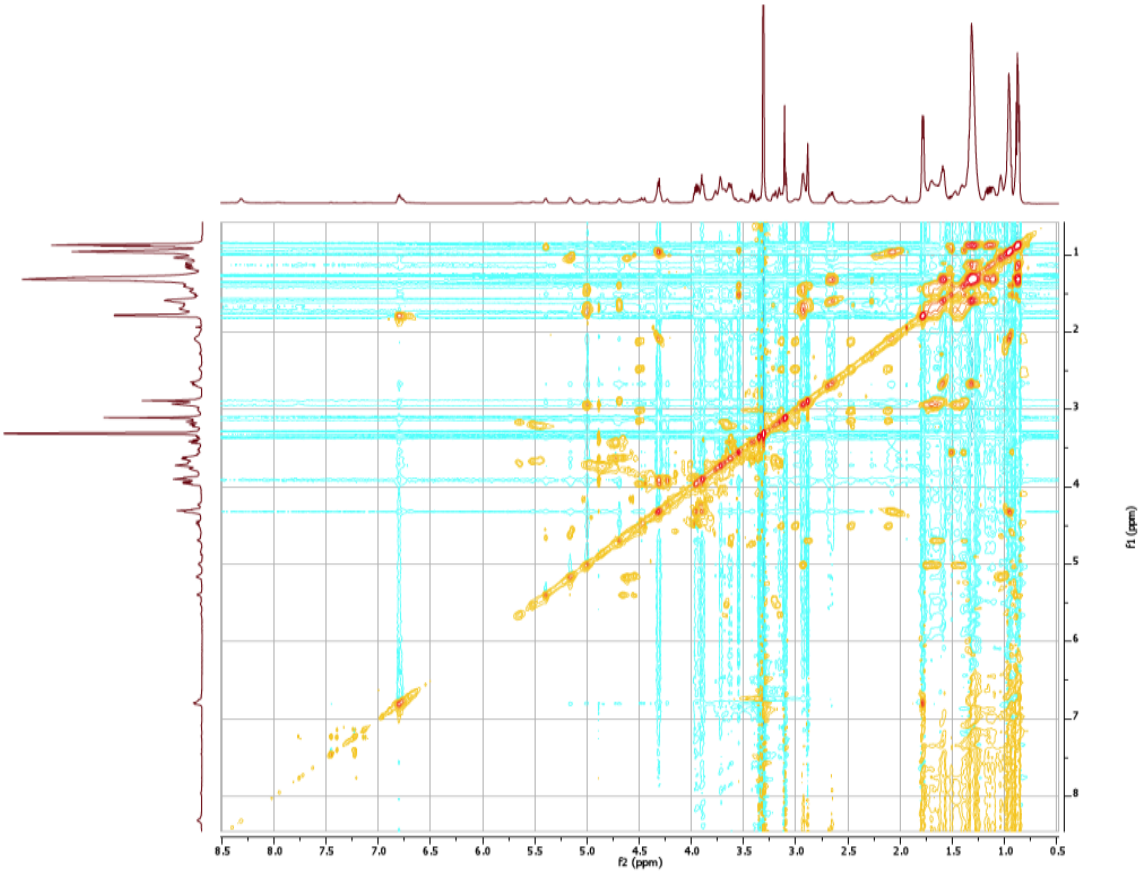


Figure S4.18: **TOCSY (600 MHz, CD_3OD) Spectrum of Stenothricin.**

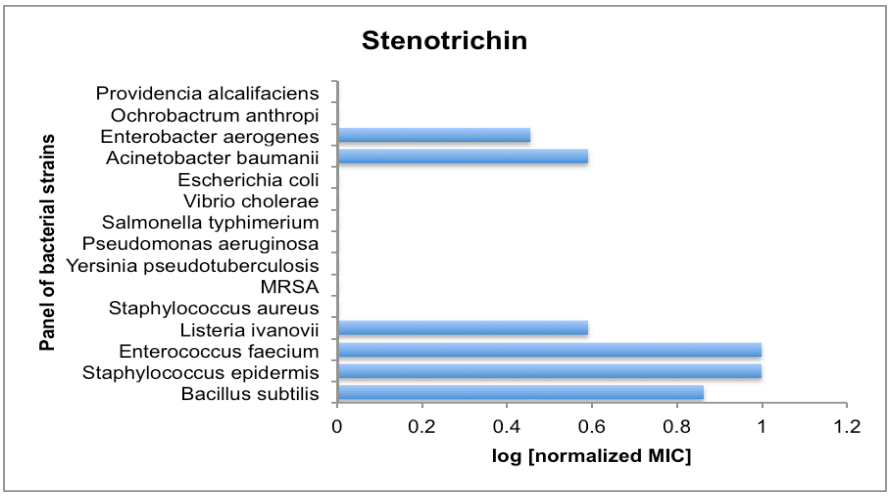


Figure S4.19: BioMAP profile of stenothricin. Stenothricin was serially diluted and screened against a panel of 15 representative pathogenic bacterial strains to obtain normalized MIC with values ranging from 0 (inactive) to 1 (most potent).

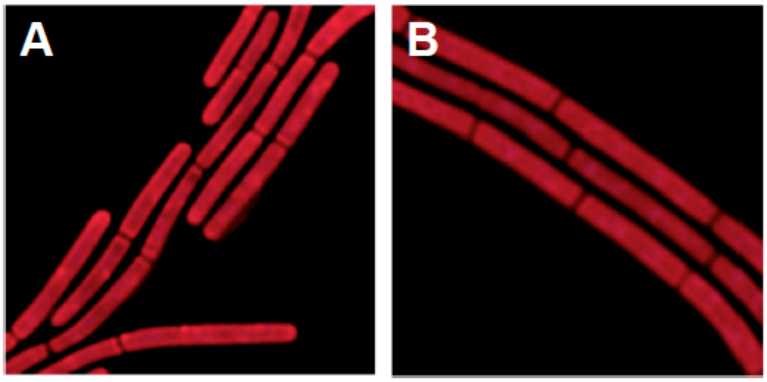


Figure S4.20: Effects of stenothricin single form and mixture of variants on *B. subtilis* cell architecture. Fluorescence micrographs revealed the effect on cell architecture of *B. subtilis* when treated with (**A**) 18 µM stenothricin D (**B**) 18 µM stenothricin mixtures of variants.

Table S4.1: Annotation of the genes involved in stenthricin biosynthetic cluster and the neighboring regions.

