METABOLOMIC AND BIOCHEMOINFORMATIC APPROACHES FOR MINING HUMAN MICROBIOTA FOR IMMUNOMODULATORY SMALL MOLECULES

By ROSTYSLAV ZVANYCH, H.BSc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University Master of Science (2014) Hamilton, Ontario (Chemical Biology)

Title: Metabolomic and Biochemoinformatic Approaches For Mining Human Microbiota For Immunomodulatory Small Molecules

Author: Rostyslav Zvanych, H.BSc. (McMaster University, Ontario, Canada)

Supervisor: Dr. Nathan A. Magarvey

Committee members: Dr. Gerard D. Wright and Dr. Michael G. Surette

Number of pages: xviii, 183

Abstract

The numerous benefits associated with natural products isolated from the environmental sources, including soil bacteria, plants and fungi, are long known and well appreciated. Interestingly, the immense number of microorganisms that reside within our bodies and whose cell counts greatly outnumber our own represents a potentially new and practically untapped reservoir of bioactive compounds. With the advent of next generation sequencing we are only now starting to realize the complexity and biological diversity of the human microbiome. With this ever-increasing flow of genomic information, more bioactive potential in these microbes can be identified. For instance, biosynthetic assembly lines responsible for the production of two largest classes of bioactive compounds, polyketides and nonribosomal peptides, can be readily identified within the microbial genomes, providing us with a view of their bioactive profiles.

In addition to the identification of biosynthetic assembly lines, the building blocks of polyketide and nonribosomal peptide products can also be accurately predicted, given the well-understood logic of assembly line operations. Nonetheless, the identification of actual products is still lagging behind. The discovery of these bioactive molecules can be achieved, however, by establishing a unique connection between genomes and molecules. Using several concrete examples, this thesis demonstrates how both metabolomic and biochemoinformatic platforms can assist in discovery of bioactive small molecules. More specifically, investigations involving three members of the human microbiome, *Streptococcus mutans, Lactobacillus plantarum* and *Pseudomonas aeruginosa*, provide

distinct examples of identification of bioactive agents and assessment of their immunomodulatory potential.

Studying the human microbiome form the angle of small molecules is critical for evaluation of microbial effects on our cells, and ultimately our health. Although this interaction is likely multifaceted and deeply interconnected with a number of other factors, identification of small molecules is an important step in demystifying their role. Studying these agents will hopefully reveal interesting principles on how microorganisms speak to human cells and how this communication could lead to therapeutic strategies or downstream mechanistic revelations.

Acknowledgements

First, I would like to recognize my supervisor and a great mentor, Dr. Nathan Magarvey for taking me on as a graduate student and fostering my scientific, as well as personal development. His forward thinking, profound scientific vision and outmost positivity fueled my everyday motivation. The invaluable skills that Nathan helped me develop during my graduate studies will surely be reflected in my professional education and future career.

My scientific development was greatly advanced thanks to my great committee members. It was a true privilege to work with Dr. Gerry Wright, whose insightful comments and experimental suggestions during committee meetings were simply invaluable. It was likewise a great fortunate to not only have Dr. Michael Surette as my committee supervisor, but to also collaborate on a number of projects.

I would also like to recognize a number of individuals who I was privileged to work with during my time at the Magarvey Lab. Dr. Xiang Li is not only a great friend, but also a phenomenal scientist whose NMR and drug discovery skills were absolutely essential in all of my research endeavors. I would also like to acknowledge Ashraf Ibrahim, who taught me pretty much everything I know about mass spectrometry and chromatographic separation to date. My transition into graduate studies and initial scientific development would not be possible if not for Dr. Morgan Wyatt, who has been incredibly instrumental and helpful from day one. It was also a great pleasure to work and collaborate with Chad Johnston, who is a multi-talented scientist, and truly one of a few individuals to demonstrate outmost dedication, passion and persistence towards science.

My acknowledgments extend to Satheeisha Tharmarajah, who was one of a kind undergraduate student who gladly and passionately assisted me in a number of my research projects. I would like to thank Ron Hermenau and Nicole Rasanayagam for their invaluable inputs as well. Lastly, I would like to thank Nikola Lukenda for pioneering some of my research and educating me about the value of good science.

My research progress was also greatly assisted by inputs of Vi Dang and I am truly grateful for her expertise in metabolomics and mass spectrometry. Similarly, I would like to thank Inga Kireeva for her patience and expertise in mass spectrometry, which were truly priceless. It was also a great pleasure to work with Dr. Waliul Khan and Janice Kim, whose collaborative inputs in my research were invaluable. I would also like to thank the Surette Lab and especially Dr. Matthew Workentine and Dr. Steve Bernier who gladly provided several bacterial strains for my research and Jennifer Lau and Anne-Marie Lacroix who assisted me with anaerobic culturing. I am grateful to Breanne Cuddington and Dr. Joseph McPhee for their assistance in my research with the tissue cultures. Lastly, I would like to thank Dr. Dan Sørensen for his extensive NMR expertise and assistance in a number of experiments.

On a more personal note, I would like to thank my dear parents, Roman and Oksana Zvanych, and the rest of my family. Their love, support and words of encouragement were absolutely essential throughout my life journey, and especially so during my graduate studies.

Table of Contents

Lists of Figures

Chapter 1

Chapter 2

Chapter 4

Chapter 5

Abbreviations

A – adenylation

ACP – acyl carrier protein

AMB – L-2-amino-4-methoxy-trans-3-butenoic acid

AMP – adenosine monophosphate

AT – acyltransferase

BHI – brain heart infusion

BPC – base peak chromatogram

C – condensation

CDM – chemically defined medium

CF – cystic fibrosis

CID – collision-induced dissociation

CoA – coenzyme A

COSY – correlation spectroscopy

DH – dehydratase

DMSO – dimethyl sulfoxide

EIC – extracted ion chromatogram

ELISA – enzyme-linked immunosorbent assay

ER – enoylreductase

ESI – electrospray ionization

FBS – fetal bovine serum

G-CSF – granulocyte colony-stimulating factor

GNP – Genomes to Natural Products

HHQ – 4-hydroxy-2-heptylquinoline

HMBC – heteronuclear multiple-bond correlation spectroscopy

HMP – Human Microbiome Project

HPLC – high-performance liquid chromatography

HRMS – high-resolution mass spectrometry

HSL – homoserine lactone

HSQC – heteronuclear single-bond correlation spectroscopy

IFN-γ – interferon-gamma

IL – interleukin

iSNAP – informatic Search algorithm for NAtural Products

KS – ketosynthase

LB – Luria-Bertani broth

LC-MS – liquid chromatography-mass spectrometry

LPS - lipopolysaccharide

MCP-1 – monocyte chemoattractant protein-1

MO – monooxygenase

MPLC – medium pressure liquid chromatography

MRM – multiple reaction monitoring

MRS – de Man, Rogosa and Sharpe

MS – mass spectrometry

MS/MS – tandem mass spectrometry

NADH – nicotinamide adenine dinucleotide

NF-κB – nuclear factor-kappa B

NMR – nuclear magnetic resonance (spectroscopy)

NRP – nonribosomal peptide

NRPS – nonribosomal peptide synthetase

OPLS-DA – orthogonal partial least squares discriminant analysis

PBS – phosphate buffered saline

PCA – principal component analysis

PK – polyketide

PKS – polyketide synthase

PQS – Pseudomonas quinolone signal

Re – reductase

RPMI – Roswell Park Memorial Institute medium

SMILES – simplified molecular-input line-entry system

T – thiolation

TE – thioesterase

THYE – Todd Hewitt Yeast Extract

TNF-α – tumor necrosis factor-alpha

TSB – tryptic soy broth

YNB – yeast nitrogen base

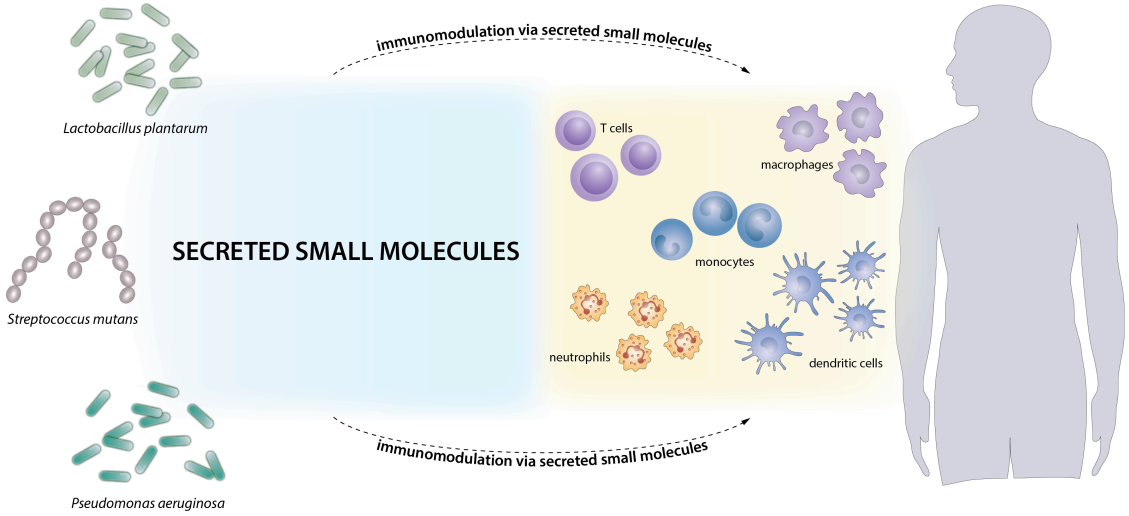
YPD – yeast extract-peptone-dextrose

Declaration of Academic Achievement

This thesis is formatted for review as a sandwich thesis. Specific details regarding each research contribution are described within each chapter preface.

Chapter 1. Introduction

1.1 Thesis Context



The human microbiome offers a unique medium for mining biologically active small molecules. Much of the current research has focused on understanding the immunomodulatory effects of microbial cell wall components, however, the roles of secreted small molecules produced by these microbes remain largely enigmatic. Modern advances in genome sequencing made it possible to not only catalogue microbes, but to also reveal their bioactive potential. Although secreted metabolomes of most of these microbes have not been looked at extensively, the inference about the production of two major classes of bioactive small molecules, polyketides and nonribosomal peptides, can be made from their genomes. However, despite the rapidly expanding wealth of genomic information and the predictive power of bioinformatic algorithms, the identification of these bioactive agents is still lagging behind. Therefore, this is an exciting area to be

explored, especially with regards to understanding the interactions of secreted small molecules with our cells.

The central hypothesis that binds this body of work is that the human microbiota modulates human cells by secreting unique metabolites that possess immunomodulatory properties; identification and characterization of these agents will reveal new chemistry and assaying them will uncover intriguing immunomodulatory activities.

1.2 The Human Microbiome

Since the moment we are born our bodies are continuously exposed to an immense number of microorganisms including bacteria, fungi and viruses. Over the years, some of these microorganisms, upon contact with our skin, gut or oral cavity, acquire residency within our bodies. The core collective of these microorganisms is known as the human microbiota or the human microbiome, which also accounts for the biomolecules that these microbes produce[1]. Although the majority of microbiota resides in the gut, microbes can be found on every surface of our bodies, and their grand number is simply remarkable. Recent projections from sampling and metagenomic profiling studies suggest that the microbial cohort of cells outnumbers our own by ten-fold[2,3]. Given the sheer size of their community, these microbes are thought to have coevolved to overcome host protective barriers to avoid recognition[4]. The diversity of functional roles that these organisms undertake is of intense study in relation to their infectious, as well as beneficial properties. Nonetheless, despite the large number and diversity of the human microbiota, the effects of these microbes are still largely uncertain due to the nascent nature of this research field.

Much research has focused on the bacteria that cause infectious diseases. These include skin pathogens such as *Staphylococcus aureus*; bacteria associated with lung infections such as *Pseudomonas aeruginosa*; bacteria indigenous to the gut such as *Helicobacter pylori*; bacteria associated with dental caries such as *Streptococcus mutans* and many others. Much less in known, however, of the collective of microbes that do not cause disease, and the agents they create to affect human health. Evidence supporting the beneficial functions of microbes is long-known, and in several instances organisms have been used as probiotics to dampen or combat diseases such as Crohn's and colitis[5-8]. Organisms in this category include Lactobacilli, Bifidobacteria and Streptococci and repurposing these seemingly beneficial microbes as therapeutic modalities has been established[9,10]. However, the molecular mechanisms that underlie their beneficial effects are also lesser understood.

1.3 Evidence for Immunomodulation

Although the research is still in its emerging stages, efforts have been made to link some of the microbial agents to their roles in interacting with our cells. In the case of pathogens, immunomodulation is one of their approaches to circumvent our immunity, both innate and acquired[4]. For example, *Pseudomonas aeruginosa* secretes a series of proteinaceous toxins that manipulate immunity by disrupting signaling cascades associated with the immune response[11]. In the context of pathogenesis and virulence, proteins and known small molecules have been the focal points with respect to understanding how these organisms cause disease.

Although sparse, there is some evidence that suggests the involvement of small molecules in immunomodulation. For instance, acyl homoserine lactones, molecules involved in cell-to-cell signaling commonly known as quorum sensing, have been shown to interact with various eukaryotic cells and modulate immune responses[12-17]. In particular, one of such molecules, 3-oxo-C_{12}-HSL, was shown to inhibit the activation of lipopolysaccharide-stimulated macrophages by tuning down the secretion of pro-inflammatory cytokines such as tumor necrosis factor alpha and interleukin-12[18]. Other research studies using lymphocytes indicated the involvement of 3-oxo-C_{12}-HSL in anti-proliferative and cytokine release-modulating effects[19,20]. In addition, its immunomodulatory activity in lymphocytes was further confirmed by the observation of its anti-inflammatory and immunosuppressive properties[21].

In select cases, studies have revealed that small molecule fragments of either cell walls or cell membranes may contribute to the observed immunomodulatory activities of microbes[22-24]. More specifically, the zwitterionic polysaccharides produced by the symbiotic bacteria have been shown to promote health by demonstrating unique anti-inflammatory properties that protect the gut from abscess[23]. The mechanism of such protection is related to their ability to modulate cytokine secretion, in particular interleukin-10, induced secretion of which has been shown to provide protection against inflammation[25,26]. A key player among zwitterionic polysaccharides is polysaccharide A, which is considered the most immunodominant and has been used as a tool to explain the evolved ability of commensal bacteria to induce protective immune responses and provide protection against inflammation[27]. Nonetheless, the affects of secreted small

molecules produced by commensal bacteria and other members of the human microbiota are largely poorly understood.

1.4 The Human Microbiome Project

Realizing the important implications of select representatives of the human microbiota on our health, international research efforts have been put forward to delineate the effects of microbial communities on our cells. Much of these efforts hinge on the research advancements made by the Human Microbiome Project (HMP) that launched in 2007[28]. In fact, the field of human microbiome research expanded exponentially starting from the early 2000's, as may be inferred from the number of "human microbiome" counts on PubMed (**Figure 1.4**). The main goal of looking into the microbiome was to understand what functional roles these microbes may have and how they influence our

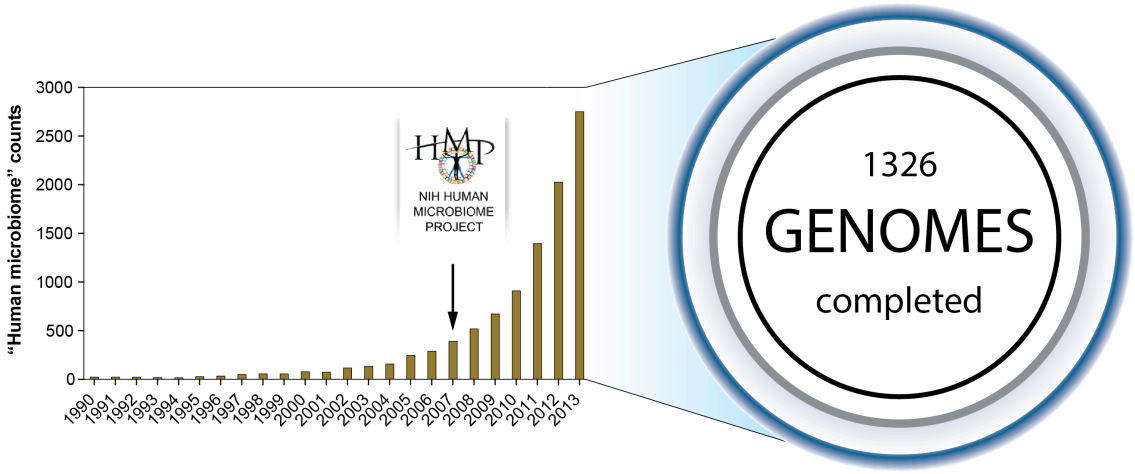


Figure 1.4. The exponentially increasing number of "human microbiome" counts on PubMed. The rapid increase is largely due to the advances made by the Human Microbiome Project that launched in 2007. One of the major achievements is the collection of 1326 completed genomes (http://www.hmpdacc.org).

health or cause disease. The goal of the HMP was precisely that – to establish the healthy human microbiome make-up and use it as a reference in comparative studies.

One way to catalogue microbes is to sequence their genomes. Interestingly, one of the major accomplishments over the past few years has been an increasing rate of genome sequencing. This increase in building the genomic wealth was largely due to the advent of next generation sequencing that put microbial genome sequencing on an upward spiral of progress[29]. As of early 2014, efforts made by the HMP led to the completion of 1326 reference microbial genomes from the human microbiome (**Figure 1.4**), and more are in the queue.

1.5 Polyketides and Nonribosomal Peptides

Due to the nascent nature of human microbiome research, the immunomodulatory effects of metabolic endproducts of specialized pathways, such as those associated with secondary metabolism, have not been extensively examined in its members. Polyketides and nonribosomal peptides are important metabolic endproducts and represent two major classes of bioactive compounds whose highly privileged scaffolds have been appreciated for many decades. Their bioactive profiles have been utilized for targeting a multitude of human diseases. Some examples of polyketides are epothilone B, nystatin, erythromycin and rapamycin[30-33]. The latter is a well-known and powerful immunosuppressant[34]. Examples of nonribosomal peptides include penicillin, vancomycin, bleomycin and cyclosporin A[35-38]. The latter is also a well-known immunomodulatory agent[39].

The information about production of these privileged classes of bioactive compounds can be inferred from the genome of their corresponding producers using well-

developed bioinformatic algorithms. The understanding of biosynthetic logic used by both polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) systems allows for making predictions of the final biosynthetic products. In fact, a number of bioinformatic programs have been designed to locate biosynthetic assembly lines and predict their respective products based on the known affinities of the acyltransferase (PKS) or adenylation (NRPS) domains. Such programs include SBSPKS, SMURF, ClustScan, CLUSEAN, NRPSpredictor2, NP.searcher and many others[40-45]. One of the most recently developed software suites, AntiSMASH, allows for entire genome analysis with complete annotation of domain organization, including automatic prediction of the chemical structure of final products based on the monomer selectivity[46].

Polyketides are synthesized by the multimodular enzymatic assembly lines, known as polyketide synthases (PKSs), with every module adding more monomer units and introducing further chemical complexity[47]. The most simplistic module organization consists of three domains – ketosynthase (KS), acyltransferase (AT) and thiolation (T) domains, often connected in KS-AT-T sequence (**Figure 1.5**). The functional role of KS is to form C-C bonds during the elongation event by attaching additional malonyl or methylmalonyl units that have been introduced by the AT domain. The role of T domain is that of a carrier protein, and is to provide a scaffold on which the growing polyketide products are assembled. The termination event is often carried out by the thioesterase (TE) domain, where the water molecule or an intramolecular nucleophilic attack releases the final products from the assembly line. In addition, a number of tailoring enzymes may

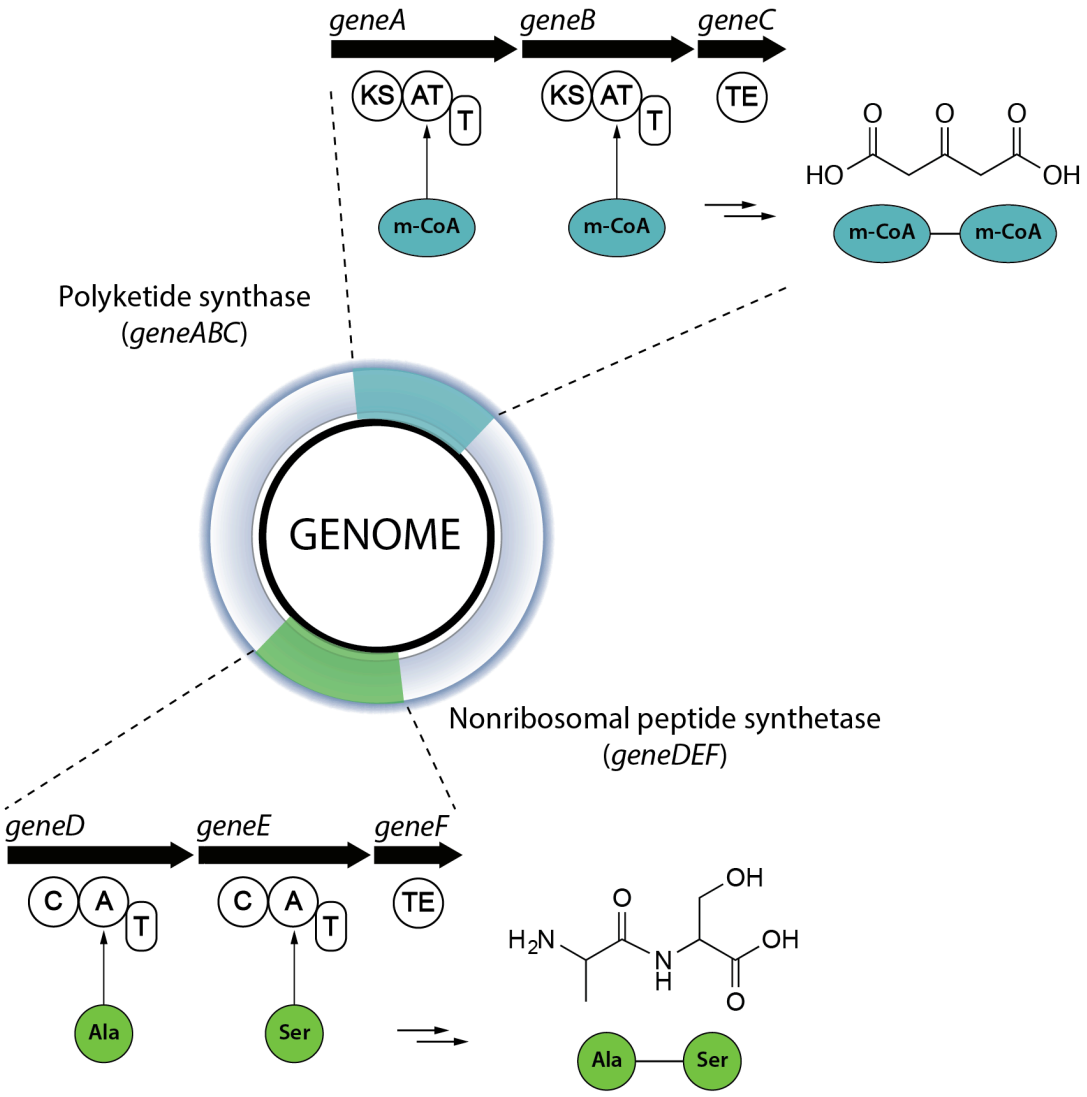


Figure 1.5. The biosynthetic logic of polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs). The KS domain in PKSs forms the C-C bond between the two malonyl-CoA (m-CoA) units that were introduced by the AT domain. The C domain in NRPSs forms a peptide bond between the two amino acids (e.g. alanine and serine), adenylated by the A domain. The roles of T and TE domains in both systems are to provide a scaffold on which the growing product is assembled and facilitate the liberation of the final product, respectively.

be present in a PKS cluster including ketoreductase (KR), dehydratase (DH) and enoylreductase (ER), which introduce further chemical complexity.

Nonribosomal peptides are synthesised similarly to polyketides, however, the multimodular enzymatic assembly lines, known as nonribosomal peptide synthetases (NRPSs), are made out of different enzymes and use different building blocks[47]. Similarly to PKSs, the most simplistic NRPS module consists of three domains – condensation (C), adenylation (A) and thiolation (T) domains, often connected in C-A-T sequence (**Figure 1.5**). The number of building blocks used by the NRPSs by far exceeds that of PKSs. In addition to 20 proteinogenic amino acids, there is a suit of over 500 non-proteinogenic amino acids, including aryl acids and α-hydroxy acids, amounting to an incredible downstream structural diversity[47,48]. Similarly to the logic employed in ribosomal protein synthesis, the adenylation domain activates the amino acid building blocks by conversion to aminoacyl-AMP and catalyzes their transfer to the T domain. Lastly, condensation (C) domain executes the actual peptide bond formation. Similarly to the PKSs, the release of final nonribosomal products from the assembly line is often achieved by a nucleophilic attack of a water molecule or an intramolecular cyclization, facilitated by the TE domain. Additional complexity may be introduced to the nonribosomal peptide products by tailoring enzymes such as epimerase (E) and methyltransferase (MT).

Although the bioinformatic predictions are easily accessible and the identification of biosynthetic machineries is possible on genome-wide level, identification of the actual products is still a major hurdle. The challenges associated with their identification are

multifactorial and are often complicated by the inaccurate predictions due to enzymatic promiscuity and post-assembly modifications. Therefore, finding the most useful means of selectively identifying these agents is of critical importance.

1.6 Metabolomic mining

The growing filed of metabolomics, that is the study of metabolites and their unique profiles, offers a number of exclusive modes to connect genomes to molecules. There are two widely recognized branches of metabolomics – targeted and untargeted[49]. In targeted metabolomics, a specific list of known metabolites is measured for a particular, often quantitative, purpose. In untargeted metabolomics, on the other hand, a global view of metabolites is considered and analyzed without bias to a particular compound class. When studying products of bacterial biosynthetic assembly lines, the wild-type strains (**Figure 1.6.1A**) and the strains with deactivated biosynthetic genes (**Figure 1.6.1B**) are often cultured in replicate and upon extraction are analyzed using mass spectrometry. Liquid chromatography coupled to mass spectrometry (LC-MS) is the preferred method for sample analysis in metabolomics, especially instruments with the time of flight mass analyzers[50,51]. Due to the intrinsic complexity of metabolomes, untargeted field of metabolomics relies heavily on the powerful statistical analyses such as multivariate analyses like orthogonal partial least squares discriminant analysis (OPLS-DA), often used for identification of biomarkers and principal component analysis (PCA), utilized for identification of outliers and assessment of biological variation[52]. The scores plot in PCA graphically visualizes variation between the groups under investigation and the tightness of clustering among replicate samples based on the two most significant

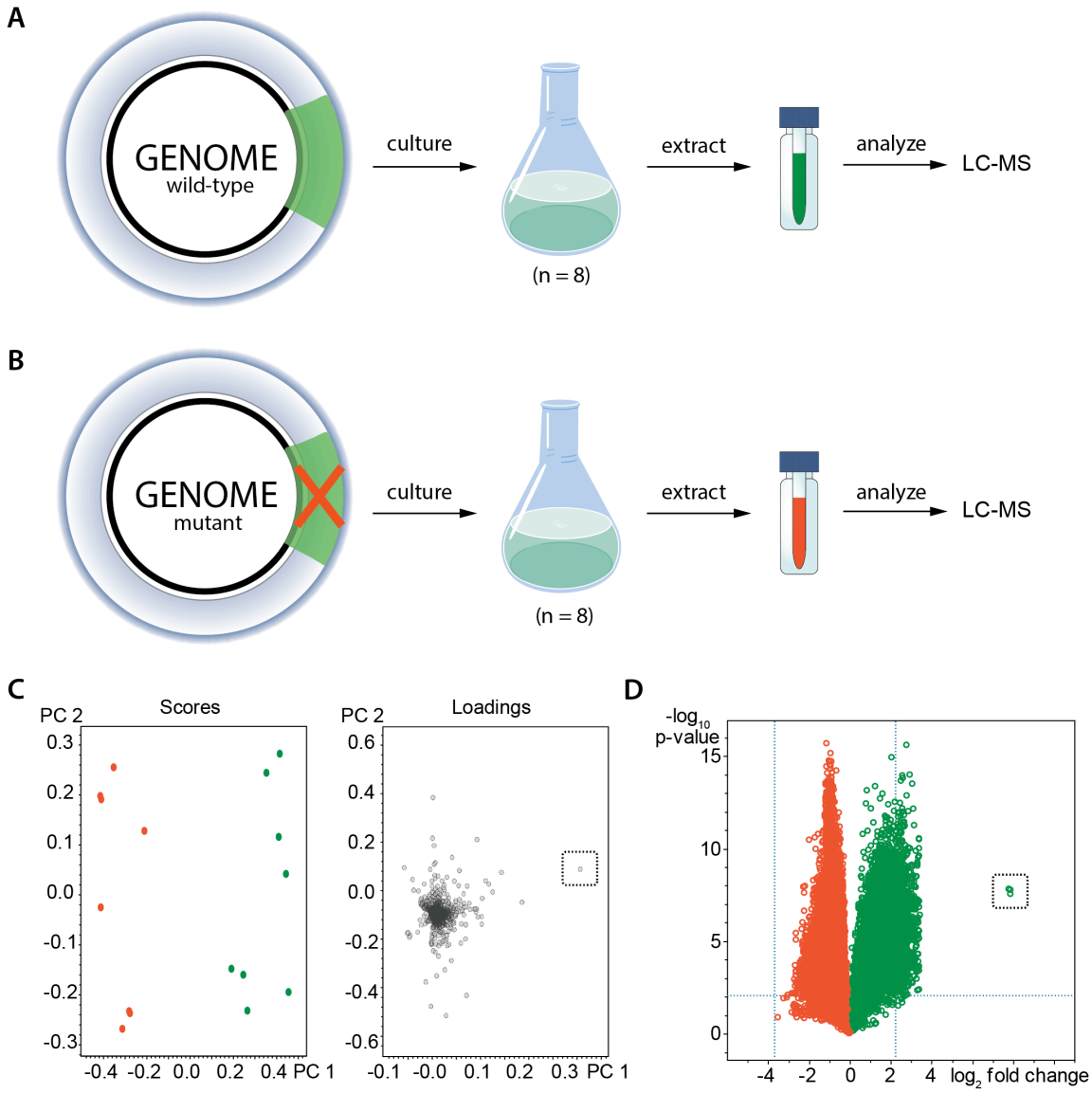


Figure 1.6.1. Multivariate and univariate analyses for metabolomic mining. **A** – an organism harbouring a functional biosynthetic cluster is cultured in replicate, extracted and analyzed using LC-MS; **B** – an organism harbouring a deactivated biosynthetic cluster is cultured in replicate, extracted and analyzed using LC-MS; **C** – typical principal component analysis (scores and loadings plots) output from the analysis of samples from **A** and **B**. **D** – volcano plot constructed from LC-MS data obtained for **A** and **B** samples.

principal components (**Figure 1.6.1C**). The loadings plot in PCA indicates what metabolite features (ions with unique *m/z* value and LC retention time) actually cause the observed separation in the scores plot. The core utility of these multivariate analyses is to identify which metabolite features make sample sets different and contribute the most to the observed variation. However, the utility of these multivariate analyses extends beyond the comparisons of different data sets. They can also be used to identify subtle differences within similar data sets, as the case is for bacterial cultures at different growth phases. Oftentimes, the differences between samples are not limited to one or two agents and to gain a systems view of these differences, univariate analysis such as volcano plot can be utilized[52,53]. The volcano plot is extensively useful for systems analyses, as it relies on the student's *t*-test and offers virtually unbiased representation of the differences based on the statistical significance (*p*-value) and the ratio of ion intensities (fold change)[52]. Therefore, unlike PCA, unique metabolite features of lower abundance can also be identified. Using the *p*-value and fold change cut-offs one can rapidly determine which metabolite features pertain to a particular group of replicate samples (**Figure 1.6.1D**).

The improved predictive power of modern bioinformatic software allows for accurate prediction of building blocks used in biosynthesis of NRP and PK products (**Figure 1.6.2A**). In the case of NRP product, knowing the building blocks allows for feeding isotopically labeled amino acids, incorporation of which can be traced using a mass spectrometer. More specifically, both ^{13}C (labeled) and ^{12}C (unlabeled) amino acids can be used directly for supplementing bacterial media (**Figure 1.6.2B**). The extracts of the corresponding cultures can then be analyzed using LC-MS and chromatographic

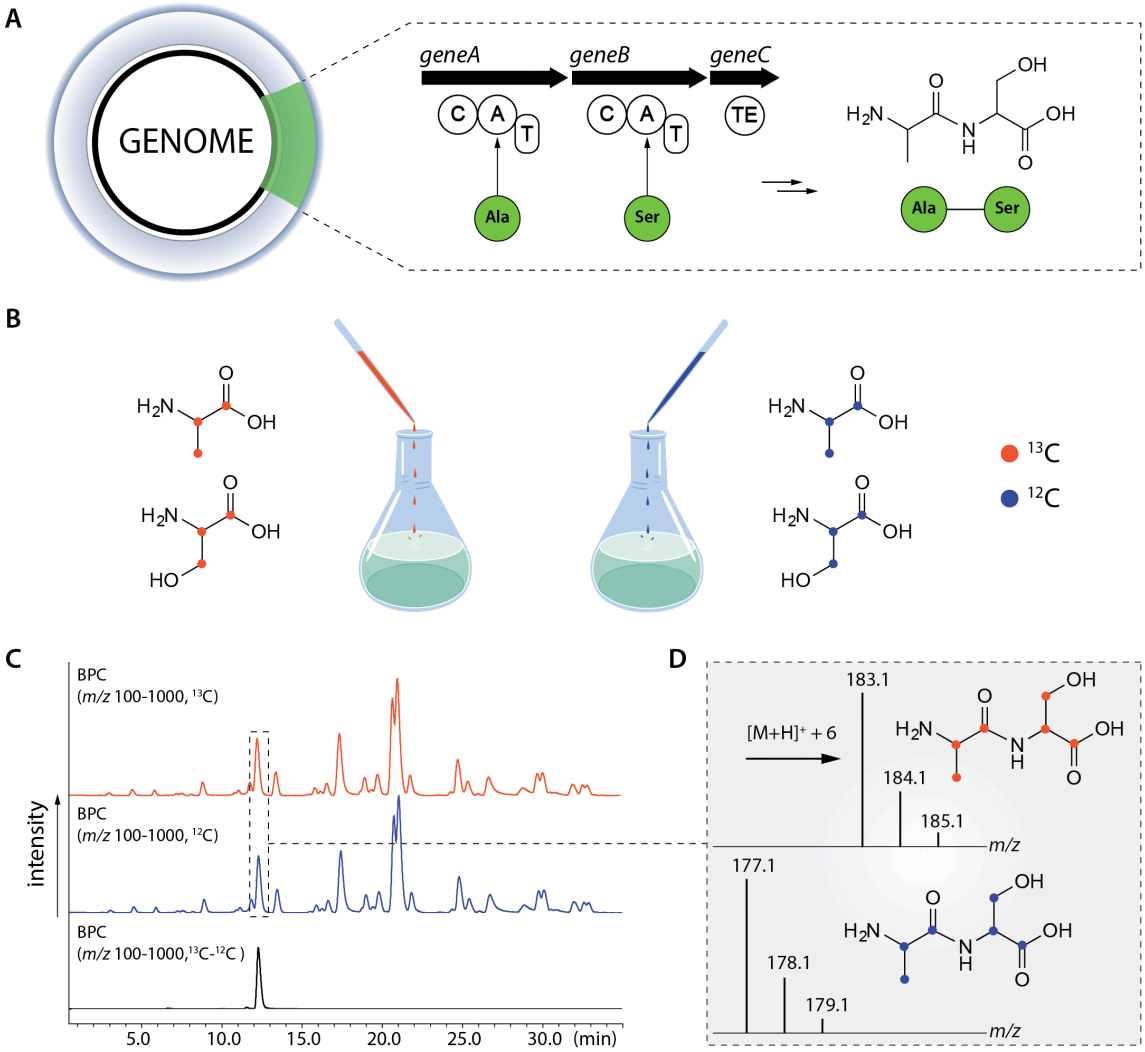


Figure 1.6.2. Chromatographic subtraction coupled to isotopic labeling experiments. **A** – monomer units can be predicted from the genome due to known affinity of the loading enzymes. **B** – predicted labeled (^{13}C) and unlabeled (^{12}C) building blocks are supplemented into the bacterial growth medium. **C** – chromatographic subtraction of the LC-MS data using respective (i.e. labeled and unlabeled) bacterial extracts readily identifies biosynthetic products that incorporate isotopically labeled monomer units.

subtraction processing can identify the biosynthetic products that incorporated the labeled amino acids (**Figure 1.6.2C**). Similar analysis can be performed using polyketide building blocks for the identification of polyketide biosynthetic products. However, the success of such experiments is highly dependent on a number of factors, such as the composition of unlabeled growth medium, nutrient requirements of selected organisms, growth conditions and conditions for the expression of particular polyketides or nonribosomal peptides. Oftentimes, the labeled building blocks get incorporated into a multitude of other biosynthetic products, ultimately creating a number of false positives that complicate the analysis. Utilizing chromatographic subtraction that compares LC-MS data of the wild-type strain and the NRPS mutant extracts minimizes the number of possible undesirable differences. However, the feeding experiments may serve the purpose of validating the identified hits.

1.7 Bioinformatic mining

With an increasing predictive power of modern bioinformatic software that greatly assists the identification of new metabolites, a need for a bioinformatic platform capable of identification of the actual metabolites became apparent. Realizing the potential of moving from genomes directly to molecules and to address the void of lacking software, an informatic search algorithm for natural products (iSNAP) was developed[54]. Briefly, the algorithm is designed to identify related compounds based on their fragmentation profiles observed in the mass spectrometer under collision-induced dissociation (CID), which for nonribosomal peptides arise from the cleavage of the amide bonds. To identify related compounds, the seed structure can be supplied using either the amino acid sequence

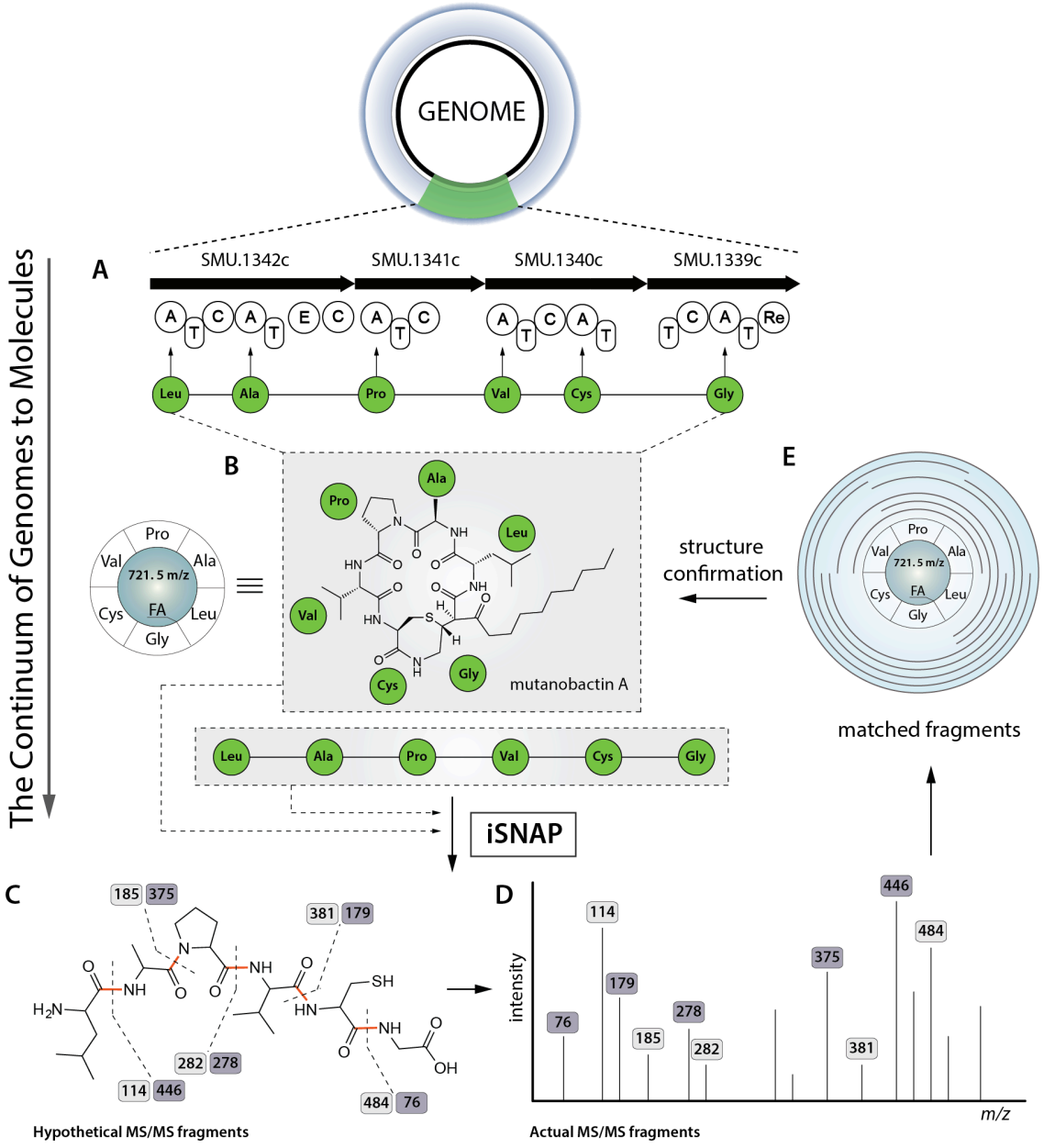


Figure 1.7.1. Connecting genes to molecules with iSNAP. **A** – monomer units can be predicted directly from the genome and used as seeds; **B** – known metabolites can also be used as seed structures; **C** – creation of *in silico* fragments by iSNAP; **D** – comparison to the actual MS/MS spectrum; **E** – report of matched fragments and their statistical scores.

predicted directly from the genome (**Figure 1.7.1A**) or the known compounds, like mutanobactin A, as an example (**Figure 1.7.1B**). The supplied seed structures then undergo a hypothetical *in silico* fragmentation within iSNAP, generating all possible fragments (**Figure 1.7.1C**). The iSNAP algorithm then compares these hypothetical fragments to the actual MS/MS spectrum that has also been supplied (**Figure 1.7.1D**). The related compounds in the mass spectrum of bacterial extract are then identified based on the statistical significance of their matched fragments (**Figure 1.7.1E**).

To explore the structural diversity of hits identified by iSNAP a combination of MS/MS and isotopically labeled precursor-feeding experiments may be carried out. Having the knowledge of building blocks expected in these hits and relying on the above experiments one may infer what biosynthetic changes have occurred in these related compounds. When elucidated using MS/MS and isotopically labeled feeding experiments, the elemental composition of these compounds may be confirmed by high-resolution mass spectrometry, eliminating the need for their isolation.

1.8 *Streptococcus mutans*

Streptococcus mutans is a Gram-positive human pathogen and is commonly recognized as a primary causative agent in dental caries[55]. Recent literature precedent suggests that it might also be involved in immunomodulation through the unique cell wall molecules that act as adhesins and modulins (e.g. rhamnose-glucose polymers and lipoteichoic acids)[56,57]. Although these surface agents are well-known, other mediating forms of communication between *S. mutans* and the host cells may exist and could be carried out through the means of non-acid secreted agents. Interestingly, *S. mutans* is a

rich source of non-bacteriocin bioactive secondary metabolites, whose role has not been interrogated with respect to immunomodulation. The genome of *S. mutans* contains a hybrid biosynthetic machinery that consists of both NRPS and PKS modules. It is not only their presence, but also the unique arrangement and encoding properties that make these gene sets of a particular interest. Recently, Wang et al. identified a small secreted lipopeptide, mutanobactin A and three of its variants, to arise from this cluster[58]. Their preliminary bioactivity assessment demonstrated that the mutanobactins lack appreciable antibacterial activity and instead act to alter the morphological plasticity of *Candida albicans*. More specifically, these compounds were shown to blunt the hyphal formation in *C. albicans*, a step critical to its infectivity[59]. In this way, these molecules have an activity that is intriguing with respect to their possible therapeutic utility in blocking virulent stage transitioning. Although the action of these complex lipopeptides is noteworthy, a number of simple molecules such as fatty acids have been shown to demonstrate the same activity[60]. This suggests that perhaps this activity is distant from their possible naturally evolved functions. Given the biological environment of *S. mutans*, the biosynthetic products that it creates may be involved in signalling that extends into the eukaryotic niche.

1.9 *Lactobacillus plantarum*

Lactobacillus plantarum is a bacterium well-known to many for its probiotic properties and potential therapeutic implications[9,61]. When studying its probiotic affects, the majority of the focus is often spent on the cellular level, overlooking a rich pool of secondary metabolites and small molecules produced by this bacterium. As a

consequence, only few bioactive agents have been identified. Examples of these include lipoteichoic acid – bioactive effector molecule with immunomodulatory properties[62], plantaracin – a peptide with antimicrobial properties[63] and capsular polysaccharides – molecules capable of suppressing pro-inflammatory responses[64]. Nonetheless, the complete picture of the probiotic agents has not been mapped out completely and since it is very likely that the overall beneficial activity of *L. plantarum* and other beneficial bacteria is a combination of all agents, there is a discovery gap to be filled. Interestingly, there is some concrete evidence that the probiotic activity of *L. plantarum* strongly correlates with the stage of its growth. Relatively recent clinical study was investigating the affects of *L. plantarum* cells when fed to patients. Remarkably, the activity was only detected from the cells that were in their stationary phase of growth and not in the mid-log phase[65]. Research efforts from the Magarvey lab directed to seeking answers on the metabolomic level, employed a principal component analysis to explore the metabolomic changes of *L. plantarum* as it transitions from the mid-log to the stationary phase of growth. The most significant outliers identified from this analysis were the three dipeptides comprised of a pyroglutamic acid unit and a corresponding amino acid – phenylalanine, named, leucine and isoleucine. The most abundant phenylalanine variant of these compounds was isolated and tested for its immunomodulatory activity *in vivo*, on mice stimulated with lipopolysaccharide. Analysis of the dendritic cells isolated from the mice spleens revealed that the production of interferon gamma, cytokine known to be associated with a number of auto-inflammatory and autoimmune diseases[66], was significantly suppressed when compared to the control. It is likely that more analogs of

pyroglutamic acid dipeptides may be produced by *L. plantarum* that also possess immunomodulatory properties, and that their activity is highly dependent on the amino acid unit attached.

1.10 *Pseudomonas aeruginosa*

A Gram-negative bacterium *P. aeruginosa* is not only an ideal source of diverse chemistry, but is also associated with increased morbidity and mortality in CF patients[67]. Interestingly, the entirety of molecules and mechanisms associated with *P. aeruginosa* overcoming human immunity to inflict damage has not been revealed. However, some early indicators are known, and particularly those for the pleiotropic small molecules such as pyocyanin, that disrupts the NADH oxidase that creates reactive oxygen species[68]. Nonetheless, *P. aeruginosa* is a highly metabolically diverse microorganism and in order to more effectively address the role of small molecules in immunomodulation a full survey of its secreted metabolome needs to be interrogated for unique chemistry and assayed to accurately probe their functions. The predominant classes of regulatory compounds produced by *P. aeruginosa* are the quinolones and quinolone-like compounds. The roles of these compounds in cell-to-cell signalling, commonly referred to as quorum sensing, are well-understood and have been associated with a variety of multicellular behaviours such as biofilm formation, virulence and pathogenicity[69]. Apart from being involved in the above processes, molecules such as Pseudomonas quinolone signal (PQS) and 4-hydroxy-2-heptylquinoline (HHQ) were shown to down-regulate the innate immune responses in mice[70]. Interestingly, however, *P. aeruginosa* produces a variety of other PQS-like compounds[71], and their impacts on the human immune system

(i.e. macrophages, neutrophils) have not yet been studied. It is likely that these molecules may also have immunomodulatory activities and other PQS variants may exist that have not yet been identified, but are produced by *P. aeruginosa*.

Another class of bioactive compounds produced by *P. aeruginosa* are the nonribosomal peptides. Examples of those are pyochelin and pyoverdine – also key players in quorum sensing and virulence[72,73]. Similarly to other nonribosomal peptides, the biosynthesis of these two is encoded in the genome of *P. aeruginosa* and can be predicted using the logic of the assembly lines for this class of compounds[74,75]. Interestingly, within the genome of *P. aeruginosa*, there is a putative nonribosomal peptide synthase (NRPS) that can be readily identified by *in silico* genome analysis and for which the biosynthetic products have not yet been identified. Interestingly, this bimodular system with predicted affinities for serine and proline is strongly reminiscent of the one found in the NRPS cluster responsible for the production of serine-proline-containing lipocyclocarbamate SB-253514, produced by *P. fluorescens*. This molecule is a selective inhibitor of lipoprotein-associated phospholipase A2 (Lp-PLA2), and was an inspiration for synthesis of the first-in-class phospholipase inhibitor darapladib, currently in phase III clinical trials for treatment of atherosclerosis[76]. Efforts have been made to identify and confirm the biosynthetic origin of these potent immunomodulins[77]. Based on the strong resemblance, it is likely that the product of this putative NRPS may also possess similar or related immunomodulatory activities and to test this its identification and characterization needs to be first pursued.

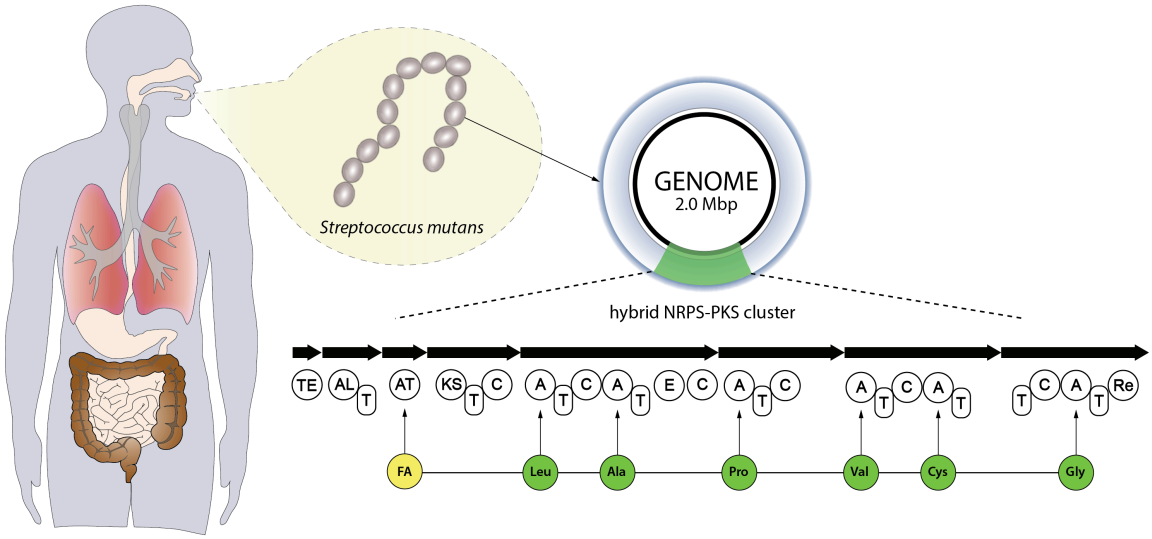
1.11 Thesis Overview

Human microbiome is a complex network of diverse variety of microorganisms that in various ways can interact with the human cells. One of the ways, an interaction through the means of small molecules is the pillar of this thesis, focusing in great detail on the immunomodulatory activities of these agents. Our currently limited understanding of their actions is a great framework for new and exciting discoveries, setting the stage for exploration of the means microbes may use to modulate our immune system.