Gene	Protein	Size [aa]	Predicted function	Closest homolog (Similarity/Identity) [%/%]
StenA	EFE73293.1	514	Multidrug resistance efflux pump	YP_001827483.1| major facilitator superfamily permease [Streptomyces griseus subsp. griseus NBRC 13350] (99/96)
StenB	EFE73294.1	189	TetR-family transcriptional regulator	ZP_09399712.1| TetR family transcriptional regulator [Streptomyces sp. W007] (99/95)
StenC	EFE73295.1	721	putative helicase	ZP_09399711.1| putative helicase [Streptomyces sp. W007] (93/89)
StenD	EFE73296.1	343	oxidoreductase	YP_001827480.1| oxidoreductase [Streptomyces griseus subsp. griseus NBRC 13350] (97/97)
StenE	EFE73297.1	475	argininosuccinate lyase (ArgH)	ZP_09399709.1| argininosuccinate lyase [Streptomyces sp. W007] (99/98)
StenF	EFE73298.1	398	argininosuccinate synthase (ArgG)	ZP_09399708.1| argininosuccinate synthase [Streptomyces sp. W007] (99/99)
StenG	EFE73299.1	462	hypothetical protein	YP_004801446.1| hypothetical protein SACTE_0977 [Streptomyces sp. SirexAA-E] (62/55)
StenH	EFE73300.1	236	secreted protein/L,D-transpeptidase	ZP_08239687.1| putative secreted protein [Streptomyces griseus XylebKG-1] (94/90)
StenI	EFE73301.1	178	arginine repressor (ArgR)	ZP_09399706.1| arginine repressor [Streptomyces sp. W007] (100/99)
StenJ	EFE73302.1	403	N2-acetyl-L-ornithine:2-oxoglutarate aminotransferase (ArgD)	YP_001827475.1| acetylornithine aminotransferase [Streptomyces griseus subsp. griseus NBRC 13350] (96/95)
StenK	EFE73303.1	314	N-acetylglutamate kinase (ArgB)	ZP_09399704.1| acetylglutamate kinase [Streptomyces sp. W007]
StenL	EFE73304.1	384	N2-acetyl-L-ornithine:L-glutamate N-acetyltransferase (ArgJ)	ZP_09399703.1| bifunctional ornithine acetyltransferase/N-acetylglutamate synthase protein [Streptomyces sp. W007] (98/96)
StenM	EFE73305.1	341	N-acetyl-gamma-glutamylphosphate reductase (ArgC)	YP_001827472.1| N-acetyl-gamma-glutamyl-phosphate reductase [Streptomyces griseus subsp. griseus NBRC 13350] (97/95)
StenN	EFE73306.1	335	cysteine synthase/2,3-diaminopropionate biosynthesis protein (SbnA)	ZP_09402162.1| putative cysteine synthase [Streptomyces sp. W007] (86/77)
StenO	EFE73307.1	351	ornithine cyclodeaminase/2,3-diaminopropionate biosynthesis protein (SbnB)	ZP_09402163.1| putative ornithine cyclodeaminase [Streptomyces sp. W007] (88/80)
StenP	EFE73308.1	1184	NRPS (A-T-CAS)	ZP_08236917.1| amino acid adenylation domain protein [Streptomyces griseus XylebKG-1] (75/69)
StenQ	EFE73309.1	271	type II thioesterase	AEW95033.1| thioesterase [Streptomyces cattleya NRRL 8057 = DSM 46488] (61/46)
StenR	EFE73310.1	75	MbtH-like protein	ZP_09798514.1| hypothetical protein GOTRE_049_00170 [Gordonia terrae NBRC 100016] (83/71)
StenS	EFE73312.1	6082	NRPS	ABD65957.1| nonribosomal peptide synthetase [Streptomyces fungicidicus] (63/50)
StenT	EFE73313.1	3883	NRPS	ZP_08287828.1| non-ribosomal peptide synthetase [Streptomyces griseoaurantiacus M045] (69/56)
StenU	EFE73314.1	431	cysteate synthase	ADI04248.1| threonine synthase [Streptomyces bingchenggensis BCW-1] (76/65)

Table S4.2: Summary of NMR data (in CD_3OD) for stenothricin.

Residue	Position	δ_C,[a] type	δ_H (*J* in Hz)	HMBC	COSY	TOCSY
N-Me-Gly-9	1	168.2, C	–			
	2a	52.9, CH_2[b]	4.46, d (18.6)	1, 3'	2b	
	2b	52.9, CH_2[b]	3.95,[c] m	1, 3'	2a	
	3	36.6, CH_3	3.11, s	4		
	3' rotomer	35.8, CH_3	2.89, s	2, 4		
Lys-8	4	173.2, C	–			
	5	49.1, CH	5.01, m		6	6-9
	5' rotomer	49.7, CH	4.69, t (5.4)	4, 6', 7, 10	6'	6',7,8
	6	32.6, CH_2	1.75,[d] m		5	5, 7-9
	6' rotomer	33.9, CH_2	1.65,[d] m		5'	5', 7-9
	7	22.9, CH_2	1.46,[d] m			5,9
	8	27.8, CH_2	1.66,[d] m		9	5,9
	9	40.2, CH_2	2.93, q (6.2)	7, 8	8	6-8
Ser-7	10	171.4, C				
	11	56.1, CH	4.84, m			
	12	63.5, CH_2	3.72, m	10, 11		
Val-6	13	172.6, C	–			
	14	60.1, CH	4.31, m	13, 15-18	15	15, 16/17
	15	32.6, CH	2.08, m		14, 16/17	14, 16/17
	16	19,[e] CH_3	0.96,[f] m	14, 15, 17	15	14
	17	19,[e] CH_3	0.96,[f] m	14-16	15	14
Dhb-5	18	165.2, C	–			
	19	129.9, C	–			
	20	135.1, CH	6.80, q (6.6)	18, 19, 21	21	20
	21	13.1, CH_3	1.77, d (6.6)	18-20, 22	20	21
Dpr-4	22	169.9	–			
	23	52.4, CH	4.73, m		24a/b	24a/b
	24a	40.5, CH_2[b]	3.63,[g] m	22, 23	23, 24b	23, 24b
	24b	40.5, CH_2[b]	3.42, dd (12.9, 9.9)	22, 23	23, 24a	23, 24a
Ser-3	25	173.3, C	–			
	26	59.8, CH	4.23, m		27a/b	27a/b
	27a	62.0, CH_2[b]	3.95[c], m	25, *26 weak*	26, 27b	26, 27b
	27b	62.0, CH_2[b]	3.90, m	25, *26 weak*	26, 27a	26, 27a

Table S4.2: Summary of NMR data (in CD_3OD) for stenothricin. Continued.

Residue	Position	δ_C,[a] type	δ_H (J in Hz)	HMBC	COSY	TOCSY
Thr-2	28					
	29	52.4, CH	4.65, m		30, 31, 30', 31'	30, 31, 30', 31'
	30	72.8, CH	5.39, m	1	29, 31	29, 31
	30' rotomer	73.1, CH	5.17, m		29, 31r	29, 31r
	31	22.8, CH_3	0.88, m		29, 30	29, 30
	31' rotomer	16.2, CH_3	1.04, m		29, 30r	29, 30r
Cys-1	32	172.0, C	–			
	33	50.6, CH	5.52, m		34a	34a/b
	34a	52.5, CH_2[b]	3.18, m	32, 33	33	33, 34b
	34b	52.5, CH_2[b]	3.67,[g] m		33, 34a	33, 34a
Acyl Tail	35					
	36					
	37	208.0, C	–			
	38	43.8, CH_2	2.65, m	37, 39, 40[h]	39	39, 40-45a[i]
	39	24.2, CH_2	1.59,[d] m	37, 38, 40[h]	38, 40-45a[i]	38, 40-45a[i]
	40	30[j], CH_2	1.32,[k] m			
	41	30[j], CH_2	1.32,[k] m			
	42	30[j], CH_2	1.32,[k] m			
	43	30[j], CH_2	1.32,[k] m			
	44	28.1, CH_2	1.29,[l] m			
	45a	37.6, CH_2[b]	1.31,[l] m			
	45b	37.6, CH_2[b]	1.10,[m] m	44, 46		40-46,[i] 48/49[i]
	46	35.5, CH	1.31,[l] m			
	47	30.4, CH_2	1.15,[m] m	45, 46, 48, 49	48/49[i]	40-46,[i] 48/49[i]
	48	11.5, CH_3	0.88,[n] m			
	49	19.4, CH_3	0.88,[n] m			

[a]Carbon shifts extrapolated from HSQC and HMBC spectra. [b]Diastereotopic species. [c,d,g,l,m]Overlapping proton species, assignment by HSQC. [e,j]Carbon species unresolvable by HSQC and HMBC. [f,k,n]Proton species unresolvable by both ^{1}H-NMR and HSQC. [h]HMBC assignment tentative because of unresolved carbon species. [i]COSY and TOCSY assignment tentative because of unresolved proton species.

4.4 Material and Methods

4.1.1 Strains and materials

Streptomyces roseosporus NRRL 15998 was acquired though Michael Fischbach from the Broad institute, MIT/Harvard, MA, USA. *Bacillus subtilis* PY79 was originated from Richard Losick lab and was acquired via Kit Pogliano lab. Strains used for BioMAP screening were described in Weng et al., in press[37]. All chemicals used for ISP2 media were purchased from Sigma-Aldrich. Organic solvents were purchased from J. T. Baker.

4.1.2 Bacterial metabolite extraction

One single ISP2 agar plate was inoculated with *S. roseosporus* starter culture by 4 parallel streaks. The plates were incubated for 10 d at 28 °C. The agar was sliced into small pieces and then put in a 50 ml centrifuge tube and covered with equal amount of Milli-Q water and n-butanol or methanol for 12 h at 28 °C, shaken at 225 rpm. The n-butanol layer were collected using transfer pipette, dried with rotary evaporator.

4.1.3 Data-dependent MS/MS dataset collection

MS/MS datasets were collected both with or without LC separation in-line with mass spectrometry. For LC-MS, capillary columns were prepared by drawing a 360 μm O.D., 100 μm I.D. deactivated, fused silica tubing (Agilent) with a Model P-2000 laser puller (Sutter Instruments) (Heat: 330, 325, 320; Vel, 45; Del, 125) and were packed at 600 psi to a length of about 10 cm with C18 reverse-phase resin suspended in methanol. The column was equilibrated with 95% of solvent A (water, 0.1% AcOH) and loaded with 10 μl (10 ng/μl in 10% CH_3CN) of bacterial butanol extract by flowing 95% of

solvent A and 5% of solvent B (CH_3CN, 0.1% AcOH) at 200 µl/min for 15 mins. A gradient was established with a time-varying solvent mixture [(min, % of solvent A): (20, 95), (30, 40), (75, 5)] and directly electrosprayed into the LTQ-FT MS inlet (source voltage, 1.8 kV; capillary temperature, 180 C). The first scan was a high resolution broadband scan. The subsequent six scans were low resolution data-dependent on the first scan. In each data-dependent scan, the top intensity ions excluded the ones in exclusion list were selected to be fragmentated by CID which generated hundreds of fragmentation spectra collected as individual data events. The resulting .RAW files were converted to .mzXML using the program ReAdW (tools.proteomecenter.org).

4.1.4 General LTQ/FT-ICR MS and MS/MS procedure for isolated molecules.

For the IT and FT MS data acquisition, each compound was dissolved in spray solvent 50:50 MeOH/H_2O containing 1% formic acid, and underwent nano-electrospray ionization on a biversa nanomate (Advion Biosystems, Ithaca, NY) using a back pressure of 0.3-0.5 p.s.i. and the spray voltage of 1.4 -1.5 kV. MS and MS/MS spectra were acquired on a 6.42 T Finnigan LTQ-FTICR MS or a Finnigan LTQ-MS (Thermo-Electron Corporation, San Jose, CA) running Tune Plus software version 1.0 and Xcalibur software version 1.4 SR1. The instrument was first auto-tuned on the *m*/*z* value of the ion to be fragmented. Then, the ions were isolated by the linear ion trap and fragmented by collision induced dissociation (CID) (isolation window: 3 *m/z*; collision energy: 30).

4.1.5 Purification of stenothricin

10L ISP2 (4 g yeast extract, 10 g malt extract, 4 g dextrose, per liter) liquid media were inoculated with *S. roseosporus* starter culture and were incubated for 48 hours at 28 °C, shaken at 225 rpm. 300 mL of n-butanol were used to extract per liquid of bacterial culture, dried with rotary evaporator, re-dissolved and then fractioned via size exclusion using a Sephadex LH-20 column (25 cm x 30 mm) using MeOH at a flow rate of 0.5 mL/min. Each fraction (1 mL) was analyzed by MALDI MS and the fractions contained stenothricin were collected. Further purification were done by multiple runs of HPLC (C-18, 25 cm x 10 mm) with a sets of different solvent systems and gradient setting. Stenothricin crude extract were first fractioned by running a H_2O, MeCN gradient from 40% solvent B to 95% solvent B in 25 minutes with flow rate 2mL/min whereas Solvent A is H_2O containing 0.1% TFA; solvent B is MeCN containing 0.1% TFA (aq). Fractions contain stenothricin were collected, and were combined into three vials based on the three major variants stenothricin B, F, G. These three vials each was dominated by one of the structural variant were dried, re-suspended and subjected to a second run of HPLC were done by running same gradient as above but changed solvent A to isopropanol/MeCN = 7:3 containing 0.1% TFA; and solvent B is 0.1% TFA (aq). Fractions that contain each of these three stenothricin variants were collected separately and were further clean up with additional runs of HPLC with gradients adjustments customized for each analog. Purified stenothricin were lyophilized and stored at -80ºC before using for bioassay and structural elucidation.

4.1.6 MALDI-TOF MS analysis

MALDI-TOF MS was used to obtain the general metabolic output as well as to detect target molecules in crude extracts and fractions of gel filtration and HPLC. The sample was mixed 1:1 with a saturated solution of Universal MALDI matrix (Sigma-Aldrich) in 78 % acetonitrile containing 0.1 % TFA and spotted on a Bruker MSP 96 anchor plate. The sample was dried and inserted into the Microflex mass spectrometer (Bruker Daltonics). Mass spectra were obtained with the FlexControl method as used for MALDI-imaging and a single spot acquisition of 80 shots. Single spot MALDI-TOF MS data was analyzed by FlexAnalysis software.

4.1.7 NMR measurement

50 µg stenothricin D and G were dissolved in 40 µL of CD_3OD for NMR acquisition. ^{1}H-NMR spectra were recorded on Bruker Avance III 600 MHz NMR with 1.7 mm Micro-CryoProbe at 298 K, with standard pulse sequences provided by Bruker. The data was analyzed using the Topspin 2.1 software.