Four related research projects, consisting of one published scientific article and three manuscripts prepared for submission for publication, expand metabolomic and biochemoinformatic approaches that may be used for not only identification and isolation of immunomodulatory agents, but also other bioactive small molecules. The interplay between microbes from the human microbiome and our cells is multilayered and highly interconnected, and require a concrete understanding of what exactly microorganisms produce. The ultimate goal of mining selected bacteria from the human microbiome is to find the small bioactive molecules and characterize their actions. This clinically relevant information could help explain some of the symptoms associated with the fluctuations in the populations of their respective producers. Additionally, these findings will shine the light on the principles and mechanisms utilized by these microorganisms as means to communicate with the human cells.

Chapter 2. Systems Biosynthesis of Secondary Metabolic Pathways Within the Oral Human Microbiome Member *Streptococcus mutans*

2.1 Chapter Preface



Streptococcus mutans is a member of the human oral microbiome that is noted for its interaction with the human host, and specifically for its contribution to dental caries. The production of acids is a property shared by many fermentative non-cariogenic bacilli and is a significant factor in the degradation of enamel and dentine. Exclusive to *S. mutans* are the other salient cariogenic features such as unique cell wall molecules that act as adhesins and modulins. Although these surface agents are well-known, other mediating forms of communication between *S. mutans* and host cells may exist and could be carried out through the means of non-acid secreted agents. Despite the potential therapeutic implications that could arise from such studies, research focused on specialized branches of *S. mutans* metabolism has not been under extensive investigation. Apart from adhesins and modulins, another differentiating factor for *S. mutans* is the unique hybrid

nonribosomal peptide synthetase-polyketide synthase (NRPS-PKS) biosynthetic machinery that can be identified within its genome. Preliminary studies revealed that this assembly line gives rise to four complex lipopeptides, mutanobactin A and three of its analogs. Although their bioactivity assessment revealed notable involvement in cross-kingdom signaling with another human microbiome member, *Candida albicans*, the implications of these molecules in modulating eukaryotic biology have not been investigated. Therefore, to explore the complete metabolomic output of this unique assembly line and establish the effects of select agents on eukaryotic biology, I carried out system-wide in-depth analysis by utilizing a combination of metabolomic and chemoinformatic platforms.

The following chapter is formatted as a manuscript that is prepared for submission for publication, in which I am the lead author. I conducted the experimental design, performed all experiments (except for the ones stated below), interpreted results and wrote the manuscript. Nikola Lukenda pioneered some of this work, including mutanamide identification and the initial fatty acid feeding experiments that validated its biosynthetic origin. Dr. Xiang Li elucidated the structure of mutanamide, including the determination of its absolute stereochemistry. Janice J. Kim assisted in cytokine analysis and Satheeisha Tharmarajah assisted in fatty acid feeding experiments. Dr. Nathan Magarvey provided ideas and guidance for the progression of this research.

2.2 Abstract

Streptococcus mutans, a Gram-positive human commensal and pathogen, is commonly recognized as a primary causative agent in dental caries. Metabolic activity of this strain results in the creation of acids and secreted products are recognized as pathogenic factors and agents that promote immunomodulation by stimulating the release of pro-inflammatory cytokines. Products of secondary metabolic pathways of microorganisms from the human microbiome are increasingly investigated for their immunomodulatory functions. In this study, we sought to explore the metabolomic output of nonribosomal peptide pathways within the model *S. mutans* strain, *S. mutans* UA159, using a systems metabolomic approach to gain in-depth analysis on products created by this organism and probe these molecules for their immunomodulatory function. Comparative metabolomics and biosynthetic studies using wild-type and nonribosomal peptide deletion strains (within the mutanobactin biosynthetic locus), precursor feedings (fatty acid derivatives) led to the identification of 58 metabolites, 13 of which were structurally elucidated. In addition to these, an assembly line derailment product, mutanamide, was also identified and used to assess immunomodulatory properties of mutanobactins and actions relating to their previously reported functions describing hyphal inhibitory profiles in *Candida albicans*. The results of this study demonstrate both the complexity and the divergent roles of products stemming from this unique biosynthetic assembly line.

2.3 Introduction

Bacteria often create bioactive products of intriguing chemical complexity and activity from biosynthetic assembly lines. Small molecule products from these often multimodular assembly lines are frequently studied for their impacts on the producers and other microbes, as well as for their direct impacts on eukaryotic biology. Some examples of microbial polyketide (PK) and nonribosomal peptide (NRP) products from environmental bacteria with notable clinically relevant implications in eukaryotic biology are nystatin (PK), epothilone B (PK-NRP), cyclosporine (NRP) and rapamycin (PK-NRP)[1-4]. The latter two are noted immunomodulatory agents used to chemically control human immune pathways via altering cell communication and ultimately functions such as cytokine release (cyclosporine – IL-2, rapamycin – IL-2, IL-10)[5-7]. Interrogation of these bacterial assembly lines and the PK and NRP products created from them is becoming increasingly more possible through metabolomic studies and facile biological testing methods such as multiplexed cytokine release assays to measure their cellular effects on eukaryotes.

The human microbiome is a fascinating place to explore new biosynthetic products because of their possible evolved capacities towards human cellular targets and implications towards human disease. *Streptococcus mutans* UA159 bacterium is an interesting case, as it has a clear ability to function as a commensal, but also causes disease such as dental caries[8]. Similarly to other Gram-positive bacteria, *S. mutans* is known to create complex peptidic products of nonribosomal origin, mutanobactin A and three of its analogs[9,10]. Initial bioactivity assessment of mutanobactins demonstrated that

they lack appreciable antibacterial activity, but do alter the physiology of another oral microbiome member, *Candida albicans*. More specifically, these lipopeptides were shown to blunt the hyphal formation in *C. albicans*, a critical step to its infectivity[11]. The implications of these molecules in affecting eukaryotic cells, such as macrophages that are involved in the interaction between *S. mutans* and its host, have not been investigated, however. The biosynthetic machinery of mutanobactins features a hybrid nonribosomal peptide synthetase-polyketide synthase (NRPS-PKS) assembly line that is one of the few assembly lines that lead to the creation of specialized thioether ring architecture. Moreover, cross-comparative analysis indicates that these NRPS and PKS gene sets readily distinguish *S. mutans* from the other oral microbes[12]. Not only their presence, but also the unique arrangement and encoding properties make these gene sets of a particular interest. In this work we sought to develop a systems approach to biosynthetic assembly lines and used *S. mutans* as a case in point of how to explore biosynthesis in-depth with a platform we previously published called iSNAP[13]. We make additions to this approach and demonstrate that the NRPS-PKS assembly from *S. mutans* is much more complicated by revealing the entirety of metabolites it can create and demonstrating the immunomodulatory activity of mutanobactins.

Systematic evaluation of secondary metabolism on the whole is still in its infancy, as common reductionist approaches are aimed at investigating one or two agents at a time. However, it is widely known that these assembly line systems have an imparted enzymatic promiscuity and naturally create a diverse suite of endproducts which may conceptually have alternate evolved functions for its producing organism[14]. Few have

defined the metabolomic approach to systematically interrogate metabolic outputs of biosynthetic clusters to define variability, infer structural changes and capture small molecule diversity. Thus, we sought to define a systems approach that would capture the evolved flexibility, encompass a metabolomic framework to further query flexibility via precursor directed biosynthetic feedings, map the biosynthetic variation on metabolomic level and facilitate structural assignment of new and minor constituents. Here we present a systematic analysis of biosynthetic assembly line products within the human microbiome organism, *S. mutans*, from a chemical and bioactivity perspectives to shed light on the diversity of products and activities arising from these microbiome chemistries.

2.4 Results and Discussion

To define the chemical diversity arising from the NRPS-PKS assembly line of *S. mutans*, we employed a metabolomic analysis with a functional and a mutated NRPS cluster. A univariate analysis was performed on the LC-MS data acquired for the replicate samples from both sets of strains and plotted as a volcano plot to visualize the unique metabolite features (**Figure 2.4.1**). From the volcano plot, the shear metabolomic complexity associated with the functional biosynthetic machinery illustrates why *S. mutans* is a great example to impart a systems biosynthetic metabolomic approach. Given the peptidic nature of four reported mutanobactins, we sought to make a connection between the metabolite features from the volcano plot and these compounds by analyzing the culture extracts of *S. mutans* using the iSNAP platform. To ensure accurate identification, we first defined that iSNAP was able to accurately identify mutanobactin A

in the culture extracts of *S. mutans* using the mutanobactin A seed structure that was inputted into the iSNAP database. Using a tight precursor *m/z* mass tolerance of ± 1.0 Da, both mutanobactin A and its isomer, mutanobactin C, were identified exclusively with their corresponding retention times at 28.6 and 31.7 min (data not shown). Mutanobactin B was also accurately identified at its corresponding retention time at 35.6 min (data not shown). Due to its low abundance in the culture extract, mutanobactin D was not detected.

To explore the observed diversity of metabolite features observed from the volcano plot, the precursor *m/z* mass tolerance was expanded to ± 400.0 Da. Using the chemical structures of mutanobactin A, B and D as seed structures, a total of 69 hits were identified by iSNAP (**Figure 2.4.2**). More specifically, the iSNAP analysis using mutanobactin A seed structure yielded 42 unique hits, whereas mutanobactin B and D seed structures yielded 17 and 10 unique hits, respectively. Although the total number of identified unique hits using three mutanobactin seed structures was 69, their structural similarity gave rise to some degree of overlap in the identified hits. Accounted for this overlap, the total of 58 unique hits identified by iSNAP were compiled. This provided an interesting exploration point to go forward and further probe the chemical identity of these molecules. To validate these 58 hits, we performed a feeding experiment using stable isotopically labeled amino acids, which were expected to be incorporated in these molecules at the conserved positions found in the mutanobactin A-D[10]. These amino acids were proline, alanine and leucine. In three separate feeding experiments incorporation of isotopically labeled proline, alanine and leucine was observed in 44, 42 and 42 hits,

respectively (**Figure 2.4.2**). This conservation in labeling let us define which hits were real and which ones were plausible false positives. Some of the hits including 12-14, 23-25, 41-43 and 47-48 were indeed false positives, as they lacked incorporation of all three

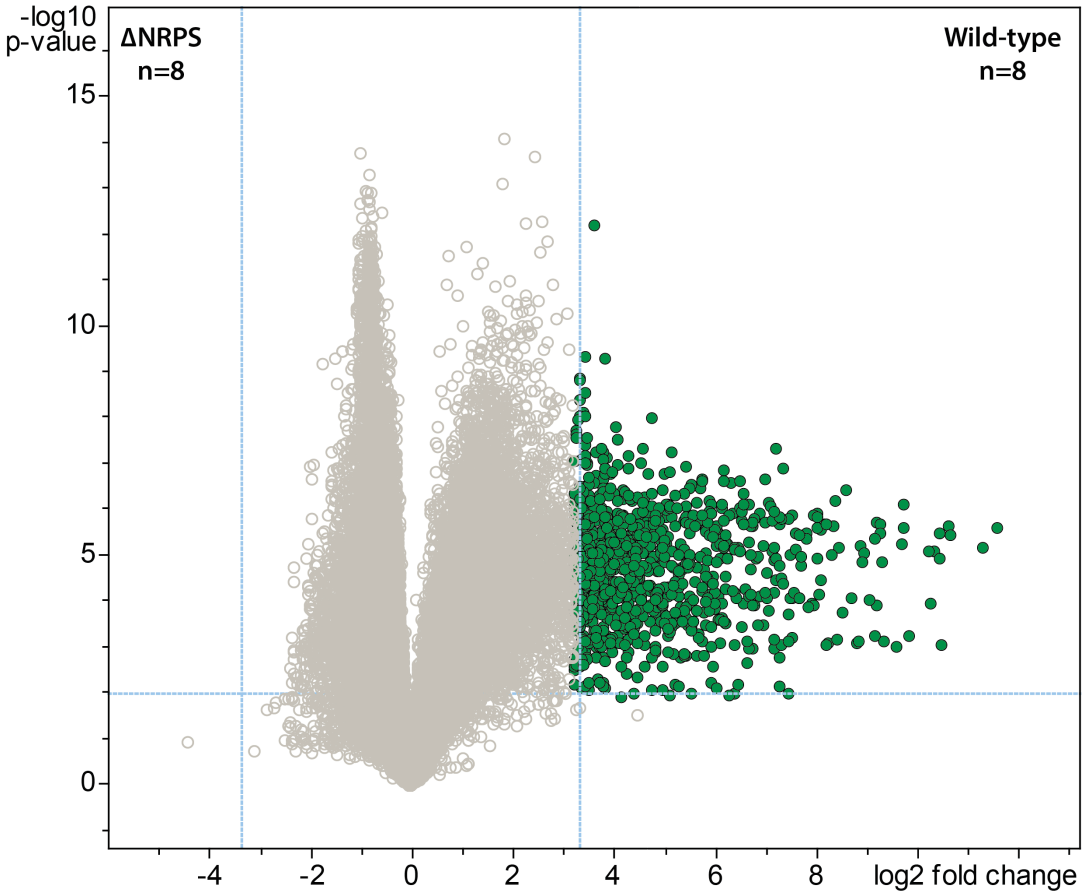


Figure 2.4.1. Volcano plot displaying metabolomic diversity arising from the NRPS-PKS biosynthetic machinery of *Streptococcus mutans*. Vertical and horizontal dashed blue lines indicate the *p*-value ($p = 0.05$) and fold-change (fold change = 5) cut-offs, respectively. Metabolite features (black circles) in the right top sector were considered statistically significant and pertain to wild-type, specifically due to the presence of the functional NRPS-PKS biosynthetic machinery.

labeled amino acids. However, in other instances, some hits were plausibly true, however, incorporation of some labeled amino acids was not definitive due to their intrinsically low abundance. Nonetheless, the ability of iSNAP to identify them as hits illustrated that it was very applicable and the overall fidelity of this platform to define a systems approach was strong.

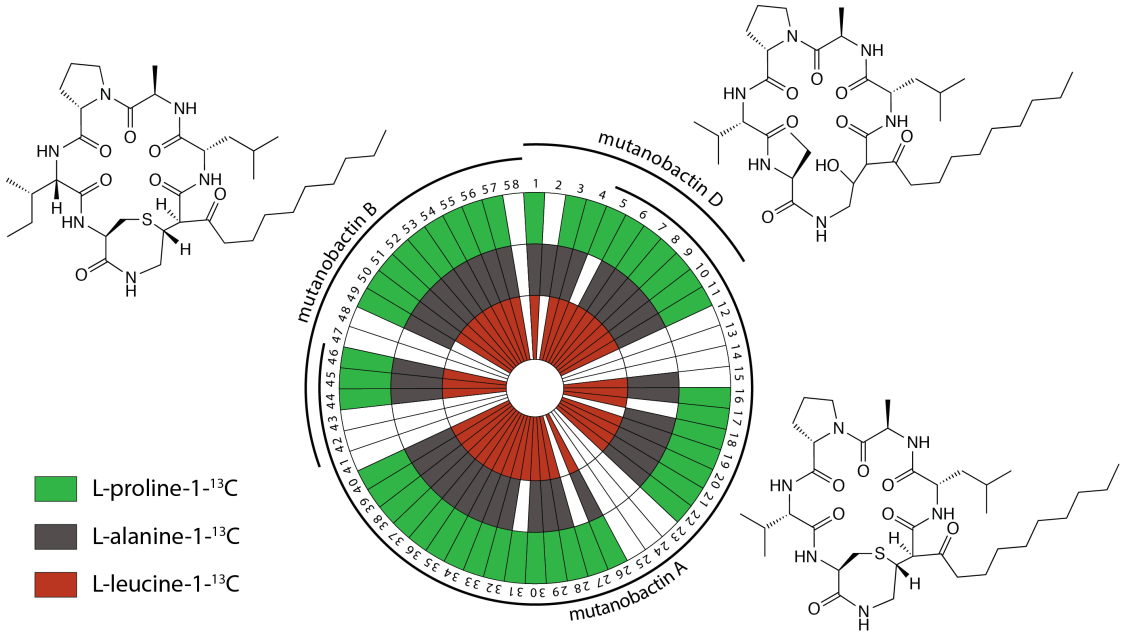


Figure 2.4.2. Targeted identification of biosynthetic products arising from the NRPS-PKS cluster of *Streptococcus mutans* using iSNAP. Metabolite features identified with iSNAP using mutanobactin A, B and D seed structures were validated with ^{13}C-labeled amino acids. The white sectors indicate no observable incorporation.

To explore the structural diversity of some of the hits that were identified by iSNAP a combination of detailed MS/MS and isotopically labeled precursor feeding experiments was carried out. To better investigate the fragmentation pathway that

mutanobactin A undergoes under collision-induced dissociation (CID) in positive ionization mode and obtain structural information of the fragments, seven separate feeding experiments were carried out using isotopically labeled leucine, alanine, proline, valine, cysteine and glycine and acetate. After assigning the dominant fragments of mutanobactin A (**Supplementary Figure S2.6.1**), structures of selected 13 most abundant compounds from the list of iSNAP hits were elucidated using further MS/MS and isotopically labeled precursor feeding experiments, and confirmed by high-resolution mass spectrometry (HRMS, **Supplementary Figure S2.6.2-14**). The existence of mutanobactin B2, an intramolecular cyclization isomer of mutanobactin B was expected (**Supplementary Figure S2.6.2**), as such isomer of mutanobactin A, mutanobactin C, has been reported[10]. The biosynthetic patterns for known lipid-baring molecules suggest slight enzymatic promiscuity, reflected in the ability to incorporate acyl tails of varying length[15]. The variants of mutanobactin A and B, mutanobactin E and F, are also produced and contain twelve-carbon long acyl tail (**Supplementary Figure S2.6.3-4**). Interestingly, variants of mutanobactin A and B, mutanobactin G and H, have also been identified and contain alanine for serine substitution (**Supplementary Figure S2.6.5-6**). In addition to HRMS experiments, the identities of mutanobactin G and H were validated using isotopically labeled serine. In both instances, only serine incorporation was observed and no alanine (data not shown). Semi-cyclized analogs of mutanobactin A and B also exist and were identified as mutanobactin I and J, respectively (**Supplementary Figure S2.6.7-8**). Their existence was anticipated, given the stereochemical variability of cyclized biosynthetic products. In addition to the identified mutanobactin analogs, larger

biosynthetic products were identified by iSNAP as mutanobactin-like compounds, named mutanolins. Further analysis revealed that mutanolins were indeed of mutanobactin ilk, and their increased molecular weight was due to the incorporation of additional amino acid monomers. For example, mutanolin A and B were found to have almost identical fragmentation profiles as mutanobactin A and B, and both incorporated an additional cysteine unit, which was also in agreement with the observable cysteine loss in their corresponding MS/MS spectra (**Supplementary Figure S2.6.9-10**). This biosynthetic curiosity was unexpected, because there is no additional cysteine adenylation domain in the NRPS portion of this cluster. Perhaps, this is also related to malfunctioning of enzymes responsible for the intramolecular cyclization, which exposes the unfinished products to the reactive neighboring cysteine, which then nucleophilically attacks the mutanobactin intermediate prior liberation from the assembly line. Interestingly, the existence of even larger mutanolins was observed. The complied analysis of their structures suggested that mutanolin C and D contain an additional glutamic acid that is seemingly attached via a peptide bond to the additional cysteine of mutanolin A and B, respectively (**Supplementary Figure S2.6.11-12**). Together with cysteine, glutamic acid is the second most essential amino acid required for growth of *S. mutans*[16]. Perhaps mutanolins act as a temporary storage of essential amino acids, as their relative abundance seems to decrease over time (data not shown). Lastly, the largest mutanolins, mutanolin E and F, were elucidated to contain a glycine monomer between cysteine and glutamic acid or mutanolin C and D, respectively (**Supplementary Figure S2.6.13-14**).

Although again unexpected, the sequence of the connected monomers is in keeping with the original sequence observed in mutanobactins, where cysteine follows glycine.

Having elucidated a number of mutanobactin analogs, it became evident which sites were prone to substitution. In addition, the identification of mutanolins was even more surprising and suggested that this hybrid NRPS-PKS assembly line does not only display wide substrate flexibility, but also has seemingly low fidelity, generating these two classes of biosynthetic products (**Supplementary Figure S2.6.15**). Despite such vast and unexpected variation, guided by the iSNAP analysis these products were nonetheless identified, demystifying some of the vast diversity observed in the volcano plot. It is noteworthy that some molecular features displayed in the volcano plot were much smaller than mutanobactins or mutanolins, which explained the likely reason why iSNAP was not able to detect them (**Supplementary Figure S2.6.16**). Since these molecular features were exclusive to the wild-type strain, we reasoned that perhaps they were the spuriously generated derailment products from the biosynthetic assembly line. Detailed analysis revealed that the clustering observed in the 3D volcano plot around *m/z* 400.0 was due to one predominant ion – *m/z* 399.3. Driven by curiosity we set out to define this possibly truncated mutanobactin cluster-derived product. In doing so, we isolated mutanamide (**Supplementary Figure S2.6.17A**). The chemical structure of mutanamide is interesting, as similarly to mutanobactins it contains the decanoic acid moiety attached to the first few amino acids – leucine and alanine (**Supplementary Figure S2.6.21-27**). To validate its biosynthetic origin, feeding experiments were performed using decanoic acid analogs to test for their incorporation in both mutanamide and mutanobactin A (**Supplementary**

Figure S2.6.18). As expected, both compounds incorporated 8-nonenoic and 7-octenoic fatty acids and their identity was confirmed with MS/MS analysis. It was curious what caused the actual derailment of mutanamide, so we decided to determine the absolute stereochemistry of its two amino acids (**Supplementary Figure S2.6.17B, S2.6.19**). Based on the obtained results, the derailment seemingly arisen from the incorporation of L-alanine and failure to epimerize to D-alanine, which is observed in all four previously reported mutanobactins[10]. The opposite stereochemistry likely precluded the assembly line form going downstream. Therefore, the systems approach utilized to unravel the biosynthetic potential of NRPS-PKS cluster in *S. mutans* was not only successful at defining the sites of variable amino acid incorporation, but also led to the identification of

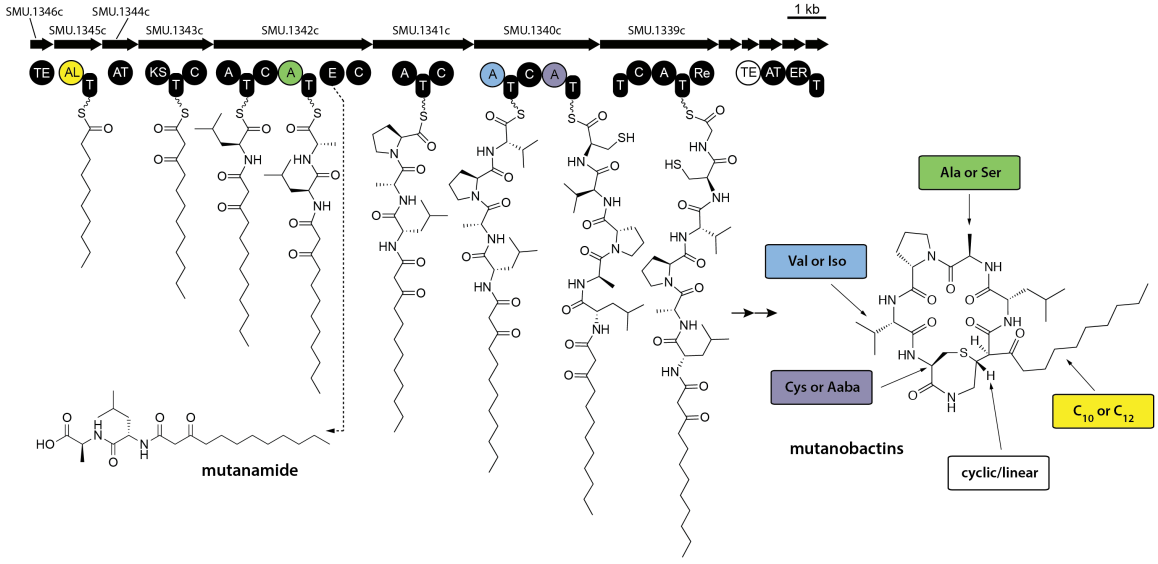


Figure 2.4.3. Biosynthetic assembly line that gives rise to mutanobactins. Mutanamide is a derailment product liberated from the assembly line due to the likely malfunctioning of epimerase. The diversity of mutanobactins is granted by the enzymatic flexibility in the colour-coded adenylation and thioesterase domains.

the truncated intermediate with opposing stereochemistry (**Figure 2.4.3**). The latter revealed the mechanism that must be in place in the assembly line to block the progression of the stereo-altered products, which are perhaps inactive with regards to their natural biological function.

The desired stereochemistry is critical for natural products, especially in the complex structures like mutanobactins, and it is not surprising that the assembly lines would have some way to divert these unintended compounds so that they are not processed any further. However, this does not mean that these products are entirely biologically inactive. Due to the predominant acyl tail moiety in mutanamide, we noted

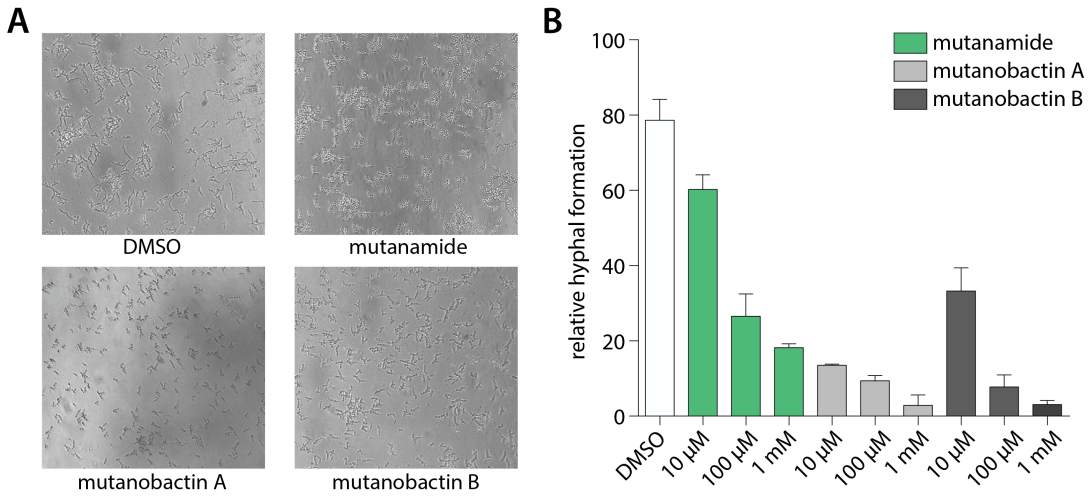


Figure 2.4.4. Bioactivity of mutanamide and mutanobactin A and B. **A** – inhibition of hyphal formation in *C. albicans* when treated with mutanamide and mutanobactins A and B (DMSO, 100 µM mutanamide, 100 µM mutanobactin A and 100 µM mutanobactin B). **B** – relative hyphal formation when treated with mutanamide, mutanobactins A and B.

that like the other fatty acids that have been previously tested to define activity against yeast signaling, mutanamide may also possess such activity[17]. In addition, since this was the original interesting activity assigned to the mutanobactins, we sought that mutanamide would be the ultimate product to test whether the true function of mutanobactins is actually the reported one. Thus, we tested mutanamide and mutanobactin A and B side by side in a yeast hyphal formation assay (**Figure 2.4.4**). The ability of mutanamide to blunt hyphal formation was much the same as that of mutanobactin A and B, indicating that the functional moiety of these molecules is indeed the acyl tail. Despite their lipopeptide nature, both mutanamide and mutanobactin A had no affect on the proliferation of HT-29

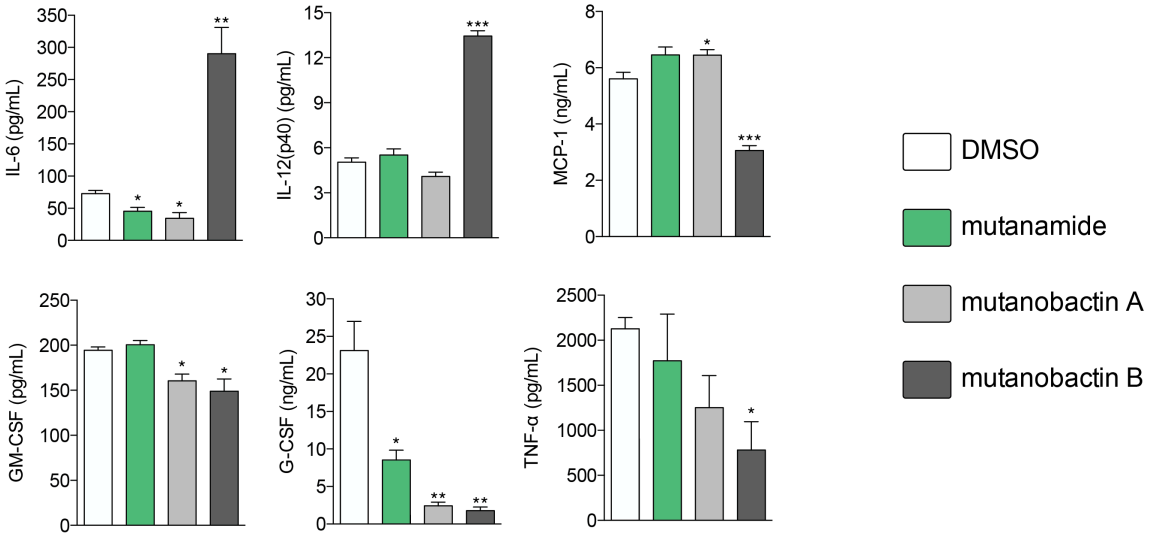


Figure 2.4.5. Distinct immunomodulatory activity of mutanobactins. The bar graphs represent the levels of secreted cytokines measured in the supernatant of LPS-stimulated RAW264.7 macrophages upon treatment with 100 μM mutanamide, mutanobactin A and B. Data are presented as mean ± SEM, asterisks (*) indicate significant differences between treatments (*$p<0.05$, **$p<0.01$, ***$p<0.001$).

cells (**Supplementary Figure S2.6.20**). Therefore, we reasoned that the peptidic core that is so exclusive to the human microbiome member may have another activity.

Given the biological habitat of *S. mutans* and having identified mutanamide, we sought to test whether mutanobactins display any immunomodulatory activity and to determine whether the peptidic core is divergent from the acyl tail. To assess the immunomodulatory potential of mutanobactin A and B versus mutanamide, we chose RAW264.7 macrophage cell line and the assay that would read out multiple cytokines that would be produced by the immune cells upon activation. Compared to mutanamide, the affects of mutanobactins, especially mutanobactin B, were much more pronounced (**Figure 2.4.5**). Interestingly, in same instances, mutanobactin B caused a significant increase in the pro-inflammatory cytokines, such as IL-6 and IL-12 and in others a decrease, as the case is for the other four measured cytokines. Although more research needs to be conducted to establish the most dominant immunomodulatory effect of mutanobactins, this preliminary data are significant. More specifically, these data are indicative of that perhaps the true or likely role of mutanobactins is a means of manipulating the immune responses of the human cells, rather than having a divergent activity against another organism like *C. albicans*.

From this complete systems approach to studying metabolic pathways it is evident that these types of biosynthetic assembly lines generate a multitude of products and we were able to create a platform to assign what those products are and to define what selective changes have gone on to impart their divergent chemistries. It also allowed for a

systems investigation of the assembly line and the ability to define what combinations and permutations are naturally possible. In addition, the importance of stereochemistry was highlighted by the discovery of stereochemistry-altered derailment product, mutanamide. More importantly, having identified and isolated this truncated product and use it as a comparative control we were able to define the roles of mutanobactins in more depth.

We present this entirety and highlight the importance of integrating both metabolomic and chemoinformatic platforms such as iSNAP to gain a systems view of the products that would arise from the assembly line. We are curious to see what this platform and systems metabolomics biosynthetic analysis can offer for other assembly lines to define the roles of those chemicals and the diversity of structures that occur naturally and define them. In future work we hope that this is a mechanism that can be deployed in the system metabolomics of these other natural biosynthetic clusters and specifically those from the human microbiome that are of strategic importance with regards to defining the roles of their bioactive compounds in human biology.

2.5 Materials and Methods

2.5.1 Bacterial strains and reagents

Streptococcus mutans UA159 strain was used for all experiments in this study and *Candida albicans* (ATCC # 90028) was used for hyphal inhibition assays. Mutant strain of *S. mutans* UA159 with a complete deletion of biosynthetic genes (generated using in-frame allelic exchange, substituting the biosynthetic operon for an erythromycin resistance cassette) was constructed by Martha Cordova and generously provided by Dr. Dennis G. Cvitkovitch. Nonribosomal synthase mutant (ΔNRPS, ΔSmu. 1334-1348) was confirmed using PCR and primers specific to SMU_1340 (forward: AGCCAGCGTCATCTTCTGTCACC, reverse: AGGGAGTGCAGGTTGGCGGA). Unless otherwise stated, all reagents and chemicals were purchased from Sigma Aldrich.

2.5.2 Fermentation, processing and analysis of *S. mutans* extracts

Streptococcus mutans strains were grown in Todd-Hewitt broth supplemented with 3 g/L yeast extract (THYE broth) at 37°C with 5% CO_2 without shaking. For the metabolomic comparison of the wild-type and the ΔNRPS mutant, the overnight cultures were inoculated (1:100) into 50 mL of THYE medium (n = 8) in 250 mL flasks and grown for 12 hours. The supernatants were extracted using liquid-liquid partitioning with ethyl acetate (2:1 v/v, 3x) and the resulting organic phase was concentrated *in vacuo*, resuspended in methanol:water (1:1 v/v) and analyzed directly by Bruker micrOTOF II mass spectrometer (positive ESI, scanning *m/z* 100-1000) with an Agilent 1200 series HPLC with Ascentis Express C18 column (150 mm x 2.1 mm, 2.7 μm, Sigma) and A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) as the mobile

phases at 0.25 mL/min. The solvent gradient was 0-5 min, 5% A; 5-45 min, linear to 100% A; 45-55 min, isocratic at 100% A, 55-60 min, linear to 5% A; 60-65 min, isocratic at 5% A. Cell extracts of *S. mutans* grown in a chemically defined medium (CDM) were used for the studies involving mutanolin and mutanobactin analog structure elucidations using MS/MS and isotopically labeled amino acids. The chemically defined medium was composed of five 0.22 μm filter-sterilized stock solutions that were added aseptically (per 1 L): 100 mL of 10x amino acid solution (per 2 L: L-glutamic acid, 50.8 g; L-cysteine, 4 g; L-leucine, 3 g; L-lysine, 3 g; L-arginine, 4.2 g; L-proline, 0.12 g; L-glycine, 0.08 g; adenine, 0.18 g; solubility was enhanced with HCl, then readjusted to pH = 7), 20 mL of 50x salt1 solution (per 2 L: K_2HPO_4, 207.25 g; KH_2PO_4, 292.6 g; NH_4Cl, 53.6 g; urea, 30.0 g; sodium pyruvate, 60.5 g), 20 mL of 50x salt2 solution (per 2 L: $MgCl_2$, 56.16 g; $MnCl_2$ tetrahydrate, 1.98 g; $FeSO_4$ heptahydrate, 2.78 g; $CaCl_2$, 11.08 g), 10 mL of 100x vitamin solution (per 500 mL: riboflavin, 50 mg; thiamine hydrochloride, 25 mg; nicotinic acid, 50 mg, para-aminobenzoic acid, 5 mg; pyridoxine hydrochloride, 50 mg; calcium pantothenate, 5 mg; folic acid, 5 mg) and 10 mL of 100x glucose solution (per 1L: glucose or dextrose, 500 g).

For the isotopically labeled amino acid experiments, the overnight THYE cultures of wild-type *S. mutans* were inoculated (1:100) into 50 mL of CDM (n = 2), directly supplemented (50 mg/50 mL) with one of glycine-2-^{13}C, glycine-$^{13}C_2$, L-valine-1-^{13}C, L-serine-3-^{13}C, L-proline-1-^{13}C, L-leucine-1-^{13}C, L-cysteine-$^{13}C_3$, ^{15}N, L-alanine-1-^{13}C or sodium acetate-1-^{13}C in 250 mL flasks and grown for 15 hours at 37°C with 5% CO_2 without shaking. In addition to isotopically labeled amino acids, 8-nonenoic and 7-

octenoic fatty acids were also used directly as a separate supplementation (20 µL/100 mL of CDM). The cell pellets were extracted by vortexing for 5 min with 4 mL of solvent mixture containing methanol:ethanol:water (2:1:1 v/v) with 50 mg of 0.5 mm glass beads. Following centrifugation, samples were analyzed for incorporation using LC-MS as described before, but with mass spectrometer scanning *m/z* 100-2000 range.

The MS/MS experiments were carried out using Bruker amaZon X ion trap mass spectrometer (operating in auto-MS(2) positive ESI mode, scanning *m/z* 100-1800) coupled with a Dionex UltiMate 3000 HPLC, with an Ascentis Express C18 column (150 mm × 4.6 mm, 2.7µ, Sigma Aldrich) and A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) as the mobile phases at 1.2 mL/min. The solvent gradient was 0-5 min, 5% A; 5-7 min, linear to 40% A; 7-30 min, linear to 65% A, 30-35 min, linear to 100% A, 35-38 min isocratic at 100% A, 38-40 min, linear to 5% A; 40-45 min, isocratic at 5% A.

2.5.3 Metabolomic and chemoinformatic analyses of *S. mutans* extracts

The volcano plot was generated using the *t*-test model in the Bruker Daltonics ProfileAnalysis 2.1, by making pairwise comparisons between the wild-type and the mutant strains. More specifically, the bucket table for the *t*-test model was computed using rectangular bucketing parameters defined as $\Delta Rt = 0.1$ min, with a kernel size of 0.05 min, and $\Delta m/z = 0.5$ with a kernel size of *m/z* 0.15, and normalized by the sum of bucket values in the analysis. The fold change and *p*-value cut-offs used to construct the volcano plot were 5 and 0.05, respectively.

To construct a 3D volcano plot, the data were imported from Bruker Daltonics ProfileAnalysis 2.1 into JMP 10.0.0 and the 1187 imported metabolite features were plotted using scatter plot 3D function, with nonpar density contour option selected. The data files obtained from the MS/MS experiments using unlabeled cell extract samples were converted to mzXML format and used for iSNAP analysis. For testing detection of mutanobactin A and B, the following parameters were utilized: precursor mass tolerance, ± 1 Da; precursor charge, 1-2; search mode, precise search; P1 and P2 score threshold, 20; database, built-in NRP database and user-defined compounds (SMILES of mutanobactin A or B, accordingly), number of amide cleavages, 1-5; number of cyclized thioether cleavages, 0-1; theoretical fragment *m/z* tolerance, ± 0.25 Da; theoretical fragment charge, 1-2. For the detection of mutanobactin-like compounds, the following parameters were utilized: precursor mass tolerance, ± 400 Da; precursor charge, 1-2; search mode, precise search; P1 and P2 score threshold, 10; database, built-in NRP database and user-defined compounds (one of mutanobactin A, B and D SMILES), number of amide cleavages, 1-5; number of cyclized thioether cleavages, 0-1; theoretical fragment *m/z* tolerance, ± 0.25 Da; theoretical fragment charge, 1-2. The hits obtained using all three seed structures were compiled and checked manually for the incorporation of isotopically labeled amino acids using data files from the feeding experiments. The accurate mass measurements of identified mutanobactin analogs and mutanolins were obtained using Bruker micrOTOF II mass spectrometer with internal calibration within Bruker Compass DataAnalysis 4.1 SR1 software, using mutanobactin A and spiked-in bradykinin and iturin A internal standards.

2.5.4 Isolation of biosynthetic products and absolute stereochemistry analysis of mutanamide

For the isolation of biosynthetic products, including mutanamide and mutanobactin A and B, ten 2.5 L THYE batches inoculated with 50 mL of wild-type *S. mutans* overnight seed culture were grown in 2.8 L Fernbach flasks at 37°C with 5% CO_2 without shaking for 12 hours. The supernatants were extracted using liquid-liquid partitioning with ethyl acetate (2:1 v/v, 3x) and the resulting organic phase concentrated *in vacuo*. The crude extract was resuspended in methanol and fractionated using a Combiflash® Rf reverse phase MPLC system (Teledyne Technologies Inc.) with a 50 g RediSep® Rf Gold HP C18 flash column and 40 mL/min flow rate. The following program of solvent A (acetonitrile) and B (water) was employed: 0-3 min, 5% A; 3-25 min, a linear gradient to 100% A; 26-30 min, isocratic 100% A; 30-35 min, linear gradient to 5% A. Resulting fractions were screened for the presence of compounds of interest by direct infusion using Bruker amaZon X ion trap mass spectrometer operating in positive ESI mode, scanning *m/z* 100-1000. Fractions that contained the compounds of interest were combined and subjected to preparatory HPLC-ESI-MS (Dionex UltiMate 3000 HPLC system with 160-600 nm variable UV detection connected in line with Bruker amaZon X ion trap mass spectrometer operating in positive ESI mode, and Phenomenex Luna 5μ C18(2) (250 mm x 15 mm, 4.6 μm) with 10 mL/min flow rate). Optimal separation was achieved using the following A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) solvent gradient: 0-5 min, 5% A; 5-10 min, linear gradient to 55% A; 10-28 min, linear gradient to 60% A; 28-29 min, linear gradient

to 100% A; 29-35 min, isocratic at 100 % A, 35-39 min, linear to 5% A; 39-44 min, isocratic at 5% A. High-resolution mass spectral data for mutanamide were obtained using a Thermo Scientific LTQ Orbitrap XL. NMR experiments were performed using a Bruker Avance 700 MHz spectrometer with samples dissolved in 100% methanol-d_4.

The absolute stereochemistry of mutanamide was determined using Marfey's reagent for chiral amino acid analysis[18]. Briefly, approximately 0.5 mg of mutanamide were hydrolyzed with 1 mL of 6N HCl in an ACE high-pressure tube at 110°C for 16 h. The hydrolysate was evaporated to dryness and resuspended in 100 µL of water. To each 50 µL of above solution, 10 µL of 1M $NaHCO_3$ and 100 µL of 1% L- or D-FDLA (1-fluoro-2,4-dinitrophenyl-5-leucinamide, solution in acetone) were added, respectively. The reaction vials were placed in a water bath at 40°C for 1 h and then were cooled to room temperature, quenched with 10 µL of 1N HCl, and evaporated to dryness. The residue was dissolved in 200 µL of methanol and analyzed on HPLC-UV system (Waters Alliance e2695 with Waters 2998 PAD set at 340 nm) together with the L-leucine, D-leucine, L-alanine and DL-alanine standards. The separation was achieved using with the following A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) solvent gradient: 0-5 min, 20% A; 5-65 min, linear gradient to 60% A; 60-65 min, linear gradient to 100% A; 65-70 min, isocratic at 100 % A, 70-75 min, linear to 20% A; 75-80 min, isocratic at 20% A.

2.5.5 Hyphal inhibition in *C. albicans* and HT-29 cell proliferation assay

Candida albicans pre-cultures were grown in YNB medium without amino acids and with ammonium sulfate (Bioshop® Canada Inc., YNB406.100), supplemented with

0.1% maltose and 0.2% glucose at 30°C with agitation at 200 rpm for 16 h. To assay hyphal formation, pre-cultures were used to inoculate hyphal promoting medium (1:100 dilution), consisting of YNB with 2.5 mM N-acetylglucosamine in a 96-well plate. Mutanobactin A, B and mutanamide were dissolved in DMSO and added at the noted concentrations. Importantly, no more than 1 µL of DMSO was added to the 100 µL cultures to avoid secondary effects. Cultures were grown under noted *C. albicans* conditions and were examined using light microscopy after 4 hours to assess cell morphology. More specifically, relative hyphal formation was calculated as the number of hyphal cells divided by the total number of counted cells. The examination of *C. albicans* cell morphology was performed using Nikon Eclipse TS100 microscope at a total magnification of 200x and iPhone 4S was used to capture the microscopy images. HT-29 cells (ATCC No. HTB-38TM) were routinely cultured in RPMI 1640 medium, supplemented with 10% (w/v) fetal bovine serum (FBS, PAA laboratories, A15-751), 2 mM L-glutamine (VWR, BDH), at 37°C in a 5% CO_2 incubator, keeping the passage below ten.

To assay antiproliferative affects of mutanamide and mutanobactin A and B, cells were subcultured 1:8 and incubated in 96-well plate (BD Biosciences) with a total volume of 100 µL, treated with 1 µL of either each compound dissolved in DMSO (indicated 10 nM - 100 µM range) and monitored using CellTiter 96® AQueous non-radioactive cell proliferation assay (Promega), following manufacturer's recommendations. DMSO was used as a negative control and the absorbance at 490 nm was measured after 1 h and 15 h of incubation at 37°C in the 5% CO_2 incubator.

2.5.6 Multiplexed cytokine profiling

RAW 264.7 macrophage cell line (ATCC # TIB-71) was cultured in RPMI 1640 medium, supplemented with 10% (w/v) fetal bovine serum (FBS, PAA laboratories, A15-751), 2 mM L-glutamine (VWR, BDH), at 37°C in a 5% CO_2 incubator, keeping the passage below ten. Cells were subcultured 1:8 and incubated in 24-well plates (BD Biosciences) with a total volume of 2 mL. After reaching 50% confluence (~48 hours), cells were stimulated with 100 ng/mL LPS (L3012-5mg, Sigma) and treated with dissolved in DMSO mutanobactin A, B and mutanamide (20 µL, 100 µM final concentration). After 16 hours of incubation, supernatants were collected and stored at -20°C. For analysis of the wide range of cytokines, samples were analyzed directly using the Bio-Plex Luminex system with a Bio-Plex Pro™ mouse cytokine 23-plex assay kit (Bio-Rad, M60-009RDPD). Fluorescence data were acquired and analyzed by Bio-Plex Manager software (version 5.0, Bio-Rad Laboratories), following the manufacturer's recommendations.

2.5.7 Statistical analysis

All data are presented as mean ± SEM. Unpaired *t*-test (for parametric data) was performed using GraphPad Prism version 6.0b for Mac OS X (GraphPad Software, La Jolla, CA). An associated p-value <0.05 was considered statistically significant.

2.6 Supplementary Figures and Legends (Figure S2.6.1-27)

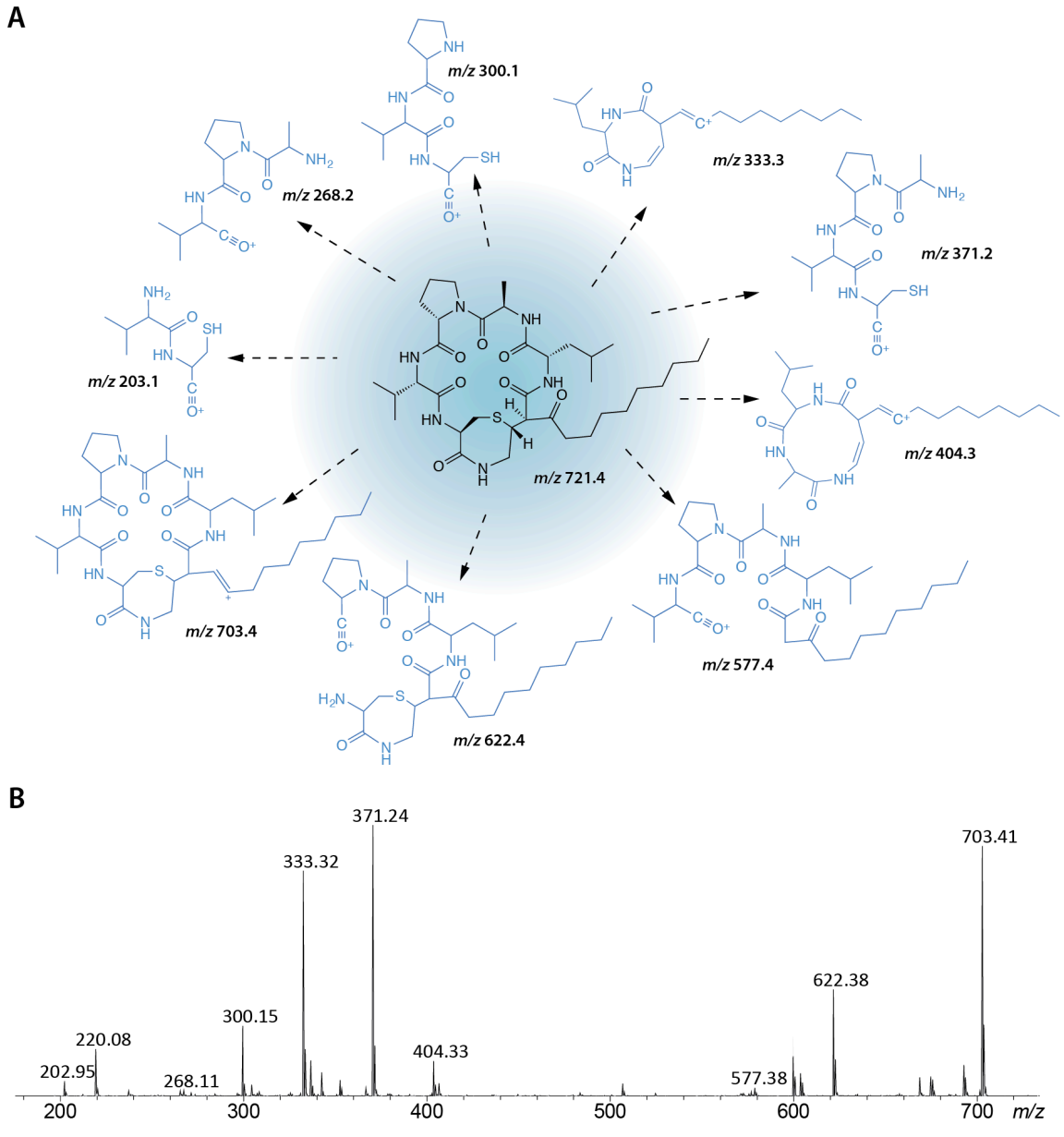


Figure S2.6.1. Fragmentation profile of mutanobactin A. **A** – fragments generated by CID have been identified and validated by incorporation of isotopically labeled amino acids and acetate. **B** – the actual MS/MS spectrum of mutanobactin A.

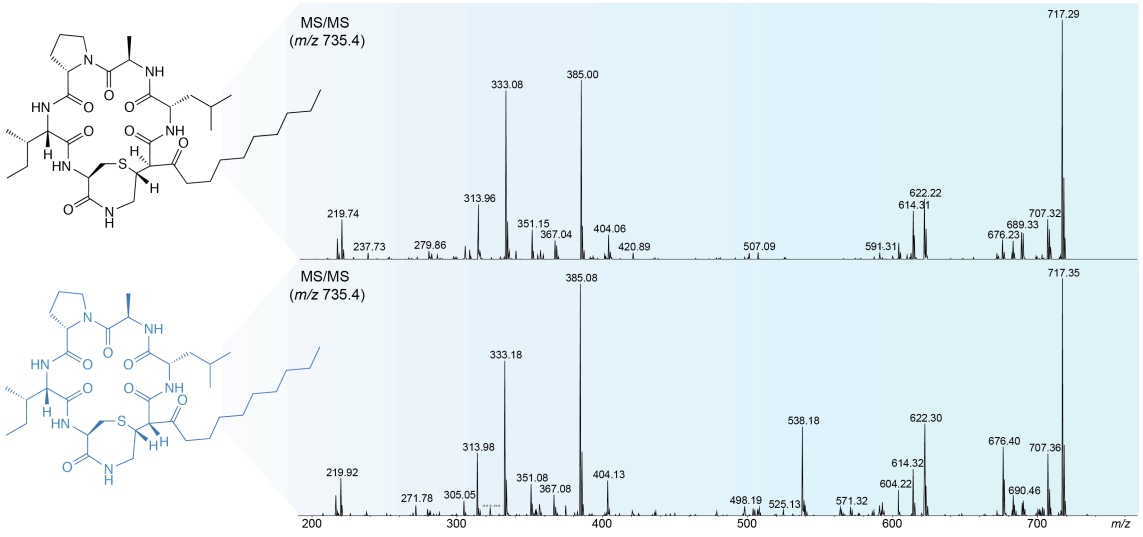


Figure S2.6.2. The fragmentation profiles of mutanobactin B and its proposed isomer. The two compounds have distinct retention times (25.7 and 30.5 min), but identical MS/MS profiles. The structure indicated in blue is the structure of the proposed isomer of mutanobactin B (mutanobactin B2). The elemental composition of mutanobactin B2 was validated using HRMS (*m/z* 735.44735 calculated, *m/z* 735.44929 observed, 2.6 ppm).