Phylogenetic analysis of C domains

Following the analysis done by Rausch *et al.*[2] and Imker *et al.*[3], a subset of condensation domain sequences were collected to represent the six condensation families (heterocyclization, epimerization, dual condensation/epimerization (dual), condensation of L amino acids to L amino acids (L to L), and condensation of D amino acids to L amino acids (D to L), and starter). Sequences were annotated with the accession number and C domain type and then aligned using ClustalW2.

4.1.8 Effect of stenothricin on *B subtilis* cell growth curve

The effect of stenothricin and other antimicrobials including nisin, daptomycin, vancomycin and Triton X-100 on *B. subtilis* cell growth was performed using 96 well microtiter plates. A 2 mL overnight culture in LB media was centrifuged at 6000g for 10 minutes and supernatants discarded. The cell pellets were resuspended using 2 mL of ISP2 media. OD_{595} of the resuspended cells were measured (ELx808 Ultra Microplate Reader, Bio-TEK Instruments), and the final OD_{595} was adjusted to 0.03 with ISP2 media. 100 μL diluted culture with indicated working concentrations of stenothricin or other anitmicrobials were aliquoted into each well. The plate was shaken at 37^{o}C, 120 rpm. OD_{595} were measured and recorded at each time point.

4.1.9 BioMAP

200 μg of stenothricins were subjected to BioMAP profiling and to compare with the antibiotics in training set[37]. Antibiotics for the training set includes 72 FDA-approved antibiotics to cover all of the major compound classes currently available in the clinic, including 16 classes that function by different modes of action[37]. Alphanumeric labels were given to each drug to identify derivatives from the same antibiotic class. Screening plates were generated by serially diluting antibiotics in DMSO (2-fold dilutions, final screening concentrations 100 mM–3 nM). Saturated overnight cultures of pathogenic strains were diluted 1:1000 and dispensed into sterile clear polypropylene 384 well plates (30 ml screening volume). DMSO solutions of test compounds (300 nl) were pinned into each well at t0 using a high-throughput pinning robot (Perkin Elmer Janus MDT). After compound addition, screening plates were stacked in an automated plate reader/shaker

(Perkin Elmer EnVision) and OD600 reading was collected every hour for 24 hr. The resulting growth curves for each dilution series were used to determine MIC values for all test compounds and extracts. BioMAP profiles were plotted according to normalized MIC values as described in Wong et al[37].

4.1.10 Fluorescence microscopy

The effects of stenothricin and other antimicrobials on individual *B. subtilis* cells were investigated in 15 μL cultures prepared in the following manner. Cultures were grown in LB media to an OD_{600} of 0.3, centrifuged, resuspended in 1/10 the volume and 14.25 μL of concentrated cells were added to 1.7 mL microcentrifuge tubes. At t = 0, indicated working concentrations of DMSO, stenothricin, or other anitmicrobials (in 10% DMSO) was added to different aliquots of cells. The tubes were capped and incubated at 37°C in a roller. Samples were collected for imaging every 30 minutes. 2 μL of cells were added to 0.5 μL of a stain mix containing 30 μg/mL FM 4-64, 2.5 μM Sytox Green and 1 μg/mL DAPI prepared in 1X T-base. Cells were immobilized with Poly-L-Lysine and imaged on an Applied Precision Spectris Microscope. Images were deconvoluted and the medial focal planes shown.

Reference

1. Newman, D.J., Cragg, G.M. & Snader, K.M. Natural products as sources of new drugs over the period 1981-2002. *Journal of Natural Products.* **66**, 1022-1037 (2003).

2. Demain, A.L. Antibiotics: natural products essential to human health. *Med Res Rev.* **29**, 821-842 (2009).

3. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the last 25 years. *Journal of Natural Products.* **70**, 461-477 (2007).

4. Klevens, R.M. et al. Invasive methicillin-resistant Staphylococcus aureus infections in the United States. *JAMA.* **298**, 1763-1771 (2007).

5. Chambers, H.F. & Deleo, F.R. Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat Rev Microbiol.* **7**, 629-641 (2009).

6. Fischbach, M.A. & Walsh, C.T. Antibiotics for emerging pathogens. *Science.* **325**, 1089-1093 (2009).

7. Lewis, K. Antibiotics: Recover the lost art of drug discovery. *Nature.* **485**, 439-440 (2012).

8. Li, J.W. & Vederas, J.C. Drug discovery and natural products: end of an era or an endless frontier? *Science.* **325**, 161-165 (2009).

9. Walsh, C. Where will new antibiotics come from? *Nat Rev Microbiol.* **1**, 65-70 (2003).

10. Liu, W.T. et al. Interpretation of Tandem Mass Spectra Obtained from Cyclic Nonribosomal Peptides. *Analytical Chemistry.* **81**, 4200-4209 (2009).

11. Ng, J. et al. Dereplication and de novo sequencing of nonribosomal peptides. *Nature Methods.* **6**, 596-U565 (2009).

12. Kersten, R.D. et al. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol.* **7**, 794-802 (2011).

13. Liu, W.T., Kersten, R.D., Yang, Y.L., Moore, B.S. & Dorrestein, P.C. Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycin in Streptomyces roseosporus. *J Am Chem Soc.* **133**, 18010-18013 (2011).

14. Watrous, J. et al. Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A* **109**, E1743-1752 (2012).

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

16. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature.* **422**, 198-207 (2003).

17. Patti, G.J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol.* **13**, 263-269 (2012).

18. Shannon, P. et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research.* **13**, 2498-2504 (2003).

19. Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* **27**, 431-432 (2011).

20. Debono, M. et al. A21978C, a complex of new acidic peptide antibiotics: isolation, chemistry, and mass spectral structure elucidation. *J Antibiot (Tokyo).* **40**, 761-777 (1987).

21. Baltz, R.H. Biosynthesis and genetic engineering of lipopeptide antibiotics related to daptomycin. *Curr Top Med Chem.* **8**, 618-638 (2008).

22. Miao, V. et al. Daptomycin biosynthesis in Streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology.* **151**, 1507-1523 (2005).

23. Jin, X. et al. Biosynthesis of new lipopentapeptides by an engineered strain of Streptomyces sp. *Biotechnol Lett.* **34**, 2283-2289 (2012).

24. Kaysser, L. et al. Identification of a napsamycin biosynthesis gene cluster by genome mining. *Chembiochem.* **12**, 477-487 (2011).

25. Medema, M.H. et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research.* **39**, W339-W346 (2011).

26. Li, M.H.T., Ung, P.M.U., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D.H. Automated genome mining for natural products. *Bmc Bioinformatics.* **10** (2009).

27. Rottig, M. et al. NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res* **39**, W362-367. (2011).

28. Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol.* **6**, 493-505 (1999).

29. Konig, W.A., Engelfried, C., Hagenmaier, H. & Kneifel, H. Structure of the peptide antibiotic Stenothricin. *Justus Liebigs Annalen der Chemie.* **1976**, 2011-2020 (1976).

30. Hasenbohler, A., Kneifel, H., Konig, W.A., Zahner, H. & Zeiler, H.J. Stenothricin, a new inhibitor of the bacterial cell wall synthesis. *Arch Microbiol.* **99**, 307-321 (1974).

31. Rinken, M., Lehmann, W.D. & Konig, W.A. The Structure of Stenothricin - Revision of a Previous Structure Assignment. *Liebigs Annalen Der Chemie.* 1672-1684 (1984).

32. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D.H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *Bmc Evolutionary Biology.* **7**, 78-92 (2007).

33. Crawford, J.M., Portmann, C., Kontnik, R., Walsh, C.T. & Clardy, J. NRPS substrate promiscuity diversifies the xenematides. *Org Lett.* **13**, 5144-5147 (2011).

34. Pirri, G., Giuliani, A., Nicoletto, S.F., Pizzuto, L. & Rinaldi, A.C. Lipopeptides as anti-infectives: a practical perspective. *Central European Journal of Biology.* **4**, 258-273 (2009).

35. Baltz, R.H., Miao, V. & Wrigley, S.K. Natural products to drugs: daptomycin and related lipopeptide antibiotics. *Nat Prod Rep.* **22**, 717-741 (2005).

36. Martin, N.I. et al. Isolation, structural characterization, and properties of mattacin (polymyxin M), a cyclic peptide antibiotic produced by Paenibacillus kobensis M. *J Biol Chem.* **278**, 13124-13132 (2003).

37. Wong, W.R., Oliver, A.G. & Linington, R. Development of Antibiotic Mode of Action Profile Screening for the Classification and Discovery of Natural Product Antibiotics. *Chemistry & Biology.* **19**, 1483-1495 (2012).

38. Peleg, A.Y., Seifert, H. & Paterson, D.L. Acinetobacter baumannii: emergence of a successful pathogen. *Clin Microbiol Rev.* **21**, 538-582 (2008).

39. Maragakis, L.L. & Perl, T.M. Acinetobacter baumannii: epidemiology, antimicrobial resistance, and treatment options. *Clin Infect Dis.* **46**, 1254-1263 (2008).

40. Dijkshoorn, L., Nemec, A. & Seifert, H. An increasing threat in hospitals: multidrug-resistant Acinetobacter baumannii. *Nat Rev Microbiol.* **5**, 939-951 (2007).

41. Lorang, J. & King, R.W. Cytological profiling: providing more haystacks for chemists' needles. *Genome Biol.* **6**, 228 (2005).

42. Perlman, Z.E. et al. Multidimensional drug profiling by automated microscopy. *Science.* **306**, 1194-1198 (2004).

43. Lamsa, A., Liu, W.T., Dorrestein, P.C. & Pogliano, K. The Bacillus subtilis cannibalism toxin SDP collapses the proton motive force and induces autolysis. *Molecular Microbiology.* **84**, 486-500 (2012).

44. Steenbergen, J.N., Alder, J., Thorne, G.M. & Tally, F.P. Daptomycin: a lipopeptide antibiotic for the treatment of serious Gram-positive infections. *J Antimicrob Chemother.* **55**, 283-288 (2005).

45. Paetzel, M., Goodall, J.J., Kania, M., Dalbey, R.E. & Page, M.G. Crystallographic and biophysical analysis of a bacterial signal peptidase in complex with a lipopeptide-based inhibitor. *J Biol Chem.* **279**, 30781-30790 (2004).

46. Ngoka, L.C. & Gross, M.L. Multistep tandem mass spectrometry for sequencing cyclic peptides in an ion-trap mass spectrometer. *J Am Soc Mass Spectrom.* **10**, 732-746 (1999).

47. Leao, P.N. et al. Synergistic allelochemicals from a freshwater cyanobacterium. *Proc Natl Acad Sci U S A.* **107**, 11183-11188 (2010).

48. McLafferty, F.W. Mass Spectrometric Analysis. Molecular Rearrangements. *Analytical Chemistry.* **31**, 82-87 (1959).

49. Chill, L., Kashman, Y. & Schleyer, M. Oriamide, a new cytotoxic cyclic peptide containing a novel amino acid from the marine sponge Theonella sp. *Tetrahedron.* **53**, 16147-16152 (1997).

50. Laird, D.W. et al. Halogenated cyclic peptides isolated from the sponge Corticium sp. *Journal of Natural Products.* **70**, 741-746 (2007).

51. Graham, D.E., Taylor, S.M., Wolf, R.Z. & Namboori, S.C. Convergent evolution of coenzyme M biosynthesis in the Methanosarcinales: cysteate synthase evolved from an ancestral threonine synthase. *Biochem J.* **424**, 467-478 (2009).

52. Thomas, M.G., Chan, Y.A. & Ozanick, S.G. Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob Agents Chemother.* **47**, 2823-2830 (2003).

53. Felnagle, E.A., Rondon, M.R., Berti, A.D., Crosby, H.A. & Thomas, M.G. Identification of the biosynthetic gene cluster and an additional gene for resistance to

the antituberculosis drug capreomycin. *Applied and Environmental Microbiology*. **73**, 4162-4170 (2007).

54. Zhang, W. et al. Nine enzymes are required for assembly of the pacidamycin group of peptidyl nucleoside antibiotics. *J Am Chem Soc.* **133**, 5240-5243 (2011).

Chapter 4, in full, is the material that is being prepared for publication. Wei-Ting Liu, Anne Lamsa, Weng Ruh Wong, Paul D. Boudreau, Roland Kersten, Peng Yao, Wilna Moree, Brendan M. Duggan, Bradley S. Moore, William H. Gerwick, Roger G. Linington, Kit Pogliano, Pieter C. Dorrestein. The thesis author was the primary investigator and author of this paper.

Chapter V

Future directions – A proposal

Gaining molecular insight into fecal transplantation for the treatment of *Clostridium difficile*-associated diarrhea

5.1 Abstract

Clostridium difficile-associated diarrhea continues to be a significant health threat in hospitalized patients because of the high recurrence rates and frequent failures to respond to multiple courses of antibiotics. Fecal transplantation is an alternative and highly effective approach for recurrent, refractory *C. difficile* infections (CDI) with a better than 90% success rate and no reported detrimental side effects. Despite its remarkable success, the molecular mechanisms that underlie fecal transplantation are poorly understood. The aim of this study is to develop new methodologies that incorporate advanced genomic and metabolic approaches in order to characterize the molecular mechanisms that mediate the beneficial effects of fecal transplantation. Specifically, DNA sequencing and mass spectrometry techniques will be used to monitor changes in microbiota genomic (species composition) and metabolic (chemical composition) profiles before and during the restoration process associated with fecal transplantation. These insights will lead to the identification of previously unknown metabolites, possibly leading to new avenues of treatment for *C. difficile* infections. The approaches developed in this study could be widely applied to other important investigations of human microbiota associated diseases/disorders, such as Crohn's diseases, obesity, and diabetes.