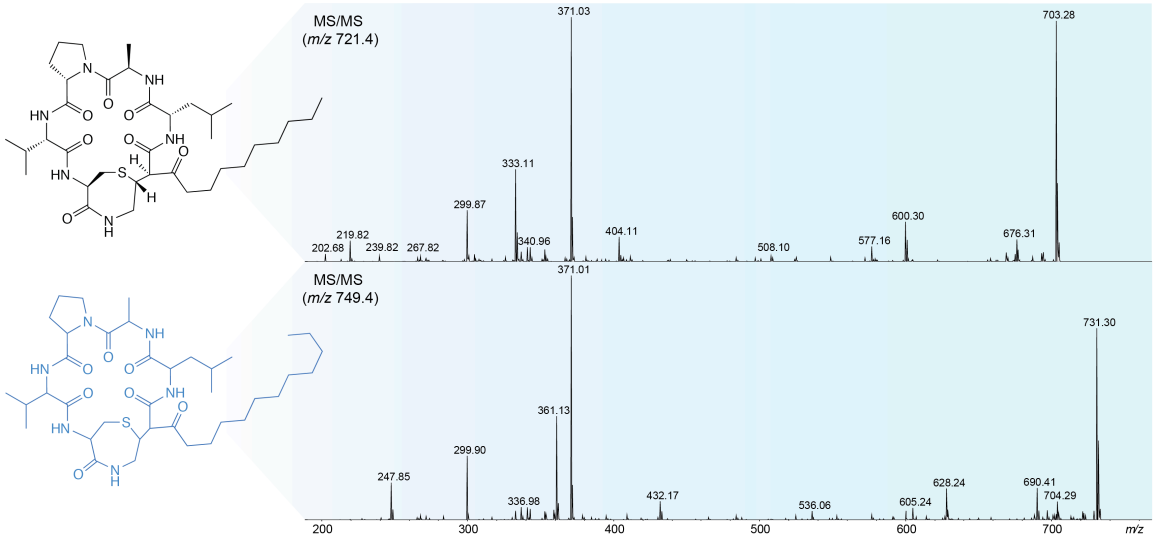


Figure S2.6.3. The fragmentation profiles of mutanobactin A and mutanobactin E. The structure of mutanobactin E (in blue) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 749.46300 calculated, *m/z* 749.46314 observed, 0.2 ppm).

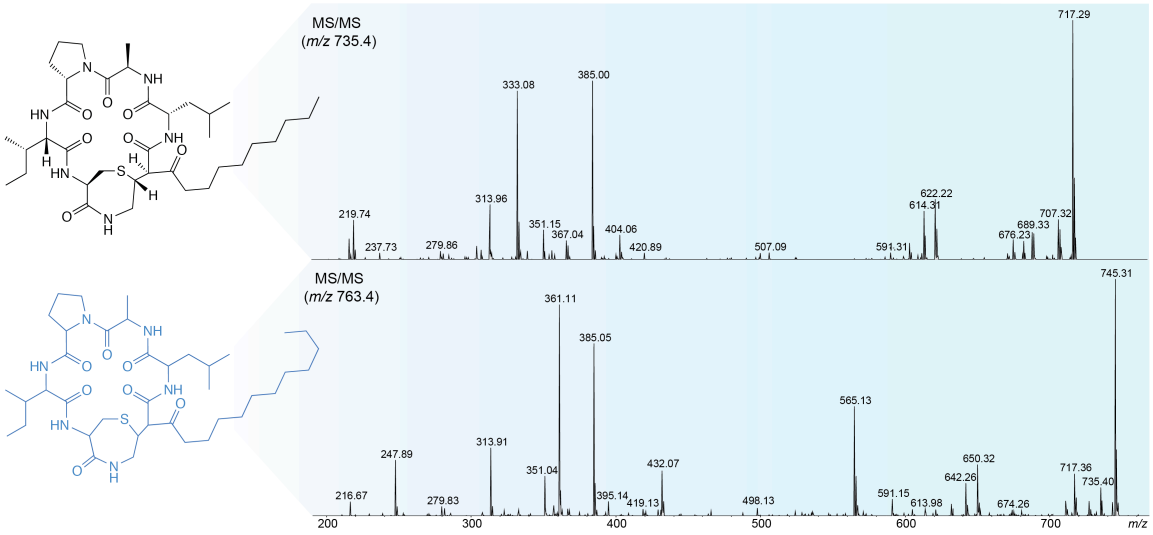


Figure S2.6.4. The fragmentation profiles of mutanobactin A and mutanobactin F. The structure of mutanobactin F (in blue) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 763.47865 calculated, *m/z* 763.47872 observed, 0.1 ppm).

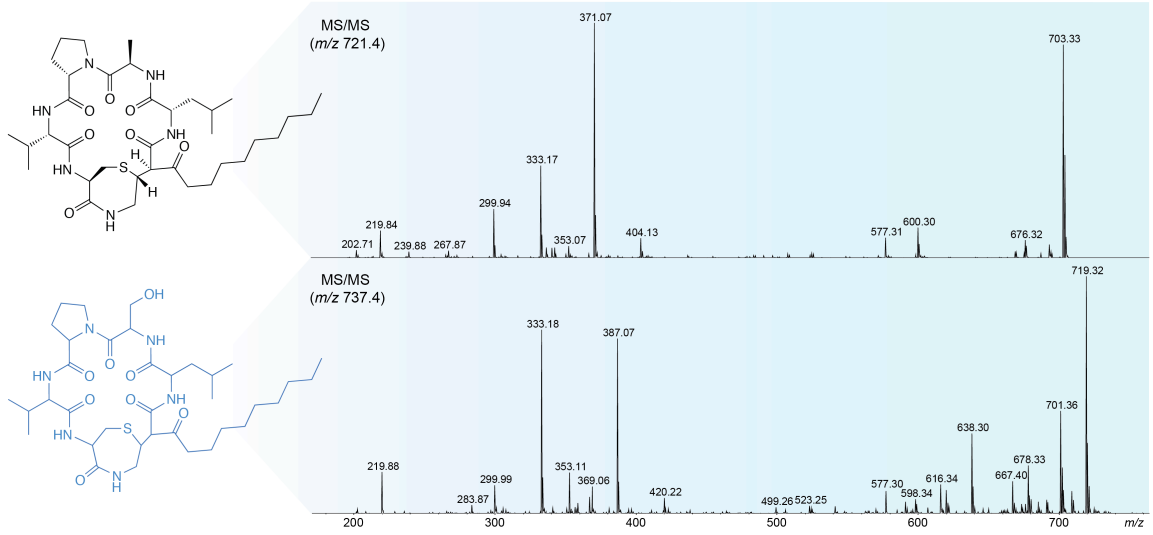


Figure S2.6.5. The fragmentation profiles of mutanobactin A and mutanobactin G. The structure of mutanobactin G (in blue) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 737.4266 calculated, *m/z* 737.4280 observed, 1.9 ppm).

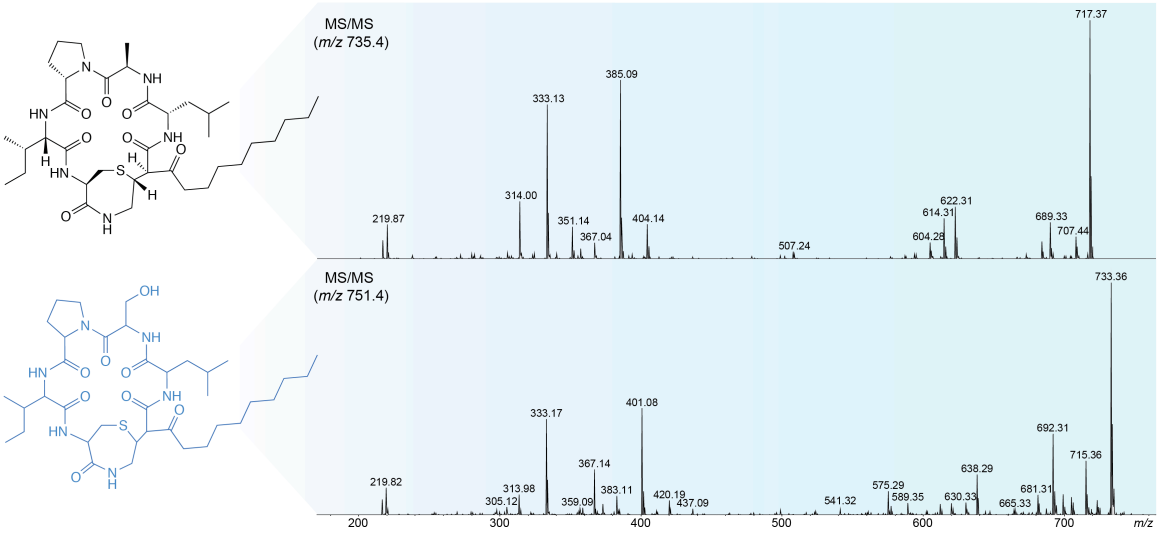


Figure S2.6.6. The fragmentation profiles of mutanobactin A and mutanobactin H. The structure of mutanobactin H (in blue) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 751.44226 calculated, *m/z* 751.44384 observed, 2.1 ppm).

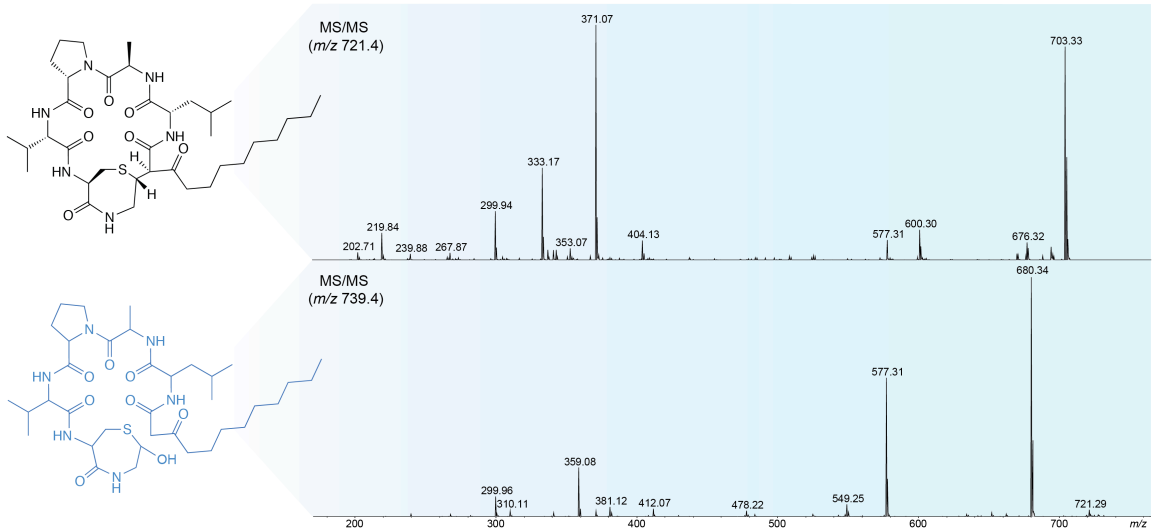


Figure S2.6.7. The fragmentation profiles of mutanobactin A and mutanobactin I. The structure of mutanobactin I (in blue) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 739.44226 calculated, *m/z* 739.44243 observed, 0.2 ppm).

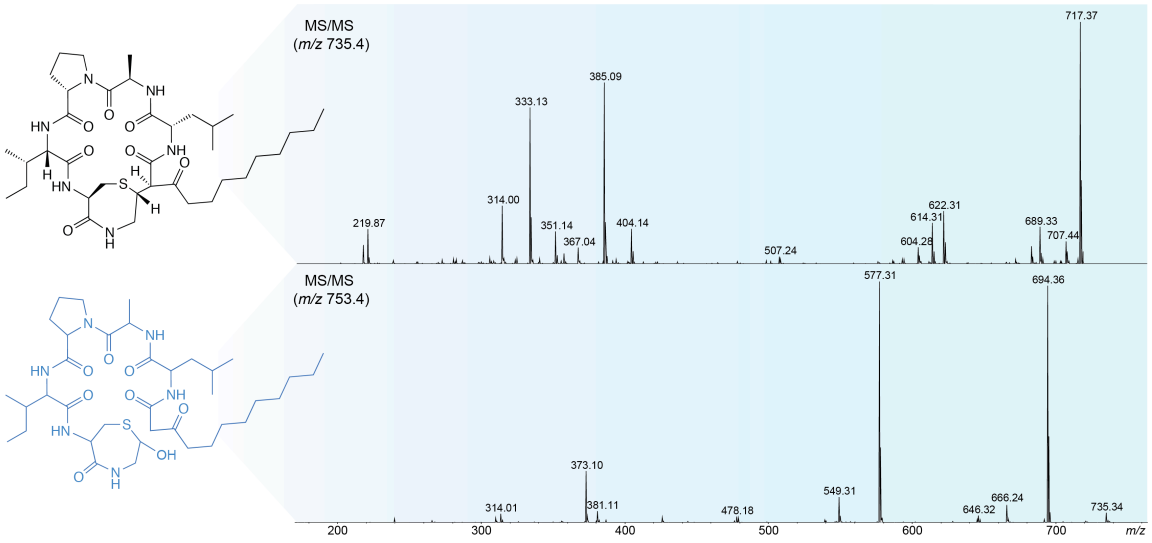


Figure S2.6.8. The fragmentation profiles of mutanobactin A and mutanobactin J. The structure of mutanobactin J (in blue) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 753.45791 calculated, *m/z* 753.45800 observed, 0.1 ppm).

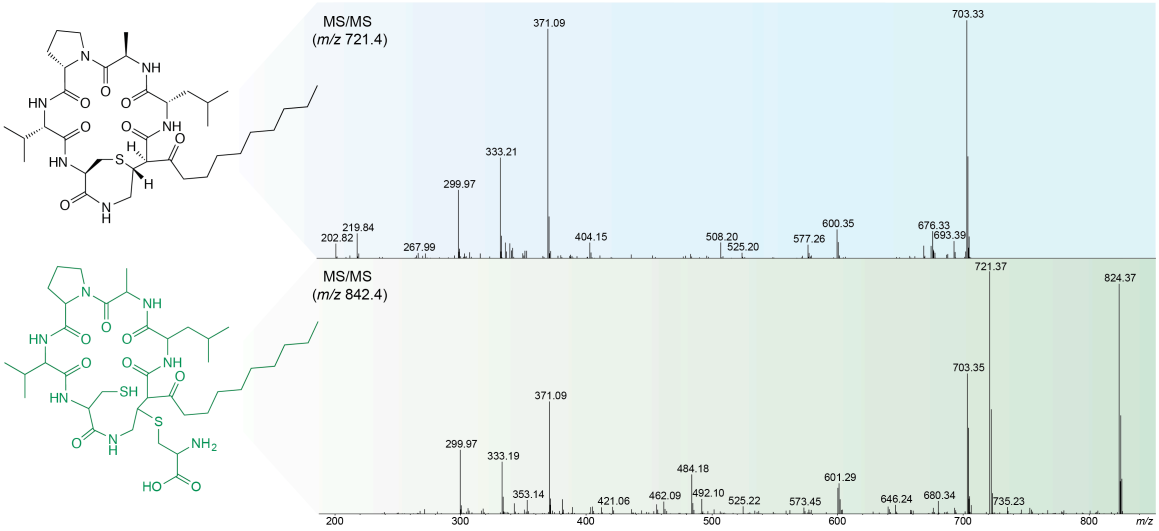


Figure S2.6.9. The fragmentation profiles of mutanobactin A and mutanolin A. The structure of mutanolin A (in green) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 842.4514 calculated, *m/z* 842.4511 observed, 0.4 ppm).

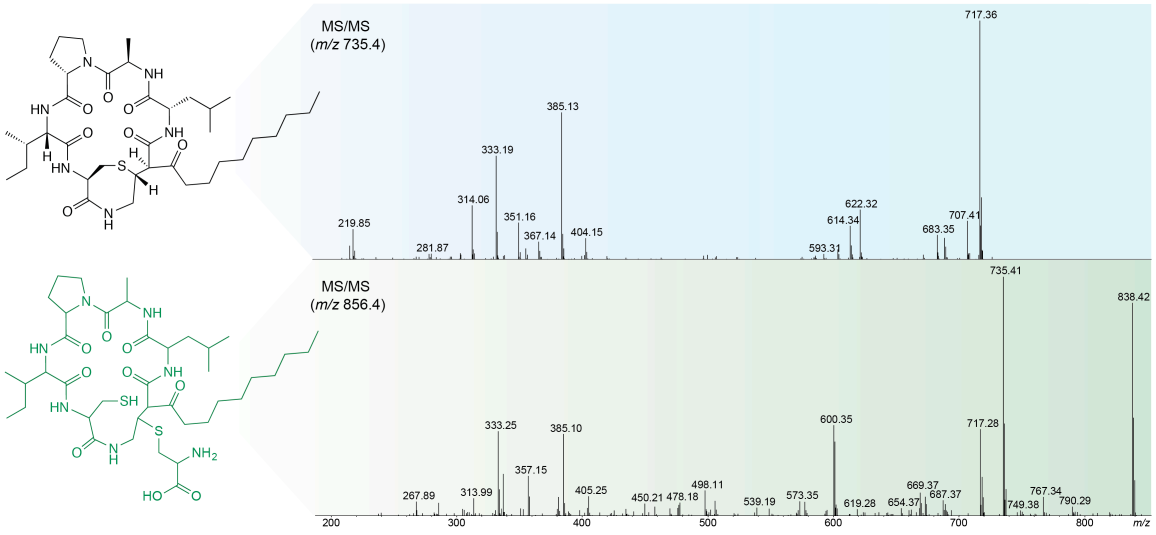


Figure S2.6.10. The fragmentation profiles of mutanobactin B and mutanolin B. The structure of mutanolin B (in green) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 856.4671 calculated, *m/z* 856.4670 observed, 0.1 ppm).

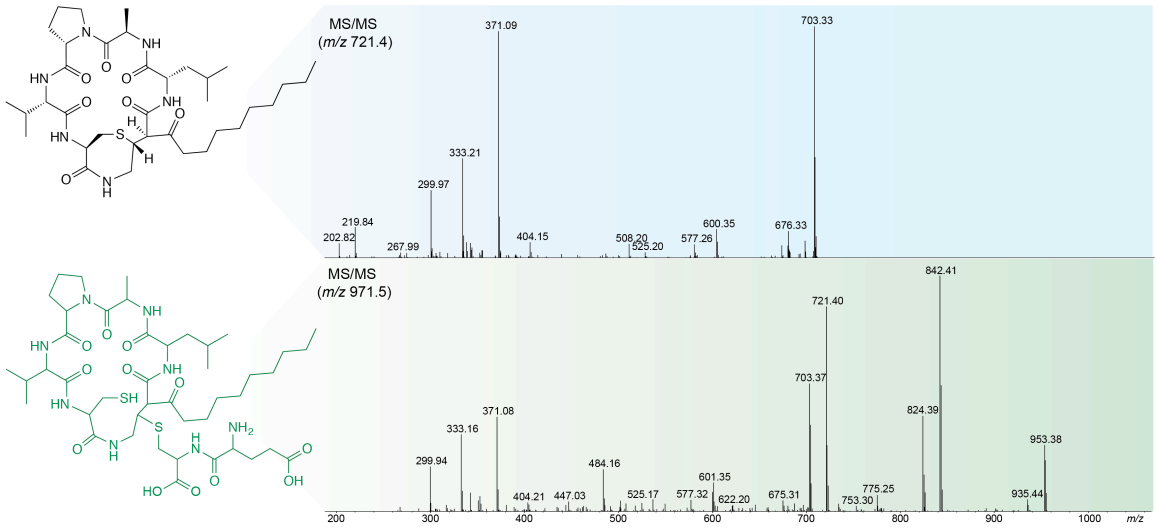


Figure S2.6.11. The fragmentation profiles of mutanobactin A and mutanolin C. The structure of mutanolin C (in green) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 971.4940 calculated, *m/z* 971.4945 observed, 0.5 ppm).

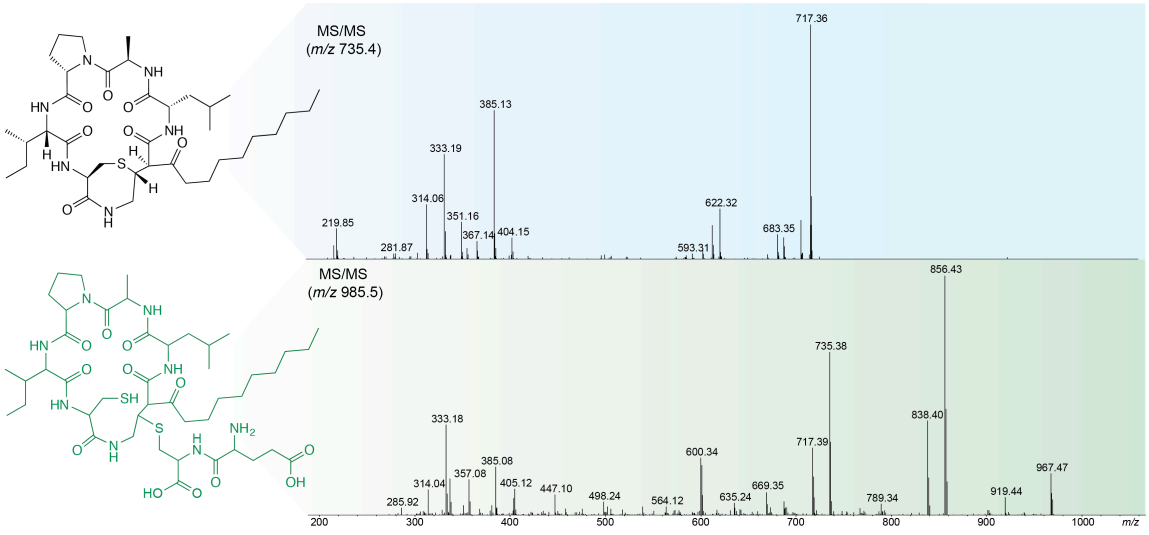


Figure S2.6.12. The fragmentation profiles of mutanobactin B and mutanolin D. The structure of mutanolin D (in green) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 985.50969 calculated, *m/z* 985.50951 observed, 0.2 ppm).

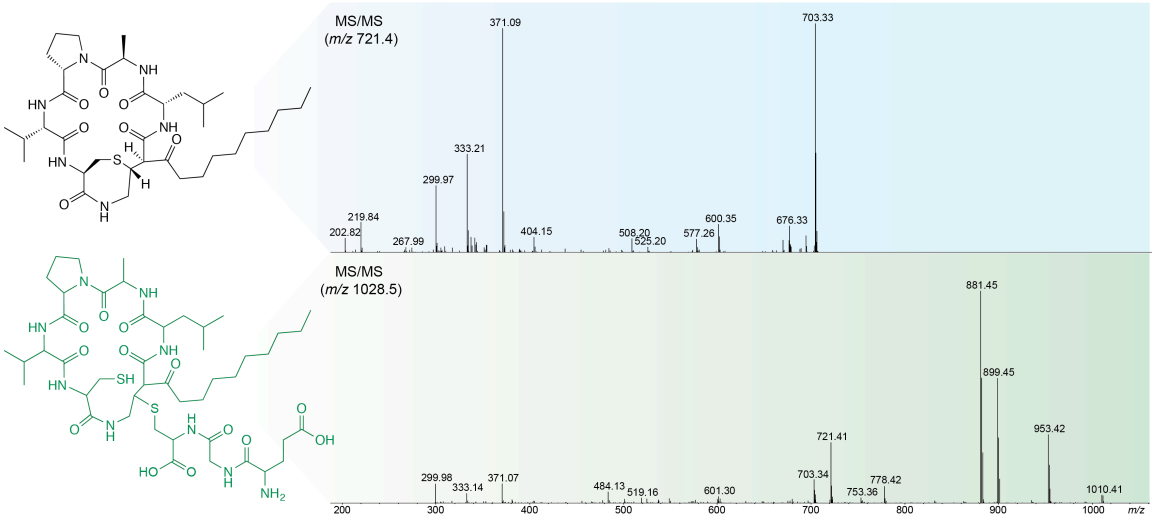


Figure S2.6.13. The fragmentation profiles of mutanobactin A and mutanolin E. The structure of mutanolin E (in green) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 1028.51550 calculated, *m/z* 1028.51530 observed, 0.2 ppm).

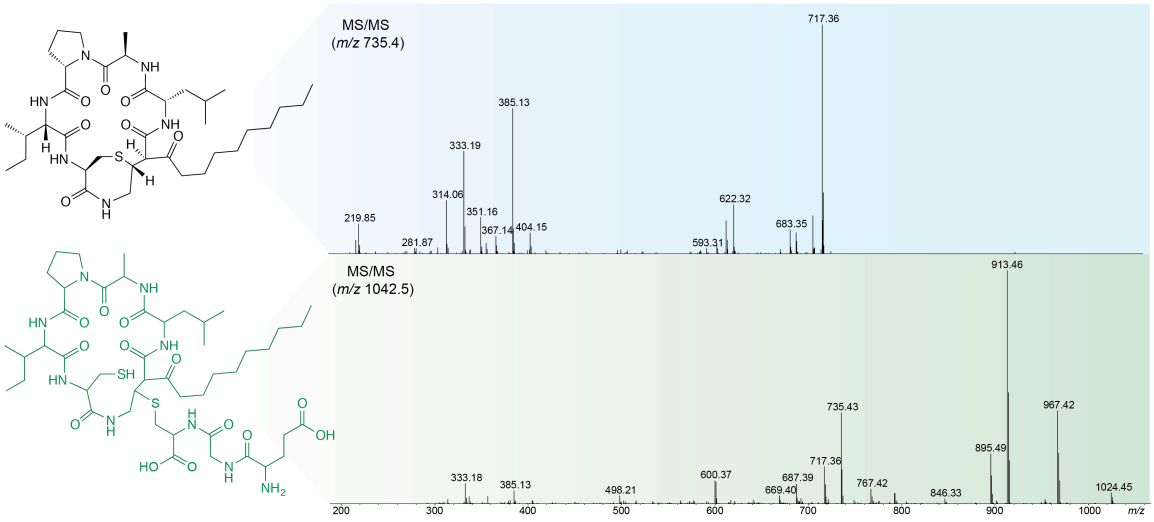


Figure S2.6.14. The fragmentation profiles of mutanobactin B and mutanolin F. The structure of mutanolin F (in green) was elucidated using the combination of MS/MS and incorporation of isotopically labeled building blocks. The elemental composition of the proposed structure was validated using HRMS (*m/z* 1042.53115 calculated, *m/z* 1042.52969 observed, 1.4 ppm).

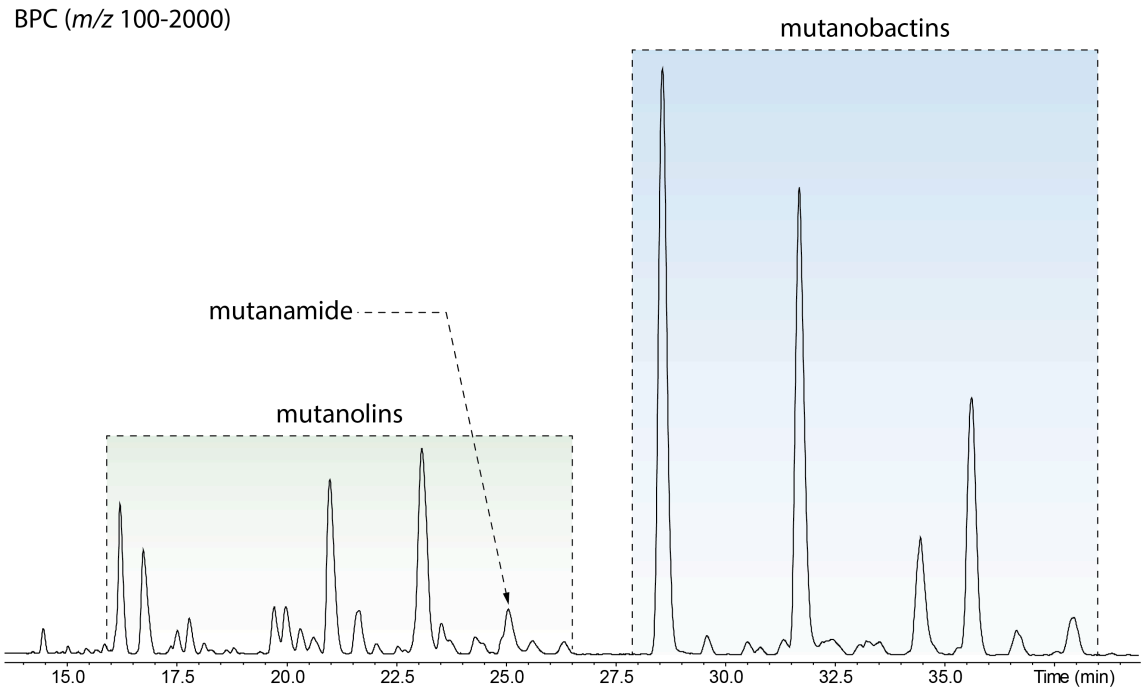


Figure S2.6.15. Base peak chromatogram of cell extract of *S. mutans* UA159. The approximate regions where mutanobactins, mutanamide and mutanolins elute are indicated.

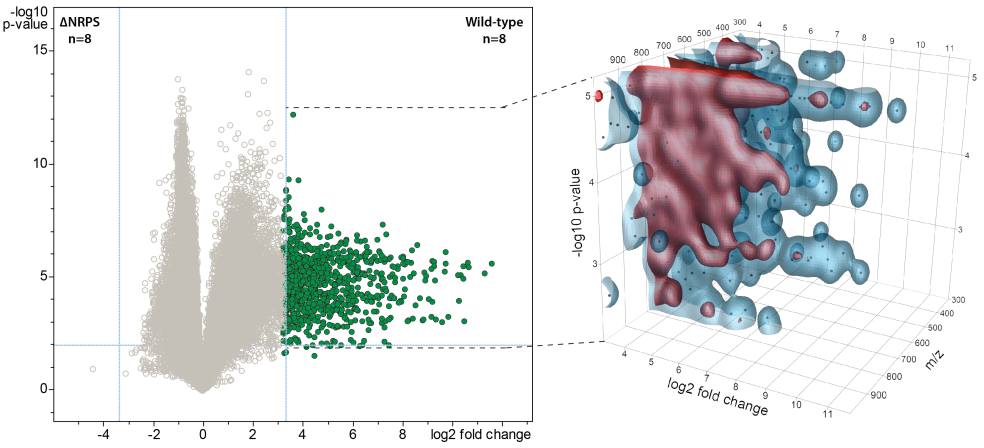


Figure S2.6.16. Clustering of metabolite features from the volcano plot. When *m/z* dimension is added to the volcano plot, molecular features pertaining to the wild-type strain (green circles) separate based on their corresponding *m/z* values. Clustering of the molecular features is represented by the contour plot, where red indicates dense clustering; two major clusters can be identified – around *m/z* 730 and *m/z* 400.

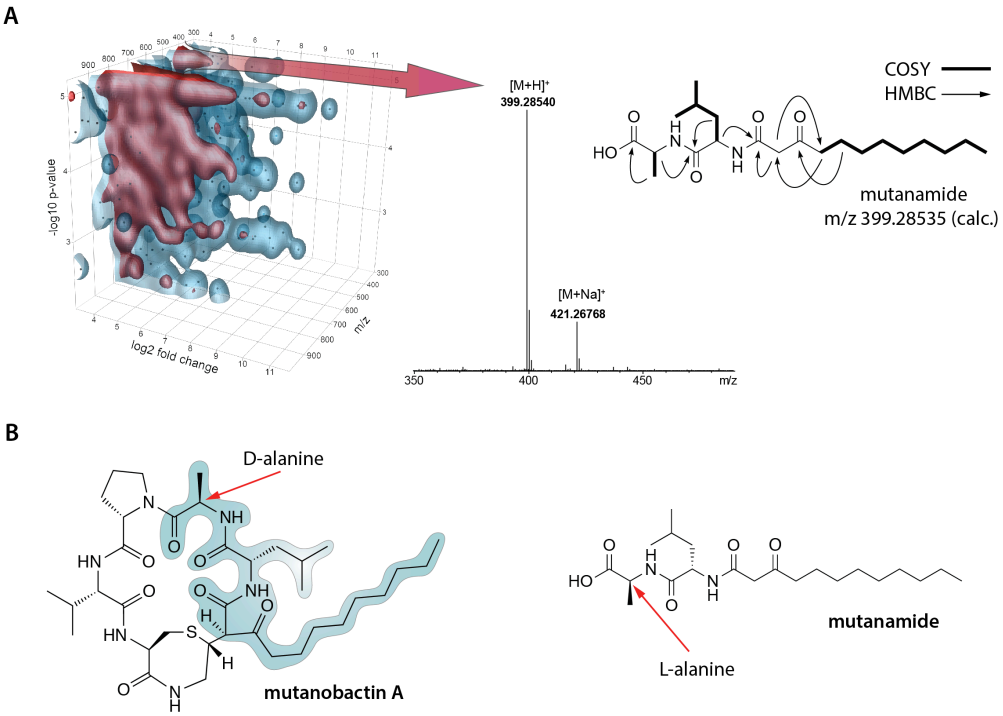


Figure S2.6.17. Identification and structure determination of mutanamide. **A** - the unexpected clustering around *m/z* 400 was due to *m/z* 399.3 (mutanamide), which was isolated and whose structure was elucidated. **B** – the chemical structure of mutanamide strongly resembles mutanobactin A, however, the latter contains D-alanine, whereas mutanamide – L-alanine.

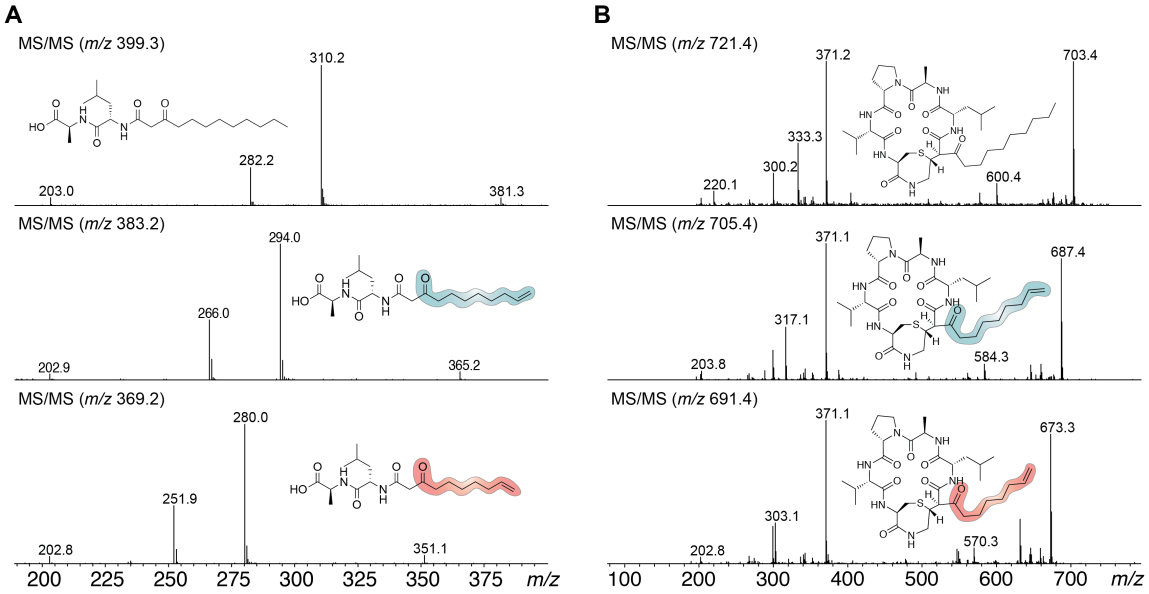


Figure S2.6.18. Fatty acid incorporation by mutanamide and mutanobactin A. **A** – MS/MS analysis of original mutanamide (*m/z* 399.3) and two analogs (*m/z* 383.2 and *m/z* 369.2) that incorporated 8-nonenoic and 7-octenoic fatty acids, respectively. **B** – MS/MS analysis of original mutanobactin (*m/z* 721.4) and two analogs (*m/z* 705.4 and *m/z* 691.4) that incorporated 8-nonenoic and 7-octenoic fatty acids, respectively.

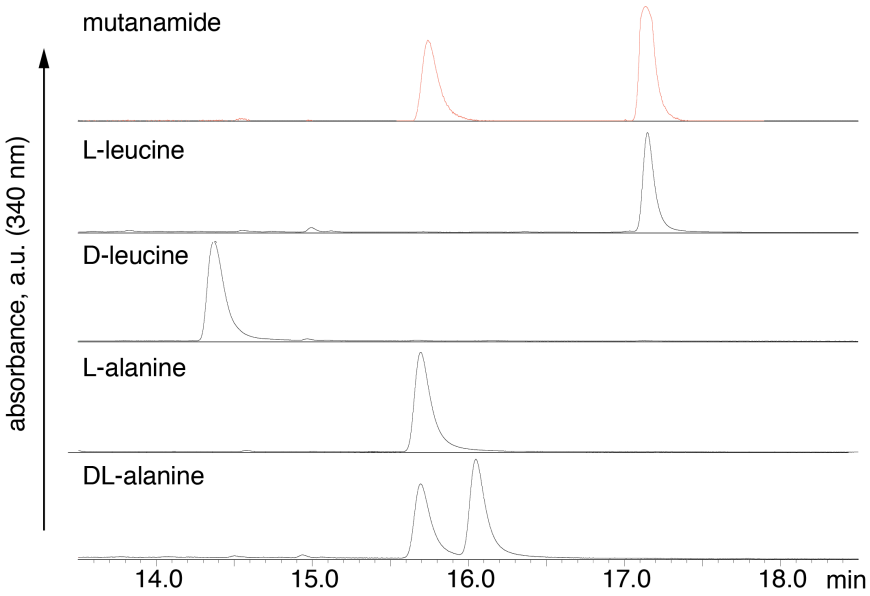


Figure S2.6.19. Results of Marfey's reaction to determine the exact stereochemistry of mutanamide. Mutanamide was fully hydrolyzed, reacted with Marfey's reagent and run on the HPLC-UV together with the L-leucine, D-leucine, L-alanine and DL-alanine standards.

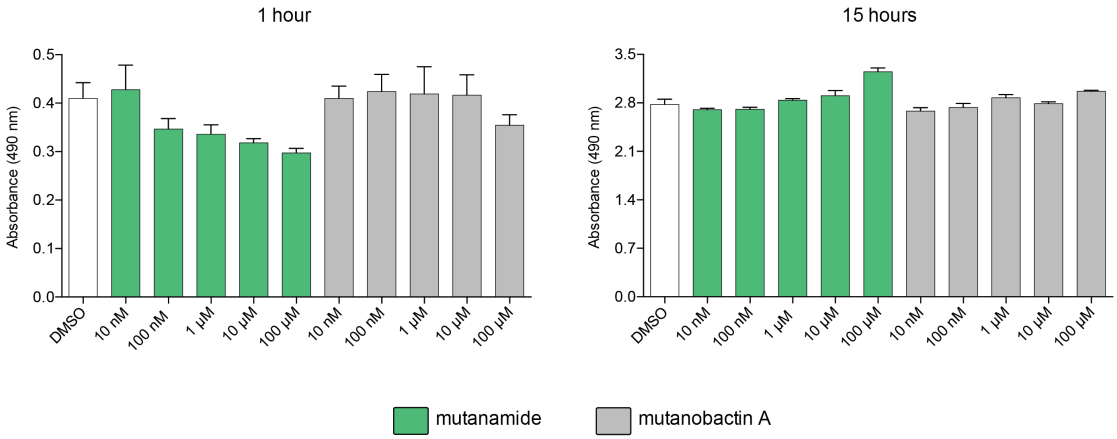


Figure S2.6.20. Antiproliferative activity of mutanamide and mutanobactin A in HT-29 cells. Both mutanamide and mutanobactin A did not have any significant effect on the growth of HT-29 cells after 1 or 15 hours.

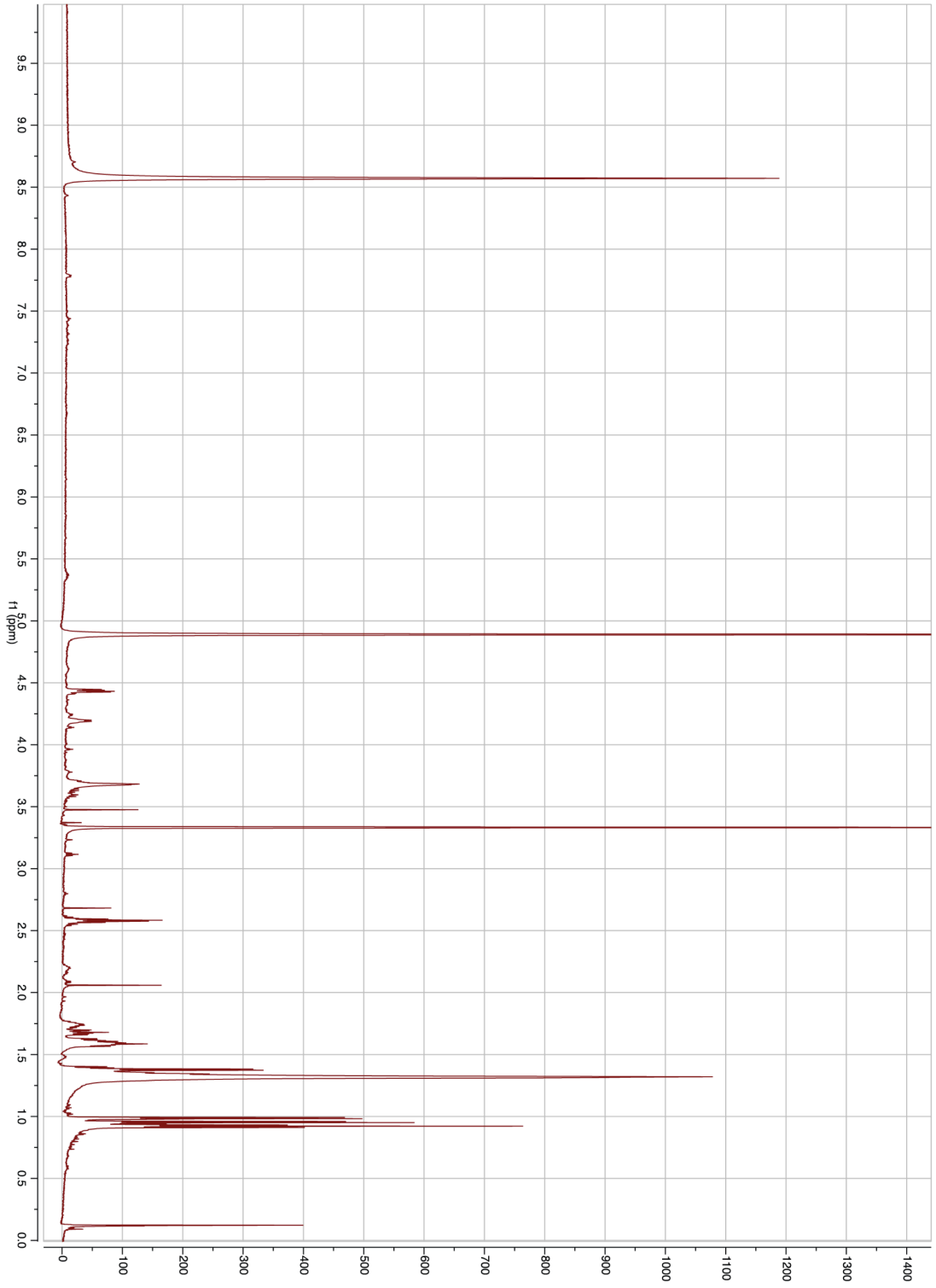


Figure S2.6.21. ^{1}H NMR spectrum of mutanamide in 100% methanol-d_4.

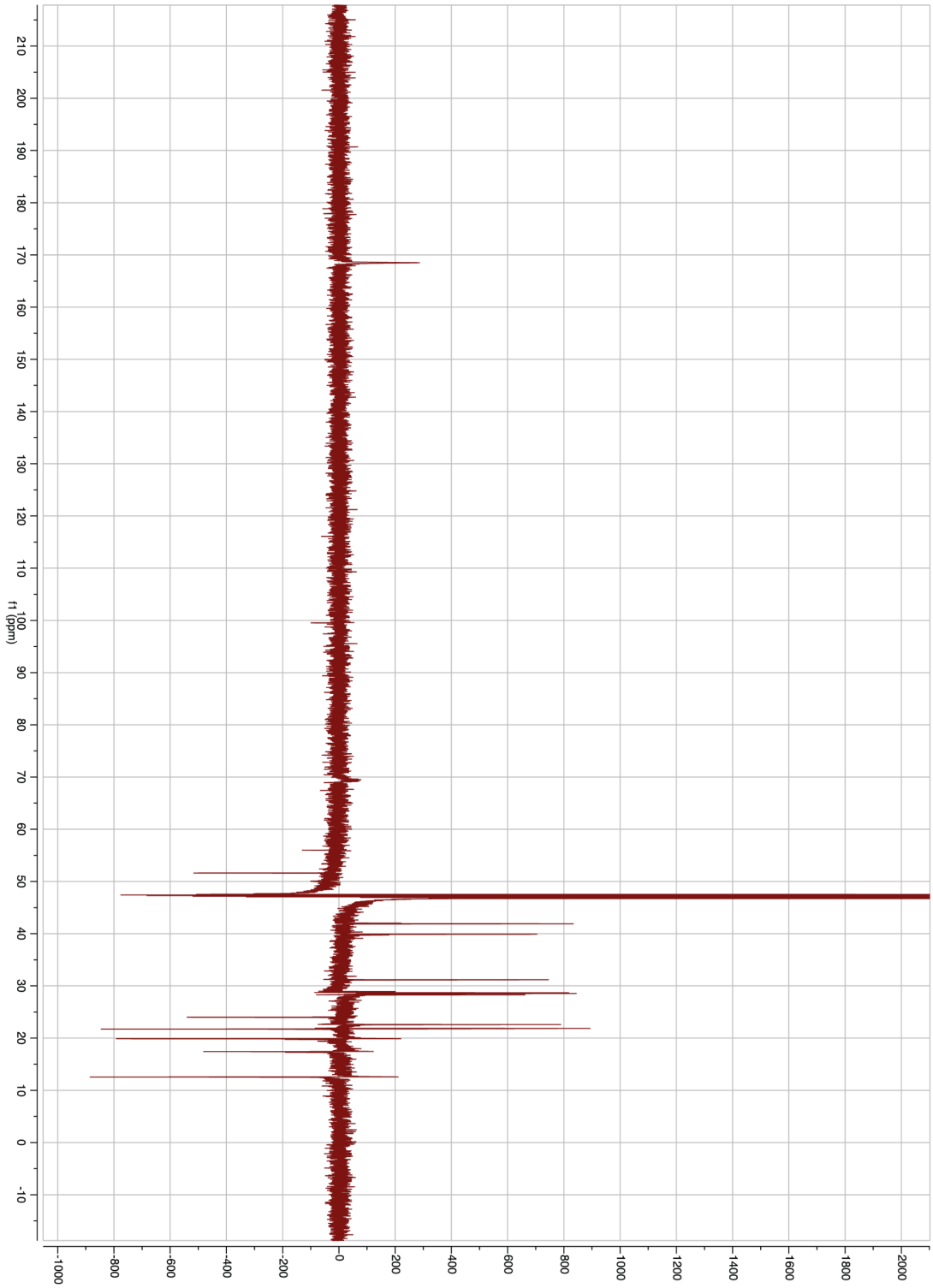


Figure S2.6.22. ^{13}C DEPTq135 NMR spectrum of mutanamide in 100% methanol-d_4.

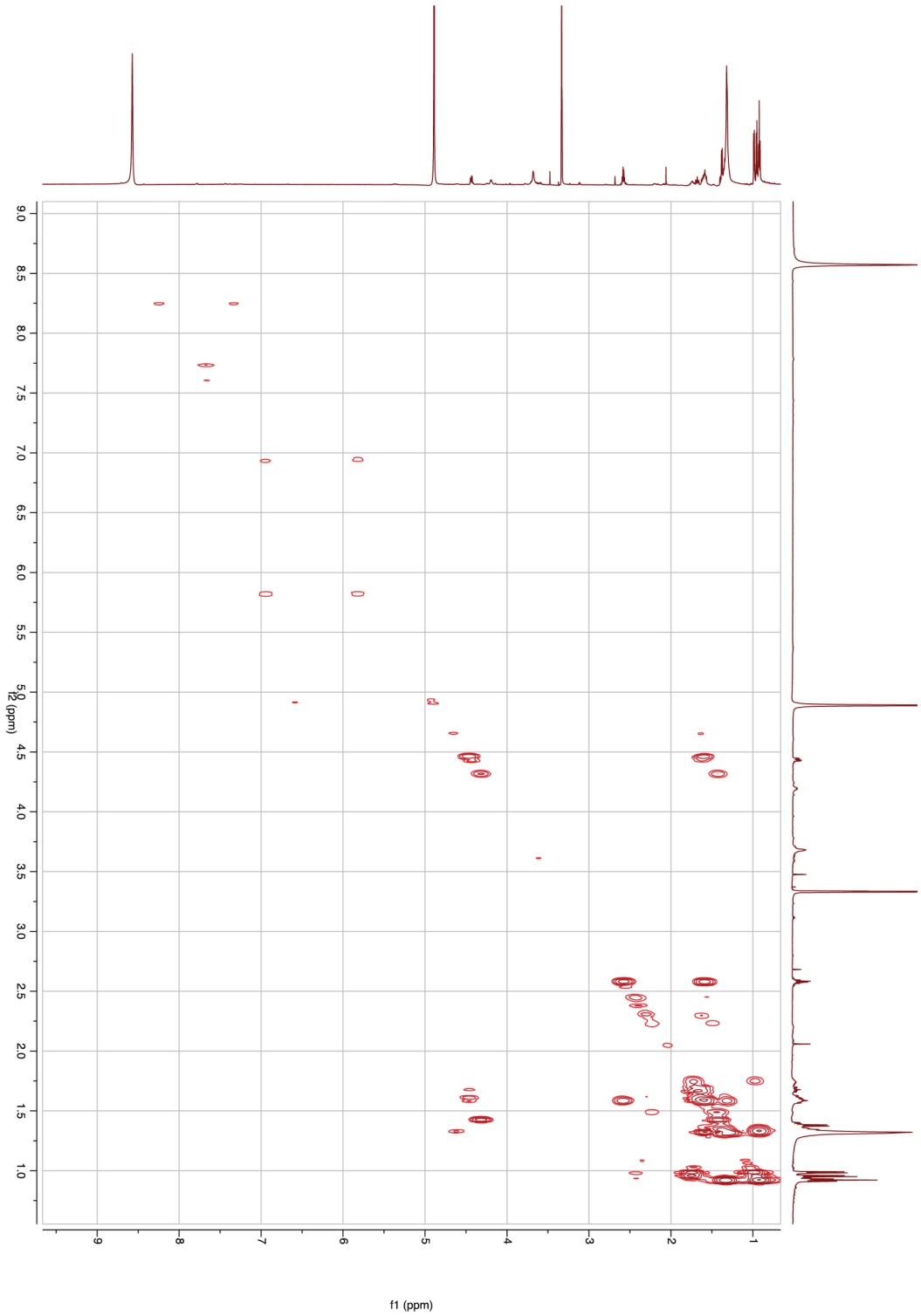


Figure S2.6.23. 2D COSY spectrum of mutanamide in 100% methanol-d_4.