5.2 Background and Significance

Clostridium difficile infection (CDI), one of the major causes of infectious diarrhea and colitis, is an emerging health threat in hospitalized patients, with an estimated ~500,000 cases per year in the United States, causing 15,000-20,000 deaths; comparable to the number of deaths attributable to HIV in the US[1]. While certain antibiotics such as metronidazole and vancomycin provide clinical benefit, challenges in effectively treating CDI include clinical resistance and high likelihood of reinfection due to residual *C. difficile* spores that persist in the intestinal tract. When standard antibiotic treatment fails, fecal transplantation is an alternative and highly effective approach with an ~90% success rate[2-4]. Fecal transplantation is performed by transferring the stool from a healthy individual into a recipient, in order to restore the gut microbiota to a health-associated state in which *Bacteroidetes* and *Firmicutes* species, which are often missing in CDI, are re-established, and *C. difficile* is eradicated[5]. However, fecal transplantation is considered a last resort for the treatment of recurrent CDI, as this approach has been met with reluctance from patients and physicians due to concerns over its safety and the fact that it is a poorly characterized therapeutic and an unconventional treatment. Furthermore, while this practice has been performed since 1958[6], the molecular mechanisms that underlie its effects remain elusive, hindering the development of more precise approaches. With the ongoing advances in genome sequencing technologies, computational methods, as well as proteomic and metabolic (i.e. metabolomics and secondary metabolite profiling) analyses, a more complete understanding of the human microbiome is now possible[7-13], creating new opportunities for more well-defined and directed therapies[14,15]. **We hypothesize that the molecular mechanisms behind fecal**

transplantation include the killing or inhibition of *C. difficile* by specific chemical factors produced by a "healthy" commensal gut microbiota (Fig. 5.1). To address this hypothesis, we aim to develop new methodologies that incorporate advanced genomic and metabolic techniques and knowledge in order to characterize the molecular mechanisms that underlie fecal transplantation. A recent clinical review that followed 94 patients after fecal transplantation revealed that although ~90% of patients eventually had resolution of their diarrhea, some responded sooner (within 3 days) and others responded later (up to 2 months)[16]. Thus, by comparing the genomic and chemical profiles of the gut microbiota associated with these different classes of responders, we expect to identify putative beneficial effectors of fecal transplants, which may facilitate the development of new therapeutic agents to treat *C. difficile* infections.

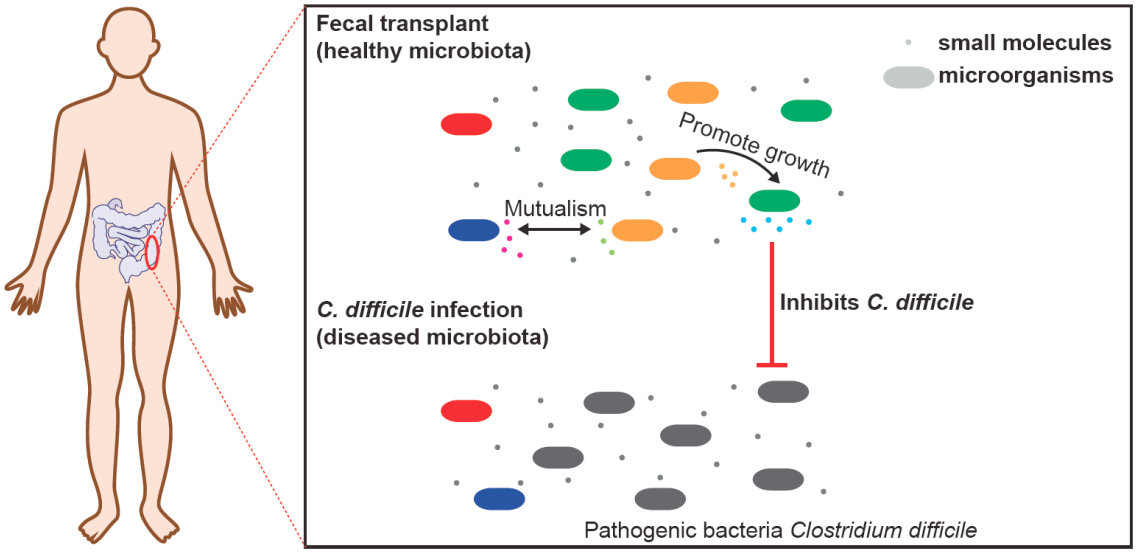


Figure 5.1: Outline of the hypothesized molecular model behind fecal transplantation. One possible molecular mechanism of fecal transplantation is the killing or inhibition of *C. difficile* by specific chemical factors produced by the microbiota introduced through fecal transplantation. Specifically, microorganisms interact via small molecules to establish relationships such as commensalism, mutualism, and antagonism. Interactions are critical for homeostasis in the gut ecosystem.

5.3 Specific Aims

In order to understand the mechanisms by which fecal transplants have a beneficial effect, we seek to identify important strains and key small molecules. Towards this end, we will monitor changes in microbiota genomic (species composition) and metabolic (chemical composition) profiles before and during the restoration process associated with fecal transplantation. Three specific aims are delineated as follows:

Aim 1: Genomic profiling to reveal key microorganisms associated with the beneficial effects of fecal transplantation. 16S rDNA and community metagenomic sequencing will be used to determine the species composition, and changes in microbial community composition over time, as well as to compare microbiota temporal compositional variation between patients that respond differently in terms of the timing of resolution of diarrhea.

1a. Fecal samples from patients prior and after the fecal transplantation, as well as a portion of the fecal transplant, will be collected, prepared, and subjected to DNA sequencing.

1b. Results from Aim 1a will be analyzed to reveal microbiota shifts over time, and to compare the microbiota succession profiles between patients that respond to fecal transplantation differently.

Aim 2: Metabolic profiling to reveal key molecules that are responsible for a healthy gut microbiota. Mass spectrometry and cutting edge bioinformatics approaches will be

used to compare the chemical profiles of gut microbiota, and to reveal key molecules that lead to or maintain a healthy gut microbiota.

2a. From the same fecal samples as in Aim 1, mass spectrometric data sets will be collected to reveal chemical profiles.

2b. Results from Aim 2a will be processed and compared to reveal potential key molecules (e.g., molecules whose patterns of abundance presage and predict resolution of diarrhea).

Aim 3: Identify and verify the bioactivity of some key molecules. Based on results from Aim 2, we will select molecules of interest, identify them, and assess their bioactivity against *C. difficile*.

3a. Identity molecules of interest.

3b. Isolate and test bioactivity.

Training potential. This proposal requires extensive inter-connections between genomic and metabolic expertise. The host laboratory Dr. David Relman, Stanford University, has strong expertise in clinical medicine and microbiology and is among the pioneers in genomic analysis of the human microbiome, and co-mentor Dr. Michael Fischbach at UCSF is a leading expert in genome mining and natural product (secondary metabolites) chemistry and biosynthesis. Under their mentorship, I will have access to clinical samples as well as cutting edge tools and knowledge necessary to investigate the microbial metabolic interactions that control *C. difficile*.

5.4 Research Strategy

Approach Summary. The overall goal of this proposal is to identify bioactive molecules that control *C. difficile*. There are three key steps in this investigation (**Fig. 5.2**). First, DNA sequencing techniques will be used to reveal key microbial players and their partial genomes. Second, tandem mass spectrometry will be used to obtain chemical profiles revealing small molecules that may be involved in clinical improvement. Both genomic and metabolic datasets will be collected over multiple time points in the same individuals relative to the fecal transplantation. Third, with the aid of temporal information revealed in steps 1 and 2, molecules of interest will be correlated to possible producing strains based on the biosynthetic potential that can be revealed from bacteria genomes. Identification of candidate key molecules, along with their producing strains, will allow subsequent isolation, purification, and structural characterization with evaluations of bioactivity.

Aim 1: Genomic profiling to reveal key microorganisms associated with the beneficial effects of fecal transplantation. Within this aim, we (through clinical collaborators) will recruit patients who are suffering from recalcitrant, recurrent *C. difficile* infection, for whom fecal transplantation is indicated and for which they have agreed. Fecal samples will be collected daily or weekly before and after the procedure and immediately frozen. Samples will be subjected to DNA sequencing for characterization of microbiota taxonomic composition and functional potential. Patterns will be compared between patients that respond differently in terms of timing and degree of resolution of diarrhea.

1a. Determine the microbiota composition. Daily fecal samples from patients one week before, and two weeks after the procedure, and weekly until three months after the procedure, will be collected and immediately frozen. A portion of the fecal transplant materials from the healthy donors will also be collected for analysis. Each sample will be processed and subjected to 16S rDNA amplicon sequencing to determine the microbial species composition, and shotgun metagenomic sequencing to obtain partial genomes that can serve as reference datasets providing vital information of their genetic potential that will be utilized in Aim 3.

1b: Compare the microbiota succession profiles between patients that respond to fecal transplantation differently. Differences in patterns of microbiota succession between individuals that respond to fecal transplantation differently in terms of resolution of diarrhea will be assessed. Multiple kinds of datasets from the same samples over time will enable the correlation of genotype (gut microbiota succession) to phenotype (clinical responses). This will provide insight into why and how individuals respond to fecal transplantation differently.

Expected results. Within this specific aim, we expect to address several important issues for which our current understanding is incomplete, including microbiota taxonomic compositional similarities between patients post-procedure and the healthy donors, differences in a patient's gut microbiota composition before and after a procedure, the succession of the microbiota during the restoration process, and the difference in microbiota succession patterns between multiple patients. The results from Aim 1 will define the key microbial players within the gut ecosystem and would provide a foundation for the investigation of metabolic interactions within specific consortiums.

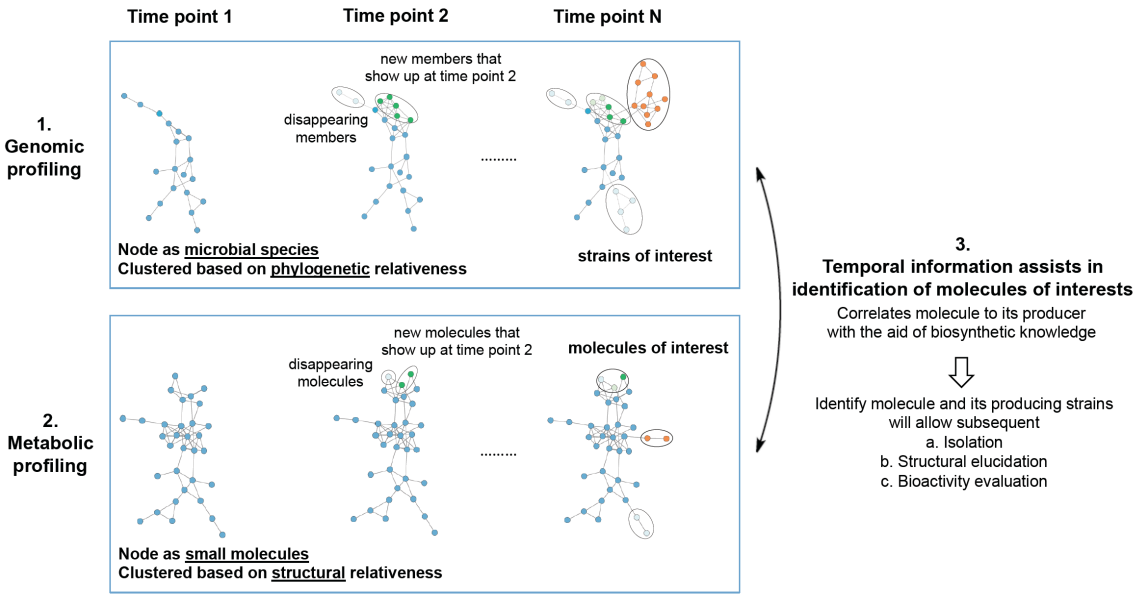


Figure 5.2: A schematic overview of the experimental approach. Genomic and metabolic datasets will be collected over time before and after the fecal transplantation procedure. Then, with the aid of temporal information, molecules of interest will be identified and correlated with possible producing strains based on the biosynthetic potential that can be revealed from bacteria genomes. Identification of candidate key molecules, along with their producing strains, will allow subsequent isolation, purification, and full structural characterization, as well as evaluation of bioactivity.

Aim 2: Metabolic profiling to reveal key molecules that are responsible for a healthy gut microbiota. It has long been appreciated that microbes can interact via metabolic exchange[13,17,18]. Mass spectrometry (MS) techniques and cutting edge bioinformatics approaches will be used to compare the metabolic profiles of gut microbiota at various time points relative to the transplantation procedure and between patients with different responses to reveal candidate molecules that lead to or maintain a healthy gut microbiota.

2a. Obtain chemical profiles using mass spectrometry techniques. From the same fecal samples described in Aim 1, approximately 1 gram of fecal material will be treated with ethyl acetate, butanol, water, and methanol to extract metabolites of differing

polarities. Extracts will be dried and re-dissolved in solvents suitable for electrospray ionization mass spectrometry, and extensive molecular information will be gathered using data dependent fragmentation (MS/MS). When appropriate, simple desalting procedures or liquid chromatography separations can be performed prior to MS analysis for data quality improvement.

2b. Metabolic profiling to reveal key molecules that are responsible for a healthy gut microbiota. Results from Aim 2a will be organized using a newly developed molecular MS/MS network approach, in which molecules with structural similarities are revealed based on the alignment and clustering of mass spectrometric fragmentation patterns[19]. The benefits of adopting this approach include being able to organize the metabolic profiles of samples based on their chemical structural components, straightforward visualization of global molecular information from a single sample, and the ability to display and interconnect data derived from different samples. With clustering of shared metabolites (both identical and similarly structured) from different samples, the unique signals that are pertinent to specific datasets (e.g., those molecules correlated with time to resolution of diarrhea) will stand out, which will drastically shorten the time it takes to pinpoint candidate key molecules that may be responsible for establishing or maintaining a healthy gut microbiota.

Expected results. Within specific Aim 2, we expect to map the changes in metabolic profile of the gut microbiota over the course of the restoration process, and also to compare the metabolic profiles of multiple patients. We expect that the investigation of metabolic profiles will reveal several candidate molecules that may be

responsible for inhibiting *C. difficile* growth (antibiotics) or vital metabolic exchange factors that are essential for establishing and/or modulating healthy microbiota.

Aim 3: Identify and verify the bioactivity of some key molecules. Within this aim, we will adopt new genome mining approaches that will connect genomic and metabolic information. These approaches will be based on knowledge of the biosynthetic machinery of the interacting microbes, as well as on characterization of the structural components of the molecules of interest derived from Aim 2. Following putative identifications we will verify the potential bioactivity of select molecules.