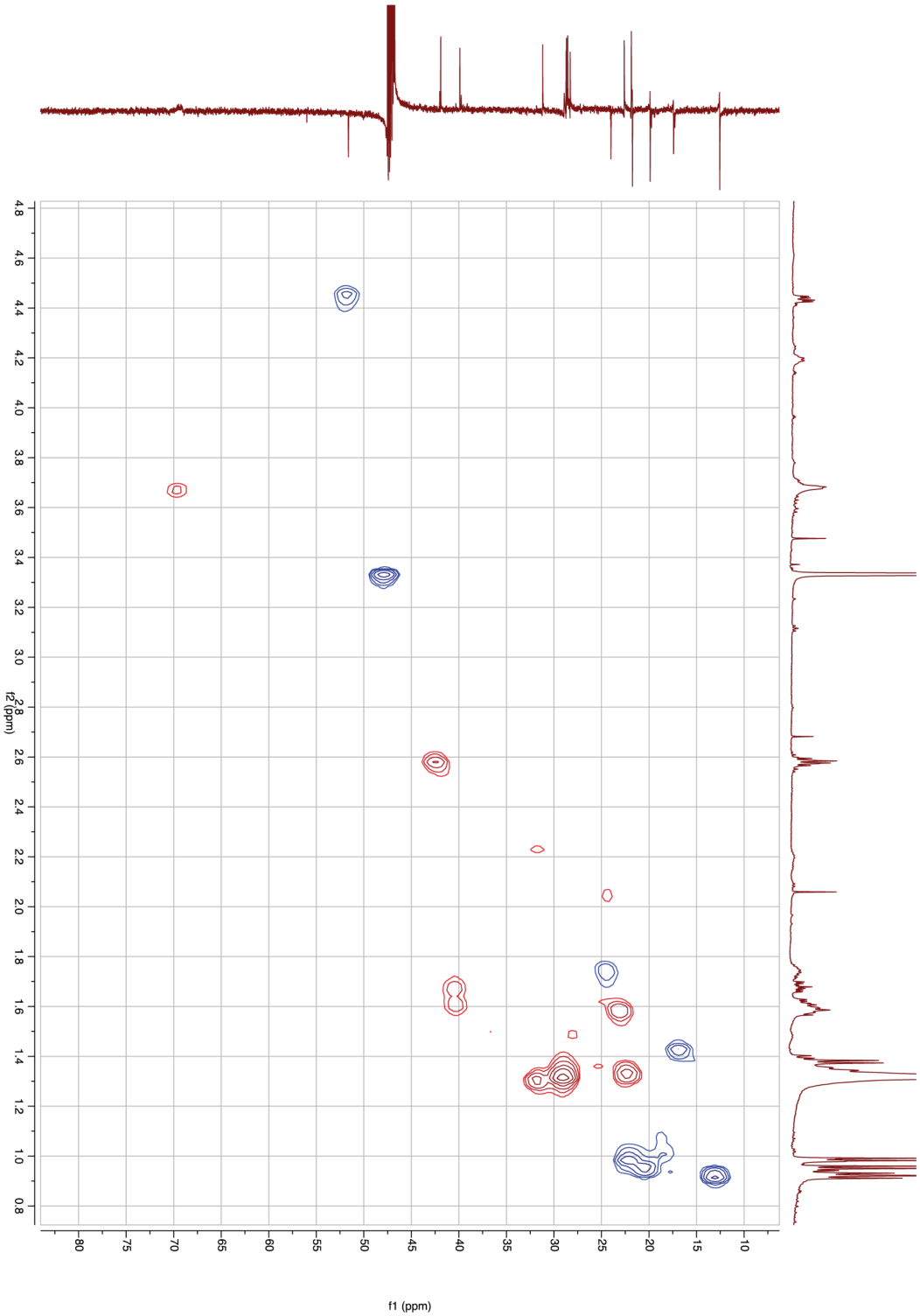


Figure S2.6.24. 2D HSQC spectrum of mutanamide in 100% methanol-d_4.

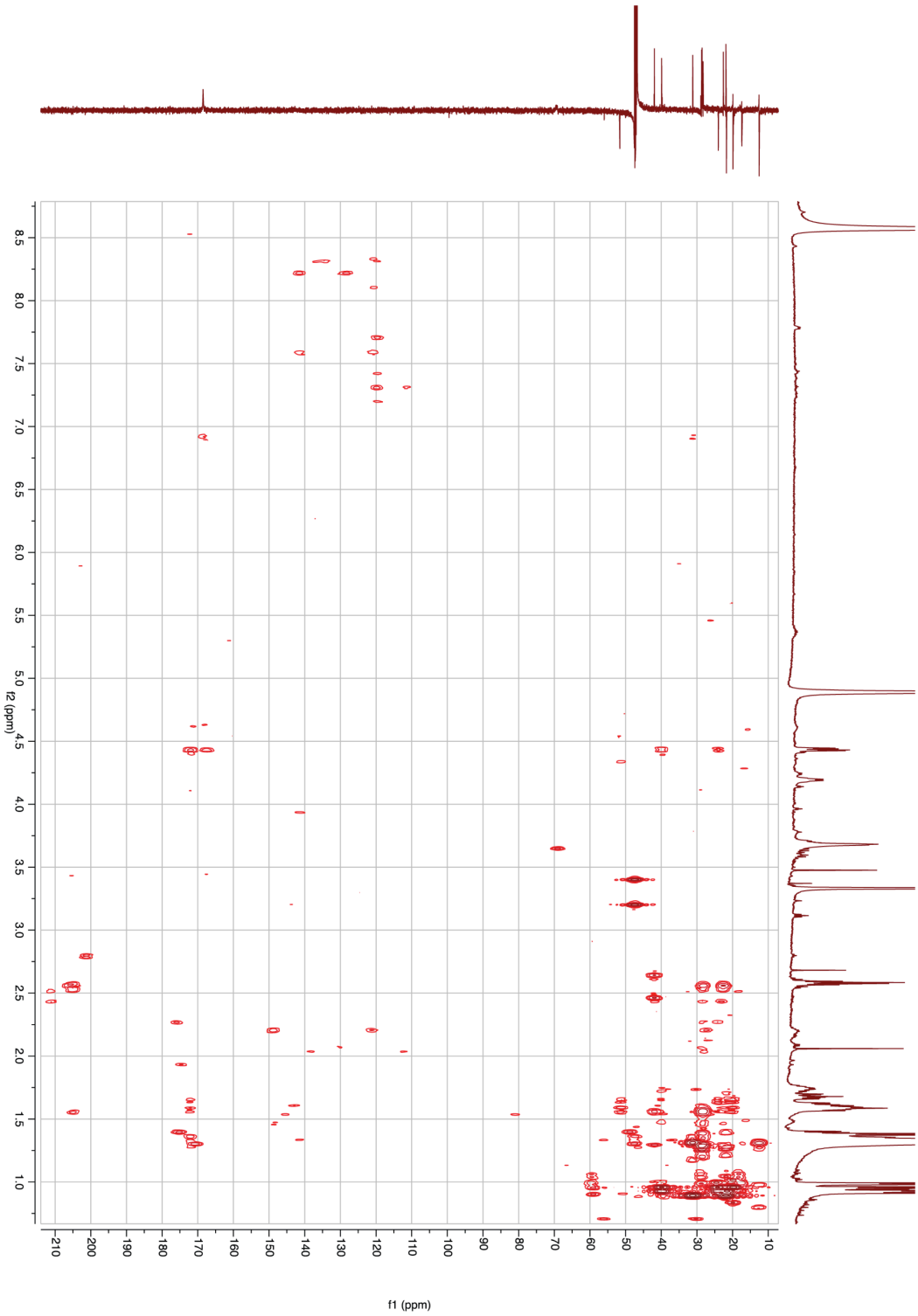


Figure S2.6.25. 2D HMBC spectrum of mutanamide in 100% methanol-d_4.

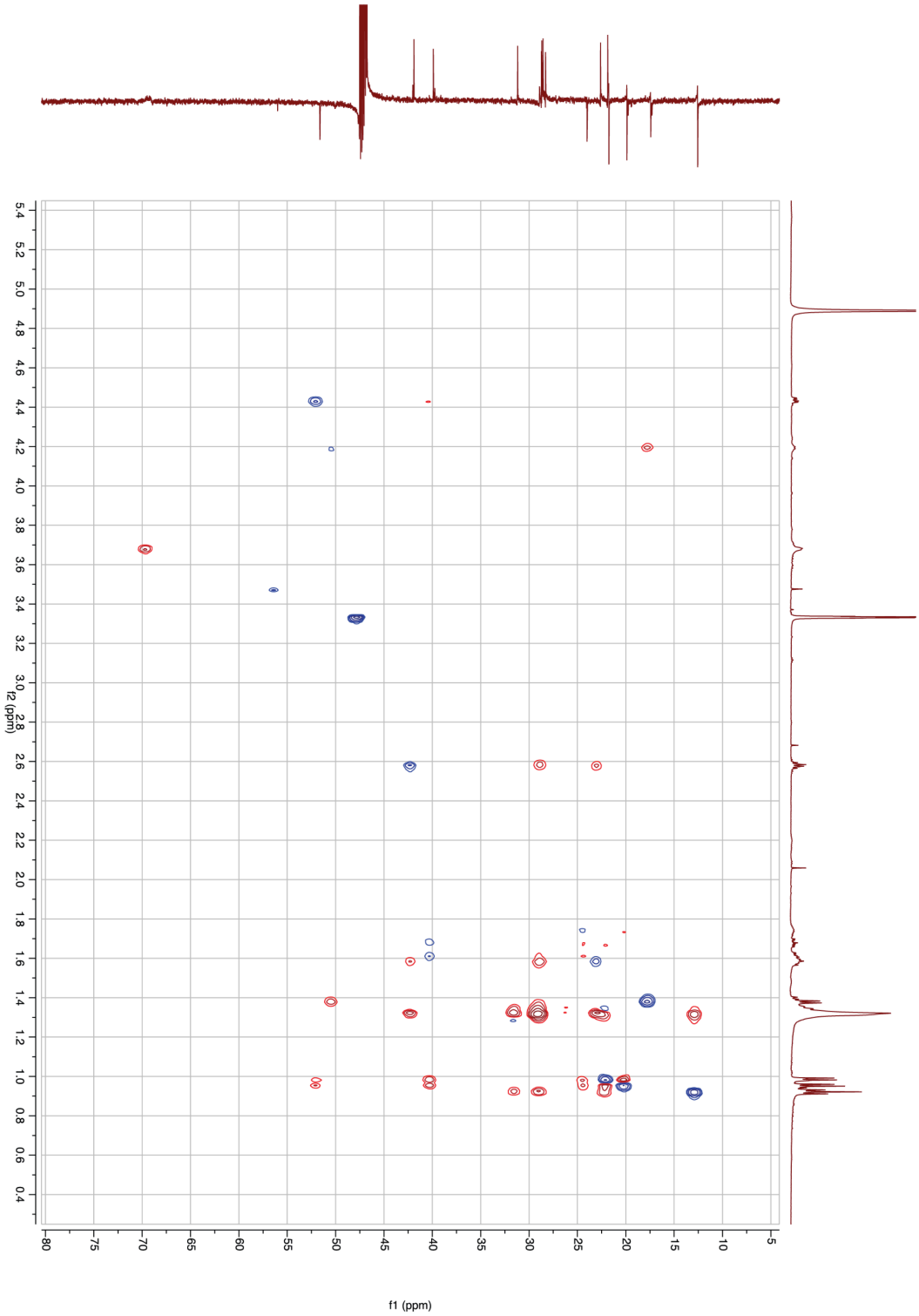


Figure S2.6.26. 2D HSQC-TOCSY spectrum of mutanamide in 100% methanol-d_4.

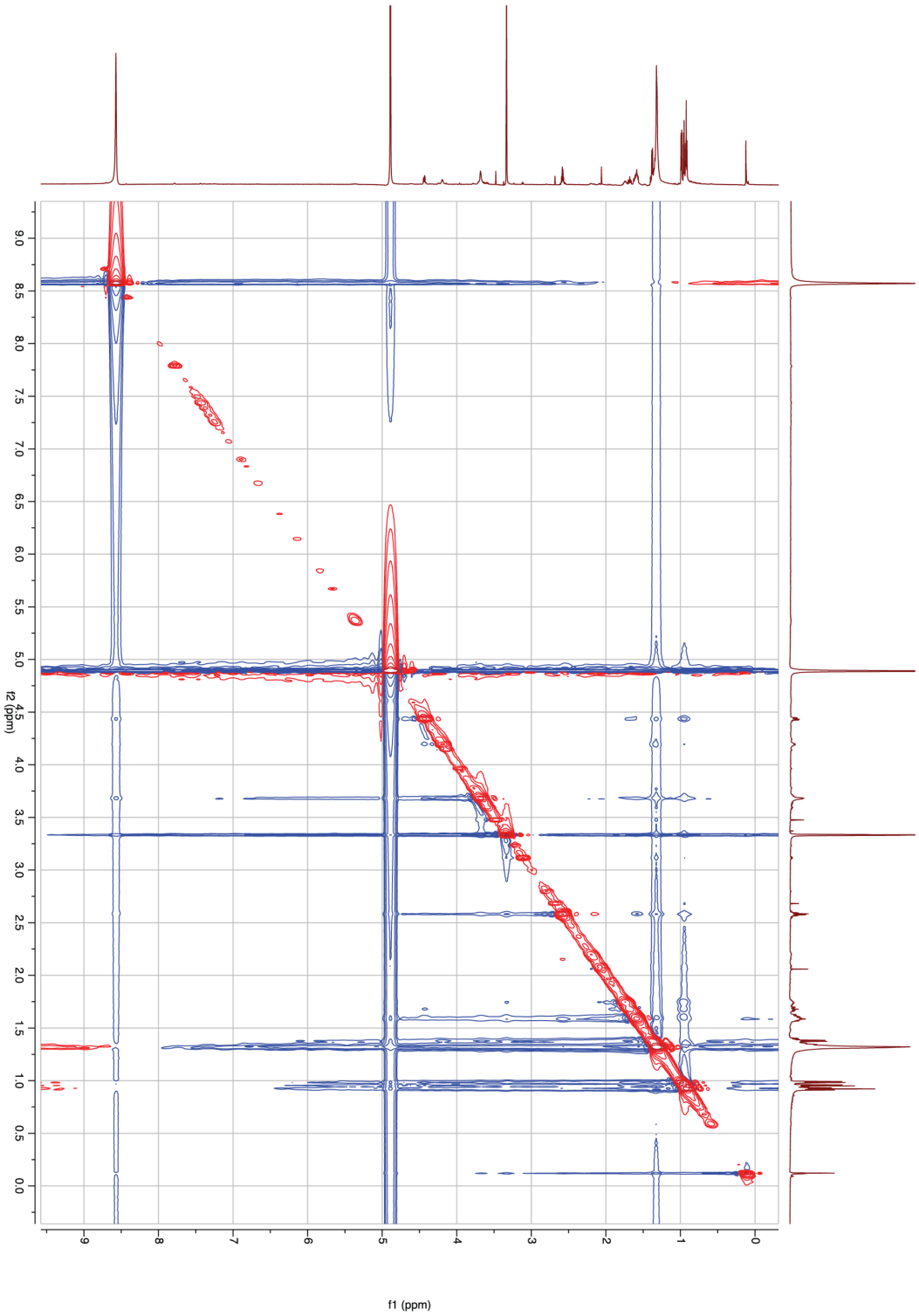


Figure S2.6.27. 2D NOESY spectrum of mutanamide in 100% methanol-d_4.

2.7 References

1. Martins, H. P. R., da Silva, M. C., Paiva, L. C. F., Svidzinski, T. I. E. & Consolaro, M. E. L. Efficacy of fluconazole and nystatin in the treatment of vaginal Candida species. *Acta Derm. Venereol.* **92,** 78–82 (2012).
2. Starks, C. M., Zhou, Y., Liu, F. & Licari, P. J. Isolation and characterization of new epothilone analogues from recombinant Myxococcus xanthus fermentations. *J. Nat. Prod.* **66,** 1313–1317 (2003).
3. Shaw, L. M. Advances in cyclosporine pharmacology, measurement, and therapeutic monitoring. *Clin. Chem.* **35,** 1299–1308 (1989).
4. Vézina, C., Kudelski, A. & Sehgal, S. N. Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *The Journal of Antibiotics* **28,** 721–726 (1975).
5. Ferrini, S., Moretta, A., Biassoni, R., Nicolin, A. & Moretta, L. Cyclosporin-A inhibits IL-2 production by all human T-cell clones having this function, independent of the T4/T8 phenotype or the coexpression of cytolytic activity. *Clin. Immunol. Immunopathol.* **38,** 79–84 (1986).
6. Feuerstein, N., Huang, D. & Prystowsky, M. B. Rapamycin selectively blocks interleukin-2-induced proliferating cell nuclear antigen gene expression in T lymphocyte. Evidence for inhibition of CREB/ATF binding activities. *J. Biol. Chem.* **270,** 9454–9458 (1995).
7. Foldenauer, M. E. B., McClellan, S. A., Berger, E. A. & Hazlett, L. D. Mammalian target of rapamycin regulates IL-10 and resistance to Pseudomonas aeruginosa

corneal infection. *The Journal of Immunology* **190,** 5649–5658 (2013).

8. Ajdić, D. *et al.* Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 14434–14439 (2002).
9. Joyner, P. M. *et al.* Mutanobactin A from the human oral pathogen Streptococcus mutans is a cross-kingdom regulator of the yeast-mycelium transition. *Org. Biomol. Chem.* **8,** 5486 (2010).
10. Wang, X., Du, L., You, J., King, J. B. & Cichewicz, R. H. Fungal biofilm inhibitors from a human oral microbiome-derived bacterium. *Org. Biomol. Chem.* **10,** 2044 (2012).
11. Sudbery, P. E. Growth of Candida albicans hyphae. *Nat. Rev. Microbiol.* **9,** 737–748 (2011).
12. Wu, C. *et al.* Genomic island TnSmu2 of Streptococcus mutans harbors a nonribosomal peptide synthetase-polyketide synthase gene cluster responsible for the biosynthesis of pigments involved in oxygen and H2O2 tolerance. *Applied and Environmental Microbiology* **76,** 5815–5826 (2010).
13. Ibrahim, A. *et al.* Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 19196–19201 (2012).
14. Vanner, S. A. *et al.* Chemical and biosynthetic evolution of the antimycin-type depsipeptides. *Mol Biosyst* **9,** 2712–2719 (2013).
15. Robbel, L. & Marahiel, M. A. Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J. Biol. Chem.* **285,** 27501–27508 (2010).

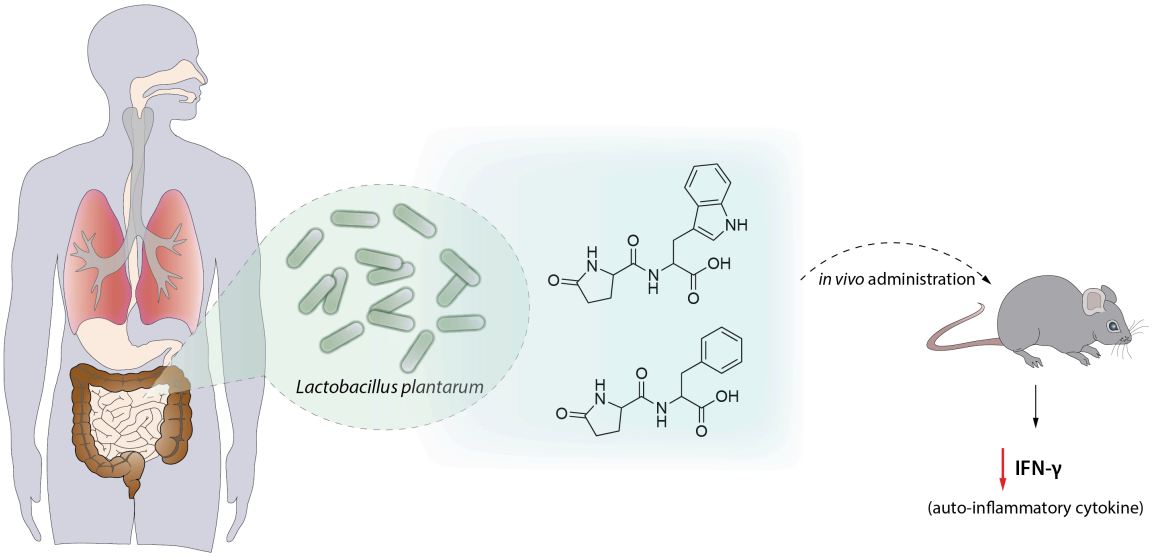
16. Terleckyj, B. & Shockman, G. D. Amino acid requirements of Streptococcus mutans and other oral streptococci. *Infection and Immunity* **11,** 656–664 (1975).

17. Shareck, J. & Belhumeur, P. Modulation of morphogenesis in Candida albicans by various small molecules. *Eukaryotic Cell* **10,** 1004–1012 (2011).

18. Bhushan, R. & Brückner, H. Marfey's reagent for chiral amino acid analysis: a review. *Amino Acids* **27,** 231–247 (2004).

Chapter 3. Small Molecule Immunomodulins From Cultures of the Human Microbiome Member *Lactobacillus plantarum*

3.1 Chapter Preface



Lactobacillus plantarum strains are noted for their presence in the human gut and the production of certain probiotic and immunomodulatory components that suggest their therapeutic use. The molecular mechanisms directing the noted activities have remained enigmatic, however, recent evidence suggests that secreted metabolites are linked to these actions. To delineate the active metabolites produced by *Lactobacillus plantarum* as it transitions from exponential to stationary phase of growth, I investigated the metabolomic changes using principal component analysis (PCA). From this analysis, similarly to the observations made by my colleague, Nikola Lukenda, I determined that the most pronounced changes were mainly due to the significant increase in the production of a small dipeptide compound, pyroglutamic acid-phenylalanine (pyro-phenylalanine).

Although a more in-depth metabolomic analysis of *L. plantarum* is required to gain full understanding of its beneficial effects, my research shines the light on the effects of secreted metabolites *in vivo*. More importantly, newly identified pyroglutamic acid dipeptides substantiate the importance of studying the secreted metabolome in greater detail. It is plausible to envision that there might be other immunomodulatory agents produced by the Lactobacilli and other probiotic microorganisms that make-up their corresponding beneficial arsenals.

The following chapter is a modified version of a previously published journal article in which Nikola Lukenda and I are the lead authors. Together with Nikola Lukenda I contributed equally to the experimental design, interpretation of results and writing of the manuscript. Janice J. Kim carried out the *in vivo* activity testing of pyro-phenylalanine and pyro-tryptophan and performed the cytokine analysis. Dr. Xiang Li elucidated the structures of four pyroglutamic acid dipeptides. Dr. Elaine O. Petrof, Dr. Waliul I. Khan and Dr. Nathan A. Magarvey provided input for the progression of this research. The citation for this publication is as follows:

Zvanych, R. *et al.* Small molecule immunomodulins from cultures of the human microbiome member *Lactobacillus plantarum*. *The Journal of Antibiotics* **67,** 85–88 (2014). Published online Nov 27, 2013.

3.2 Abstract

Lactobacillus plantarum strains are noted for their presence in the human gastrointestinal tract and are distinguished for their immunomodulatory actions and therapeutic applications. Despite the uncertainty in the underlining molecular mechanisms, recent evidence suggests that *L. plantarum* secretes immunomodulatory agents that alter immunological signaling cascades. Elaboration of these metabolic products from *L. plantarum* strain WCFS1 was demonstrated previously to correlate with the mid-log-stationary transition, perhaps consistent with secondary metabolite expression. Here, we present the metabolomic shifts revealed by principal component analysis that correspond to the mid-log-stationary transition of *L. plantarum*, and identify pyroglutamic (pyro) dipeptides within this transition as correlative with the immunomodulatory actions. Four of these (pyro- phenylalanine, pyro-leucine, pyro-isoleucine, pyro-tryptophan) were characterized and the two dominant members, pyro-phenylalanine and pyro-tryptophan, were directly interrogated for immunomodulatory activity through in vivo administration using C57BL/6 mice. Administration of these compounds resulted in decreased production of pro-inflammatory cytokine interferon (IFN)-gamma, which is of noted importance in gastrointestinal immune homeostasis.

3.3 Introduction

The microbial community which resides within the human gastrointestinal (GI) tract, collectively known as the human gut microbiota, is comprised of a diverse group of microorganisms which are thought to play a critical role in maintaining intestinal homeostasis.[1-3] Select members of the commensal GI tract microbiota have been

repurposed for therapeutic use and are currently used as agents in the probiotic mixtures such as VSL#3, and shown to confer cytoprotective and anti-inflammatory effects in vitro, in vivo and in the clinical trials.[4-6] *Lactobacillus plantarum*, a GI tract microbiome member and a constituent of VSL#3, is often used as a probiotic in its own for its beneficial properties and anti-inflammatory activity.[7-10] Its actions are thought to contribute to the amelioration of gastric ulcers and prevention of colitis-associated colorectal cancer.[8,11] Investigations of *L. plantarum* have led to a plausible concept that the organism modulates the immune system through the secretion of metabolites.[12] However, the chemical nature of these molecules is not known, nor have select agents been isolated and probed for their activities in cell or animal models.

Lactobacilli are noted for their production of modified secreted peptides, including lantibiotics, bacteriocins, and various ribosomally-derived peptides.[13-17] Similar to other bacterial secondary metabolites, modified peptides from the Lactobacilli are notably produced during the transition between exponential (mid-log) to stationary phase of growth in broth cultures.[15,18,19] In keeping with this observation, Baarlen et al. recently demonstrated that the administration of mid-log and stationary growth phase *L. plantarum* preparations differentially impacted the human immune system.[20] More specifically, only the stationary phase preparations were shown to induce the production of negative regulators of the NF-κB signaling pathway, hinting at the existence of growth phase-specific molecular entities. In agreement with the above results, Petrof et al. also observed inhibition of the NF-κB signaling pathway, however, with the *L. plantarum* supernatant alone, further substantiating the involvement of secreted agents.[12] Although these studies

were able to delineate the timing of the metabolite production, they did not focus on specific chemicals. However, they did provide a partial representation of the mechanisms by which these agents might act. Testing these chemical entities through a direct administration to whole organisms would be optimal for revealing any alterations of known markers of immune tolerance. Pertinent to the observed effects with *L. plantarum*, markers such as interferon-gamma (IFN-γ) would be an ideal measure, given the strong literature precedent of its involvement in NF-κB signaling.[21-23] Herein, we present a series of growth phase transition-dependent pyroglutamic acid dipeptides that possess anti-inflammatory activity in the in vivo setting by reducing the release of the pro-inflammatory cytokine IFN-γ.

3.4 Results and Discussion

To delineate possible active metabolites produced by *L. plantarum* as it transitions from exponential to stationary phase of growth, we investigated the metabolomic changes using principal component analysis (PCA). Apart from obtaining a graphical representation of the similarities and differences between the bacterial strains, PCA provides a means of determining the variations among very similar sample sets, as is the case for metabolomic changes through transitioning phases of growth.[24] The scores plot generated for mid-log and stationary phase samples of *L. plantarum* resulted in a distinct separation of the two groups with tight replicate clustering, indicating the presence of previously anticipated metabolomic differences between the two groups (**Figure 3.1a**). Detailed analysis of the loadings plot (**Figure 3.1b**) and the metabolite features defining the separation between the two groups identified a suite of metabolite features pertaining

to the stationary phase samples. Such analysis led to the identification of a metabolite with a mass to charge ratio of 277.1 with a distinct retention time (16.4 min, **Figure 3.1c**) and, as determined from the *t*-test analysis and correlative with the growth-phase transition, an 11-fold intensity increase. MS/MS analysis of this metabolite produced a distinctive *m/z* 165.9 ion, corresponding to a singly charged phenylalanine, and was correlative with a loss of another proteinogenic amino acid - pyroglutamic acid (**Supplementary information**, **Figure S3.6.1**). Following isolation and purification of the compound using a series of normal and reverse phase chromatography methods, conducting ^{1}H, COSY (^{1}H-^{1}H) and HMBC (^{1}H-^{13}C) NMR experiments provided sufficient information to elucidate the structure (**Supplementary information**, **Figure**

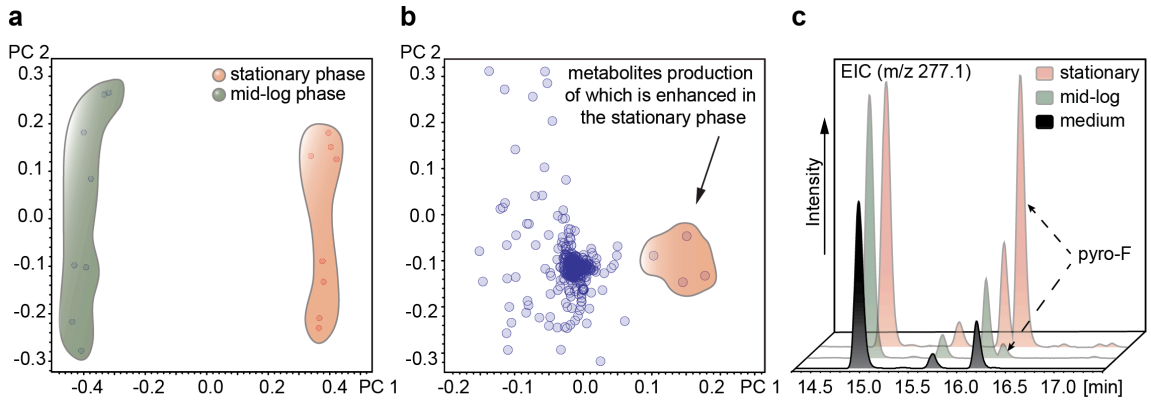


Figure 3.1. Principal component analysis of mid-log and stationary-phase extracts of *L. plantarum*. (**a**) Scores plot indicating the separation and clustering of the two groups of samples under investigation. (**b**) Loadings plot displaying the outliers leading to the separation in panel (**a**). (**c**) Extracted ion chromatogram of pyro-F (*m/z* 277.1, indicated with dashed arrows) identified from the PCA analysis and its relative intensity in the growth medium, and mid-log and stationary extracts.

S3.6.2, **S3.6.6**). From the ^{1}H NMR data, the first set of signals including a signal at 7.73 ppm (s) that corresponds to the secondary amine, 7.21 ppm (t) and 7.15 ppm (m) that correspond to the aromatic ring protons, 4.12 ppm (m) that corresponds to the α-proton, 2.96 ppm (dd) and 3.10 ppm (dd) that correspond to the methylene protons were all assigned to phenylalanine. The second set of signals including a signal at 7.90 ppm (s) corresponding to a secondary amine, together with 2.22 ppm (m), 1.98 ppm (m) and 1.79 ppm (m) signals were all corresponded to pyroglutamic acid (**Supplementary information**, **Figure S3.6.2**). The correlations obtained using COSY and HMBC experiments further supported these assignments (**Supplementary information**, **Figure S3.6.6**).

The NMR data were also in keeping with the loss of 111.0 observed in the MS/MS spectrum that provided further confidence in the assignment and our reporting of this compound as a pyroglutamic acid (pyro) phenylalanine dipeptide (pyro-F) (**Figure 3.2a**). Although pyroglutamic acid is a rare constituent in natural products, it is a part of several noted human metabolites such as thyrotropin-releasing hormone, luteinizing hormone-releasing hormone and gastrin I.[25] Its free form has also been previously described within several Lactobacilli, and although the enzymatic machinery directing cyclization of the presumed precursor has not been demonstrated, the origins of this amino acid would seemingly come from the cyclization of glutamic acid.[26,27] Since several of the other metabolites identified within the PCA plots were close in mass to pyro-F, we reasoned that these other molecules may be related and may likewise contain the pyroglutamic acid moiety. Thus, we interrogated these metabolites using tandem mass

spectrometry to reveal fragment losses consistent with pyroglutamic acid (*m/z* 111.0) (**Supplementary information**, **Figure S3.6.1**). Using this strategy, three other dipeptide analogs of pyro-F were identified – pyro-tryptophan (pyro-W), pyro-leucine (pyro-L) and pyro-isoleucine (pyro-I) (**Figure 3.2a**) and their structures were also elucidated using 1H, COSY and HMBC NMR experiments (**Supplementary information**, **Figure S3.6.3-6**) and confirmed with MS/MS. It is noteworthy that in the case of all four dipeptide compounds, pyroglutamic acid was conjugated to a hydrophobic amino acid, perhaps in keeping with a shared mechanism directing their synthesis.

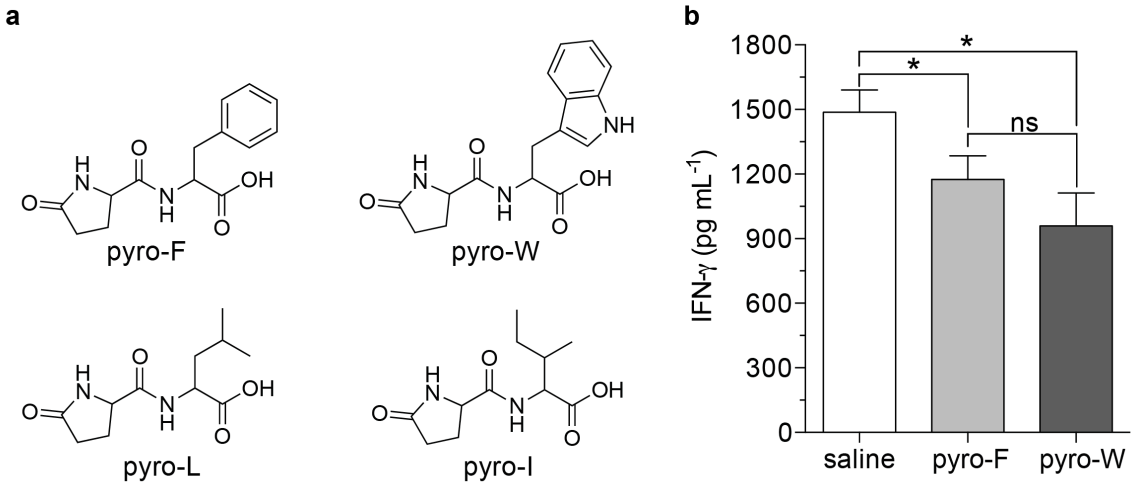


Figure 3.2. Chemical structures of pyro-dipeptides from *L. plantarum* and *in vivo* biological activity of pyro-F and pyro-W. (**a**) Chemical structures of pyro-F, pyro-W, pyro-L and pyro-I. (**b**) IFN-g concentration in supernatants of spleen cell cultures with addition of lipopolysaccharide derived from mice treated with pyro-F, pyro-W or saline. Data are presented as mean±s.e.m. *P<0.05 denotes significant differences between groups; ns, not significant.

To probe the immunomodulatory activity of the isolated compounds, two dominant members, pyro-F and pyro-W, were individually tested in vivo by intraperitoneal administration of 100 μg of pyro-F, pyro-W, or saline vehicle for three consecutive days. On day three, spleens were removed and spleen cells were cultured. After 24 hours of culture, spleen cells obtained from pyro-F and pyro-W-treated mice were found to produce significantly lower levels of IFN-γ in comparison to saline-treated controls (**Figure 3.2b**). Despite the structural difference between pyro-F and pyro-W, no significant difference in IFN-γ production was seen between these groups. Given the pro-inflammatory nature of this cytokine, down-regulation of IFN-γ in vivo indicates the anti-inflammatory potential of these compounds. In addition, these results are also in agreement with the previous studies, where the pronounced beneficial effects were noted increasingly in the stationary phase.[20]

Although a more in-depth metabolomic analysis of *L. plantarum* is required to gain full understanding of its beneficial effects, our data shed light on the effects of secreted metabolites *in vivo*. Newly identified pyroglutamic acid dipeptides substantiate the importance of studying the secreted metabolome of *L. plantarum* in greater detail. It is plausible to envision that there are other immunomodulatory agents produced by the Lactobacilli and other microorganisms that make-up their corresponding beneficial arsenals. Our findings reveal some interesting principles on one of the many ways a beneficial member from the human microbiome might communicate with human cells and could possibly lead to the downstream therapeutic strategies or mechanistic revelations.

3.5 Materials and Methods

3.5.1 General

NMR experiments were performed using a Bruker Avance 700 MHz spectrometer with all samples dissolved in methanol-d_4 99.8% (Sigma Aldrich, Oakville, ON, Canada). High resolution electrospray ionization mass spectrometry (HRESIMS) was conducted using a linear ion trap coupled to an Orbitrap mass spectrometer (LTQ/Orbitrap Thermo Fisher Scientific, Ottawa, ON, Canada). Samples for the multivariate analysis were analyzed using a Bruker micrOTOF II mass spectrometer (operating in positive ESI mode, with scanning range of *m/z* 100–1000) coupled with an Agilent 1200 series HPLC, using an Ascentis Express C18 column (150 mm × 2.1 mm, 2.7µ, Sigma Aldrich) with acetonitrile (0.1% formic acid) and water (0.1% formic acid) as the mobile phase with solvent flow at 0.25 mL min^{-1}. The crude extract of *L. plantarum* was fractionated using a Combiflash® Rf normal phase MPLC system (Teledyne Technologies Inc., Lincoln, NE, USA) with a 24 g RediSep® Rf silica Flash column. Final purification was carried out using a Bruker amaZon X ion trap mass spectrometer (operating in positive ESI mode, with scanning range of 100–1000 m/z) coupled with a Dionex UltiMate 3000 HPLC, using 250 × 15 mm Phenomenex Luna 5µ C18(2) 100Å column with acetonitrile (0.1% formic acid) and water (0.1% formic acid) as the mobile phase with a flow rate of 8 mL min^{-1}. The MS/MS experiments were carried out on the above LC-MS system operating in positive ESI auto MS/MS mode and with an Ascentis Express C18 column (150 mm × 4.6 mm, 2.7µ, Sigma Aldrich) and 1 mL min^{-1} flow rate. Multivariate analysis was performed on the resulting data sets using Bruker Daltonics Profile Analysis 2.1 software.

Pairwise comparisons between mid-log and stationary cultures using principal component analysis (PCA) and *t*-test models were performed using rectangular bucketing. Bucket parameters were defined as ΔRt = 0.05 min, with a kernel size of 0.008 min, and Δ*m/z*= 0.5 with a kernel size of *m/z* 0.150. Retention time and *m/z* ranges examined were between 5.0–40.0 min and *m/z* 150.0–1000.0, respectively.

3.5.2 Bacterial Material and Culture

Lactobacillus plantarum WCFS1 was kindly provided by Dr. Colin Ingham at the Wageningen University, Netherlands. Culturing was performed in de Man, Rogosa and Sharpe (MRS) medium (BD sciences) at 37°C without shaking. Cultures were grown to either mid-log (OD_{600} = 1.0) or stationary phase (overnight growth). For metabolomic analysis, 50 mL cultures of *L. plantarum* were grown to either mid-log (OD_{600}=1; 6 hours) or stationary phase (until OD_{600} stabilized at 1.8; 48 hours).

3.5.3 Extraction and Analysis

L. plantarum stationary phase cultures (n = 8) were extracted using liquid-liquid partitioning (ethyl acetate 2:1, twice) with the resulting organic phase concentrated *in vacuo*. Dried material was reconstituted in 100% methanol. Optimal separation for multivariate analysis samples was achieved using a gradient of 5% acetonitrile for 5 min, ramping to 100% acetonitrile by 45 min, holding at 100% for 10 min, then returning to 5% acetonitrile by 60 min, and re-equilibrating at 5% acetonitrile until 65 min. Pyroglutamic dipeptides eluted after 15 minutes. Optimal separation for MS/MS analysis samples was achieved using 5% acetonitrile for 5 min, ramping to 60% acetonitrile by 23 min, and reaching 100% acetonitrile by 25 min, holding at 100% till 28 min and then

returning to 5% acetonitrile by 30 min, and re-equilibrating at 5% acetonitrile until 35 min. Pyroglutamic dipeptides eluted after 14 min.

3.5.4 Extraction and Isolation

To isolate compounds identified by the multivariate analysis, 6 × 1 L fermentations of wild-type *L. plantarum* were grown to stationary phase and processed by liquid-liquid extraction (ethyl acetate 2:1, twice). The organic phase was concentrated *in vacuo*, loaded directly on to the Flash column and the following program of solvent A (hexanes) and B (ethyl acetate) was employed: 0–5 min, 0% B; 5–12 min, 25% B; 12–22 min, 50% B; 22–30 100% B. This was followed by a dichloromethane (solvent A)/methanol (solvent B) program as follows: 0–7 min, 0% B; 7–15 min, 25% B; 15–22 min 50% B; 22–30 min 75% B; 30–40 min, 100% B. A total of 9 fractions were generated and the 50% methanol/dichloromethane fraction was further interrogated by the LC-MS. Pure compounds were isolated using the following gradient: 0–5 min, 5% B; 5–23 min ramping to 45% B; 23–25 min, ramping to 100% B; 25–29 min, holding at 100% B; 29–30.5 min, returning to 5% B; and 30.5–35 min, 5% B.

3.5.5 Animals

Male (8–10 week old) C57BL/6 mice (Taconic) were kept in sterilized, filter-topped cages under specific pathogen-free conditions and fed autoclaved food. All experiments were approved by the animal ethics committee of McMaster University and conducted under the Canadian guidelines for animal research.

3.5.6 Cytokine Assay

Male C57BL/6 (Taconic, Cambridge City, IN, USA) mice were injected intra-peritoneally daily with 100 μg of either pyro-F or pyro-W for 3 days (n = 7). Controls received saline only. Spleens were removed from mice and collected in Hank's balanced salt solution containing antibiotics. A single cell suspension was prepared by straining spleens through a 100-μm nylon mesh filter (BD Falcon, Mississauga, ON, Canada). The resulting cell suspension was centrifuged and the cell pellet was re-suspended in lysis medium to remove red blood cells, washed with Hank's solution, and centrifuged (10 minutes at 453×g). The supernatant was removed and the remaining cell pellet was re-suspended in RPMI 1640 supplemented with 10% fetal bovine serum, 5 mM L-glutamine, 100 U mL^{-1} penicillin/streptomycin, 100 μg mL^{-1} streptomycin, and 50 mM 2-mercaptoethanol (Invitrogen Life Technologies, Burlington, ON, Canada). Cells were placed in 24-well plates at a concentration of 1×10^6 cells per well. Lipopolysaccharide (LPS, Sigma-Aldrich) was added to a final concentration of 100 ng mL^{-1}. Cells were placed in a 37°C humidified incubator at 5% CO_2. After 24 hours, the supernatant was collected, centrifuged to remove any cells and IFN-γ levels were measured using a commercially available ELISA kit according to the manufacturer's instructions (Quantikine Murine, R&D Systems, Minneapolis, MN).

3.5.7 Statistical analysis

All data are presented as mean ± s.e.m. Unpaired *t*-test (for parametric data) or Mann-Whitney *U*-test (for non-parametric data) was performed using GraphPad Prism

version 6.0b for Mac OS X (GraphPad Software, La Jolla, CA). An associated p value <0.05 was considered statistically significant.

3.6 Supplementary Figures and Legends (Figure S3.6.1-6)

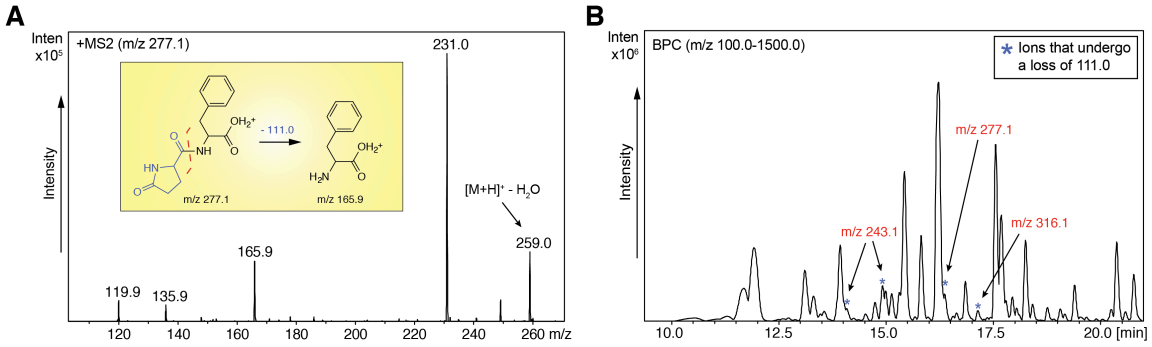


Figure S3.6.1. Identification of analogs of pyro-F using tandem mass spectrometry. (**A**): the fragmentation profile of 1 (m/z 277.1) and a diagnostic loss of 111.0 corresponding to the loss of pyroglutamic acid. (**B**): the base peak chromatogram showing the compounds that undergo the loss of 111.0 upon fragmentation; m/z 277.1 – pyro-F, m/z 316.1 – pyro-W, m/z 243.1 – pyro-L and pyro-I.

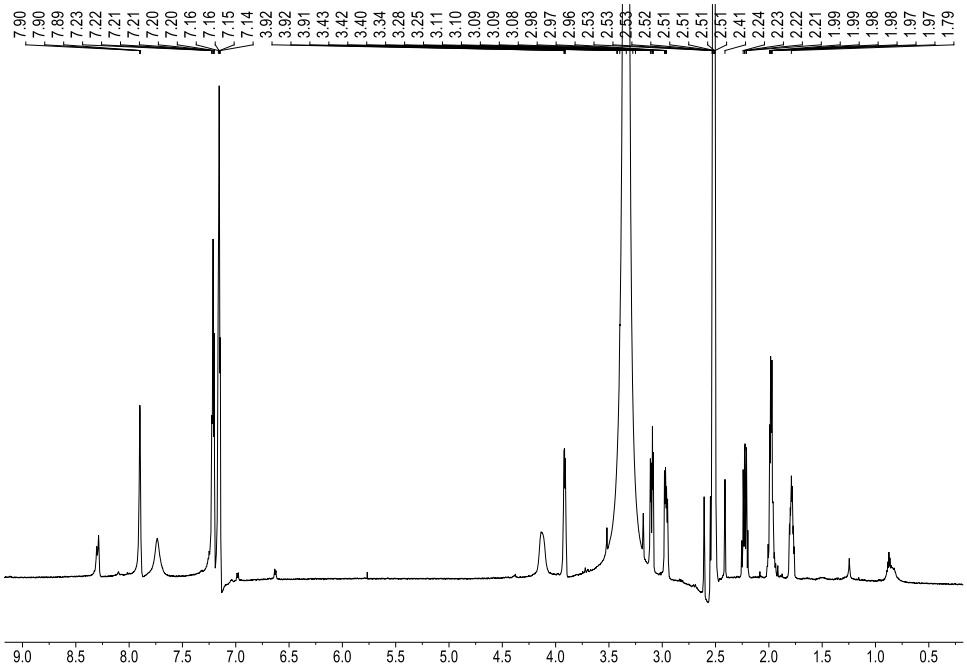


Figure S3.6.2. ^{1}H NMR (700 MHz, CD_3OD) spectrum of pyroglutamic acid phenylalanine (pyro-F).

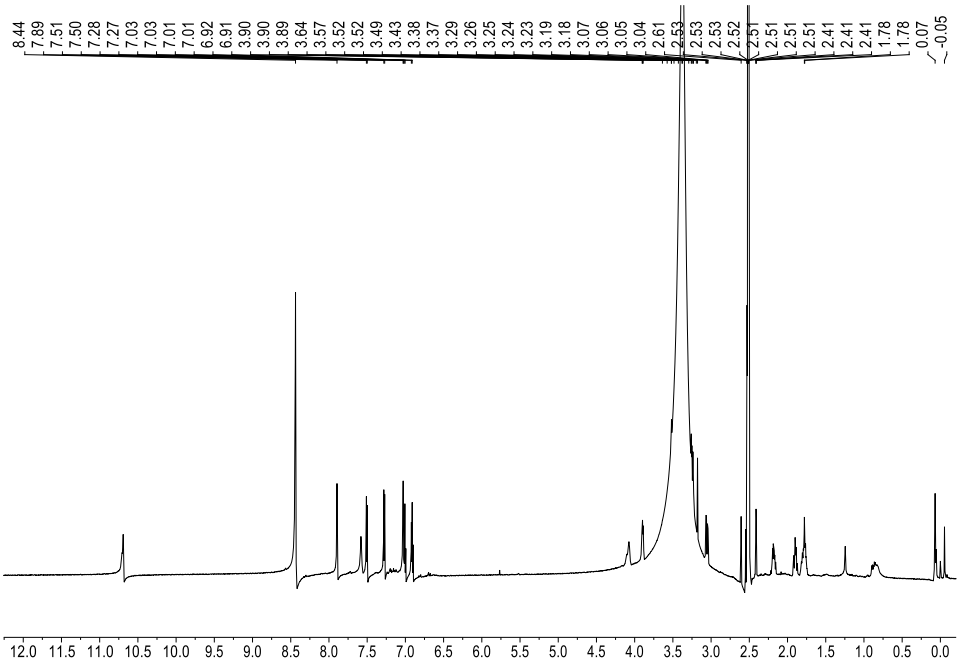


Figure S3.6.3. ^{1}H NMR (700 MHz, CD_3OD) spectrum of pyroglutamic acid tryptophan (pyro-W).

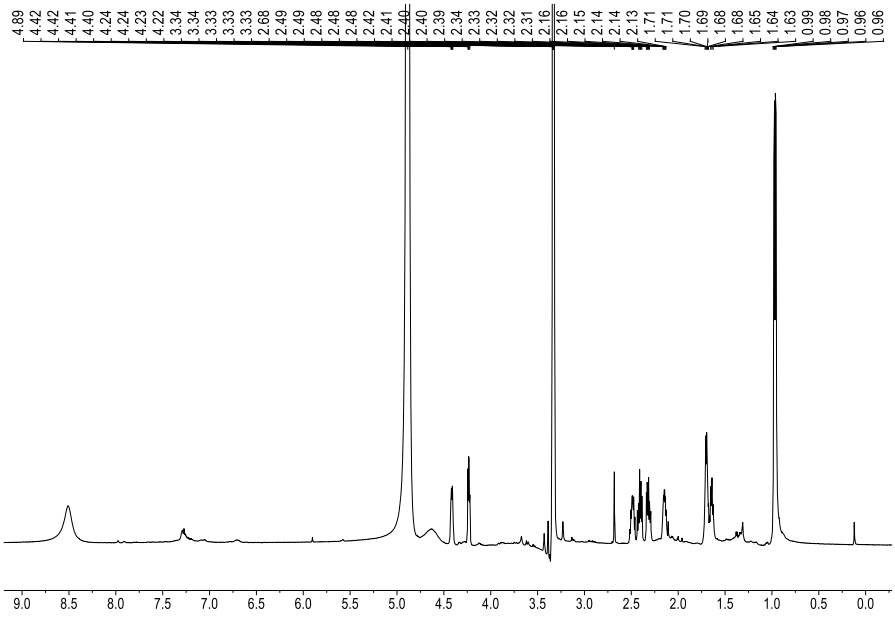


Figure S3.6.4. ^{1}H NMR (700 MHz, CD_3OD) spectrum of pyroglutamic acid leucine (pyro-L).