3a. Identify molecules of interest. In this aim, we will adopt an innovative approach that is based on connecting the chemical features of a molecule of interest, as revealed from MS/MS, to its corresponding biosynthetic gene cluster in order to reveal the producing strains. Interesting molecules suggested from Aim 2 will be subjected to MS/MS to reveal diagnostic fragmentation patterns. For example, the MS/MS fragmentation patterns of peptides reveal fragments that correspond to the loss of specific amino acids, while MS/MS of compounds containing sugars typically give losses of 164 or 178 Da, and the loss of phosphate typically results in a mass loss of 80 Da. Many bioactive molecules in the human microbiome are secondary metabolites and are synthesized by a series of genes that are clustered nearby within the genome (so called biosynthetic gene cluster)[20,21]. Over the past few decades, knowledge of secondary metabolite biosynthesis has gradually accumulated. This knowledge now allows us to predict molecular structural features based solely on the composition of biosynthetic gene clusters found in the bacterial genome and to match these predicted structural features to the structural subunits inferred from the diagnostic mass losses derived from MS

fragmentation experiments. In this regard, the structural information derived from MS fragmentation will be used to "mine" many publicly accessible genomes, but more specifically strains that are phylogenetically related to the candidate species as suggested in Aim 1. If a search database is too large, thus yielding ambiguous matches, publicly available strains that are closely related to candidate strains can be obtained, cultured, and subjected to MS/MS molecular profiling. The resulting MS/MS data can be networked along with selected MS/MS spectra of a molecule of interest in order to reveal which strains may produce identical or similar molecules to the molecule of interest. The information provided by this type of experiment can be confirmed by verifying that the metagenome of a species from Aim 1 contains the biosynthetic potential to produce such a molecule. Recent work has demonstrated that this strategy, which utilizes publicly available genomic sequence repositories of known microorganisms to search against a network of MS/MS fragmentation spectra from an unsequenced organism, can successfully lead to the genome mining of small molecules from microorganisms that do not yet have genome sequence information (Dorrestein lab, unpublished). In the worst-case scenario, if the "borrowing genome" approach described above fails due to a lack of known strains that possess a biosynthetic gene cluster similar to that which encodes the molecule of interest, then the metagenome sequence data collected from Aim 1 will be utilized to serve as the searching database. This strategy will require high quality (partially) re-constructed genomes, and thus will be far more challenging, yet it can provide an alternative approach.

3b. Isolate and verify the bioactivity. Identifying the molecules of interest and the corresponding producing strain in Aim 3a will enable the design of specific culturing

conditions to isolate the targeted strain from fecal samples. Since the main challenge of characterizing metabolites in human derived samples is the difficulty in obtaining unlimited amount of materials for isolation purposes, obtaining the producing strain will enable subsequent large-scale cultivation and isolation of molecules of interests. This will allow full structural characterization if there are novel molecules identified, and it will provide specific molecules to test against *C. difficile,* or for evaluation of other functions, as suggested in Aim 2, to verify their specific activity.

Expected Results. Expected outcomes of Aim 3 include the structural and biochemical characterization of metabolites that affect *C. difficile* growth, morphology and survival, as well as other vital metabolic exchange factors that are important in establishing or modulating health microbiota.

5.5 Future Directions

Incorporating both genomic and metabolic strategies for investigating the human gut microbiome will provide unique insight into the molecular mechanisms that mediate the beneficial effects of fecal transplantation. This insight will lead to the identification of previously unknown metabolites, possibly leading to new avenues of treatment for *C. difficile* infections. While this project focuses on fecal transplantation and *C. difficile*-associated diarrhea, a similar approach could be undertaken to investigate other important human microbiota associated diseases/disorders (i.e., Crohn's diseases, obesity, diabetes), or to study effects of nutritional and other perturbations on microbiota, or to decipher the process of primary assembly of the microbial communities in infants (from none to adult-like microbiota). The methods discussed in this proposal will provide vital genomic and metabolic insights into the healthy microbiota restoration process associated with fecal transplantation, but success of any facet of this proposal opens the door to new methods for both academic and clinical research.

References

1. Rupnik, M., Wilcox, M.H. & Gerding, D.N. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol.* **7**, 526-536 (2009).

2. Khoruts, A. & Sadowsky, M.J. Therapeutic transplantation of the distal gut microbiota. *Mucosal Immunol.* **4**, 4-7 (2011).

3. van Nood, E., Speelman, P., Kuijper, E.J. & Keller, J.J. Struggling with recurrent Clostridium difficile infections: is donor faeces the solution? *Euro Surveill.* **14** (2009).

4. Gough, E., Shaikh, H. & Manges, A.R. Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent Clostridium difficile infection. *Clin Infect Dis.* **53**, 994-1002 (2011).

5. Khoruts, A., Dicksved, J., Jansson, J.K. & Sadowsky, M.J. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol.* **44**, 354-360 (2010).

6. Eiseman, B., Silen, W., Bascom, G.S. & Kauvar, A.J. Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. *Surgery.* **44**, 854-859 (1958).

7. Sonnenburg, J.L., Chen, C.T. & Gordon, J.I. Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. *PLoS Biol.* **4**, e413 (2006).

8. Wikoff, W.R. et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A* **106**, 3698-3703 (2009).

9. Sellitto, M. et al. Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One.* **7**, e33387 (2012).

10. Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature.* **486**, 207-214 (2012).

11. Human Microbiome Project, C. A framework for human microbiome research. *Nature.* **486**, 215-221 (2012).

12. Relman, D.A. Microbiology: Learning about who we are. *Nature.* **486**, 194-195 (2012).

13. Yang, J.Y., Karr, J.R., Watrous, J.D. & Dorrestein, P.C. Integrating '-omics' and natural product discovery platforms to investigate metabolic exchange in microbiomes. *Curr Opin Chem Biol.* **15**, 79-87 (2011).

14. Sonnenburg, J.L. & Fischbach, M.A. Community health care: therapeutic opportunities in the human microbiome. *Sci Transl Med.* **3**, 78ps12 (2011).

15. Lemon, K.P., Armitage, G.C., Relman, D.A. & Fischbach, M.A. Microbiota-targeted therapies: an ecological perspective. *Sci Transl Med.* **4**, 137rv135 (2012).

16. Brandt, L.J. et al. Long-term follow-up of colonoscopic fecal microbiota transplant for recurrent Clostridium difficile infection. *Am J Gastroenterol.* **107**, 1079-1087 (2012).

17. Nicholson, J.K. et al. Host-gut microbiota metabolic interactions. *Science.* **336**, 1262-1267 (2012).

18. Rath, C.M. & Dorrestein, P.C. The bacterial chemical repertoire mediates metabolic exchange within gut microbiomes. *Curr Opin Microbiol.* **15**, 147-154 (2012).

19. Watrous, J. et al. Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A* **109**, E1743-1752 (2012).

20. Fischbach, M.A. & Walsh, C.T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev.* **106**, 3468-3496 (2006).

21. Oman, T.J. & van der Donk, W.A. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol.* **6**, 9-18 (2010).