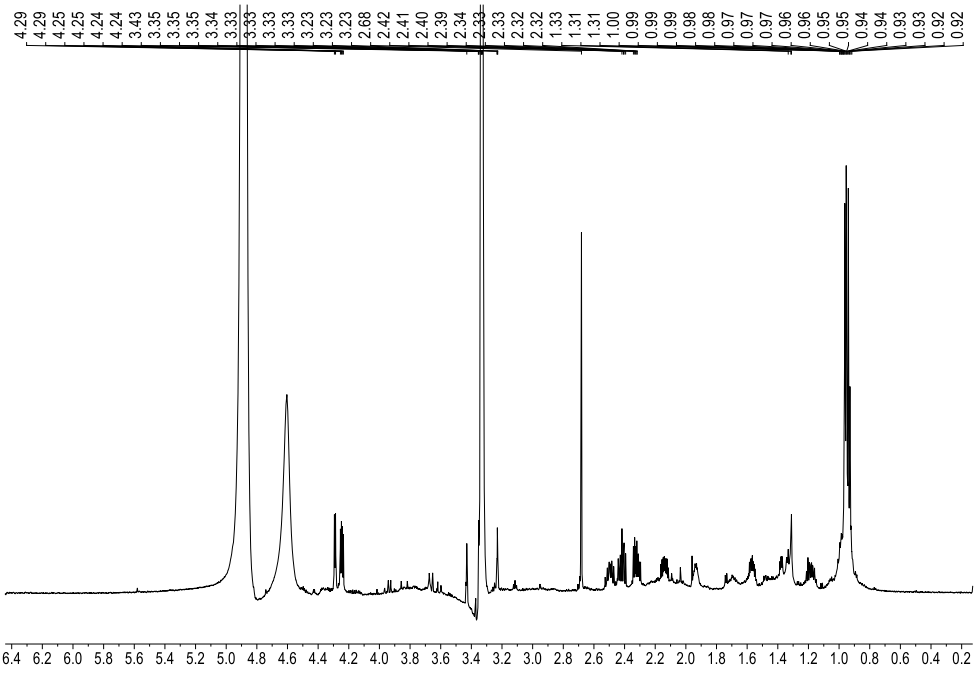


Figure S3.6.5. ^{1}H NMR (700 MHz, CD_3OD) spectrum of pyroglutamic acid isoleucine (pyro-I).

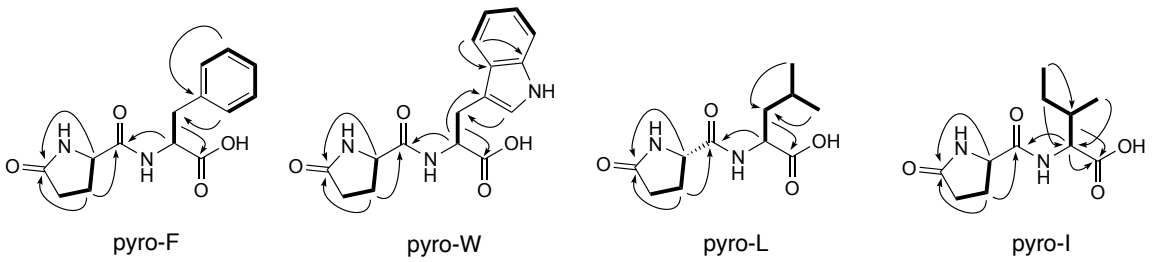


Figure S3.6.6. COSY (^{1}H-^{1}H, thick lines) and HMBC (^{1}H-^{13}C, arrows) correlations assigned for pyro-F, pyro-W, pyro-L and pyro-I.

3.7 References

1. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature.* **449**, 804–810 (2007).

2. Cho, I. & Blaser, M. J. Applications of Next-Generation Sequencing: The human microbiome: at the interface of health and disease. *Nature.* **13**, 260–270 (2012).

3. Zhu, B., Wang, X. & Li, L. Human gut microbiome: the second genome of human body. *Protein Cell.* **1**, 718–725 (2010).

4. Petrof, E. O. Probiotics and Gastrointestinal Disease: Clinical Evidence and Basic Science. *Antiinflamm Antiallergy Agents Med Chem.* **8**, 260–269 (2009).

5. Seksik, P., Dray, X., Sokol, H. & Marteau, P. Is there any place for alimentary probiotics, prebiotics or synbiotics, for patients with inflammatory bowel disease? *Mol. Nutr. Food Res.* **52**, 906–912 (2008).

6. Jonkers, D., Penders, J., Masclee, A. & Pierik, M. Probiotics in the Management of Inflammatory Bowel Disease. *Drugs.* **72**, 803–823 (2012).

7. Schultz, M. *et al.* Lactobacillus plantarum 299V in the treatment and prevention of spontaneous colitis in interleukin-10-deficient mice. *Inflammatory Bowel Diseases.* **8**, 71–80 (2002).

8. Dharmani, P., De Simone, C. & Chadee, K. The Probiotic Mixture VSL# 3 Accelerates Gastric Ulcer Healing by Stimulating Vascular Endothelial Growth Factor. *PLoS ONE.* **8**, e58671 (2013).

9. Dai, C. *et al.* VSL#3 probiotics exerts the anti-inflammatory activity via PI3k/Akt and NF-κB pathway in rat model of DSS-induced colitis. *Mol. Cell. Biochem.* **374**, 1–11 (2013).

10. Venturi, A. *et al.* Impact on the composition of the faecal ora by a new probiotic preparation: preliminary data on maintenance treatment of patients with ulcerative colitis. *Aliment Pharmacol Ther.* **13**, 1103–1108 (1999).

11. Bassaganya-Riera, J., Viladomiu, M., Pedragosa, M., De Simone, C. & Hontecillas, R. Immunoregulatory mechanisms underlying prevention of colitis-associated colorectal cancer by probiotic bacteria. *PLoS ONE.* **7**, e34676 (2012).

12. Petrof, E. O. *et al.* Bacteria-free solution derived from Lactobacillus plantarum inhibits multiple NF-kappaB pathways and inhibits proteasome function. *Inflammatory Bowel Diseases.* **15**, 1537–1547 (2009).

13. Todorov, S. D. Bacteriocins from Lactobacillus plantarum production, genetic organization and mode of action. *Braz. J. Microbiol.* **40**, 209–221 (2009).

14. Twomey, D., Ross, R. P., Ryan, M., Meaney, B. & Hill, C. Lantibiotics produced by lactic acid bacteria: structure, function and applications. *Anton. Leeuw.* **82**, 165–185 (2002).

15. Martinez, R. C. R. *et al.* Biochemical, antimicrobial and molecular characterization of a noncytotoxic bacteriocin produced by Lactobacillus plantarum ST71KS. *Food Microbiol.* **34**, 376–381 (2013).

16. Holo, H., Jeknic, Z., Daeschel, M., Stevanovic, S. & Nes, I. F. Plantaricin W from Lactobacillus plantarum belongs to a new family of two-peptide lantibiotics. *Microbiology.* **147**, 643–651 (2001).

17. Jiménez-Díaz, R., Rios-Sanchez, R. M., Desmazeaud, M., Ruiz-Barba, J. L. & Piard, J.-C. Plantaricins S and T, two new bacteriocins produced by Lactobacillus plantarum

LPCO10 isolated from a green olive fermentation. *Appl. Environ. Microbiol.* **59**, 1416–1424 (1993).

18. Flynn, S. *et al.* Characterization of the genetic locus responsible for the production of ABP-118, a novel bacteriocin produced by the probiotic bacterium Lactobacillus salivarius subsp. salivarius UCC118. *Microbiology.* **148**, 973–984 (2002).

19. Sturme, M. H. *et al.* An agr-like two-component regulatory system in Lactobacillus plantarum is involved in production of a novel cyclic peptide and regulation of adherence. *J. Bacteriol.* **187**, 5224–5235 (2005).

20. van Baarlen, P. *et al.* Differential NF- B pathways induction by Lactobacillus plantarum in the duodenum of healthy humans correlating with immune tolerance. *Proc. Natl. Acad. Sci. USA.* **106**, 2371–2376 (2009).

21. Narumi, S., Tebo, J. M., Finke, J. H. & Hamilton, T. A. IFN-gamma and IL-2 cooperatively activate NF kappa B in murine peritoneal macrophages. *J. Immunol.* **149**, 529–534 (1992).

22. Yasumoto, K. *et al.* Tumor necrosis factor alpha and interferon gamma synergistically induce interleukin 8 production in a human gastric cancer cell line through acting concurrently on AP-1 and NF-kB-like binding sites of the interleukin 8 gene. *J. Biol. Chem.* **267**, 22506–22511 (1992).

23. Paludan, S. R. Synergistic action of pro-inflammatory agents: cellular and molecular aspects. *J. Leukoc. Biol.* **67**, 18–25 (2000).

24. Al-Qadiri, H. M. *et al.* Studying of the bacterial growth phases using Fourier transform infrared spectroscopy and multivariate analysis. J. *Rapid Meth. Autom.*

Microbiol. **16**, 73–89 (2008).

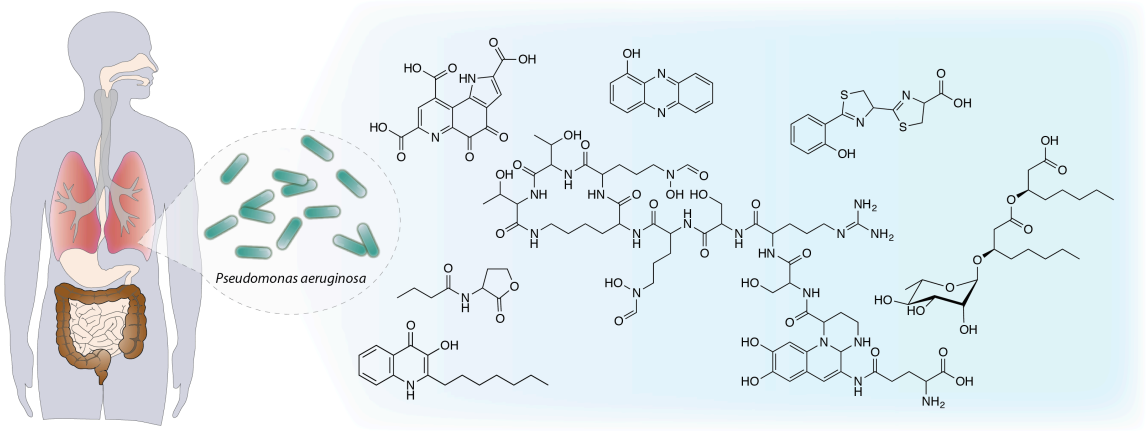
25. Suzuki, Y., Motoi, H. & Sato, K. Quantitative analysis of pyroglutamic acid in peptides. *J. Agric. Food Chem.* **47**, 3248–3251 (1999).

26. Kumar, A. & Bachhawat, A. K. Pyroglutamic acid: throwing light on a lightly studied metabolite. *Curr. Sci.* **102**, 288 (2012).

27. Mucchetti, G. *et al.* Pyroglutamic acid in cheese: presence, origin, and correlation with ripening time of Grana Padano cheese. *J. Dairy Sci.* 83, 659–665 (2000).

Chapter 4. Metabolomic Interrogation of Pseudomonad Bioactive Small Molecules

4.1 Chapter Preface



Pseudomonas aeruginosa is a noted pathogenic member of the human microbiome that is implicated in an increased morbidity and mortality rates in cystic fibrosis patients. It is also one of the most metabolically versatile bacterial organisms known to date. Although the secondary metabolome of Pseudomonads has been studied for many decades, the roles of these agents in the context of interaction with human cells are not clearly understood. There is some evidence, however, pointing at the immunomodulatory potential of some compound classes. More specifically, quinolones and acyl homoserine lactones, which are molecules deeply integrated in the cell-to-cell signaling, known as quorum sensing have been shown to affect cytokine secretion in a number of immune cells. This is significant, however, Pseudomonads craft a plethora of other chemicals, some of which are encoded in their genomes, and whose activity has not been assessed. A number of cryptic agents also exist, production of which is implied from the genome, but no products have been identified.

To gain a better understanding of the immunomodulatory effects of secreted metabolites from *P. aeruginosa*, I set out to deeply interrogate its metabolome. Using mass spectrometry techniques such as multiple reaction monitoring I was able to create a small library of quinolones and test their immunomodulatory activity in macrophages. Moreover, having access to various clinical isolates of *P. aeruginosa*, I was able to compare levels of quinolone production and investigate other metabolomic differences that may exist between these isolates and the model lab strains. This comparison yielded in the identification of a novel succinylated tetrapeptide, whose immunomodulatory potential was also assessed. Lastly, guided by the newly designed Genomes to Natural Products (GNP) platform, I was able to identify the cryptic nonribosomal peptide product from *P. aeruginosa* and validate its biosynthetic origin.

The following chapter is formatted as a manuscript that is prepared for submission for publication, in which I am the lead author. I conducted the experimental design, performed all experiments (except for the ones stated below), interpreted results and wrote the manuscript. Dr. Xiang Li elucidated the chemical structures of quinolones and LPPI-S. Michael Skinnider assisted with the GNP analysis. Janice J. Kim assisted with the cytokine profiling of quinolones and LPPI-S. Dr. Nathan Magarvey provided ideas and guidance for the progression of this research.

4.2 Abstract

Metabolomic investigations of microbial populations serve as useful tools for delineation of key features of taxonomic groups, biological processes and important biomarkers. Pseudomonads are a diverse metabolic and taxonomic group noted for their both catabolic and anabolic pathways, as well as their secreted metabolites. The complexity of their secreted secondary metabolomes is central to them being a significant source of bioactive small molecules for treating diseases, development of biocontrol agents, and also control and coordination of genetic expression (i.e. quorum sensing). In this research work we provide a systematic metabolomic investigation of the Pseudomonads, leading to defined routes of rapid identification of specific chemical groups that are known (i.e. quinolones) and ones that are unknown, but inferred from genomic sequences or phenotypic disease-differentiation in the case of transmissible strains associated with cystic fibrosis. From these investigations we have produced an approach to rapidly obtain, define and test specific groups for their biological activity assessment, and created metabolomic comparative frameworks to identify new peptides associated with select strains of *Pseudomonas* that may serve as biomarkers of infection. Lastly, these investigations led to the identification of a cryptic metabolite that is uniquely contained within the strains of *Pseudomonas aeruginosa*.

4.3 Introduction

The Pseudomonad family of Gram-negative organisms is metabolically versatile and inhabits a wide range of ecological niches – both environmental and human[1]. Diverse metabolism with specialized endproducts enables these organisms to populate diverse

niches, survive harsh environments, and coordinate their own gene expression and inter-kingdom signaling. Within a human context, *Pseudomonas aeruginosa* crafts small molecules that control its own virulence, and others implicated in modulation of the host's immune functions[2]. The collective of Pseudomonad small molecule chemistry can be divided along structural and functional classifications as lipopeptide surfactants and antibiotics (e.g. viscosins and massetolides)[3,4], chelating agents (e.g. pyoverdines, pyochelins and phenazines)[5-7], virulence-inducing small molecules such as rhamnolipids and quinolones[8,9], and the canonical quorum sensing acyl homoserine lactones[10]. Coordinated behaviour and gene expression within Pseudomonad strains is often a role fitted towards specific small molecule families, particularly those with acyl appendages including the acyl-homoserine lactones and quinolones[11]. The Pseudomonas quinolone signal (PQS) and 4-hydroxy-2-heptylquinoline (HHQ) also impact host gene expression and down-regulate the secretion of IL-6 and TNF-α in mice[12]. Interestingly however, *P. aeruginosa* produces a variety of other PQS-like compounds, but these variant PQS molecules have not been exclusively investigated for their immune system perturbations[13]. In many other instances one can envision that other metabolite variants exist and likely that entirely new classes also await discovery.

Metabolomic investigations of bacterial phyla are further uncovering the unassigned members of chemical groups and defining the chemical differences between strains. In many cases, it is appreciated that these metabolomic differences between strains serve as important biomarkers or factors for biological and niche adaptations and well reflect virulence phenotypes of pathogens. Cataloging the known natural product

chemical space of Pseudomonad metabolites would serve to assess the chemical similarity, distribution and diversity of compounds. Variations of chemical scaffolds is seemingly common in Pseudomonads, as may be inferred from investigations involving >700 natural product small molecules contained within the Dictionary of Natural Products and other literature-based chemical descriptor catalogs. As such, by comparing the entire known Pseudomonad metabolome, one may observe chemical groupings and the diversity within each of the above-mentioned chemical architectures (**Figure 4.3.1**).

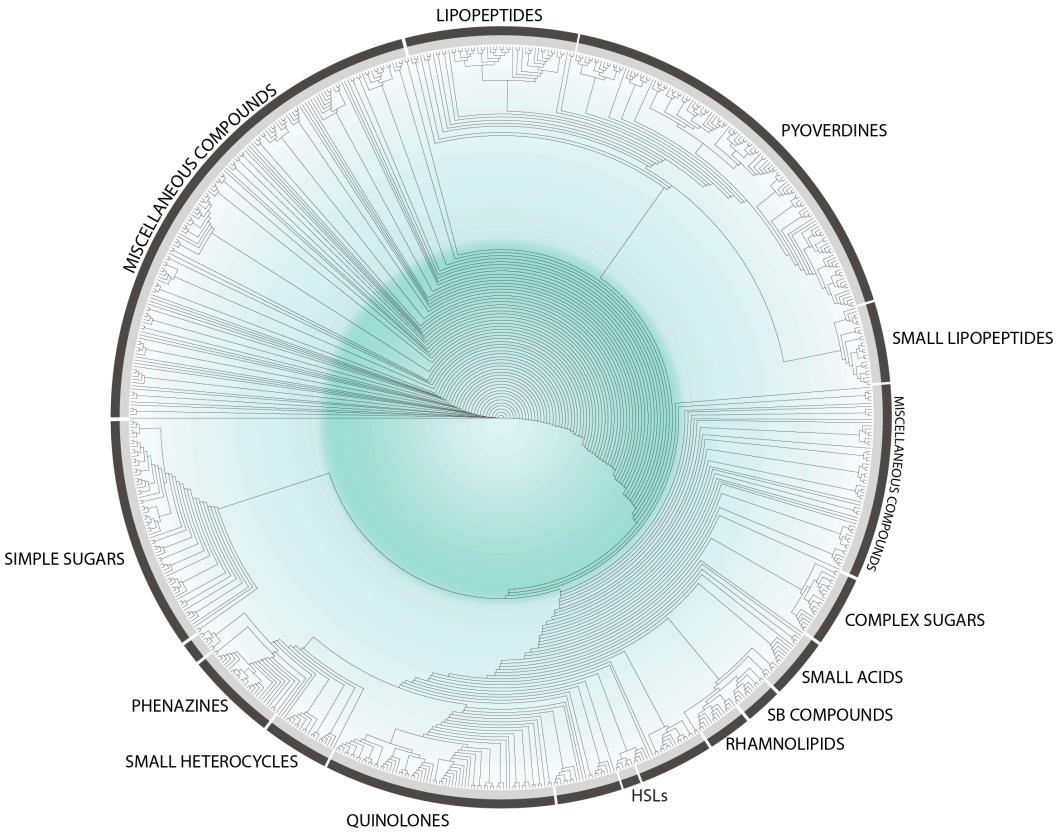


Figure 4.3.1. Circular cladogram of the known secreted metabolome of Pseudomonads. The SMILES codes of 725 Pseudomonad chemical compounds were compared using hierarchical clustering with distance matrix setting.

Varied metabolism of Pseudomonads creates alternative endproducts of categories of molecules previously identified (i.e. phenazines and quinolones) and other endproducts that have yet to be structurally elucidated. A combination of metabolomic analyses of microorganisms and the secreted metabolites found within spent media have suggested that there are many unknown products created by the Pseudomonads and other organisms[14,15]. Next generation sequencing of large cohorts of Pseudomonads including numerous human-associated *P. aeruginosa* strains, environmental organisms associated with plants such as *P. syringae*, and others found in soil such as *P. putida* have led to a defined view of the genomic capability of these diverse organisms. Facilitating the identification of new microbial metabolites is the clustering of genes (for polyketide and nonribosomal peptide assembly) within genomes whose products are not known. Polyketides and nonribosomal peptides are synthesized by the multimodular enzymatic assembly lines, known as polyketide synthases (PKs) and nonribosomal synthetases (NRPSs), respectively. More importantly, the substrate affinities of PKS and NRPS domains can be predicted, based on the conserved amino acid sequences of the acyltransferase and adenylation domains, respectively. Consequently, the chemical make-up of biosynthetic products can be predicted.

Recently we created a platform that would allow investigation of these different metabolic clusters within microbes by analyzing their genome sequences, called the Genomes to Natural Products (GNP) program. The genome search mode in GNP can accept whole genomes, condensed DNA sequences or translated protein sequences. In whole genome searches, putative biosynthetic genes are located by first identifying long,

non-overlapping open reading frames, then using these sequences to construct an interpolated context model of coding sequences[16]. This model is then used to analyze the genome and make gene predictions. Following domain prediction, the predicted natural product is output as a SMILES string. As mentioned previously, the variants of Pseudomonads frequently create chemical variety from established clusters. The ability to detect new clusters will certainly facilitate metabolomic investigation to characterize existing metabolite groups and variations that are created within different strains. This will subsequently assist the identification of new chemicals that may be important in the biology and life cycle of Pseudomonad strains. With the emergence of new strains of *P. aeruginosa* that are transmissible between cystic fibrosis patients and little genomic and phenotypic clues available for their identification, unique metabolites or the metabolite profiles of known compounds might offer some insight[17,18]. In this work we present a detailed investigation of the Pseudomonad chemistry and establish metabolomic strategy to deeply investigate known metabolite groups, define divergent metabolites between different, but related Pseudomonads, and to identify new metabolites as the result of the GNP platform, leading to the identification of new Pseudomonad chemistries.

4.4 Results and Discussion

Rapid classification of individual unknown molecules within complex microbial metabolomes can be facilitated by comparing their spectral features to known metabolite groups. Likewise, comprehensive profiling of specific chemical groups can be achieved through comparative analysis and can assist with their quick isolation and biological testing. The quinolones present an intriguing case, as their profile changes between

isolates and within differing growth situations and stress conditions[19]. There are two main structural elements that exist within this family – the cyclic core and a tail, and chemical differences exist within each element. These distinctions can also be manifested in the fragmentation patterns of these chemicals when subjected to collision-induced dissociation using mass spectrometry experiments (**Figure 4.4.1A**). Recognizing diagnostic fragments that are generated upon their fragmentation (i.e. *m/z* 159 or 175 corresponding to the remaining core) allows for the development of an efficient multiple reaction monitoring (MRM). Use of MRM is increasingly used for parallel and rapid profiling of chemicals within complex samples[20-25]. Efficient identification of chemical families promotes their selective investigation and is useful in direct monitoring of them within cultures, and within infections as biomarkers of *Pseudomonas*. Pseudomonad quinolones, PQS and HHQ are selectively produced by this group and we thus sought to utilize an MRM screen for this chemical family[26,27]. Implementation of MRM allowed for an appreciation of diversity, as well as abundance of 9 identified quinolones produced by *P. aeruginosa* (**Figure 4.4.1B**). This rapid identification allowed for their isolation (**Figure 4.4.1C**), making it possible to test them as a chemical family and examine their immunomodulatory potential by measuring cytokine expression with cultured macrophages.

One of the key mediators of the host immune system is the family of interleukin-1 (IL-1) cytokines, produced largely by macrophages and monocytes[28]. Through the action of several of its members, including well-studied IL-1α and IL-1β, IL-1 family of cytokines is responsible for a suite of early immune responses to inflammation and tissue

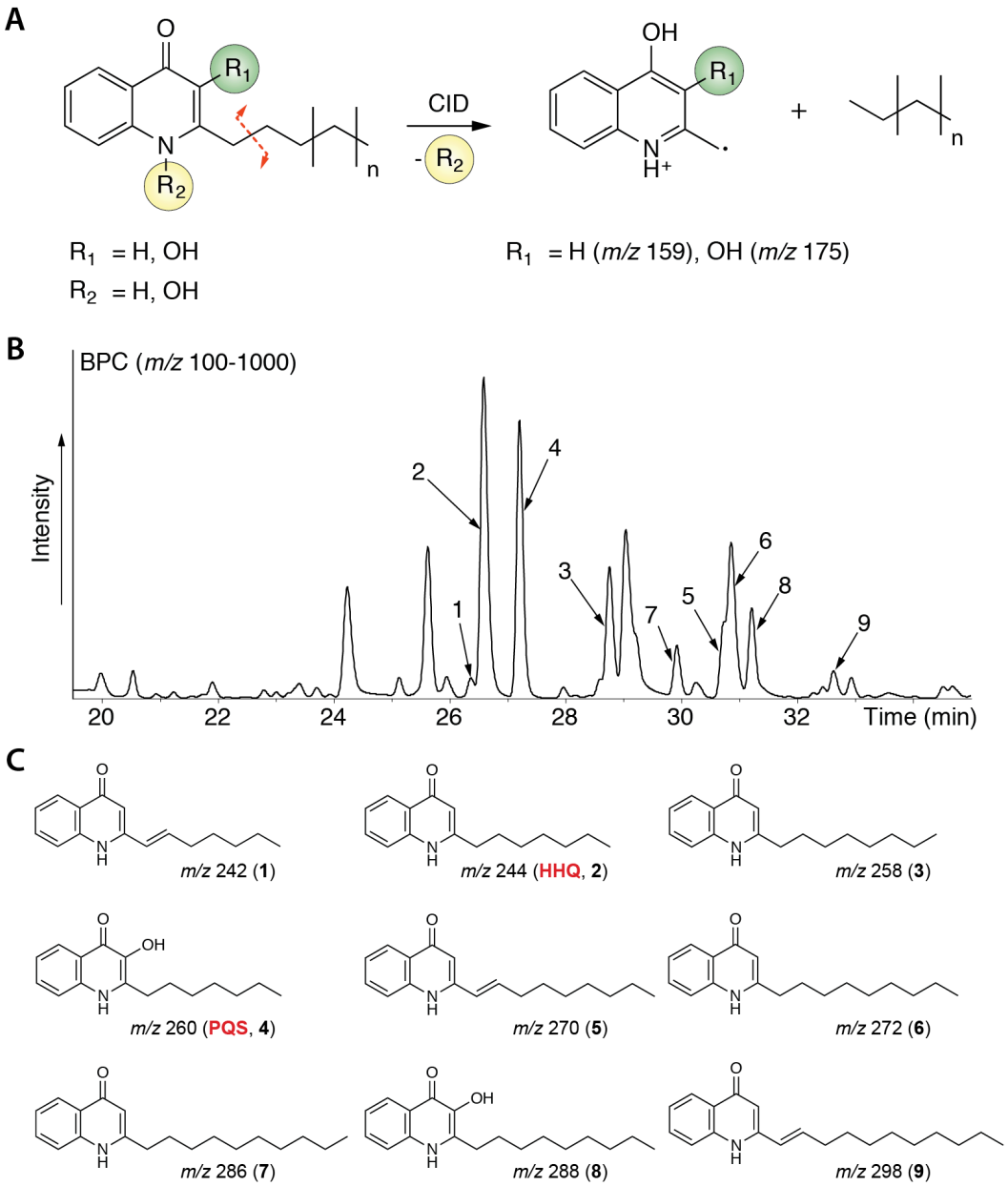


Figure 4.4.1. Targeted identification of quinolones from *Pseudomonas aeruginosa* PAO1. **A** – MRM experiment was performed on the sample of extracted culture of *P. aeruginosa* PAO1, with *m/z* 159 and 175 as diagnostic fragments to identify quinolones. **B** – base peak chromatogram with nine identified quinolones indicated. **C** – chemical structures of the isolated quinolones.

injury. Interestingly, their secretion is largely induced by the interaction with the lipopolysaccharide (LPS) component of bacterial cell walls[29]. Inhibiting the secretion of IL-1 cytokines would reverse the immune response and thus could be exploited by *P. aeruginosa*. Monitoring the impacts of quinolones on IL-1 cytokine production showed that they result in a significant decrease in both IL-1α and IL-1β secretion in LPS-induced macrophages, especially *m/z* 242 and 272 (**Figure 4.4.2**). Another potent modulator of the immune system is interlukin-6 (IL-6), and has been shown to be directly involved in the differentiation of monocytes to macrophages[30]. Likewise, testing of quinolones of IL-6 secretion levels on RAW264.7 cell line revealed an inhibitory effect of quinolones and notably, those with the shortest acyl tail such as *m/z* 242, 244 and 258. The monocyte chemoattractant protein-1 (MCP-1) is another cytokine that is considered as a potent chemotactic factor that regulates both migration and infiltration of macrophages and monocytes[31]. The secretion of MCP-1 was also decreased in quinolone-treated macrophages, with *m/z* 242 having the most pronounced effect and a general trend of decreasing the acyl tail length correlating with increased activity. The granulocyte colony-stimulating factor (G-CSF) is another key cytokine secreted by the endothelial cells, macrophages and fibroblasts to stimulate the proliferation and differentiation of myeloid progenitor cells towards neutrophilic granulocytes was also monitored[32]. From the cytokines examined GM-CSF showed the most marked decrease with *m/z* 242, 272 and 288 quinolones as most active. Lastly, tumor necrosis factor-α (TNF-α) is one of the most critical cytokines in immune system regulation and is responsible for controlling a number of cytokine releasing cascades in numerous inflammatory diseases[33].

Interestingly, macrophages are the major producers of TNF-α and respond to it with high sensitivity. Therefore, its secretion modulation is critical in controlling the immune responses. The effects of quinolones on TNF-α release were most significant in the case of *m/z* 242 and 244, as well as *m/z* 272 and 286. It is worth mentioning, however, that there is a strong interplay between the immune cells and their secreted cytokines, which complicates the process of pinpointing the exact mechanism of quinolone activity. Nonetheless, when considered together, the administration of quinolones had an overall immunosuppressive effect by decreasing the release of a number of pro-inflammatory cytokines in macrophages.

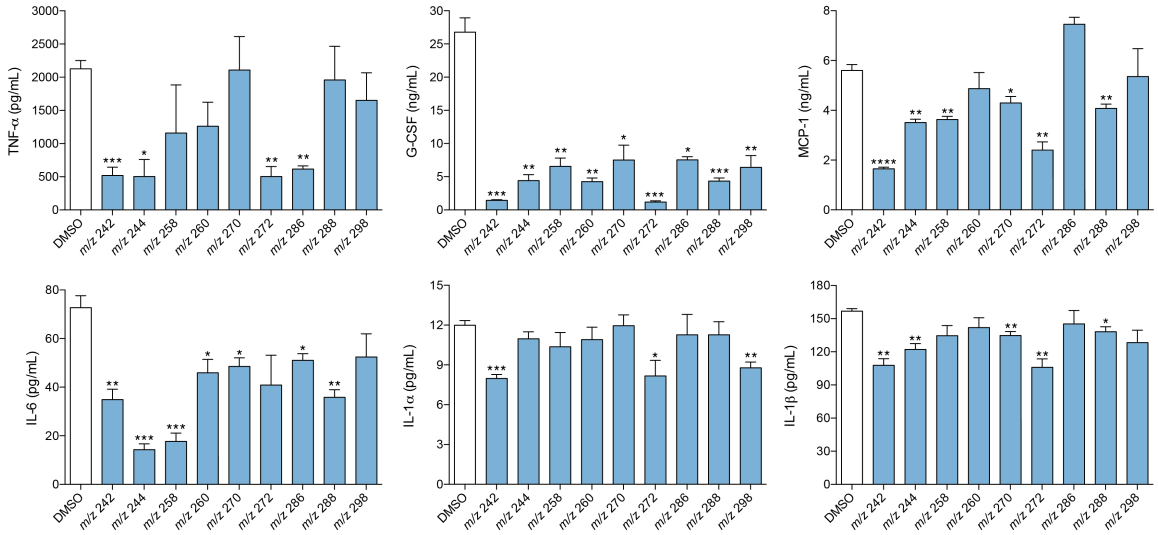


Figure 4.4.2. Levels of cytokines secreted by RAW264.7 macrophages upon treatment with quinolones. RAW264.7 macrophages stimulated with LPS were treated with nine isolated quinolones to test their ability to affect cytokine secretion. Data are presented as mean ± SEM, asterisks (*) indicate significant differences between treatments (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

Having created a strategy that is efficient and effective to investigate the impacts of these identified quinolones, we then sought to investigate the variations that may exist within this chemical family. To explore this, we investigated a number of *P. aeruginosa* strains, including the model strains such as PAO1, and those studied for their specific virulence factor production PA7 and PA14. *Pseudomonas aeruginosa* PA14 strain is known as a hyper virulent strain of *P. aeruginosa* due to its genome-encoded pathogenicity islands and high virulence factor production[34,35]. Lastly, unlike the other two, the PA7 strain of *P. aeruginosa* is a non-respiratory clinical isolate and considered as a taxonomic outlier and the least virulent opportunistic human pathogen[36]. In addition, we have been able to acquire a number of clinical isolates, including isolates from different time scales of infection from the lung sputum of a single patient, and also strains that have been acquired from clinical samples of people infected with the transmissible PES variant of *P. aeruginosa*[37]. We sought to investigate the variations of quinolones as the entirety, as well as their different abundance profiles. Appreciating the activity of distinct quinolones, we applied our metabolomic profiling to the spent cultures of each of these organisms. From this, we were able to create a heatmap of both global and relative abundances for each of these strains (**Figure 4.4.3**). We noted that compared to the other strains, PA7 strain produced very little quinolones, which might also be correlative with its lower virulence. Interestingly, the production of quinolones by the P48 strain was substantially higher, when compared to the genetically identical P590 strain that was isolated from the patient ten years later. At a global level, the overall production of *m/z* 272 and *m/z* 260 quinolones were the most abundant across all strains (**Figure 4.4.3A**).

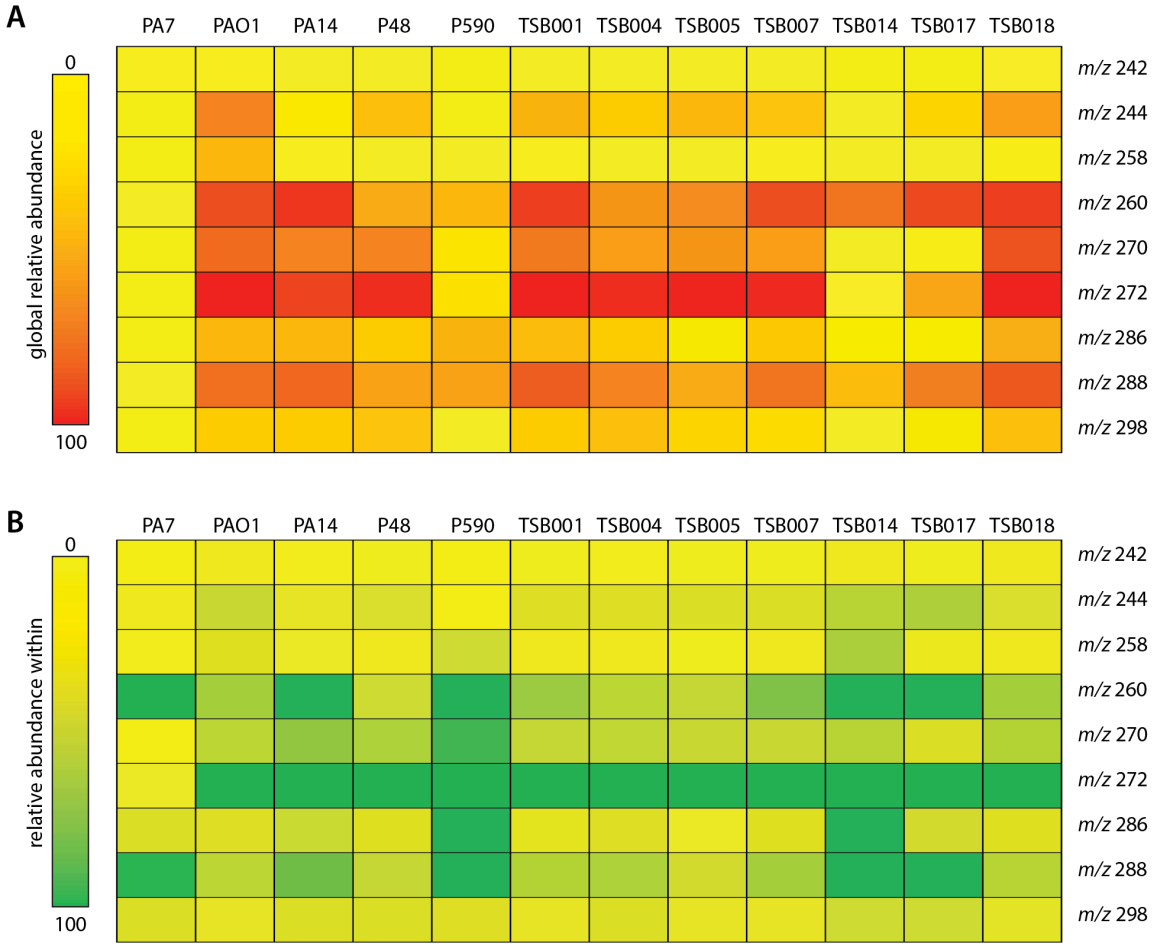


Figure 4.4.3. Heatmaps of global and relative abundances of quinolones. **A** – global relative abundance of isolated quinolones relative to the most abundant quinolone among tested strains. **B** – relative abundance of isolated quinolones within each strain relative to the most abundant quinolone produced within that strain. PA7, PAO1 and PA14 are the model clinical isolates; P590 and P48 are the genetically identical clinical strains isolated ten years apart, with the latter isolated last; TSB001-018 are the clinical isolates from a single sputum sample collected from a patient with cystic fibrosis.

Similar relative abundances of *m/z* 272 and *m/z* 260 were observed within the individual isolates as well **(Figure 4.4.3B)**. In addition to these two, *m/z* 288 was also notable within all isolates and no distinct quinolone profiles variance was observed between non-transmissible and their transmissible counterparts. We sought to investigate if any other

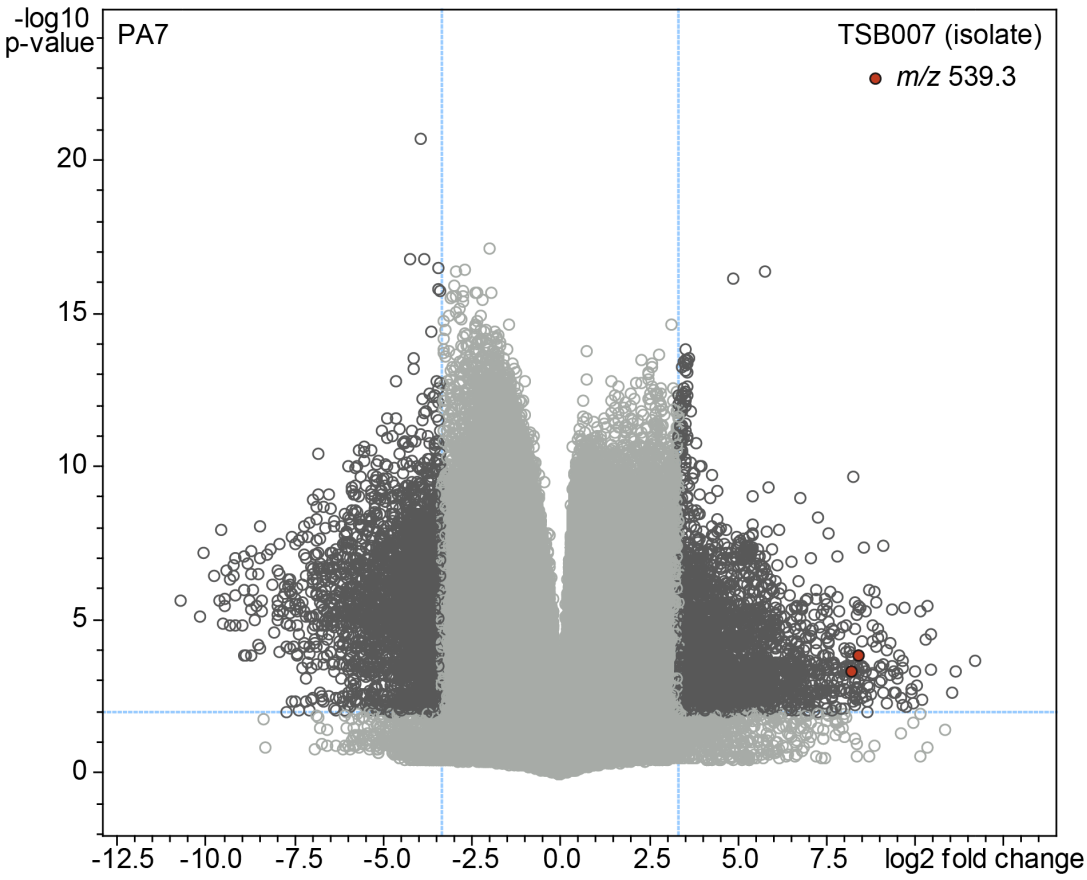


Figure 4.4.4. Volcano plot displaying the unique molecular features in TSB007 clinical isolate versus the model PA7 strain. The metabolite denoted by the red circle represents the unique, unidentified metabolite that is produced exclusively by the clinical isolate and not PA7 strain. Fold change and *p*-value cut-offs represented by the blue vertical and horizontal dash lines are 2 and 0.05, respectively.

metabolite features differentiated transmissible isolates from the non-transmissible varieties. A comparative metabolomic analysis derived from the organic culture extracts of a representative transmissible strain TSB007 versus PA7 was performed using a Student's *t*-test model to define variants and this was rendered using a volcano plot (**Figure 4.4.4**). One of the unique features revealed within the plot was the *m/z* 539.3 ion, fragmentation pattern of which indicated two consecutive proline losses, and no match to the known compounds from the constructed metabolomic cladogram of Pseudomonads was obtained. This unique metabolite was isolated and its structure was elucidated using a combination of high-resolution mass spectrometry and 1D and 2D NMR experiments (**Supplementary Figures S4.6.1-6**) and determined as an N-succinylated tetramer (LPPI-

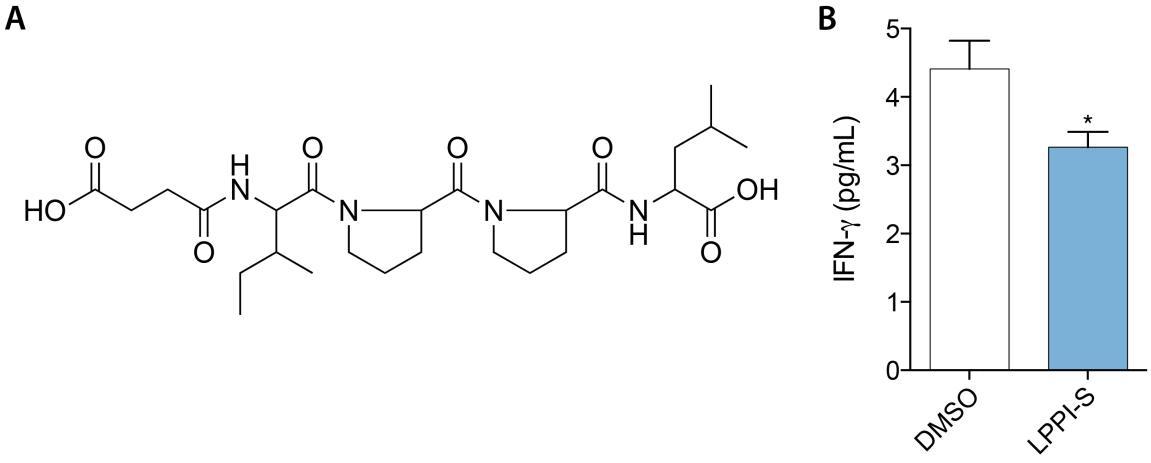


Figure 4.4.5. The chemical structure and immunomodulatory activity of the unique metabolite identified from the volcano plot. **A** – the chemical structure of the novel succinylated tetrapeptide (LPPI-S). **B** – levels of IFN-gamma secreted by RAW264.7 macrophages stimulated with LPS upon treatment with LPPI-S. Data are presented as mean ± SEM, asterisks (*) indicate significant differences between treatments (*$p<0.05$).

S) with the following sequence – leucine, proline, proline and isoleucine (**Figure 4.4.5A**). Interestingly, unlike the model strains such as PAO1 and PA14, the production of LPPI-S was exclusive to only one more of the isolates – TSB005. To reveal the likely function of this newly isolated N-succinylated small molecule, macrophage cytokine assays were performed. From this the LPPI-S was recognized to downregulate the secretion of interferon gamma (**Figure 4.4.5B**). Being an important mediator in promoting transcription of several pro-inflammatory cytokines in both monocytes and macrophages and an agent involved directly in NF-κB signaling[38-40], downregulation of interferon gamma (IFN-γ) by LPPI-S is significant. However, the link between this activity and the transmissibility of select Pseudomonads cannot be made directly from these experiments. Nonetheless, it is notable that the agent has a significant impact on the primary response driven by IFN-γ.

Next generation sequencing is rapidly expanding the genomic views of the Pseudomonads, and will produce a means to infer the metabolic differences between strains using a genomes to molecules perspective. Recently our laboratory constructed a Genomes to Natural Products (GNP) platform that scans genomes for biosynthetic clusters, compares them using structural prediction tools and chemoinformatic algorithms to define new biosynthetic machineries and predicts their endproduct chemistries. *Pseudomonas aeruginosa* alone is represented by 292 genomes either sequence or queued[41], and we sought to analyze the model *P. aeruginosa* PAO1 genome using GNP platform to define if unknown biosynthetic machineries exist and relate these to the known Pseudomonad chemistries. From this analysis, the biosynthetic loci were revealed

and their products were inferred using the predictive chemoinformatic tools embedded within the GNP platform and these were automatically compiled to make associations with the known Pseudomonad compounds. The entirety of these analyses led to the generation of predicted products from four NRPS gene clusters, and subsequent Tanimoto scores with other known chemistries were used to infer relationships and build a metabolomic cladogram (**Figure 4.4.6**). In this secondary metabolomic investigation, the GNP platform created a defined cataloging of the known compounds by accurately

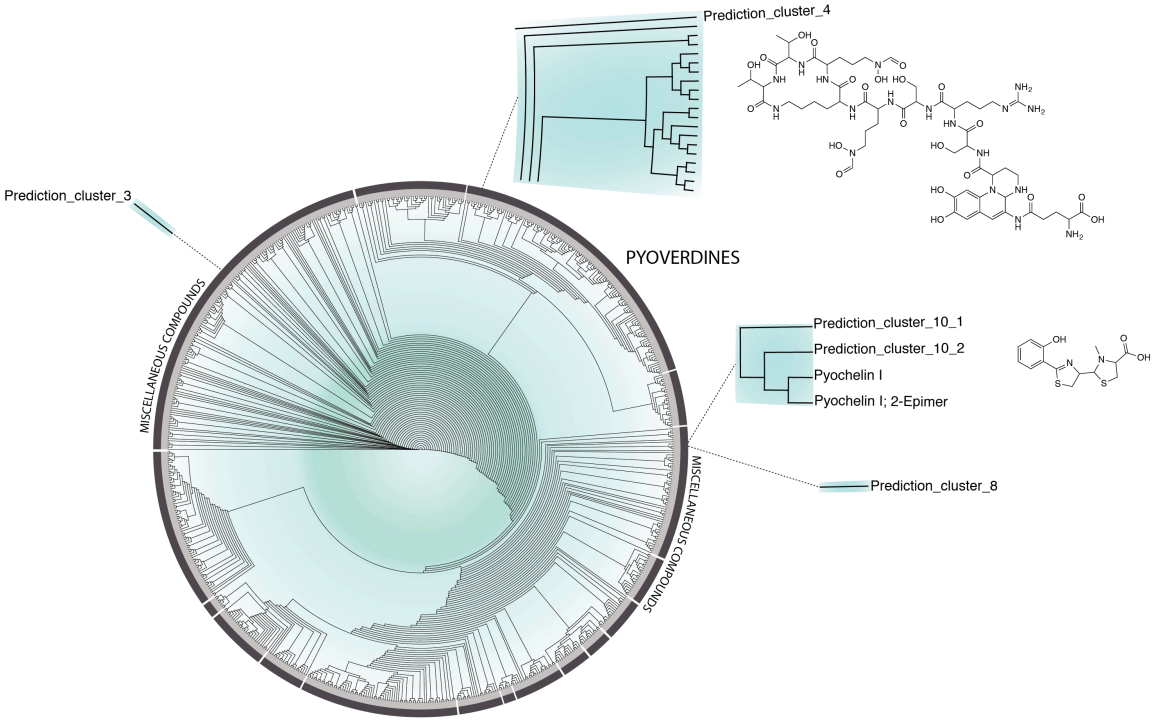


Figure 4.4.6. Connecting genes to molecules with GNP platform. Prediction of pyoverdine cluster with GNP platform (Prediction_cluster_4) localizes in the correct sector on the metabolomic tree. Pyochelin prediction (Prediction_cluster10_1/2) also localizes correctly with the known pyochelins. Both Prediction_cluster_3 and Prediction_cluster_8 appear as separate branches.

predicting pyoverdine (Prediction_cluster_4) and pyochelin (Prediction_cluster_10). Although the accuracy of these relationships can be recognized via the cladogram outputs, in two instances the inferred endproducts from two clusters did not overlay with the known natural product space (**Figure 4.4.6)**. The predicted outliers were related to the

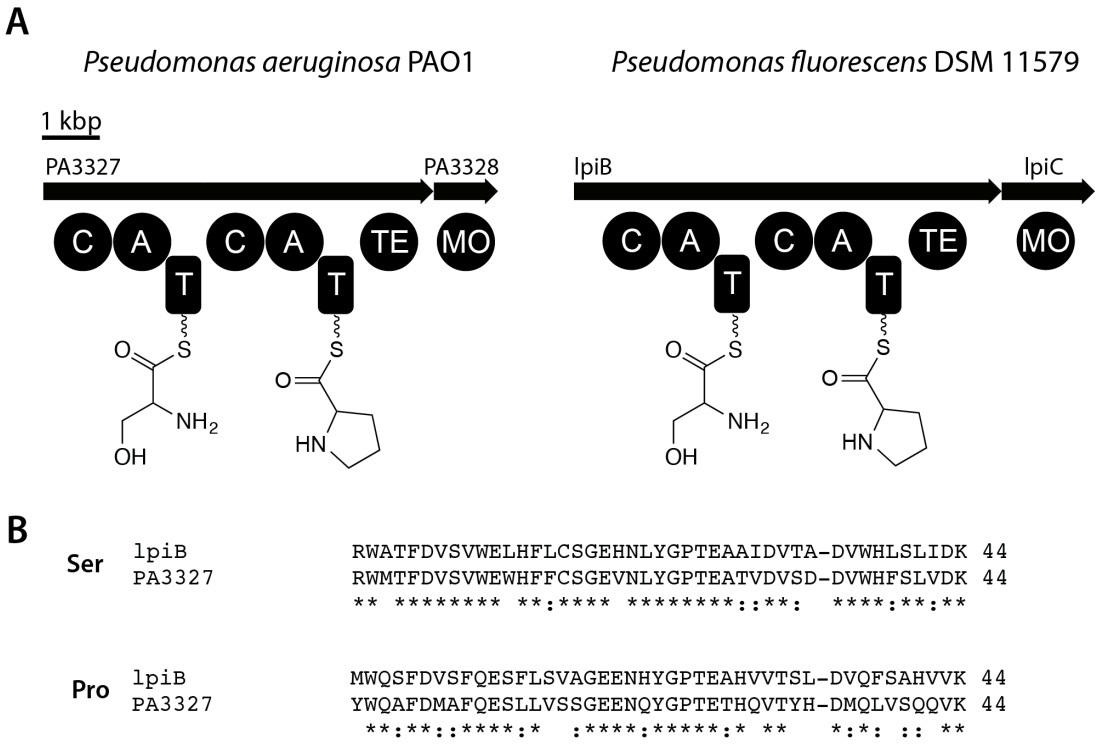


Figure 4.4.7. Comparison of the cryptic NRPS cluster from *Pseudomonas aeruginosa* PAO1 to lpiB from *Pseudomonas fluorescens* DSM 11579. **A** – the clusters with the predicted enzymatic assembly lines and the predicted substrates of the adenylation domains. C, condensation domain; A, adenylation domain; T, thiolation domain; TE, thioesterase domain. **B** – proposed amino acid sequence based on the predicted amino acids of the binding pocket of the A domains from lpiB and PA3327 genes for serine (Ser) and proline (Pro), respectively.

two cluster IDs within the outputs (Prediction_cluster_3 and Prediction_cluster_8) and their chemical differentiation was revealed by their distinct, singular branches on the Pseudomonad metabolome cladogram. A BLAST search revealed that cluster_3 has already been previously characterized to contain PA2302 and PA2305 NRPS genes, earlier thought to be responsible for the production of L-2-amino-4-methoxy-trans-3-butenoic acid (AMB), which was later reassigned to the IQS compound[42,43]. Interestingly, the products of the other cluster_8 have not yet been identified. Recently we had investigated another Pseudomonad species, *P. fluorescens* and identified the exact genes, lpiB and lpiC, that are directly involved in the creation of dipeptide bicyclic lipocyclocarbamate natural product, SB-253514[44]. During its biosynthesis, this dipeptide

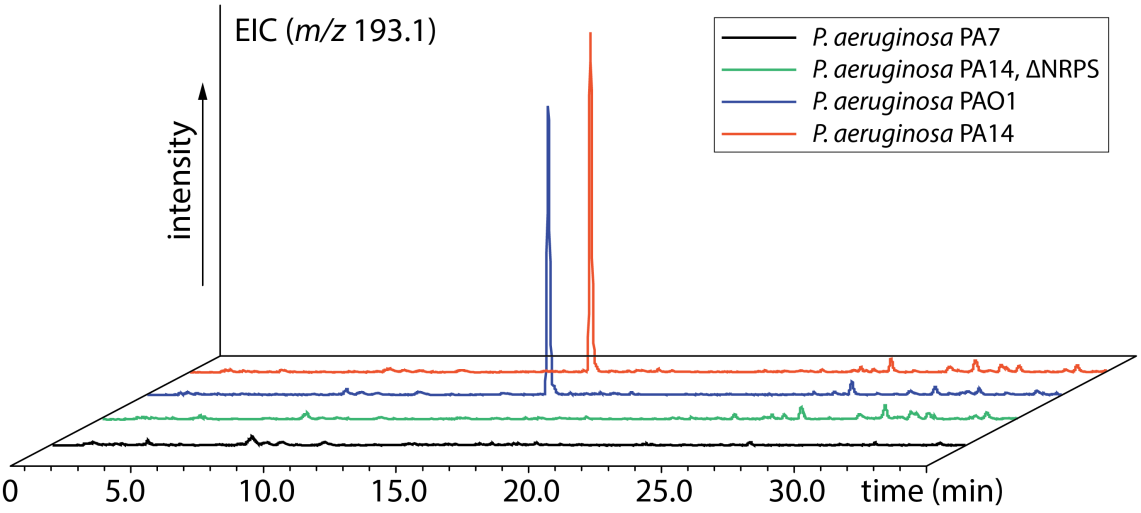


Figure 4.4.8. Extracted ion chromatogram of the metabolite related to the cryptic NRPS cluster from *Pseudomonas aeruginosa*. The identified metabolite with *m/z* 193.1 is present in both *P. aeruginosa* PAO1 and PA14 and absent in the ΔNRPS strain of *P. aeruginosa* PA14 and PA7.

compound undergoes a postulated Baeyer-Villiger reaction to insert oxygen into the SB-253514 ring system. The SB-253514 molecule exhibits nano-molar inhibition to human lipoprotein-associated phospholipase A and served as an inspiration for darapladib[45]. Interestingly, the SB-253514 cluster architecture and its amino acid specifying adenylation domains are similar to the cryptic NRPS cluster identified from *P. aeruginosa* (**Figure 4.4.7A**). However, despite the notable similarity of the amino acid sequences of their adenylation domains (**Figure 4.4.7B**), the two clusters are slightly divergent from each other, with only 36% amino acid sequence identity between lpiB and PA3327 genes. Perhaps this is because of the differences in the *Pseudomonas* strains in which these clusters are found, or perhaps because the actual chemistry is different.

The significance of this relationship also bears importance on possible biological roles of the unknown *P. aeruginosa* molecules. Inhibition of lipoprotein-associated phospholipase leads to a marked decrease in chemotactic ability of macrophages and thus any such *P. aeruginosa* secreted metabolite, related to SB-253514 would be a candidate small molecule virulence factor as a small molecule pathogen evasion strategy. Interestingly, numerous studies have suggested that the products of this cluster may be involved in virulence network of *P. aeruginosa*[46-52]. To investigate this we compared the LC-MS chromatograms of the two strains and noted no similar compounds within the SB-253514 retention time. Therefore, to rapidly identify this cryptic nonribosomal peptide metabolite we sought to investigate this using the gene knockout strains and isotopically labeled substrates identified using GNP assessment of *P. aeruginosa* genome. Metabolomic subtraction using both *P. aeruginosa* PAO1 and PA14 strains that contain

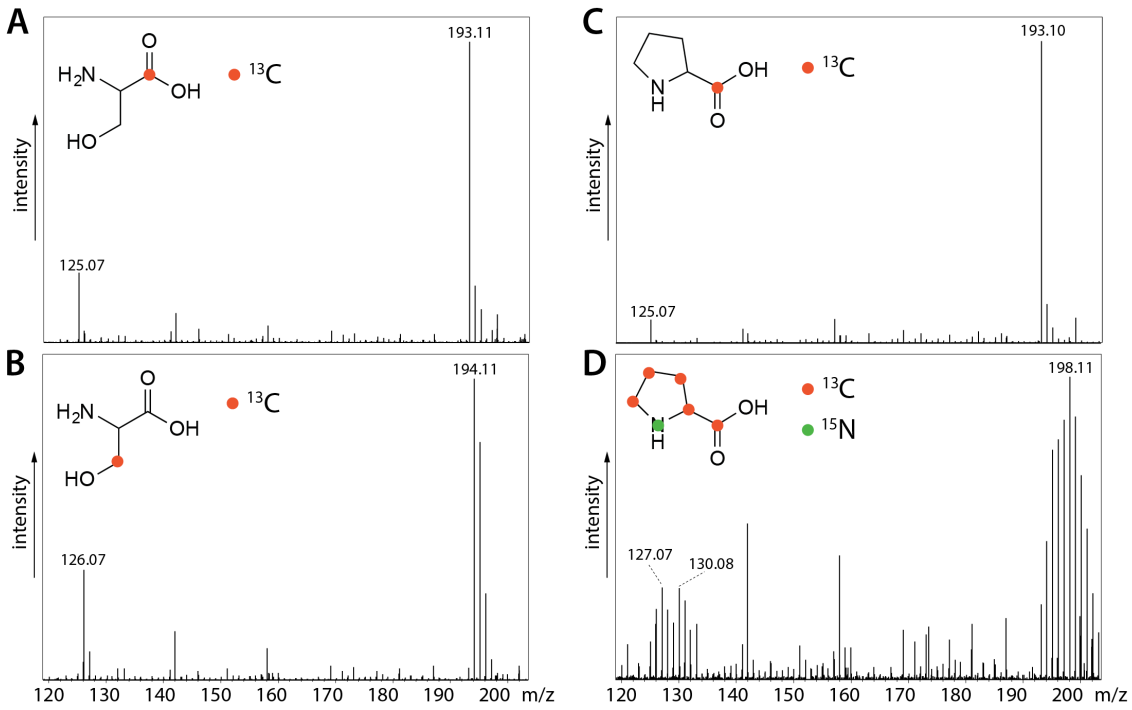


Figure 4.4.9. Incorporation of predicted isotopically labeled amino acid precursors. **A** – mass spectrum of peak corresponding to the original *m/z* 193.1 ion from the culture of *P. aeruginosa* PA14 supplemented with L-serine-1-^{13}C, label position indicated with a red circle. **B** – mass spectrum of peak corresponding to the original *m/z* 193.1 ion from the culture of *P. aeruginosa* PA14 supplemented with L-serine-3-^{13}C, label position indicated with a red circle. **C** – mass spectrum of peak corresponding to the original *m/z* 193.1 ion from the culture of *P. aeruginosa* PA14 supplemented with L-proline-1-^{13}C, label position indicated with a red circle. **D** – mass spectrum of peak corresponding to the original *m/z* 193.1 ion from the culture of *P. aeruginosa* PA14 supplemented with L-proline-$^{13}C_5$,^{15}N, label positions are indicated.

the functional NRPS and both *P. aeruginosa* PA14 ΔNRPS led to the identification of *m/z* 193.1 ion (**Figure 4.4.8**). Similarly to *P. aeruginosa* PA14 ΔNRPS strain, the strain that naturally lacks the cluster, *P. aeruginosa* PA7, did not show any production of *m/z* 193.1. The validation obtained from the feeding isotopically labeled substrates, serine and proline, further verified that this compound was arising from the NRPS cluster (**Figure 4.4.9**). Interestingly, however, unlike L-serine-3-^{13}C supplementation, the one with L-serine-1-^{13}C did not shown any incorporation of the label, suggesting a likely loss of carboxylic acid during the biosynthesis of *m/z* 193.1. Unusually, the same was true with L-proline-1-^{13}C supplementation, which likewise did not show any incorporation. However, supplementing L-proline-$^{13}C_5$, ^{15}N, which upon incorporation was expected to lead to an M + 6 shift in the mass spectrum, only resulted in an M + 5 shift, again likely due to the loss of the ^{13}C-labeled carboxylic acid. The anticipated peculiarities in this molecule are not a complete surprise, however, as the unusual chemistry observed in SB-253514 biosynthesis may also hold true with this metabolite. Its isolation in future will hopefully reveal a unique scaffold and an intriguing immunomodulatory activity.

In conclusion, this body of work describes a systematic metabolomic investigation of the Pseudomonads that led to defining the routes of rapid identification and consequent biological assessment of secreted small molecules. Such investigations resulted in selective identification of quinolones and a novel metabolite LPPI-S that was found as a differentiating feature of select clinical isolates. More importantly, our results shed light on the immunomodulatory potential of these secreted metabolites and suggests deeper analysis of their biological function in the context of virulence and infectivity of *P.*

aeruginosa. In addition, by utilizing the GNP platform we were able to make a connection between the *P. aeruginosa* genome and the downstream biosynthetic products, resulting in the identification of the cryptic metabolite. Having identified this compound, it presents a unique opportunity for testing its bioactive potential in the near future.

4.5 Materials and Methods

4.5.1 Bacterial strains and fermentation conditions

For targeted identification of quinolones, the overnight culture of *Pseudomonas aeruginosa* PAO1 (ATCC 15692) grown in the tryptic soy broth (TSB) medium (EMD, Millipore #1.05459.5007) was used to inoculate (1:100 dilution) 50 mL of TSB medium (n = 3) in 250 mL flasks and that were incubated for 72 hours at 37°C, 200 rpm. For quinolone isolation the above seed culture was used to inoculate twelve 2.8 L Fernbach flasks containing 1.5 L of TSB, which were incubated for 72 hours at 37°C, 180 rpm.

For quantification of quinolone abundance, model clinical isolate strains including *P. aeruginosa* PA7, *P. aeruginosa* PAO1 and *P. aeruginosa* PA14; genetically identical clinical strains isolated ten years apart from a cystic fibrosis patient, *P. aeruginosa* P590 and P48 (isolated last); and eight clinical isolates from a single sputum sample collected from a different patient also with cystic fibrosis, *P. aeruginosa* TSB001-018, were all grown overnight in TSB medium and were used to inoculate (1:100 dilution, normalized using OD_{600}) 50 mL of TSB medium (n = 3) in 250 mL flasks that were incubated for 72 hours at 37°C, 200 rpm.

For metabolomic analysis studies, *P. aeruginosa* TSB007 isolate and *P. aeruginosa* PA7 strain were cultured as described above, however, with increased number of replicate samples (n = 7). For the identification of cryptic nonribosomal peptide from *P. aeruginosa* PA14/PAO1, a mutant strain of *P. aeruginosa* PA14 with a gentamycin-harbouring MAR2xT7 transposon insertion at the base pair 35 of PA14_21020 gene

(ΔNRPS), obtained from a transposon mutagenesis library was cultured together with the wild-type *P. aeruginosa* PA14 using the conditions described above[53].

4.5.2 Construction of the circular cladogram of Pseudomonad metabolome

The entire database of Dictionary of Natural Products (DNP, accessed Dec 15, 2013) was screened for all compounds produced by the species of Pseudomonad class. The identified 725 compounds were compiled in a SMILES code format and clustered using ChemMine Tools software using hierarchical clustering, with a distance matrix setting (http://chemmine.ucr.edu). The downloaded results were processed by Dendroscope software using circular cladogram option. The resulting cladogram figure was exported and annotated in Adobe Illustrator CS6.

4.5.3 Identification and isolation of quinolone compounds

To identify quinolones in the cultures of *P. aeruginosa*, multiple reaction monitoring (MRM) setting was utilized using Bruker amaZon X ion trap mass spectrometer (operating in MRM, positive ESI mode, with scanning range of *m/z* 100-500) coupled with a Dionex UltiMate 3000 HPLC, with an Ascentis Express C18 column (150 mm × 4.6 mm, 2.7μ, Sigma Aldrich) and A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) as the mobile phases at 1.2 mL/min. The MRM parameters were tuned to detect any MS/MS transitions that would generate either *m/z* 159 or *m/z* 175. The separation was achieved using the following solvent gradient: 0-5 min, 5% A; 5-45 min, linear gradient to 100% A; 45-55 min, isocratic at 100 % A, 55-60 min, linear to 5% A; 60-65 min, isocratic at 5% A.

To isolate identified quinolones, following the 72-hour fermentation, the culture was extracted with Diaion® HP-20 resin (Sigma, 40 g/L) added directly to the flasks and incubated for 2 hours at 37°C, 180 rpm. The resin was harvested using vacuum filtration and washed twice with ddH_2O to remove any cellular debris. To elute the bound metabolites, resin was stirred in 1 L of methanol for 1 hour, twice. The methanolic extract was concentrated *in vacuo* and was processed by the preparatory HPLC-ESI-MS (Bruker amaZon X ion trap mass spectrometer coupled to Dionex UltiMate 3000 HPLC system with variable UV detection (160-600 nm) and Phenomenex Luna 5µ C18(2) (250 mm x 15 mm, 4.6 µm) with 10 mL/min flow rate). Optimal separation was achieved using the following A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) solvent gradient: 0-5 min, 5% A; 5-7 min, linear gradient to 50% A; 7-56 min, linear gradient to 55% A; 56-60 min, linear gradient to 100% A; 60-68 min, isocratic at 100 % A, 68-70 min, linear to 5% A; 70-75 min, isocratic at 5% A. The validate the identity of isolated quinolones, a combination of detailed MS/MS and NMR experiments with Bruker Avance 700 MHz spectrometer with samples dissolved in 99.8% methanol-d_4 was used.

4.5.4 Assessing abundance of quinolones in various strains of *P. aeruginosa*

The relative abundance of the identified quinolones in the cultures of various strains of *P. aeruginosa* was quantified using the integrated peak area of the corresponding extracted ion processed with Bruker Compass DataAnalysis 4.1 SR1 software. To calculate the global relative abundance of isolated quinolones, the averaged area values were normalized relative to the most abundant quinolone among all tested

strains. To calculate their relative abundance within each strain, the corresponding averaged area values were normalized relative to the most abundant quinolone within a particular strain. The data were then processed using Genesis 1.7.6 software to construct the heatmap plots[54].

4.5.5 Identification and isolation of LPPI-S from a clinical strain of *Pseudomonas*

The replicate cultures of *P. aeruginosa* TSB007 isolate and *P. aeruginosa* PA7 strain were extracted using liquid-liquid partitioning with ethyl acetate (2:1 v/v, 2x) and the resulting organic phase was concentrated *in vacuo*. The samples were then resuspended in methanol and analyzed directly by Bruker micrOTOF II mass spectrometer (positive ESI, scanning *m/z* 100-1000) with an Agilent 1200 series HPLC with Ascentis Express C18 column (150 mm x 2.1 mm, 2.7 µm, Sigma) and A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) as the mobile phases at 0.25 mL/min. The solvent gradient was 0-5 min, 5% A; 5-7 min, linear gradient to 50% A; 7-56 min, linear gradient to 55% A; 56-60 min, linear gradient to 100% A; 60-68 min, isocratic at 100 % A, 68-70 min, linear to 5% A; 70-75 min, isocratic at 5% A.

To analyze outliers, the volcano plot was generated using the *t*-test model in the Bruker Daltonics ProfileAnalysis 2.1, by making pairwise comparisons between the two strains. More specifically, the bucket table for the *t*-test model was computed using rectangular bucketing parameters defined as ΔR_t = 6.0 sec, with a kernel size of 3 sec, and $\Delta m/z$= 0.5 with a kernel size of *m/z* 0.1, and normalized by the sum of bucket values in the analysis. The fold change and *p*-value cut-offs used to construct the volcano plot were 2 and 0.05, respectively.

4.5.6 Identification of cryptic nonribosomal peptide in *P. aeruginosa* PA14

The entire genome of *P. aeruginosa* PA14 was analyzed using newly developed Genomes to Natural Products (GNP) platform. More specifically, its genome search mode scans entire genome FASTA files for natural product biosynthesis clusters, such as polyketides and nonribosomal peptides, and identified the enzymatic domains and generates SMILES string output of their products. From the GNP analysis, the SMILES code predictions of the final products pertaining to the NRPS clusters were added to the entire metabolome of Pseudomonads and reanalyzed using the same procedure that was applied to the construction of the circular cladogram, described above. The localization of the predictions on the cladogram was identified within Dendroscope. The amino acids of the binding pockets of A domains found within lpiB and PA3327 (identical to PA14_21020 in *P. aeruginosa* PA14) genes for serine and proline were predicted using the prediction software NRPSpredictor2[55]. These sequences were aligned using multiple sequence alignment software, ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2). Metabolomic subtraction of the wild-type and ΔNRPS *P. aeruginosa* LC-MS chromatograms was performed using Bruker Daltonics Metabolite Tools 2.0 SR4 software (MetaboliteDetect version 2.0). The incorporation of isotopically labeled amino acids was carried out using culturing protocol described above with a few modifications. Briefly, YM medium (per 1 L: bacteriological grade peptone (Amresco, J636-1KG), 5 g; yeast extract (Difco Laboratories, 212750), 3 g; malt extract (HiMedia, 61001-518), 3 g; glucose (dextrose, Caledon, 3260-1), 1 g), supplemented with one of L-serine-1-^{13}C, L-

serine-3-^{13}C, L-proline-1-^{13}C or L-proline-$^{13}C_5$, ^{15}N, added aseptically and directly to the medium (50 mg/50 mL), was used instead of TSB medium.

4.5.7 Multiplexed cytokine profiling

RAW 264.7 macrophage cell line (ATCC # TIB-71) was cultured in RPMI 1640 medium, supplemented with 10% (w/v) fetal bovine serum (FBS, PAA laboratories, A15-751), 2 mM L-glutamine (VWR, BDH), at 37°C in a 5% CO_2 incubator, keeping the passage below 10. Cells were subcultured 1:8 and incubated in 24-well plates (BD Biosciences) with a total volume of 2 mL. After reaching 50% confluence (~48 hours), cells were stimulated with 100 ng/mL LPS (L3012-5mg, Sigma) and treated with nine identified quinolones and LPPI-S (20 μL dissolved in DMSO, 100 μM final concentration). After 16 hours of incubation, supernatants were collected and stored at -20°C. For analysis of the wide range of cytokines, samples were analyzed directly using the Bio-Plex Luminex system with a Bio-Plex Pro™ mouse cytokine 23-plex assay kit (Bio-Rad, M60-009RDPD). Fluorescence data were acquired and analyzed by Bio-Plex Manager software (version 5.0, Bio-Rad Laboratories), following the manufacturer's recommendations.

4.5.8 Statistical analysis

All data are presented as mean ± SEM. Unpaired *t*-test (for parametric data) was performed using GraphPad Prism version 6.0b for Mac OS X (GraphPad Software, La Jolla, CA). An associated p value <0.05 was considered statistically significant.

4.6 Supplementary Figures and Legends (Figure S4.6.1-6)

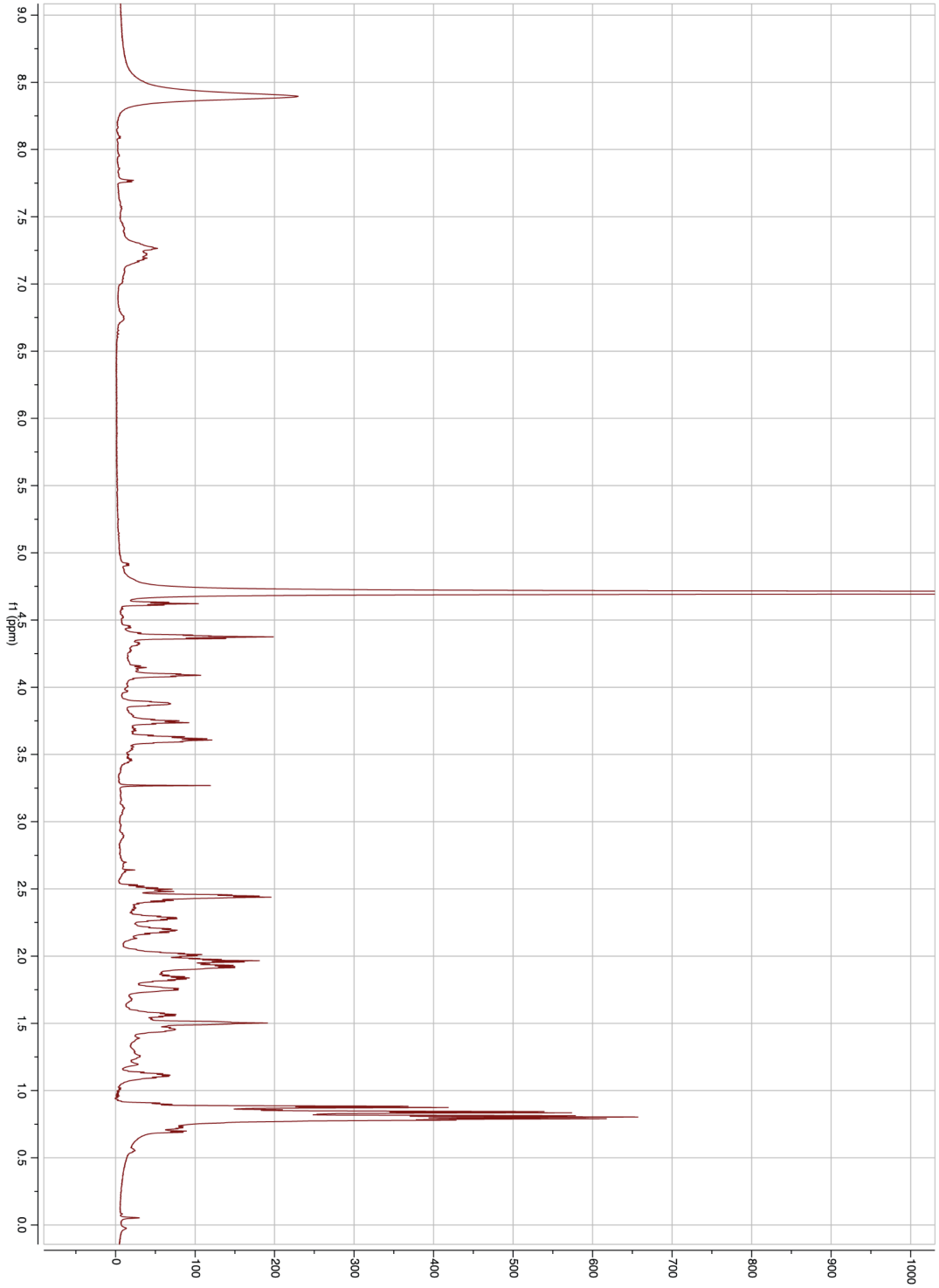


Figure S4.6.1. ^{1}H spectrum of LPPI-S in 100% D_2O.

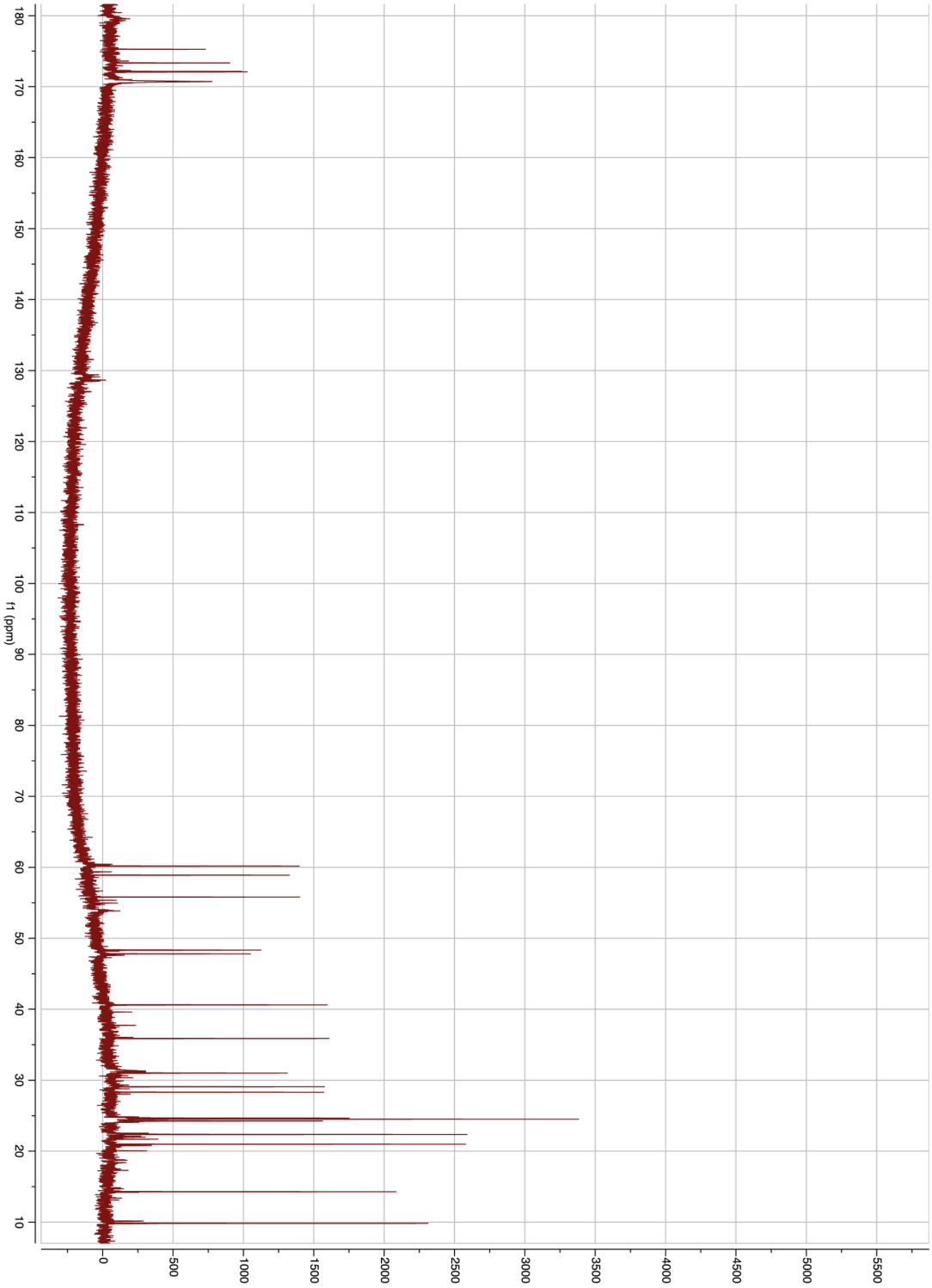


Figure S4.6.2. ^{13}C (^{1}H decoupled) spectrum of LPPI-S in 100% D_2O.

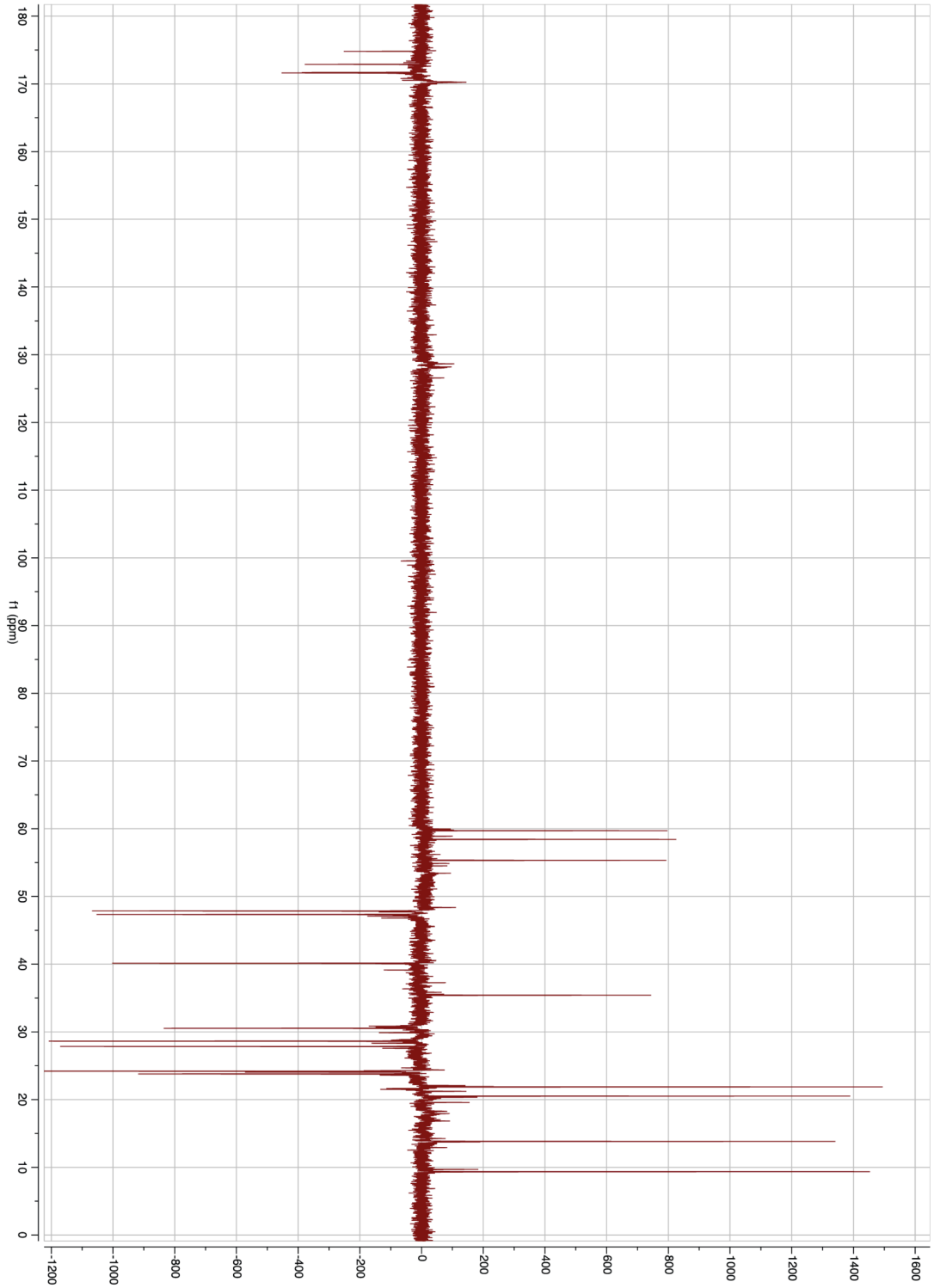


Figure S4.6.3. ^{13}C DEPTq135 spectrum of LPPI-S in 100% D_2O.

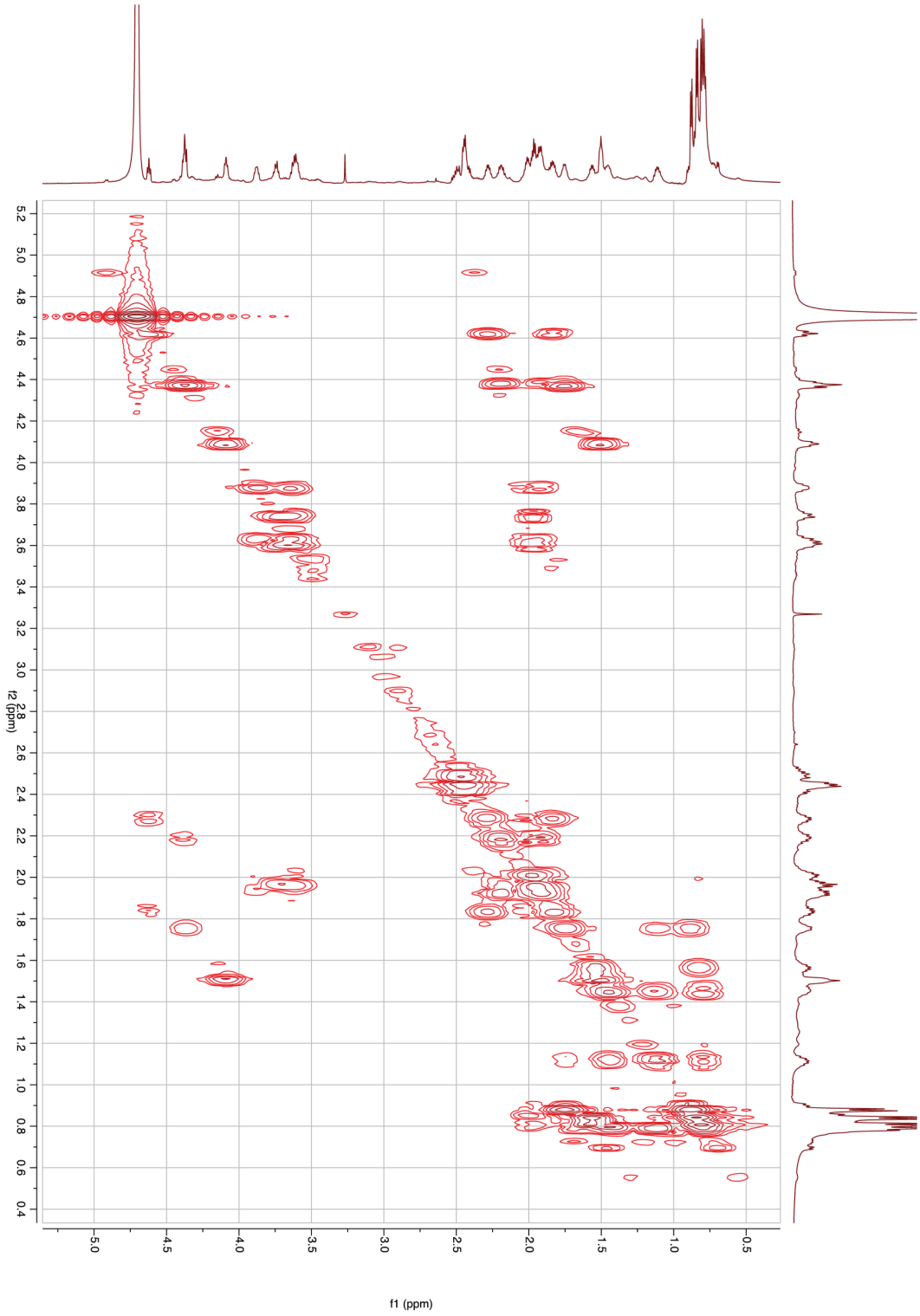


Figure S4.6.4. 2D COSY spectrum of LPPI-S in 100% D_2O.

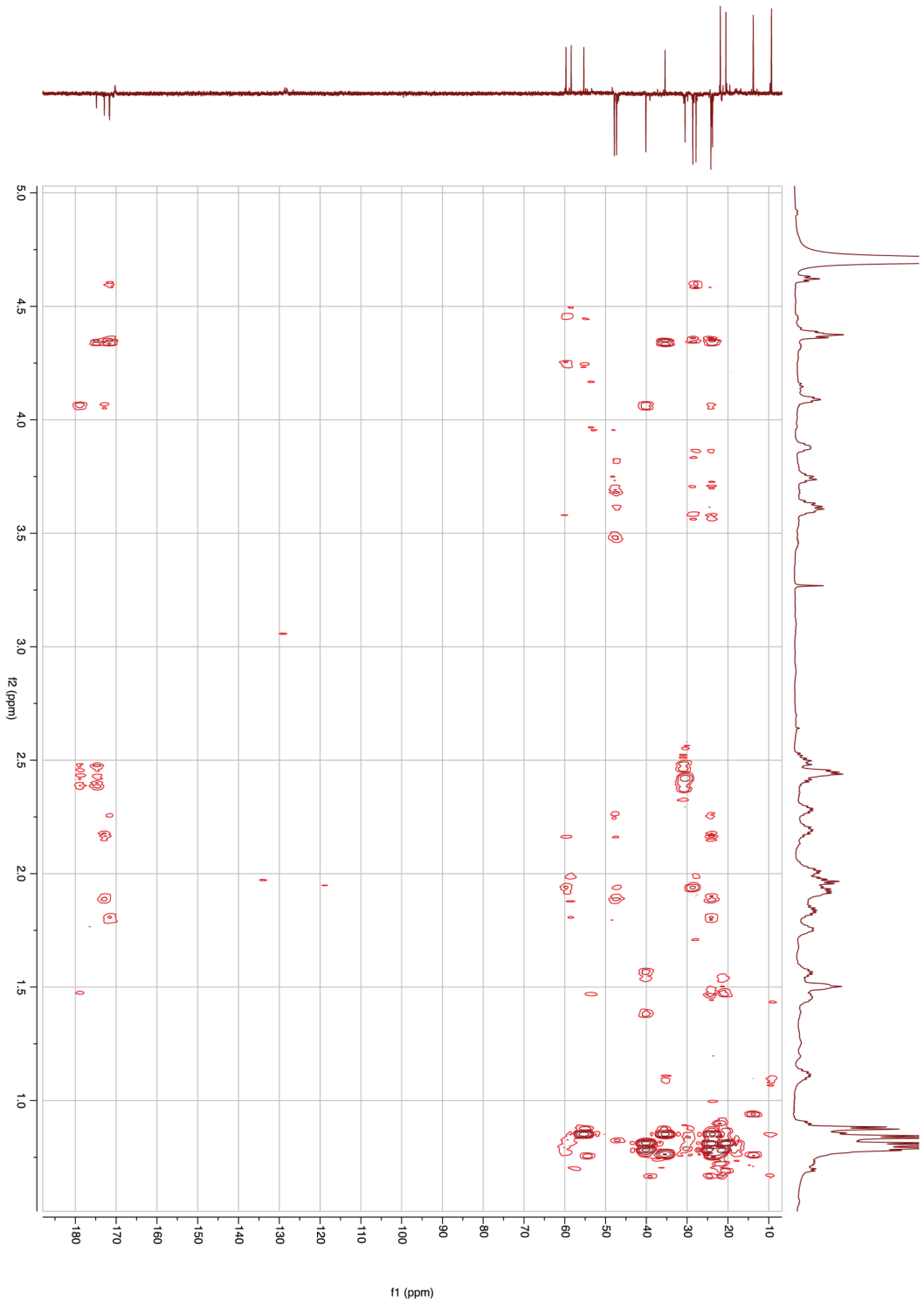


Figure S4.6.5. 2D HMBC spectrum of LPPI-S in 100% D_2O.

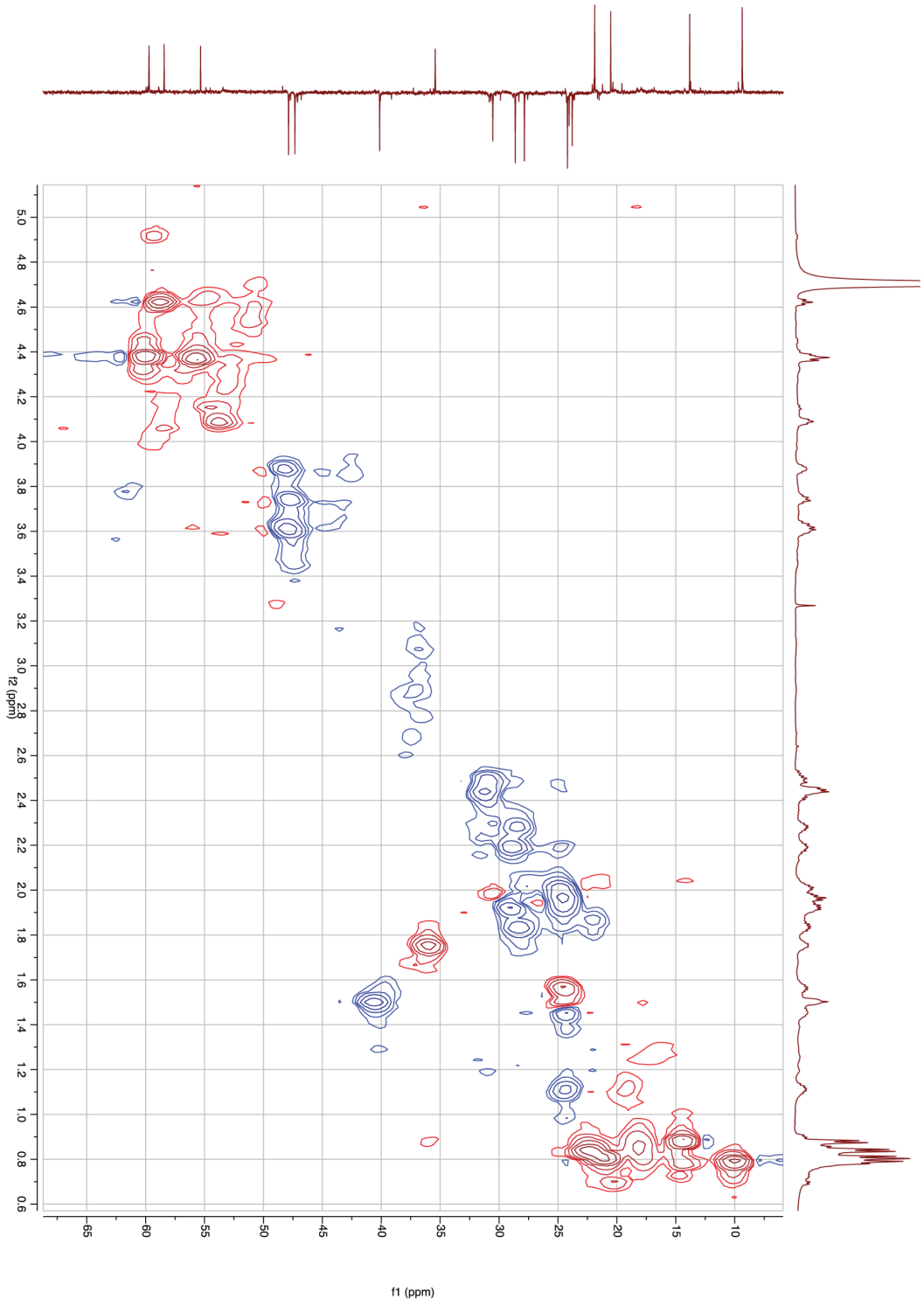


Figure S4.6.6. 2D HSQC spectrum of LPPI-S in 100% D_2O.

4.7 References

1. Spiers, A. J., Buckling, A. & Rainey, P. B. The causes of Pseudomonas diversity. *Microbiology (Reading, Engl.)* **146,** 2345–2350 (2000).
2. Reimmann, C. *et al.* The global activator GacA of Pseudomonas aeruginosa PAO positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. *Molecular Microbiology* **24,** 309–319 (1997).
3. Laycock, M. V., Hildebrand, P. D., Thibault, P., Walter, J. A. & Wright, J. L. Viscosin, a potent peptidolipid biosurfactant and phytopathogenic mediator produced by a pectolytic strain of Pseudomonas fluorescens. *Journal of agricultural and food chemistry* **39,** 483–489 (1991).
4. Gerard, J. *et al.* Massetolides A-H, antimycobacterial cyclic depsipeptides produced by two pseudomonads isolated from marine habitats. *J. Nat. Prod.* **60,** 223–229 (1997).
5. Lamont, I. L. Identification and characterization of novel pyoverdine synthesis genes in Pseudomonas aeruginosa. *Microbiology* **149,** 833–842 (2003).
6. Serino, L. *et al.* Biosynthesis of pyochelin and dihydroaeruginoic acid requires the iron-regulated pchDCBA operon in Pseudomonas aeruginosa. *J. Bacteriol.* **179,** 248–257 (1997).
7. Mavrodi, D. V., Blankenfeldt, W. & Thomashow, L. S. Phenazine compounds in fluorescent Pseudomonas spp. biosynthesis and regulation. *Annu Rev Phytopathol* **44,** 417–445 (2006).

8. Soberón-Chávez, G., Lépine, F. & Déziel, E. Production of rhamnolipids by Pseudomonas aeruginosa. *Appl. Microbiol. Biotechnol.* **68,** 718–725 (2005).

9. Dubern, J.-F. & Diggle, S. P. Quorum sensing by 2-alkyl-4-quinolones in Pseudomonas aeruginosa and other bacterial species. *Mol Biosyst* **4,** 882–888 (2008).

10. Cooley, M., Chhabra, S. R. & Williams, P. N-Acylhomoserine lactone-mediated quorum sensing: a twist in the tail and a blow for host immunity. *Chemistry & Biology* **15,** 1141–1147 (2008).

11. Diggle, S. P. *et al.* The Pseudomonas aeruginosa 4-Quinolone Signal Molecules HHQ and PQS Play Multifunctional Roles in Quorum Sensing and Iron Entrapment. *Chemistry & Biology* **14,** 87–96 (2007).

12. Kim, K. *et al.* HHQ and PQS, two Pseudomonas aeruginosa quorum-sensing molecules, down-regulate the innate immune responses through the nuclear factor-κB pathway. *Immunology* **129,** 578–588 (2009).

13. Déziel, E. *et al.* Analysis of Pseudomonas aeruginosa 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 1339–1344 (2004).

14. Loots, D. T. & Preez, Du, I. A comparison of two extraction methods for differentiating and characterising various Mycobacterium species and Pseudomonas aeruginosa using GC-MS metabolomics. *Afr. J. Microbiol. Res.* **6,** 3159–3172 (2012).

15. Krumsiek, J. *et al.* Mining the unknown: a systems approach to metabolite

identification combining genetic and metabolic information. *PLoS Genet.* **8,** e1003005 (2012).

16. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23,** 673–679 (2007).
17. Fothergill, J. L., Walshaw, M. J. & Winstanley, C. Transmissible strains of Pseudomonas aeruginosa in cystic fibrosis lung infections. *European Respiratory Journal* **40,** 227–238 (2012).
18. Scott, F. W. & Pitt, T. L. Identification and characterization of transmissible Pseudomonas aeruginosa strains in cystic fibrosis patients in England and Wales. *J. Med. Microbiol.* **53,** 609–615 (2004).
19. Guina, T. *et al.* Quantitative proteomic analysis indicates increased synthesis of a quinolone by Pseudomonas aeruginosa isolates from cystic fibrosis airways. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 2771–2776 (2003).
20. Burgess, M. W., Keshishian, H., Mani, D. R., Gillette, M. A. & Carr, S. A. Simplified and efficient quantification of low-abundance proteins at very high multiplex via targeted mass spectrometry. *Molecular & Cellular Proteomics* **13,** 1137–1149 (2014).
21. Chang, Y.-J., Chao, M.-R., Chen, S.-C., Chen, C.-H. & Chang, Y.-Z. A high-throughput method based on microwave-assisted extraction and liquid chromatography–tandem mass spectrometry for simultaneous analysis of amphetamines, ketamine, opiates, and their metabolites in hair. *Anal Bioanal Chem* **406,** 2445–2455 (2014).

22. Goto, T. *et al.* The high throughput analysis of N-methyl carbamate pesticides in fruits and vegetables by liquid chromatography electrospray ionization tandem mass spectrometry using a short column. *Analytica Chimica Acta* **555,** 225–232 (2006).

23. Yang, L. *et al.* Investigation of an enhanced resolution triple quadrupole mass spectrometer for high-throughput liquid chromatography/tandem mass spectrometry assays. *Rapid Commun. Mass Spectrom.* **16,** 2060–2066 (2002).

24. Merrill, A. H., Sullards, M. C., Allegood, J. C., Kelly, S. & Wang, E. Sphingolipidomics: high-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatography tandem mass spectrometry. *Methods* **36,** 207–224 (2005).

25. Wu, J. T., Zeng, H., Deng, Y. & Unger, S. E. High-speed liquid chromatography/tandem mass spectrometry using a monolithic column for high-throughput bioanalysis. *Rapid Commun. Mass Spectrom.* **15,** 1113–1119 (2001).

26. Fletcher, M. P., Diggle, S. P., Cámara, M. & Williams, P. Biosensor-based assays for PQS, HHQ and related 2-alkyl-4-quinolone quorum sensing signal molecules. *Nat Protoc* **2,** 1254–1262 (2007).

27. Struss, A. K. *et al.* Toward implementation of quorum sensing autoinducers as biomarkers for infectious disease states. *Anal. Chem.* **85,** 3355–3362 (2013).

28. Dinarello, C. A. Interleukin-1β, Interleukin-18, and the Interleukin-1β Converting Enzymea. *Annals of the New York Academy of Sciences* **856,** 1–11 (1998).

29. Le Feuvre, R. A., Brough, D., Iwakura, Y., Takeda, K. & Rothwell, N. J. Priming

of macrophages with lipopolysaccharide potentiates P2X7-mediated cell death via a caspase-1-dependent mechanism, independently of cytokine production. *J. Biol. Chem.* **277,** 3210–3218 (2002).

30. Chomarat, P., Banchereau, J., Davoust, J. & Palucka, A. K. IL-6 switches the differentiation of monocytes from dendritic cells to macrophages. *Nat. Immunol.* **1,** 510–514 (2000).

31. Deshmane, S. L., Kremlev, S., Amini, S. & Sawaya, B. E. Monocyte chemoattractant protein-1 (MCP-1): an overview. *J. Interferon Cytokine Res.* **29,** 313–326 (2009).

32. Morstyn, G. & Burgess, A. W. Hemopoietic growth factors: a review. *Cancer Res.* **48,** 5624–5637 (1988).

33. Parameswaran, N. & Patial, S. Tumor necrosis factor-α signaling in macrophages. *Crit. Rev. Eukaryot. Gene Expr.* **20,** 87–103 (2010).

34. He, J. *et al.* The broad host range pathogen Pseudomonas aeruginosa strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 2530–2535 (2004).

35. Mikkelsen, H., McMullan, R. & Filloux, A. The Pseudomonas aeruginosa reference strain PA14 displays increased virulence due to a mutation in ladS. *PLoS ONE* **6,** e29113 (2011).

36. Roy, P. H. *et al.* Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. *PLoS ONE* **5,** e8842 (2010).

37. Parkins, M. D. *et al.* Twenty-Five-Year Outbreak of Pseudomonas aeruginosa

Infecting Individuals with Cystic Fibrosis: Identification of the Prairie Epidemic Strain. *J. Clin. Microbiol.* **52,** 1127–1135 (2014).

38. Narumi, S., Tebo, J. M., Finke, J. H. & Hamilton, T. A. IFN-gamma and IL-2 cooperatively activate NF kappa B in murine peritoneal macrophages. *The Journal of Immunology* **149,** 529–534 (1992).
39. Yasumoto, K. *et al.* Tumor necrosis factor alpha and interferon gamma synergistically induce interleukin 8 production in a human gastric cancer cell line through acting concurrently on AP-1 and NF-kB-like binding sites of the interleukin 8 gene. *J. Biol. Chem.* **267,** 22506–22511 (1992).
40. Paludan, S. R. Synergistic action of pro-inflammatory agents: cellular and molecular aspects. *Journal of Leukocyte Biology* **67,** 18–25 (2000).
41. Pagani, I. *et al.* The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Research* **40,** D571–9 (2012).
42. Lee, X. *et al.* Identification of the biosynthetic gene cluster for the Pseudomonas aeruginosa antimetabolite L-2-amino-4-methoxy-trans-3-butenoic acid. *J. Bacteriol.* **192,** 4251–4255 (2010).
43. Lee, J. *et al.* A cell-cell communication signal integrates quorum sensing and stress response. *Nat Chem Biol* **9,** 339–343 (2013).
44. Johnston, C. W., Zvanych, R., Khyzha, N. & Magarvey, N. A. Nonribosomal assembly of natural lipocyclocarbamate lipoprotein-associated phospholipase inhibitors. *ChemBioChem* **14,** 431–435 (2013).

45. Johnson, J. L., Shi, Y., Snipes, R., Janmohamed, S. & Rolfe, T. E. Effect of darapladib treatment on endarterectomy carotid plaque lipoprotein-associated phospholipase A2 activity: a randomized, controlled trial. *PLoS ONE* (2014). doi:10.1371/journal.pone.0089034.s003

46. Hentzer, M. *et al.* Attenuation of Pseudomonas aeruginosa virulence by quorum sensing inhibitors. *The EMBO Journal* **22,** 3803–3815 (2003).

47. Schuster, M., Hawkins, A. C., Harwood, C. S. & Greenberg, E. P. The Pseudomonas aeruginosa RpoS regulon and its relationship to quorum sensing. *Molecular Microbiology* **51,** 973–985 (2004).

48. Dong, Y.-H., Zhang, X.-F., Soo, H.-M. L., Greenberg, E. P. & Zhang, L.-H. The two-component response regulator PprB modulates quorum-sensing signal production and global gene expression in Pseudomonas aeruginosa. *Molecular Microbiology* **56,** 1287–1301 (2005).

49. Dong, Y.-H., Zhang, X.-F., Xu, J.-L., Tan, A.-T. & Zhang, L.-H. VqsM, a novel AraC-type global regulator of quorum-sensing signalling and virulence in Pseudomonas aeruginosa. *Molecular Microbiology* **58,** 552–564 (2005).

50. Firoved, A. M. & Deretic, V. Microarray analysis of global gene expression in mucoid Pseudomonas aeruginosa. *J. Bacteriol.* **185,** 1071–1081 (2003).

51. LEE, B. *et al.* Mucoid Pseudomonas aeruginosa isolates maintain the biofilm formation capacity and the gene expression profiles during the chronic lung infection of CF patients. *APMIS* **119,** 263–274 (2011).

52. Schuster, M., Lostroh, C. P., Ogi, T. & Greenberg, E. P. Identification, timing, and

signal specificity of Pseudomonas aeruginosa quorum-controlled genes: a transcriptome analysis. *J. Bacteriol.* **185,** 2066–2079 (2003).

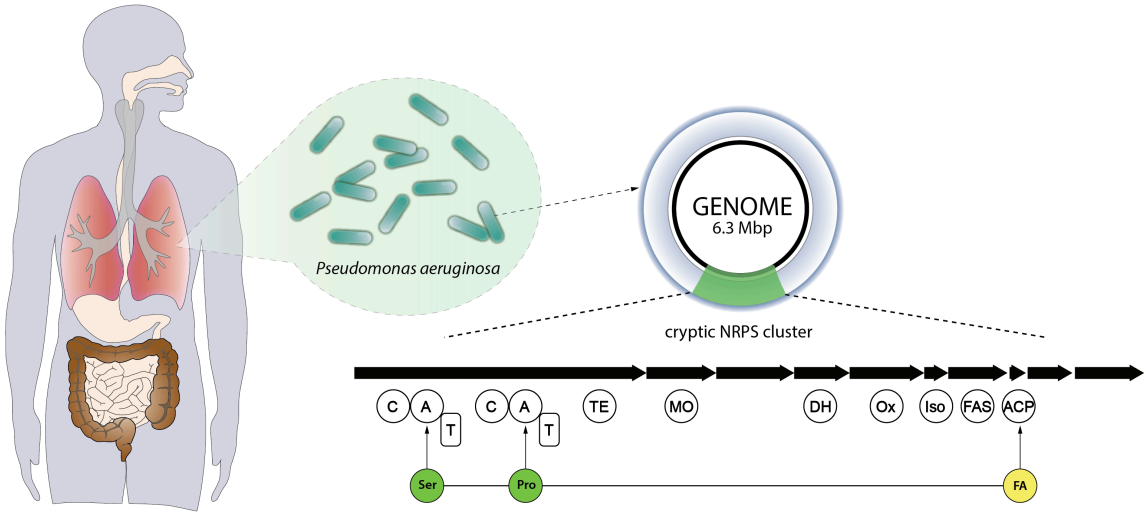
53. Liberati, N. T. *et al.* An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 2833–2838 (2006).

54. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18,** 207–208 (2002).

55. Röttig, M. *et al.* NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research* **39,** W362–7 (2011).

Chapter 5. Pseudenamide A, a Novel Nonribosomal Lipopeptide from the Pathogenic Member of the Human Microbiome *Pseudomonas aeruginosa*

5.1 Chapter Preface



The versatility of *Pseudomonas aeruginosa* metabolism and the known complexity of its secreted metabolome accompanied with the long known immune evasion potential of the organism placed this bacterium on the arena of intense investigations. Despite many intense pursuits towards metabolomic mining of the organism, it is clear that our knowledge of this organism and its' metabolic endproducts is still incomplete. Evidence supporting this assertion can be gleaned from genome data of *P. aeruginosa*, and the modern biochemoinformatic platforms that infer unknown chemistries. With elegant new algorithm and tools constructed such as Genomes to Natural Products (GNP), currently submitted for publication, in this work I was able to tackle efforts in a more refined way of making the connections between the genome and the metabolome of this microbiome member.

Evidence of immunomodulatory potential of some compound classes from *P. aeruginosa* already exists, and enriching its catalog of secreted metabolites is going to further enhance our understanding of its virulence and infectivity.

The following chapter is formatted as a manuscript that is prepared for submission for publication, in which I am the lead author. I conducted the experimental design, performed all experiments (except for structure elucidation), interpreted results and wrote the manuscript. Dr. Xiang Li elucidated the chemical structure of pseudenamide A. Dr. Nathan Magarvey provided ideas and guidance for the progression of this research.

5.2 Abstract

The chemical versatility of secreted metabolome of *Pseudomonas aeruginosa* is one of the differentiating factors that set this pathogen apart from other microorganisms. Given its pathogenicity, adaptability and noted immunomodulatory capacities, and secreted agents, *P. aeruginosa* have been extensively studied. Nonetheless, genome-inferred cryptic metabolites still exist and their function with regards to virulence and pathogenicity of *P. aeruginosa* are yet to be ascribed. In this research work, we develop a platform to selectively interrogate Pseudomonas chemistries and utilization of genomic and biosynthetic predictions using the Genomes to Natural Products (GNP) platform to present the identification and characterization of a novel acylated alkaloid, pseudenamide A of nonribosomal origin, This unique metabolite features an unprecedented chemical scaffold but with structural features that presents high similarity to other potent mediators of virulence and immunomodulatory actions of *P. aeruginosa*.

5.3 Introduction

Pseudomonads are widely recognized for their ability to construct and secrete secondary metabolites that function to modulate their own behaviours, and in certain instances impact the functioning of their hosts[1-6]. Initially, chemical interactions relied on genetic approaches and tools, but the complexity of Pseudomonad metabolome and the ample evidence pointing to numerous unknowns suggest that the discovery of new secondary metabolites and chemical-chemical, chemical-biological interactions will demand integration of omics tools (genomics and metabolomics)[7-9].

Genomic data present a unique venue to identify and reveal the bioactive potential of beneficial, pathogenic and other microbial varieties. In the case of *P. aeruginosa*, genome mining is now significantly advanced by the wealth of both 90 complete genomes, with 202 more in the queue, and the existence of defined strategies to genomically detect secondary metabolic loci (e.g. those encoded by nonribosomal peptide and polyketide biosynthetic genes)[10,11]. However, despite all of the detailed investigations, there is no single standardized platform that can compile, analyze and predict secondary metabolomic differences and selectively identify new chemistry based on this genomic backdrop[12,13]. If these distinct biosynthetic loci can be targeted, new chemical discoveries and underlining microbial chemical and biological principles may be uncovered to address the remaining gap on how organisms like *P. aeruginosa* manipulate their own biology and that of their hosts.

Certainly, the immunomodulatory needs of transient but chronic infections need to circumvent the multiple immune surveillance strategies, both innate and adaptive[14]. The ability to modulate cytokine secretion is of particular importance to microbes, as it directly tunes the immune response of the host. Although the ability of *P. aeruginosa* to achieve this intricate balance is multi-pronged, the impacts of small molecules in this are in certain cases defined. For instance, quinolones involved in quorum sensing such as the Pseudomonas quinolone signal (PQS) and 4-hydroxy-2-heptylquinoline (HHQ) have been shown to impact host gene expression and down-regulate the secretion of interleukin-6 and tumor necrosis factor-α in mice[15]. In addition, acyl homoserine lactones, molecules also implicated in quorum sensing, have been shown to interact with various eukaryotic

cells and similarly to quinolones modulate immune responses[16-25]. Secretion of these immunomodulatory molecules by *P. aeruginosa* is often accompanied by the creation of elaborate delivery mechanisms for these agents, often encapsulating them within intricate membrane vesicles[26]. The above classes of compounds, quinolones and acyl homoserine lactones, are the examples of immunomodulatory agents that obey the common chemical biosynthetic theme within Pseudomonads. There seems to be a general tendency towards the production of compounds with two basic components – a nonpolar head group and an acyl tail, with every combination serving its distinct and dedicated biological role.

On the genome level, certain patterns may also be noticed and prescribed to the observed phenotypic expression of microbes. For instance, *P. aeruginosa* PA7 is a peculiar case, as it is not only a taxonomic outlier, but also a strain that suggested to have undergone a shift from highly infectious/pathogenic to one of a chronic colonizer[27]. Detailed analysis of its genome revealed the lack of several genome-encoded virulence factors that are common to virulent strains of *P. aeruginosa*, such as *P. aeruginosa* PAO1 and PA14[28-31]. The secondary metabolome of *P. aeruginosa* PA7 is therefore a useful strain to reveal agents that correlate with pathogenicity/virulence and a focus on novel molecular entities from other Pseudomonas species. Here, we report the identification and characterization of a novel acylated alkaloid of nonribosomal peptide origin, pseudenamide A, from *P. aeruginosa* PA14 using a Genomes to Natural Products (GNP) platform to define new Pseudomonad chemistries from genomic information. Although the structure is entirely distinct from known natural products, it does follow a structural theme common to Pseudomonads bioactive compounds with a structurally unique head

group unit with an appended acyl tail unit. Uniquely, this metabolite is exclusive to all model and clinical isolate strains of *P. aeruginosa*, and absent in the lesser virulent PA7 strain. Given its Pseudomonas bioactive chemical signature (non-polar head group with acylated tail) it is postulated that pseudenamide A may be involved in quorum sensing and immunomodulation.

5.4 Results and Discussion

The cladogram network of complete Pseudomonad metabolome presents a unique venue for identification of known, related, as well as new metabolites from Pseudomonad species with the assistance of biochemoinformatic platforms. Genome assessment of *P. aeruginosa* PAO1, using one of newly emerged biochemoinformatic platforms, Genomes to Natural Products (GNP), led the identification of two unique predicted biosynthetic products that displayed no common structural motifs with the rest of metabolites in the cladogram – prediction_cluster_3 and prediction_cluster_8, pertaining to PA2302 and PA3327 NRPS genes, respectively (**Figure 5.4.1A**). In contrast to the lesser virulent strain, *P. aeruginosa* PA7, GNP assessment of its genome did not generate any predictions for these two biosynthetic clusters. Interestingly, the biosynthetic product of the PA2302 NRPS gene in *P. aeruginosa* PAO1 has already been identified and was previously assigned to a unique metabolite, L-2-amino-4-methoxy-trans-3-butenoic acid (AMB), which was later reassigned to the IQS compound[32,33]. The biosynthetic products of the PA3327 gene, on the other hand, have not yet been identified. Nonetheless, there is copious literature precedent indicating the likely involvement of its products in quorum sensing and virulence of *P. aeruginosa*[34-40]. Interestingly, this NRPS gene is part of a

complex biosynthetic cluster (PA3327-36) and can be found in all virulent strains of *P. aeruginosa*, except for the lesser virulent strain, *P. aeruginosa* PA7 (**Figure 5.4.1B**). More specifically, the GNP analysis of NRPS module (PA3327) of this cluster revealed that it encodes a dipeptide consisting of serine and proline monomer units (**Figure 5.4.1C**). In addition, further module analysis of the genes in this cluster revealed an acyl carrier protein and a short fatty acid synthase genes (PA3334 and PA3333, respectively, **Figure 5.4.1C**) that are seemingly responsible for the synthesis of a short acyl tail that is likely enzymatically attached to the serine-proline dipeptide, resulting in a lipopeptide biosynthetic product. Notable is the presence of other accessory domains such as monooxygenase and cytochrome P450 (PA3328 and PA3331, respectively, **Figure 5.4.1C**), both likely engaged in chemical modification of this product.

Chromatographic subtraction of the LC-MS data pertaining to the culture extracts of wild-type and NRPS mutant strains of *P. aeruginosa* PA14 revealed a unique ion with *m/z* 193.1 (**Supplementary Figure S5.6.1**). As expected, the extracted ion chromatogram of *m/z* 193.1 in the cluster-lacking *P. aeruginosa* PA7 strain was identical to that of the NRPS mutant. Moreover, using a panel of *P. aeruginosa* PA14 mutant strains[41], it became evident which genes were necessary for the biosynthesis of this product. There was no observable production *m/z* 193.1 in the NRPS mutant strain and its production was severely impaired in four other mutant strains (**Figure 5.4.1D**). To further probe the biosynthetic origin of this metabolite, feeding experiment was conducted using isotopically labeled acetate, serine and proline. Compared to the unlabeled acetate supplementation, this molecule incorporated two and three acetate units, *m/z* 195.1 and

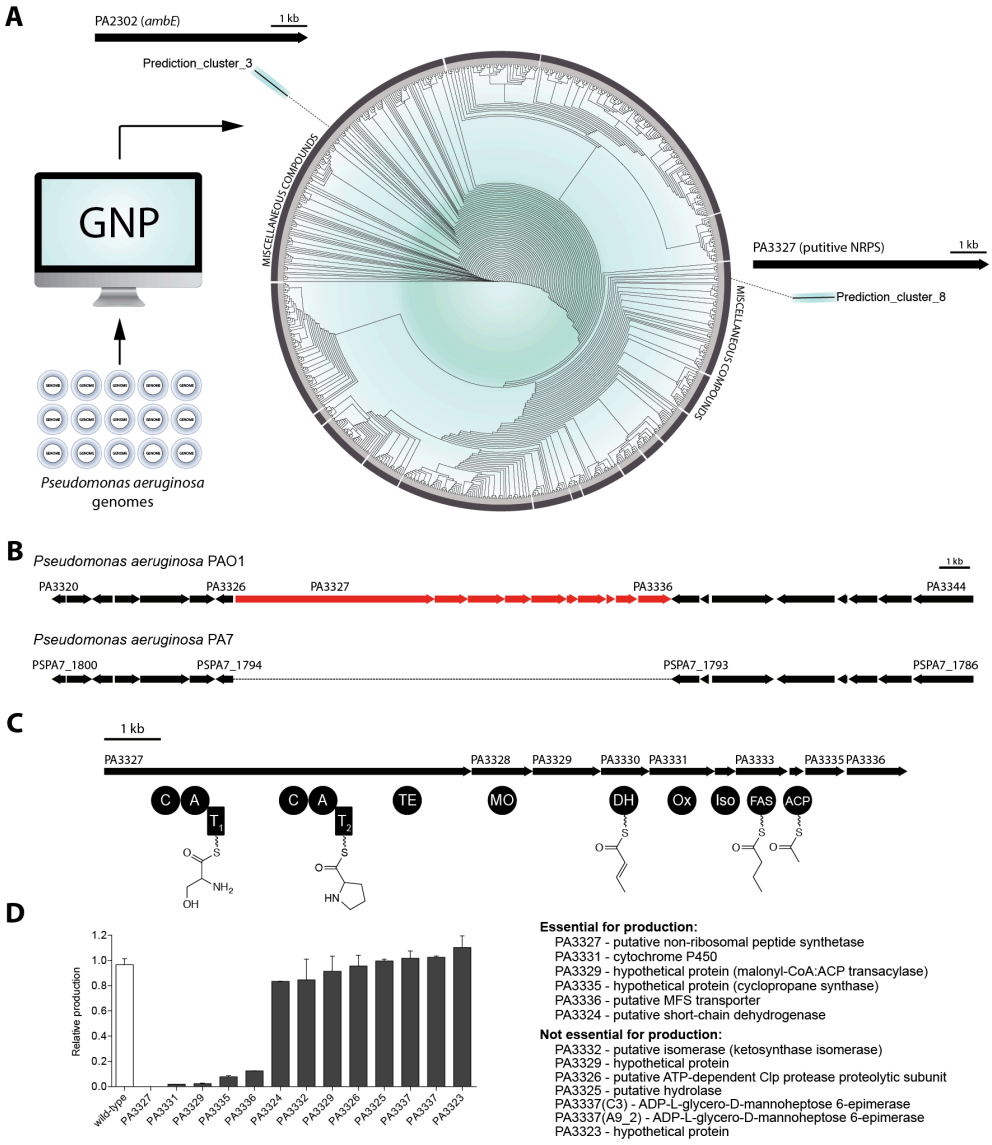


Figure 5.4.1. Identification of cryptic nonribosomal peptide from *Pseudomonas aeruginosa*. **A** – locations of unique cluster predictions identified by GNP on the metabolomic cladogram of Pseudomonads; **B** – putative NRPS cluster present in *P. aeruginosa* PAO1 is absent in the PA7 strain[27]; **C** – domain organization of the entire cluster from *P. aeruginosa* PAO1 with predicted monomer units; **D** – determination of genes essential for the production of NRPS product.

196.1, respectively, suggesting a likely short fatty acid chain that was consistent with the prediction (**Supplementary Figure S5.6.2A-B**). Surprisingly, both serine and proline with a single ^{13}C label in the respective carboxylic acid did not show any incorporation and *m/z* 193.1 was observed (**Supplementary Figure S5.6.2C, E**). However,

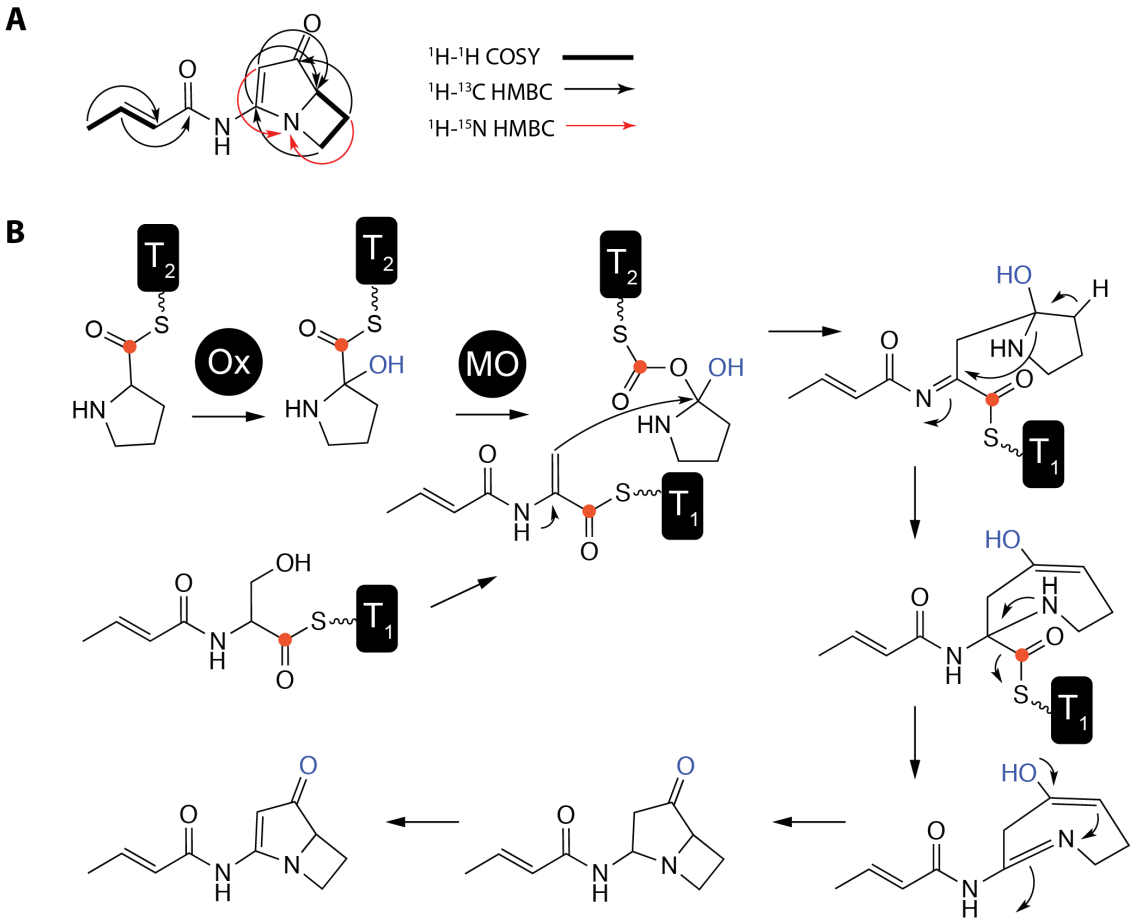


Figure 5.4.2. Chemical structure of pseudenamide A and its proposed biosynthesis. **A** – chemical structure of pseudenamide A and COSY and HMBC correlations; **B** – proposed biosynthesis of pseudenamide A, red dots indicates the labeled residues that do not get incorporated in the final structure based on the isotopically labeled feeding experiments.

supplementation of these amino acids labeled at different positions revealed near perfect incorporation (**Supplementary Figure S5.6.2D, F**). This discrepancy suggested that perhaps the two carboxylic acids were likely lost during the biosynthesis of the final product.

To reveal this peculiar biosynthetic product, *m/z* 193.1 compound was isolated, structurally elucidated and named pseudenamide A (**Supplementary Figure S5.6.3-10**). The chemical structure of this complex small molecule features an unprecedented chemical scaffold that is not only unique to Pseudomonads, but also has never been previously observed in nature (**Figure 5.4.2A**). The closest chemical that shares some similarity to pseudenamide A is jenamidine A_1 (**Supplementary Figure S5.6.11**), and the NMR chemical shifts are in keeping with the common structural elements[42,43]. As expected from the domain analysis, pseudenamide A contains a short acyl tail with an unsaturation likely arising from the action of dehydratase also present in this cluster (PA3330, **Figure 5.4.1C**). The chemical complexity of the cyclic core of pseudenamide A makes it challenging to decipher its biosynthesis. Nonetheless, its biosynthesis likely features enzymatic action of both cytochrome P450 and monooxygenase (**Figure 5.4.2B**). Similarly to the action of LpiC monooxygenase that is responsible for the dehydration of serine to dehydroalanine[44], monooxygenase present in this cluster is likely performing a similar function. Although a detailed enzymatic analysis is necessary to determine the exact assembly mechanism of pseudenamide A, atypical losses of the carboxylic acids of serine and proline are in keeping with the proposed biosynthesis.

Enzymatic promiscuity is commonly observed in biosynthetic patterns of known lipid-baring molecules and is often reflected in the ability to incorporate acyl tails of varying length[45]. The biosynthetic assembly line of pseudenamide A is also slightly promiscuous and leads to the generation of its fatty acid analogs. Knowing the structure of pseudenamide A, its fragmentation pattern was elucidated with tandem mass spectrometry. Given its structural rigidity, only two main fragments are generated under collision-induced fragmentation, corresponding to the cyclic core and acyl tail moieties that arise from cleavage of the amide bond (**Supplementary Figure S5.6.12A**). From the fragmentation analysis and the expected *m/z* value, pseudenamide B (*m/z* 207.1) was identified (**Supplementary Figure S5.6.12B**). This analog was also exclusive to the wild-type of *P. aeruginosa* only and was likely generated from the incorporation of a longer fatty acid, as the *m/z* pertaining to the cyclic core-containing fragment (*m/z* 124.9) remained unchanged. To further confirm the structures of assigned fragments, pseudenamide A was subjected to tandem high-resolution mass spectrometry and their elemental composition was validated (**Supplementary Figure S5.6.13A**). Interestingly, *P. aeruginosa* likely also creates even longer fatty acid analog, pseudenamide C (**Supplementary Figure S5.6.13B**). Although pseudenamide C has both the expected *m/z* value (*m/z* 221.1) and chromatographic retention time, its fragmentation profile was not acquired due to its low abundance.

Quorum sensing in Pseudomonads is one of the key differentiating features that surpass the cell-to-cell communication capacity of many other microbial organisms. The production of signaling molecules such as N-butanoyl-L-homoserine lactone (C4-HSL)

and N-3-oxododecanoyl-homoserine lactone (3-oxo-C12-HSL) is critical for quorum sensing, as it activates and controls a multitude of virulence factors[46]. The activation of quorum sensing and consequent production of C4-HSL and 3-oxo-C12-HSL has been shown to activate the NRPS cluster[34-40], and it is the first time that its biosynthetic

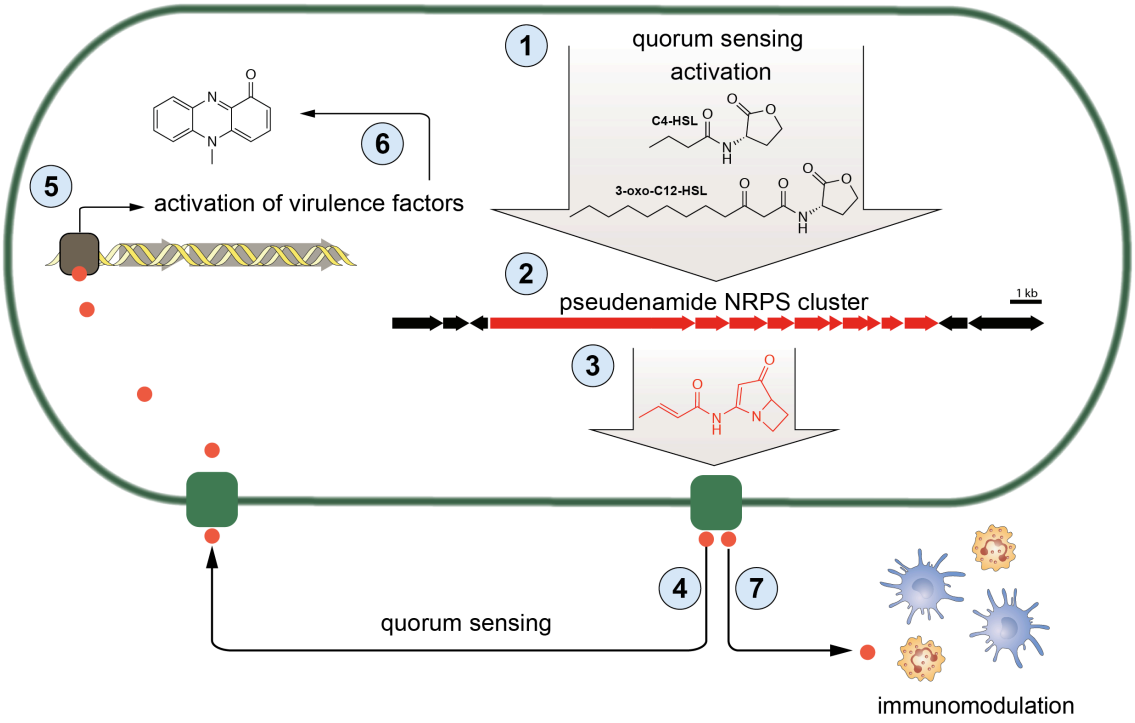


Figure 5.4.3. Pseudenamide A in the context of quorum sensing of *Pseudomonas aeruginosa*. **1** – quorum sensing activation leads to production of C4-HSL and 3-oxo-C12-HSL; **2** – activation of pseudenamide NRPS cluster[34-40]; **3** – production of pseudenamide A; **4** – likely quorum sensing activity of pseudenamide A; **5** – possible activation of known or novel virulence factors; **6** – consequent production of toxic agents like pyocyanin; **7** – likely immunomodulatory activity of pseudenamide A.

product, pseudenamide A, is being revealed (**Figure 5.4.3**). In support to regulation by cell-density signaling molecules, the production of pseudenamide A seemingly depends highly on the cell density, suggesting that it may also be involved in quorum sensing (**Supplementary Figure S5.6.14A**). Unlike PQS production, when overlaid, the relative production of pseudenamide A closely follows that of C4-HSL, production of which is cell density dependent (**Supplementary Figure S5.6.14B**). The relative production of pseudenamide A and C4-HSL do diverge, however, likely due to the lactone stability of C4-HSL that leads to its eventual disappearance. In addition, the chemical structure of pseudenamide A is highly reminiscent of C4-HSL, again suggesting its potential involvement in cell signaling and activation of known and perhaps even new virulence factors. Lastly, when tested for its antimicrobial properties, pseudenamide A did not display any notable antibacterial or antifungal activities (**Supplementary Figure S5.6.15**). Given the increasing evidence of *P. aeruginosa* metabolites having immunomodulatory properties[15,22-24], another likely function of pseudenamide A is immunomodulation and should be carefully assessed using appropriate immune cell lines, including macrophages, monocytes, dendritic cells and neutrophils.

5.5 Materials and Methods

5.5.1 Bacterial strains and fermentation conditions

The wild-type *Pseudomonas aeruginosa* PA14 strain, strains from the PA14 non-redundant set of *Pseudomonas aeruginosa* transposon insertion mutants with inactivated PA3327, PA3331, PA3329, PA3335, PA3336, PA3323, PA3324, PA3332, PA3329, PA3326, PA3325, PA3337 (C3), PA3337 (A9_2) genes; *P. aeruginosa* PA7 and *P. aeruginosa*; genetically identical clinical strains isolated ten years apart from a cystic fibrosis patient, *P. aeruginosa* P590 and P48 (isolated last); and eight clinical isolates from a single sputum sample collected from a different patient also with cystic fibrosis, *P. aeruginosa* TSB001-018, were all grown overnight in YM medium (per 1 L: bacteriological grade peptone (Amresco, J636-1KG), 5 g; yeast extract (Difco Laboratories, 212750), 3 g; malt extract (HiMedia, 61001-518), 3 g; glucose (dextrose, Caledon, 3260-1), 1 g) and were used to inoculate (1:100 dilution) 50 mL of YM medium (n = 3) in 250 mL flasks that were incubated for 12 hours at 37°C, 220 rpm. The wild-type *P. aeruginosa* PA14 was also cultured in two minimal media - M9 (Amresco, J863-500G) and asparagine proline broth (Sigma, 17129-500G), as well as, tryptic soy broth (EMD, Millipore #1.05459.5007) and LB broth (EMD, Millipore #1.10285.0500) media using the above conditions. Feeding experiments were conducted using M9 medium supplemented with L-serine-1-^{13}C, L-serine-3-^{13}C, L-proline-1-^{13}C or L-proline-$^{13}C_5$, ^{15}N, added aseptically and directly the medium (50 mg/50 mL).

5.5.2 Production and extraction optimizations

To determine the optimal production conditions, the wild-type *P. aeruginosa* PA14 was incubated in YM medium with varied temperature (28°C, 37°C and 42°C), aeration (200, 250 and 350 rpm) and volume (25 mL and 50 mL in 250 mL flask) and time (6, 12, 24, 48, 72 hours) in replicate (n = 3). Cultures were extracted using liquid-liquid extraction using ethyl acetate (2x, 2:1, v/v). The extracts were then dried *in vacuo* and resuspended in water:methanol (50:50) and analyzed using Bruker micrOTOF II mass spectrometer (positive ESI, scanning *m/z* 100-1000) with an Agilent 1200 series HPLC with Ascentis Express C18 column (150 mm x 2.1 mm, 2.7 µm, Sigma) and A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) as the mobile phases at 0.25 mL/min. The solvent gradient was 0-5 min, 5% A; 5-45 min, linear to 100% A; 45-55 min, isocratic at 100% A, 55-60 min, linear to 5% A; 60-65 min, isocratic at 5% A.

To determine the optimal extraction, the following resins were used added directly to the cultures at 5 g per 50 mL: Amberlite® IRA-67 free base (Sigma, A9960-500G), Amberlite® XAD7HP (Sigma, XAD7-500G), Diaion® HP-20 (Sigma, 13607), Diaion® HP-20SS (Sigma, 13615-U), SP Sepharose® (Sigma, S1799-500ML), silica gel (Sigma, 236772-1KG), C18 silica gel (Sigma, 97727-U), MCI GEL® CHP20P (Sigma, 13630-U). In addition, ethyl acetate extraction was performed using the bacterial cultures (2x, 2:1, v/v). Eluents were evaporated *in vacuo*, resuspended in water:methanol (50:50) and analyzed as above.

5.5.3 Extraction of pseudenamide A

Large scale fermentation was carried out by using *P. aeruginosa* PA14 overnight culture to inoculate 200 2.8 L Fernbach flasks (1:100 dilution) containing 500 mL of YM medium that were incubated for 12 hours at 28°C, 350 rpm. Cultures were extracted with 5 g Diaion® HP-20SS per 50 mL of culture, incubated for 1 hour at 28°C, 350 rpm. Resin was collected using vacuum filtration and eluted with methanol. The eluent was combined, evaporated *in vacuo*, and then extracted using liquid-liquid partitioning with ethyl acetate (2:1 v/v, 4x) and the resulting organic phase concentrated *in vacuo*. The crude extract was resuspended in water:methanol (50:50, v/v) and fractionated using a Combiflash® Rf reverse phase MPLC system (Teledyne Technologies Inc.) with a 50 g RediSep® Rf Gold HP C18 flash column and 40 mL/min flow rate. The following program of solvent A (acetonitrile) and B (water) was employed: 0-2 min, 5% A; 2-11 min, a linear gradient to 100% A; 11-13 min, isocratic at 100% A; 13-14 min, linear gradient to 5% A, 14-15 min, isocratic at 5% A. Resulting fractions were screened for the presence of compounds of interest by direct infusion using Bruker amaZon X ion trap mass spectrometer operating in positive ESI mode, scanning *m/z* 100-500. Fractions that contained the compounds of interest were combined and subjected to preparatory HPLC-ESI-MS (Dionex UltiMate 3000 HPLC system with 160-600 nm variable UV detection connected in line with Bruker amaZon X ion trap mass spectrometer operating in positive ESI mode, and Phenomenex Luna 5µ C18(2) (250 mm x 15 mm, 4.6 µm) with 10 mL/min flow rate). Optimal separation was achieved using the following A (acetonitrile), B (water) and C (methanol) solvent gradient: 0-5 min, 5% A; 5-8 min, linear gradient to

12% A; 8-28 min, linear gradient to 26% A; 28-30 min, linear gradient to 40% A; 30-36 min, isocratic at 100% C, 36 min, return to 40% A, 36-38 min, linear to 5% A; 38-45 min, isocratic at 5% A. Elution of pseudenamide A and B was observed at 18.0 and 25.7 min, respectively. High-resolution mass spectral data for pseudenamide A were obtained using a Thermo Scientific LTQ Orbitrap XL. NMR experiments were performed using a Bruker Avance 700 MHz spectrometer with samples dissolved in 100% methanol-d_4 with a hint of DCl vapour to enhance solubility.

5.5.4 Identification of analogs

The MS/MS analysis was carried out on Diaion® HP-20SS eluent of *P. aeruginosa* PA14 culture using Bruker amaZon X ion trap mass spectrometer (operating in autoMS2, positive ESI mode, with scanning range of *m/z* 100-500) coupled with a Dionex UltiMate 3000 HPLC, with an Ascentis Express C18 column (150 mm × 4.6 mm, 2.7µ, Sigma Aldrich) and A (acetonitrile with 0.1% formic acid) and B (water with 0.1% formic acid) as the mobile phases at 1.2 mL/min. The separation was achieved using the following solvent gradient: 0-5 min, 5% A; 5-45 min, linear to 100% A; 45-55 min, isocratic at 100% A, 55-60 min, linear to 5% A; 60-65 min, isocratic at 5% A.

5.5.5 Assessment of antimicrobial activity

The overnight cultures of *P. aeruginosa* PA14 (in TSB), *Staphylococcus aureus* USA300 (in TSB), *Bacillus subtilis* (in LB), *Escherichia coli* (in LB), grown at 37°C, 250 rpm; *Micrococcus luteus* (in TSB), *Candida albicans* (in YNB (Bioshop® Canada Inc., YNB406.100, supplemented with 0.1% maltose and 0.2% glucose), *Saccharomyces cerevisiae* (in YPD (yeast extract, 10 g, peptone, 20 g, dextrose, 20 g)), grown at 28°C,

250 rpm; *Lactobacillus plantarum* (in MRS (de Man, Rogosa and Sharpe medium, BD sciences), *Streptococcus mutans* (in THYE (Todd-Hewitt broth supplemented with 3 g/L yeast extract), grown at 37°C, no shaking; were used to inoculate (1:100 dilution) 100 µL of fresh respective medium with pseudenamide A (100 µM final concentration) or medium with water control and incubated using respective conditions. The optical density at 600 nm (OD_{600}) was measured to assess the density of the cultures after 15 hours.

5.6 Supplementary Figures and Legends (Figure S5.6.1-15)

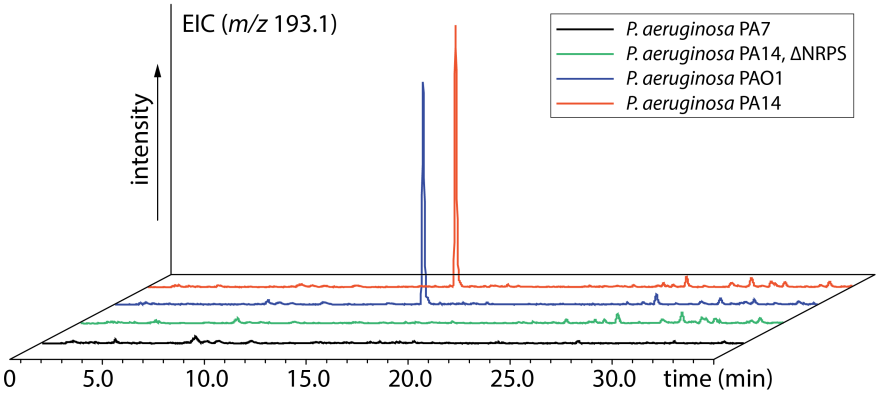


Figure S5.6.1. Extracted ion chromatogram of the metabolite related to the cryptic NRPS cluster from *Pseudomonas aeruginosa* PA14. The identified metabolite with *m/z* 193.1 is present in both *P. aeruginosa* PAO1 and PA14 and absent in the ΔNRPS strain of *P. aeruginosa* PA14 and *P. aeruginosa* PA7, strain naturally lacking the NRPS cluster.

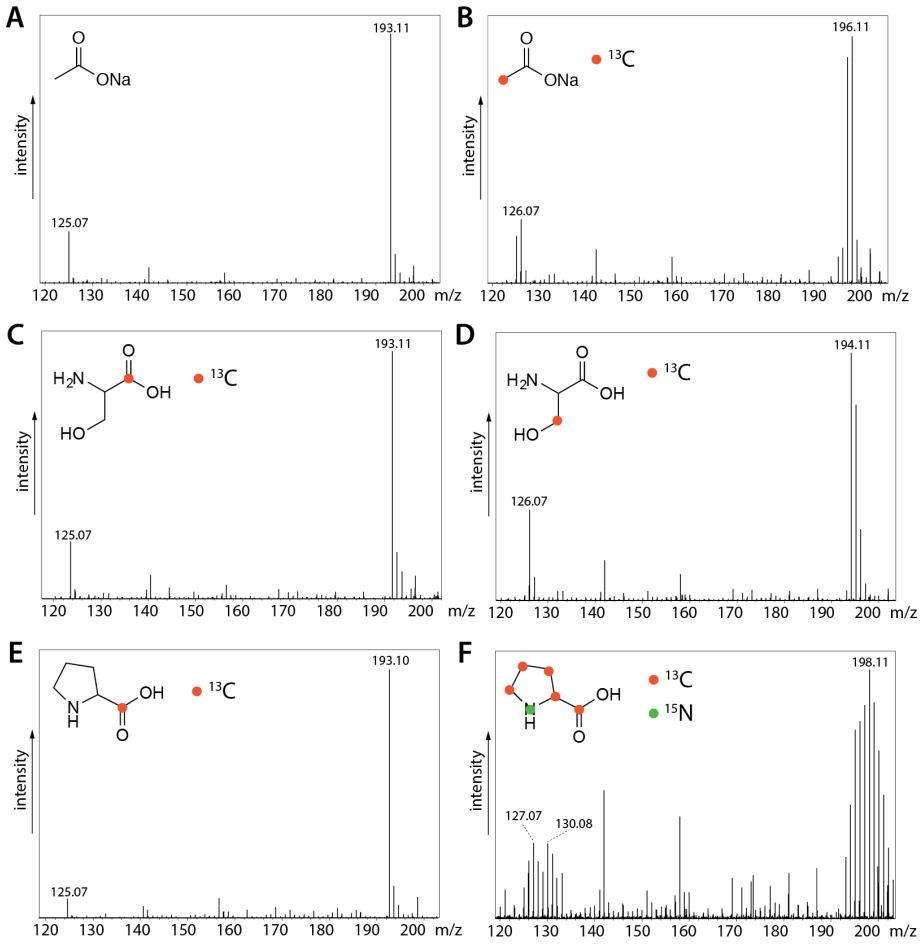


Figure S5.6.2. Incorporation of predicted isotopically labeled amino acid precursors in *P. aeruginosa* PA14. **A** – mass spectrum of peak corresponding to *m/z* 193.1 from the acetate supplemented culture. **B** – mass spectrum of peak corresponding to *m/z* 193.1 from the acetate-^{13}C supplemented culture. **C** – mass spectrum of peak corresponding to *m/z* 193.1 from the L-serine-1-^{13}C supplemented culture. **D** – mass spectrum of peak corresponding to *m/z* 193.1 from the L-serine-3-^{13}C supplemented culture. **E** – mass spectrum of peak corresponding to *m/z* 193.1 from the L-proline-1-^{13}C supplemented culture. **F** – mass spectrum of peak corresponding to *m/z* 193.1 from the L-Proline-$^{13}C_5$,^{15}N supplemented culture.

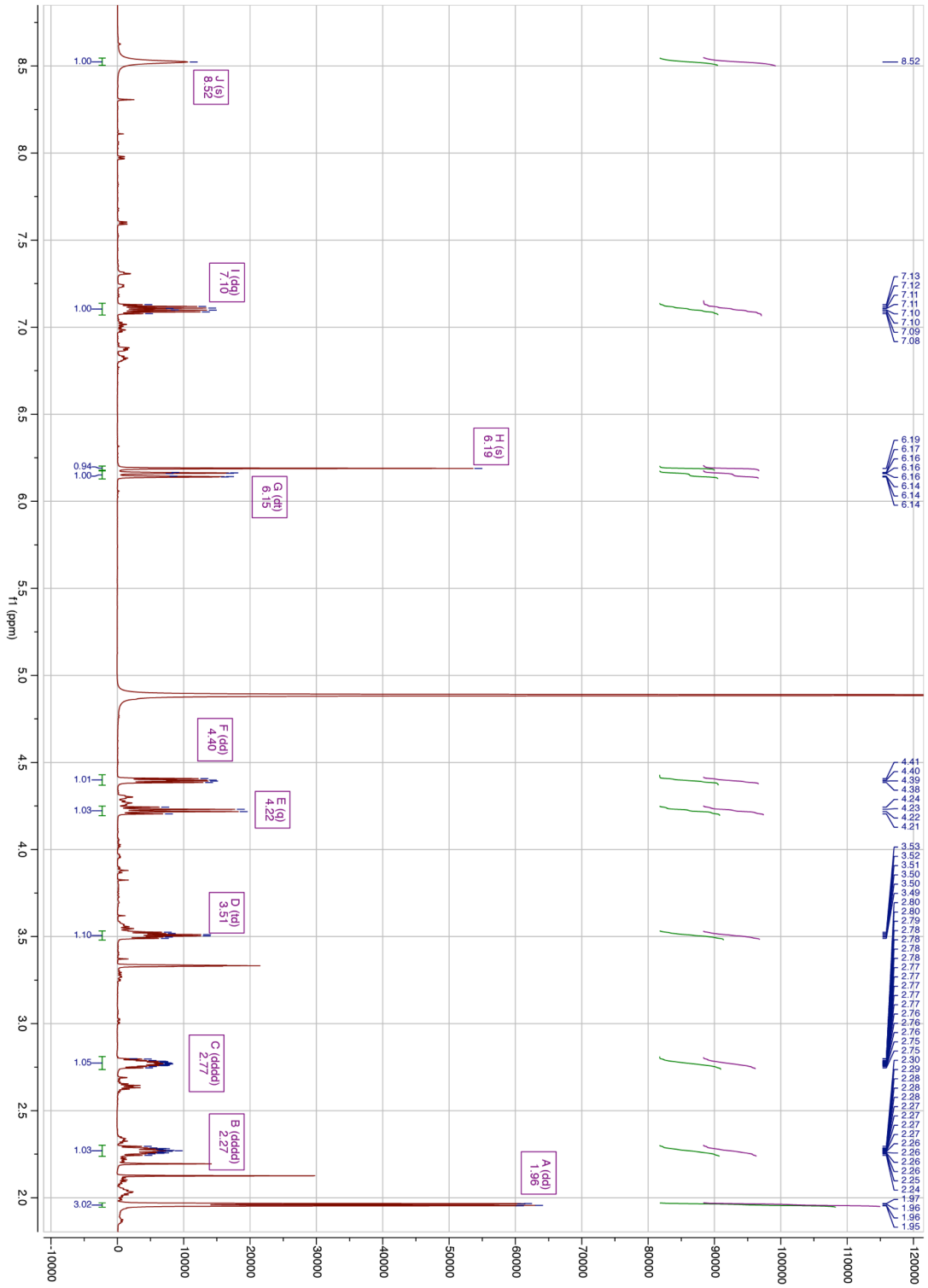


Figure S5.6.3. ^{1}H spectrum of pseudenamide A in 100% methanol-d_4.

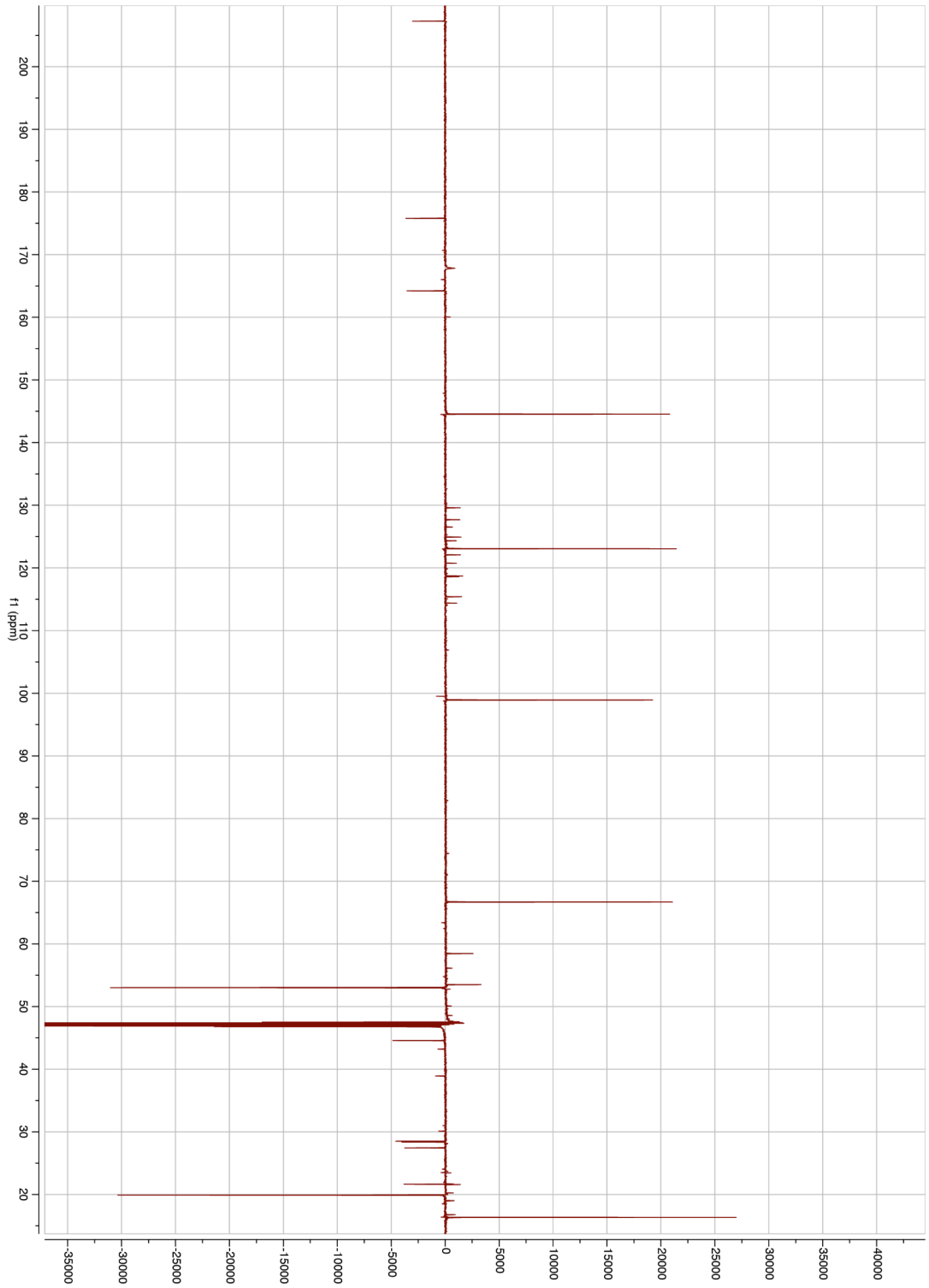


Figure S5.6.4. ^{13}C DEPTq spectrum of pseudenamide A in 100% methanol-d_4.

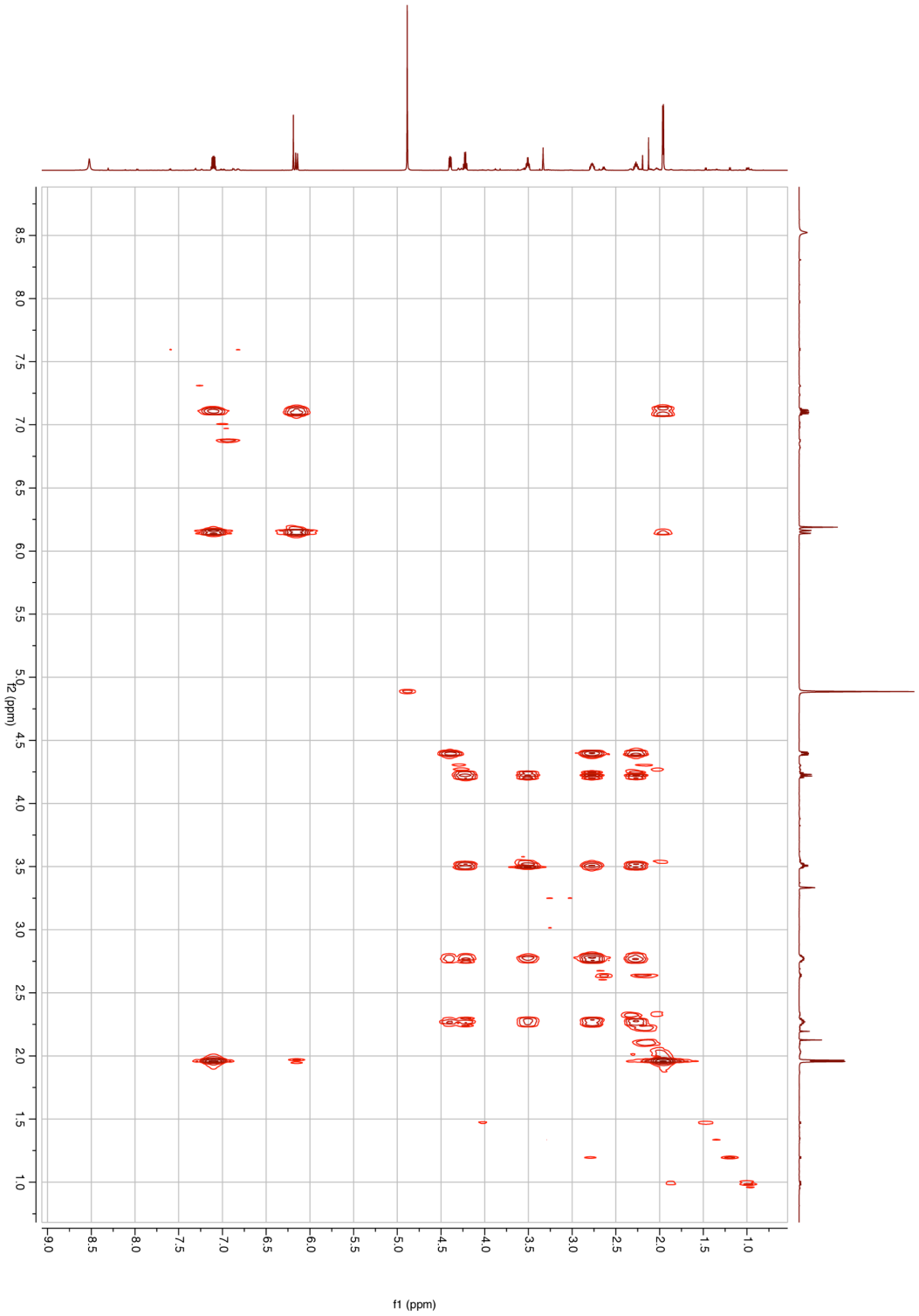


Figure S5.6.5. 2D COSY spectrum of pseudenamide A in 100% methanol-d_4.

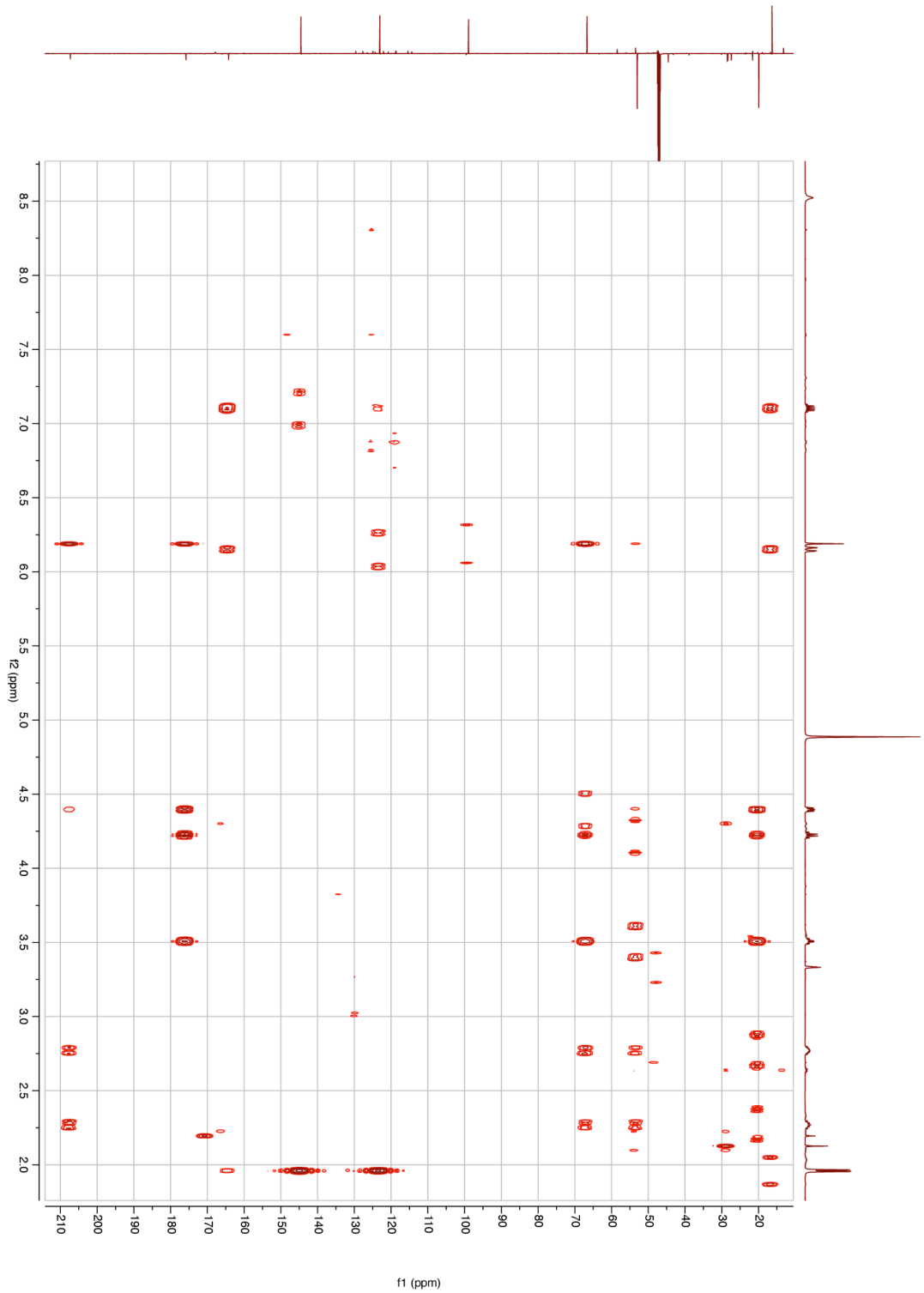


Figure S5.6.6. 2D HMBC spectrum of pseudenamide A in 100% methanol-d_4.

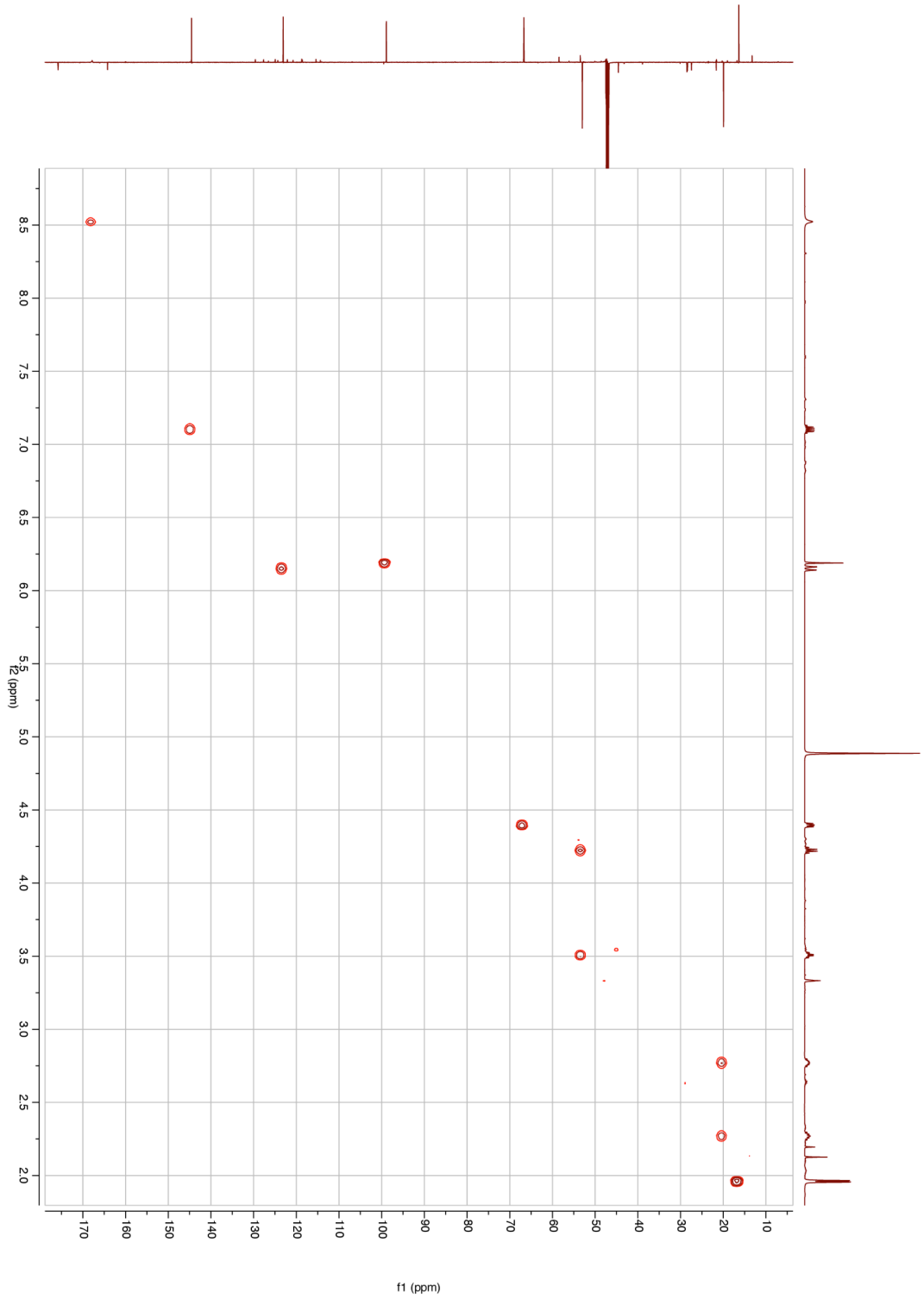


Figure S5.6.7. 2D HSQC spectrum of pseudenamide A in 100% methanol-d_4.

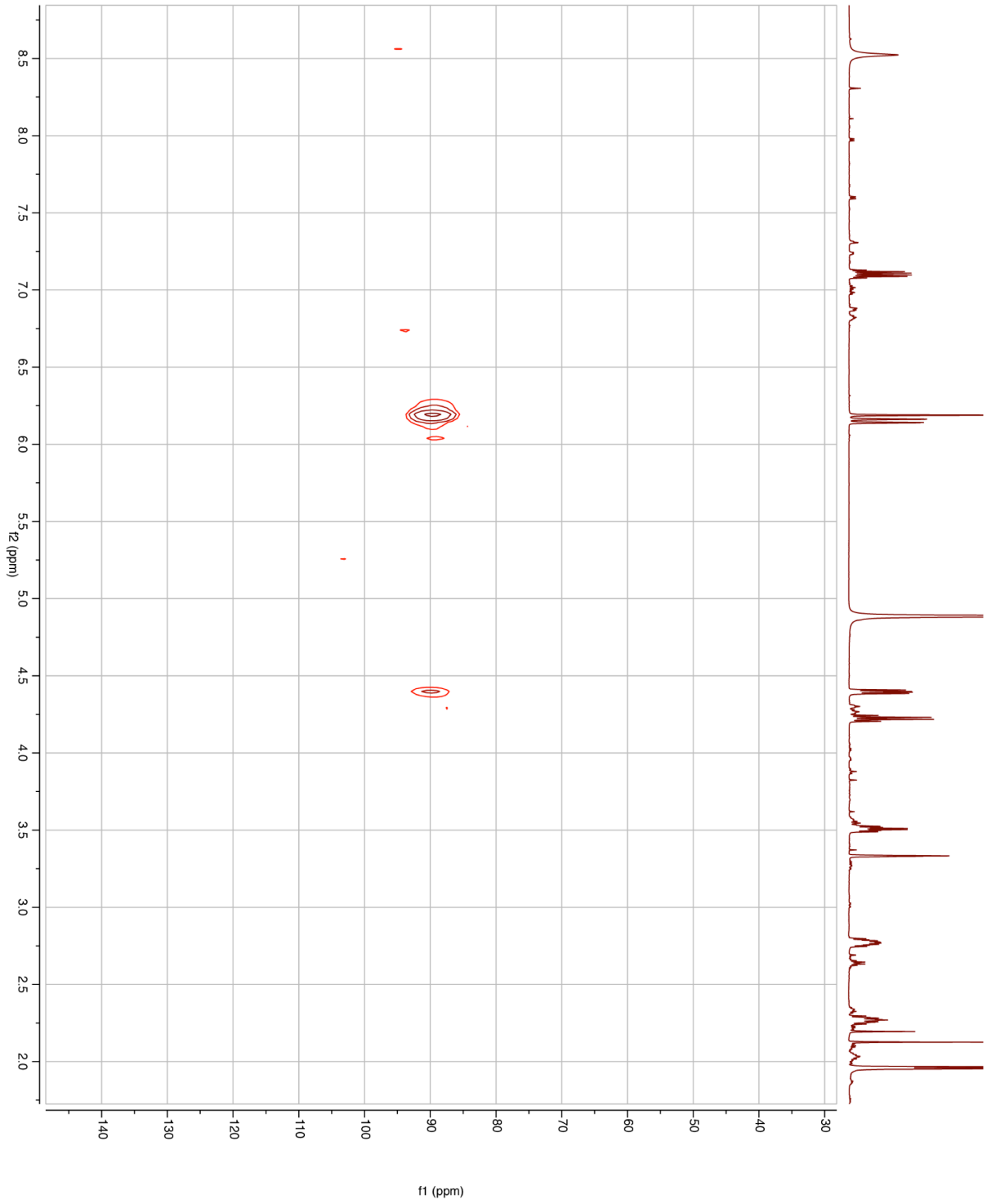


Figure S5.6.8. 2D HMBC (^{1}H-^{15}N, J_{HN} = 7 Hz) spectrum of pseudenamide A in 100% methanol-d_4.

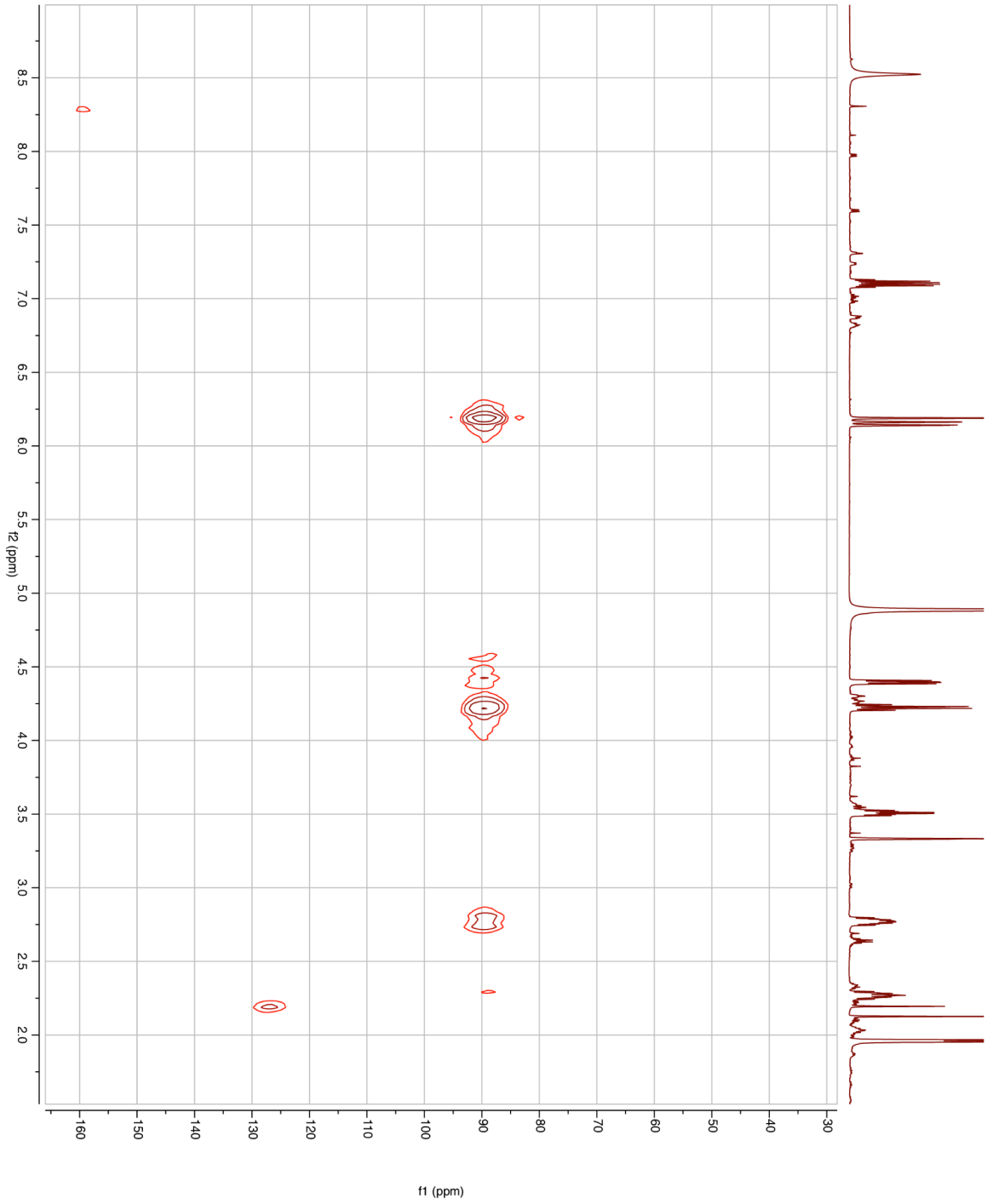


Figure S5.6.9. 2D HMBC (^{1}H-^{15}N, J_{HN} = 3 Hz) spectrum of pseudenamide A in 100% methanol-d_4.

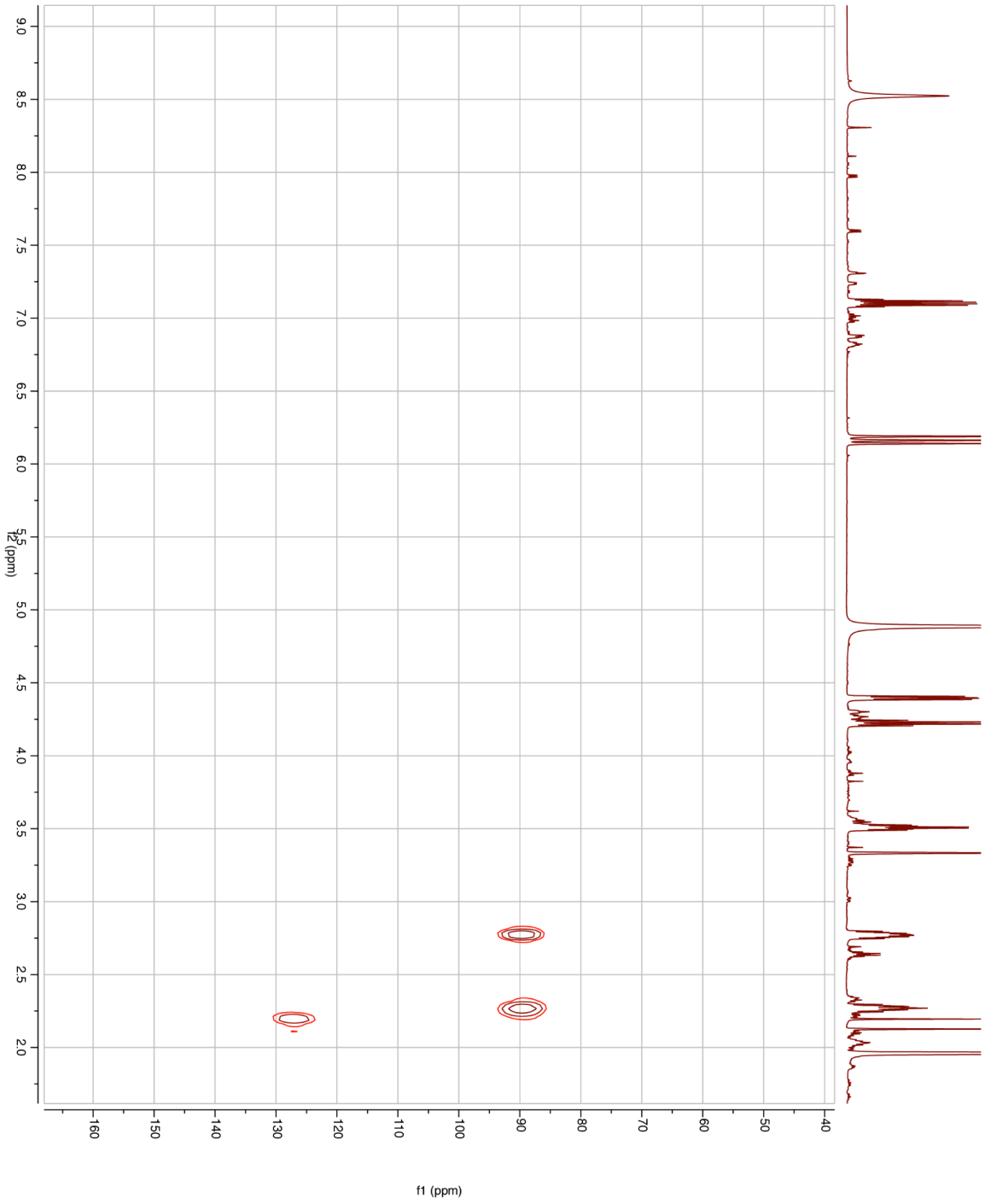


Figure S5.6.10. 2D HMBC (^{1}H-^{15}N, J_{HN} = 1.5 Hz) spectrum of pseudenamide A in 100% methanol-d_4.

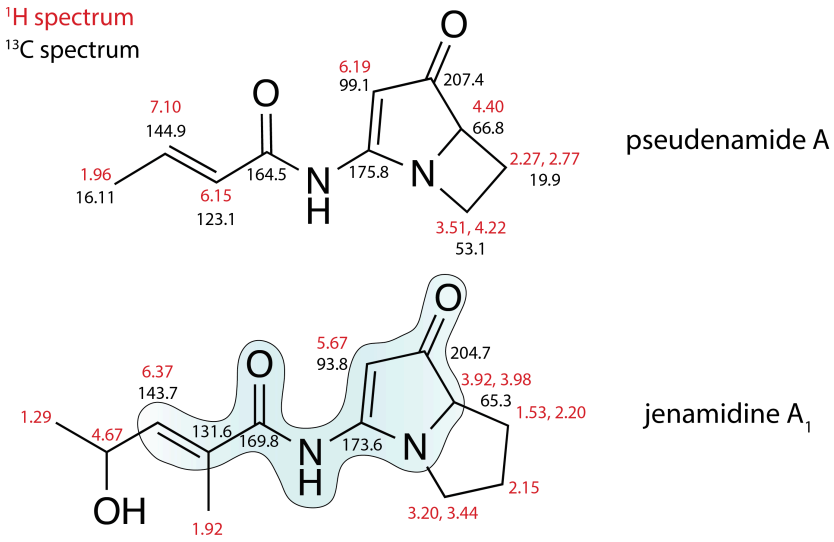


Figure S5.6.11. Comparison of ^{1}H and ^{13}C chemical shifts observed in pseudenamide A and jenamidine A_1. Structural components similar between the two compounds are highlighted in blue using jenamidine A_1 structure[42,43].

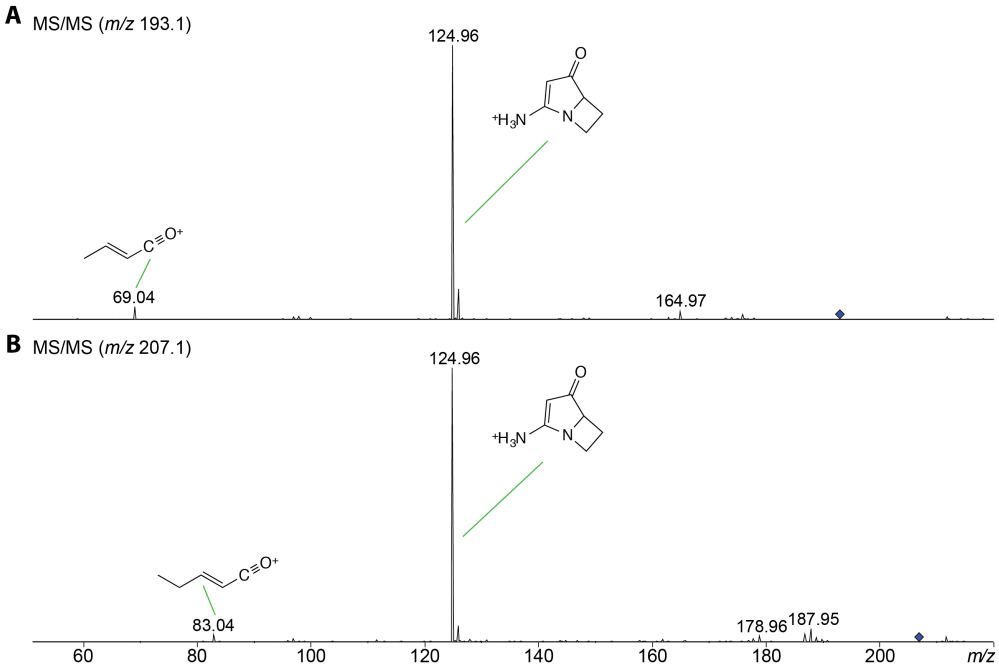


Figure S5.6.12. Fragmentation patterns of pseudenamide A and B. **A** – fragmentation of pseudenamide A; **B** – fragmentation of pseudenamide B.

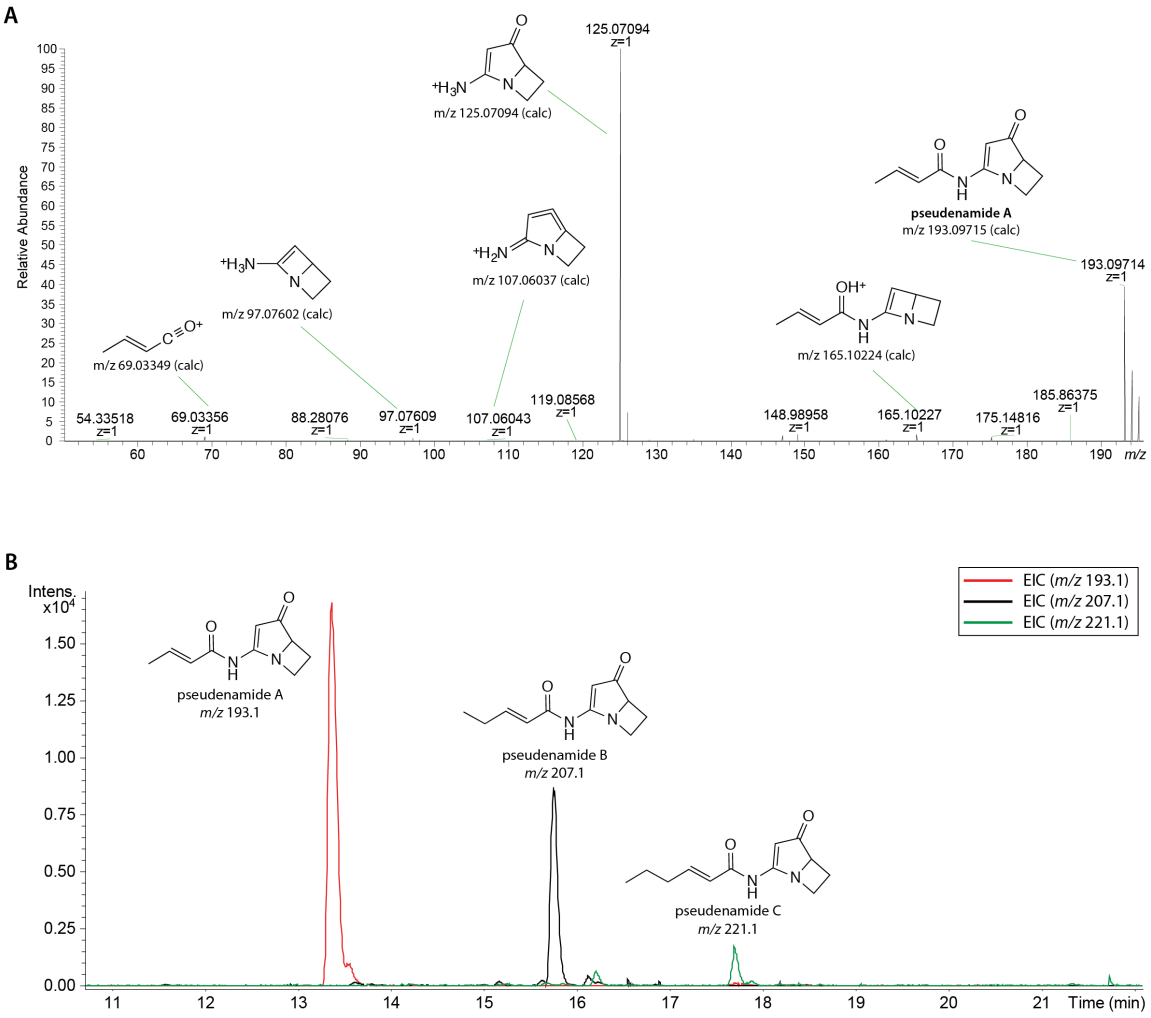


Figure S5.6.13. Structural validation of pseudenamide A and production profile of its analogs. **A** – validation of pseudenamide A fragments under CID using tandem high-resolution mass spectrometry; **B** – extracted ion chromatogram of pseudenamide A, B and C.

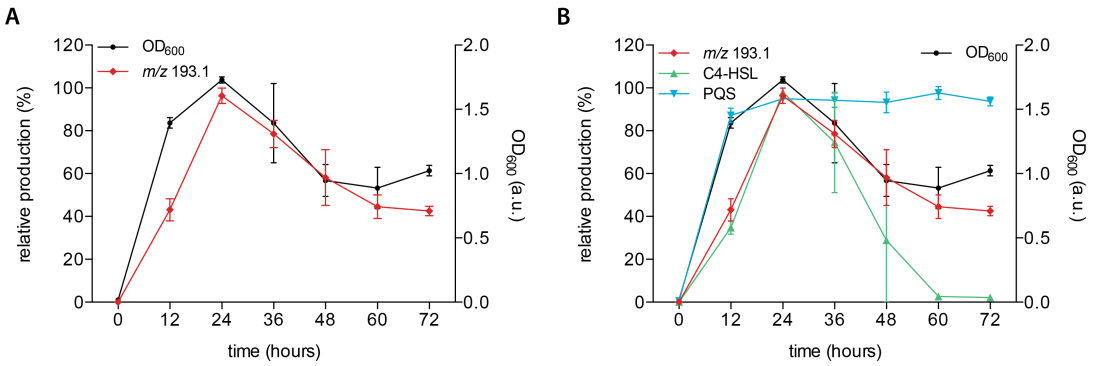


Figure S5.6.14. Time-course relative production of pseudenamide A and other quorum sensing molecules. **A** – relative production of pseudenamide A (*m/z* 193.1) along side with the optical density of the culture at 0-72 hours; **B** – overlay of relative productions of pseudenamide A (*m/z* 193.1), N-butanoyl-L-homoserine lactone (C4-HSL) and Pseudomonas quinolone signal (PQS) along side the optical density of the culture at 0-72 hours.

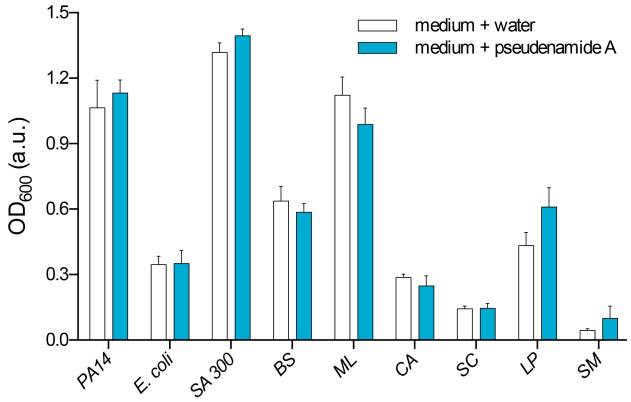


Figure S5.6.15. Antimicrobial activity of pseudenamide A. PA14 – *Pseudomonas aeruginosa* PA14; E. coli – *Escherichia coli*; SA 300 – *Staphylococcus aureus* USA300; BS – *Bacillus subtilis*; ML – *Micrococcus luteus*; CA – *Candida albicans*; SC – *Saccharomyces cerevisiae*; LP – *Lactobacillus plantarum* WCFS1; SM – *Streptococcus mutans* UA159.

5.7 References

1. Dubern, J.-F. & Diggle, S. P. Quorum sensing by 2-alkyl-4-quinolones in Pseudomonas aeruginosa and other bacterial species. *Mol Biosyst* **4,** 882–888 (2008).
2. Khajanchi, B. K., Kirtley, M. L., Brackman, S. M. & Chopra, A. K. Immunomodulatory and protective roles of quorum-sensing signaling molecules N-acyl homoserine lactones during infection of mice with Aeromonas hydrophila. *Infection and Immunity* **79,** 2646–2657 (2011).
3. Mattmann, M. E. & Blackwell, H. E. Small molecules that modulate quorum sensing and control virulence in Pseudomonas aeruginosa. *J. Org. Chem.* **75,** 6737–6746 (2010).
4. Bjarnsholt, T. *et al.* Quorum Sensing and Virulence of Pseudomonas aeruginosa during Lung Infection of Cystic Fibrosis Patients. *PLoS ONE* **5,** e10115 (2010).
5. Jarosz, L. M., Ovchinnikova, E. S., Meijler, M. M. & Krom, B. P. Microbial spy games and host response: roles of a Pseudomonas aeruginosa small molecule in communication with other species. *PLoS Pathog* **7,** e1002312 (2011).
6. Sifri, C. D. Quorum sensing: bacteria talk sense. *Clinical infectious diseases* **47,** 1070–1076 (2008).
7. Frimmersdorf, E., Horatzek, S., Pelnikevich, A., Wiehlmann, L. & Schomburg, D. How Pseudomonas aeruginosa adapts to various environments: a metabolomic approach. *Environ. Microbiol.* **12,** 1734–1747 (2010).
8. Gjersing, E. L., Herberg, J. L., Horn, J., Schaldach, C. M. & Maxwell, R. S. NMR

metabolomics of planktonic and biofilm modes of growth in Pseudomonas aeruginosa. *Anal. Chem.* **79,** 8037–8045 (2007).

9. Oberhardt, M. A., Puchałka, J., Fryer, K. E., Martins dos Santos, V. A. P. & Papin, J. A. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. *J. Bacteriol.* **190,** 2790–2803 (2008).
10. Donadio, S., Monciardini, P. & Sosio, M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* **24,** 1073–1109 (2007).
11. Letzel, A.-C., Pidot, S. J. & Hertweck, C. A genomic approach to the cryptic secondary metabolome of the anaerobic world. *Nat Prod Rep* **30,** 392–428 (2013).
12. Krug, D. & Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat Prod Rep* (2014). doi:10.1039/c3np70127a
13. Boddy, C. N. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J. Ind. Microbiol. Biotechnol.* **41,** 443–450 (2014).
14. Hornef, M. W., Wick, M. J., Rhen, M. & Normark, S. Bacterial strategies for overcoming host innate and adaptive immune responses. *Nat. Immunol.* **3,** 1033–1040 (2002).
15. Kim, K. *et al.* HHQ and PQS, two Pseudomonas aeruginosa quorum-sensing molecules, down-regulate the innate immune responses through the nuclear factor-κB pathway. *Immunology* **129,** 578–588 (2009).
16. Shiner, E. K. *et al.* Pseudomonas aeruginosa autoinducer modulates host cell

responses through calcium signalling. *Cell. Microbiol.* **8,** 1601–1610 (2006).

17. Williams, S. C. *et al.* Pseudomonas aeruginosa autoinducer enters and functions in mammalian cells. *J. Bacteriol.* **186,** 2281–2287 (2004).

18. Tateda, K. *et al.* The Pseudomonas aeruginosa autoinducer N-3-oxododecanoyl homoserine lactone accelerates apoptosis in macrophages and neutrophils. *Infection and Immunity* **71,** 5785–5793 (2003).

19. Smith, R. S., Kelly, R., Iglewski, B. H. & Phipps, R. P. The Pseudomonas autoinducer N-(3-oxododecanoyl) homoserine lactone induces cyclooxygenase-2 and prostaglandin E2 production in human lung fibroblasts: implications for inflammation. *J. Immunol.* **169,** 2636–2642 (2002).

20. Smith, R. S., Harris, S. G., Phipps, R. & Iglewski, B. The Pseudomonas aeruginosa quorum-sensing molecule N-(3-oxododecanoyl) homoserine lactone contributes to virulence and induces inflammation in vivo. *J. Bacteriol.* **184,** 1132–1139 (2002).

21. Smith, R. S. *et al.* IL-8 production in human lung fibroblasts and epithelial cells activated by the Pseudomonas autoinducer N-3-oxododecanoyl homoserine lactone is transcriptionally regulated by NF-kappa B and activator protein-2. *J. Immunol.* **167,** 366–374 (2001).

22. Telford, G. *et al.* The Pseudomonas aeruginosa quorum-sensing signal molecule N-(3-oxododecanoyl)-L-homoserine lactone has immunomodulatory activity. *Infection and Immunity* **66,** 36–42 (1998).

23. Ritchie, A. J., Yam, A. O., Tanabe, K. M., Rice, S. A. & Cooley, M. A. Modification of in vivo and in vitro T-and B-cell-mediated immune responses by

the Pseudomonas aeruginosa quorum-sensing molecule N-(3-oxododecanoyl)-L-homoserine lactone. *Infection and Immunity* **71,** 4421–4431 (2003).

24. Chhabra, S. R. *et al.* Synthetic analogues of the bacterial signal (quorum sensing) molecule N-(3-oxododecanoyl)-L-homoserine lactone as immune modulators. *J. Med. Chem.* **46,** 97–104 (2003).

25. Cooley, M., Chhabra, S. R. & Williams, P. N-Acylhomoserine lactone-mediated quorum sensing: a twist in the tail and a blow for host immunity. *Chemistry & Biology* **15,** 1141–1147 (2008).

26. Mashburn-Warren, L. M. & Whiteley, M. Special delivery: vesicle trafficking in prokaryotes. *Molecular Microbiology* **61,** 839–846 (2006).

27. Roy, P. H. *et al.* Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. *PLoS ONE* **5,** e8842 (2010).

28. Stover, C. K. *et al.* Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* **406,** 959–964 (2000).

29. Lee, D. G. *et al.* Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. *Genome Biol.* **7,** R90 (2006).

30. He, J. *et al.* The broad host range pathogen Pseudomonas aeruginosa strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 2530–2535 (2004).

31. Mikkelsen, H., McMullan, R. & Filloux, A. The Pseudomonas aeruginosa reference strain PA14 displays increased virulence due to a mutation in ladS. *PLoS ONE* **6,** e29113 (2011).

32. Lee, X. *et al.* Identification of the biosynthetic gene cluster for the Pseudomonas aeruginosa antimetabolite L-2-amino-4-methoxy-trans-3-butenoic acid. *J. Bacteriol.* **192,** 4251–4255 (2010).

33. Lee, J. *et al.* A cell-cell communication signal integrates quorum sensing and stress response. *Nat Chem Biol* **9,** 339–343 (2013).

34. Hentzer, M. *et al.* Attenuation of Pseudomonas aeruginosa virulence by quorum sensing inhibitors. *The EMBO Journal* **22,** 3803–3815 (2003).

35. Schuster, M., Hawkins, A. C., Harwood, C. S. & Greenberg, E. P. The Pseudomonas aeruginosa RpoS regulon and its relationship to quorum sensing. *Molecular Microbiology* **51,** 973–985 (2004).

36. Dong, Y.-H., Zhang, X.-F., Soo, H.-M. L., Greenberg, E. P. & Zhang, L.-H. The two-component response regulator PprB modulates quorum-sensing signal production and global gene expression in Pseudomonas aeruginosa. *Molecular Microbiology* **56,** 1287–1301 (2005).

37. Dong, Y.-H., Zhang, X.-F., Xu, J.-L., Tan, A.-T. & Zhang, L.-H. VqsM, a novel AraC-type global regulator of quorum-sensing signalling and virulence in Pseudomonas aeruginosa. *Molecular Microbiology* **58,** 552–564 (2005).

38. Firoved, A. M. & Deretic, V. Microarray analysis of global gene expression in mucoid Pseudomonas aeruginosa. *J. Bacteriol.* **185,** 1071–1081 (2003).

39. LEE, B. *et al.* Mucoid Pseudomonas aeruginosa isolates maintain the biofilm formation capacity and the gene expression profiles during the chronic lung infection of CF patients. *APMIS* **119,** 263–274 (2011).

40. Schuster, M., Lostroh, C. P., Ogi, T. & Greenberg, E. P. Identification, timing, and signal specificity of Pseudomonas aeruginosa quorum-controlled genes: a transcriptome analysis. *J. Bacteriol.* **185,** 2066–2079 (2003).

41. Liberati, N. T. *et al.* An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 2833–2838 (2006).

42. Hu, J.-F. *et al.* Jenamidines A to C: unusual alkaloids from Streptomyces sp. with specific antiproliferative properties obtained by chemical screening. *The Journal of Antibiotics* **56,** 747–754 (2003).

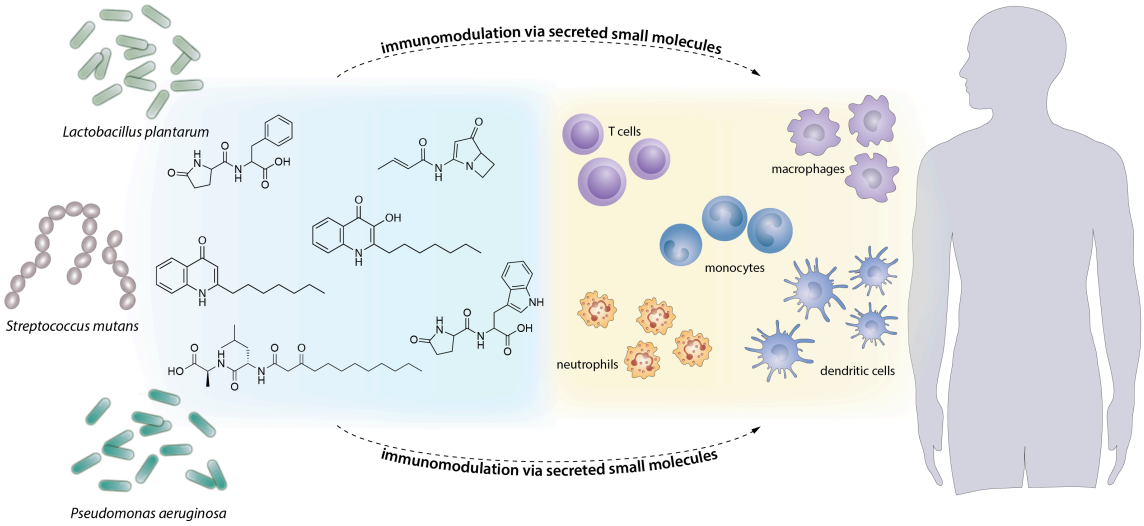
43. Duvall, J. R., Wu, F. & Snider, B. B. Structure reassignment and synthesis of Jenamidines A1/A2, synthesis of (+)-NP25302, and formal synthesis of SB-311009 analogues. *J. Org. Chem.* **71,** 8579–8590 (2006).

44. Johnston, C. W., Zvanych, R., Khyzha, N. & Magarvey, N. A. Nonribosomal assembly of natural lipocyclocarbamate lipoprotein-associated phospholipase inhibitors. *ChemBioChem* **14,** 431–435 (2013).

45. Robbel, L. & Marahiel, M. A. Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J. Biol. Chem.* **285,** 27501–27508 (2010).

46. Van Delden, C. & Iglewski, B. H. Cell-to-cell signaling and Pseudomonas aeruginosa infections. *Emerging Infect. Dis.* **4,** 551–560 (1998).

Chapter 6. Significance and Future Perspective



The majority of bioactive natural products have originated predominantly from the environmental microorganisms such as soil bacteria. They are valued for their therapeutic applications, including antibacterial, antifungal and immunomodulatory properties. Given the bioactive profiles of these molecules, their corresponding producers have been extensively mined to identify novel agents. Interestingly, with the advent of next generation sequencing we are only now starting to realize that there is actually a rich and untapped reservoir of potentially bioactive molecules residing within our bodies. This abundant reservoir is the human microbiome, which consists of both the microorganisms that reside within us and the biomolecules that they produce. The grand number of these microbes is remarkable and is ten-fold higher than the count of our own cells. Nonetheless, due to the nascent nature of this field, there has not been many discoveries made that would reveal what these microbes produce and what effects these agents may have on our cells.

The connection between the secreted metabolome of microbes from the human microbiome and immunomodulation has only recently become apparent. However, the lack of concrete evidence where select agents have been shown to modulate the immune system complicates our understanding of interaction that some microbes may engage in with our cells. Nonetheless, this nascent area of research is moving forward due to a number of advancements made in the other areas of research, including genomics, metabolomics and biochemoinformatics. Guided by genome-encoded information, some of the most valued bioactive metabolites such as polyketides and nonribosomal peptides can be identified. Moreover, utilizing various metabolomic and biochemoinformatic platforms for mining these bioactive agents makes it possible to rapidly establish the interaction link between the microbes and our cells.

My thesis uses an integrated approach in attempt to connect secreted metabolomes of select microorganisms to their biological activity with regards to immunomodulation. **Chapter 2** of my thesis is a systems biosynthesis exploration of secondary metabolic pathways within the oral human microbiome member – *Streptococcus mutans*. There is an apparent need for a system-wide analysis to fully appreciate the complexity and the interplay of various factors that ultimately shape the interaction of microbes with the human host. The global metabolomic investigation, described in this chapter, has led to the identification of a number of biosynthetic products that this bacterium creates. These also include a novel lipopeptide, mutanamide, which is a derailment product of mutanobactin biosynthesis. More importantly, a connection between some biosynthetic products and their immunomodulatory activity was established. This immunomodulatory

assessment is significant, as it opens up a stage for further biological testing to evaluate the global impact of these secreted metabolites on the immune circuitry.

Although polyketides and nonribosomal peptides are the two largest classes of bioactive small molecules, in select cases microorganism may be associated with the production of other classes of bioactive agents. **Chapter 3** of my thesis investigates the small molecule immunomodulins that were identified from the cultures of another human microbiome member – *Lactobacillus plantarum*. This chapter is based on the publication that I co-authored together with Nikola Lukenda and other researchers[78]. More specifically, this chapter describes the identification and characterization of pyroglutamic acid dipeptides, production of which was in keeping with its growth phase-dependent probiotic activity. This is a significant finding that further substantiates the importance of secreted small molecules and their immunomodulatory effects. It is possible that other agents may be produced by Lactobacilli and related probiotic species that would also correlate with their beneficial effects and future research should be focused on their identification.

There are not many bacterial organisms that are known to be as metabolically diverse as the Pseudomonads. Interestingly, however, despite the intense research efforts to identify their metabolic products, cryptic metabolites still remain and their involvement in modulation is unknown. **Chapter 4** of my thesis describes the metabolomic interrogation of bioactive small molecules from another pathogenic member of the human microbiome – *Pseudomonas aeruginosa*. This chapter showcases targeted identification of quinolones and their immunomodulatory activity in macrophages, as well as

distribution and relative abundance in model strains and several clinical isolate strains of *P. aeruginosa*. Their predominantly anti-inflammatory activity is significant and is in keeping with the postulated biological role outside of cell-to-cell communication. In addition, a discovery of novel succinylated tetrapeptide, exclusive to only two clinical isolates, is also discussed. Lastly, this chapter reveals the identification of a cryptic nonribosomal peptide product using genome-guided discovery and validation with predicted isotopically labeled precursors.

Although the metabolic versatility of Pseudomonads has been targeted by decades worth of metabolic mining, cryptic agents still remain. Identification of these metabolites requires the application of new technologies, and in cases where genomic information is available biochemoinformatic platforms may assist in their identification. **Chapter 5** of my thesis describes the identification and characterization of pseudenamide A, which is a novel nonribosomal lipopeptide from the pathogenic member of the human microbiome – *Pseudomonas aeruginosa*. Pseudenamide A features an unprecedented chemical scaffold and is very much in keeping with the biosynthetic theme observed in Pseudomonads. Similarly to quinolones and acyl homoserine lactones, molecules involved in quorum sensing and immunomodulation, pseudenamide A also contains a small cyclic core attached to a short acyl tail. Given its structural similarity and the lack of appreciable antimicrobial activity, it is highly anticipated that pseudenamide A would also engage in the similar types of activity. Future studies should be directed towards probing its immunomodulatory properties in a number of mammalian cell lines and studying its effects on gene transcription with a major focus on Pseudomonad virulence factors.

Collectively, four related research projects showcase the utility of metabolomic and biochemoinformatic approaches for identification of bioactive small molecules. Identifying these agents is critical for understanding infection in the case of pathogenic microorganisms and the probiotic effects in the case of beneficial ones. Rapid profiling of microorganisms is possible by combining both metabolomic and biochemoinformatic tools outlined in this thesis. The connection between secreted metabolome of microbes from the human microbiome and immunomodulation has only recently become apparent. The immunomodulatory activity of the identified agents is in agreement and further substantiates the hypothesis tested in this thesis. Further studies involving small molecules identified in this work will hopefully reveal some interesting principles on how microorganisms speak to human cells and how this communication could lead to therapeutic strategies or mechanistic revelations.

References

1. Morgan, X. C., Segata, N. & Huttenhower, C. Biodiversity and functional genomics in the human microbiome. *Trends Genet.* **29,** 51–58 (2013).
2. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. *Cell* **124,** 837–848 (2006).
3. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
4. Hornef, M. W., Wick, M. J., Rhen, M. & Normark, S. Bacterial strategies for overcoming host innate and adaptive immune responses. *Nat. Immunol.* **3,** 1033–1040 (2002).
5. Penders, J., Masclee, A. & Pierik, M. Probiotics in the Management of Inflammatory Bowel Disease. *Drugs* **72,** 803–823 (2012).
6. Seksik, P., Dray, X., Sokol, H. & Marteau, P. Is there any place for alimentary probiotics, prebiotics or synbiotics, for patients with inflammatory bowel disease? *Mol. Nutr. Food Res.* **52,** 906–912 (2008).
7. Petrof, E. O. Probiotics and Gastrointestinal Disease: Clinical Evidence and Basic Science. *Antiinflamm Antiallergy Agents Med Chem* **8,** 260–269 (2009).
8. Dai, C. *et al.* VSL#3 probiotics exerts the anti-inflammatory activity via PI3k/Akt and NF-κB pathway in rat model of DSS-induced colitis. *Mol. Cell. Biochem.* **374,** 1–11 (2013).
9. Fedorak, R. & Demeria, D. Probiotic bacteria in the prevention and the treatment of inflammatory bowel disease. *Gastroenterol. Clin. North Am.* **41,** 821–842

(2012).

10. Tamboli, C. P., Caucheteux, C., Cortot, A., Colombel, J.-F. & Desreumaux, P. Probiotics in inflammatory bowel disease: a critical review. *Best Practice & Research Clinical Gastroenterology* **17,** 805–820 (2003).
11. Hauser, A. R. The type III secretion system of Pseudomonas aeruginosa: infection by injection. *Nat. Rev. Microbiol.* **7,** 654–665 (2009).
12. Shiner, E. K. *et al.* Pseudomonas aeruginosa autoinducer modulates host cell responses through calcium signalling. *Cell. Microbiol.* **8,** 1601–1610 (2006).
13. Williams, S. C. *et al.* Pseudomonas aeruginosa autoinducer enters and functions in mammalian cells. *J. Bacteriol.* **186,** 2281–2287 (2004).
14. Tateda, K. *et al.* The Pseudomonas aeruginosa autoinducer N-3-oxododecanoyl homoserine lactone accelerates apoptosis in macrophages and neutrophils. *Infection and Immunity* **71,** 5785–5793 (2003).
15. Smith, R. S., Kelly, R., Iglewski, B. H. & Phipps, R. P. The Pseudomonas autoinducer N-(3-oxododecanoyl) homoserine lactone induces cyclooxygenase-2 and prostaglandin E2 production in human lung fibroblasts: implications for inflammation. *J. Immunol.* **169,** 2636–2642 (2002).
16. Smith, R. S., Harris, S. G., Phipps, R. & Iglewski, B. The Pseudomonas aeruginosa quorum-sensing molecule N-(3-oxododecanoyl) homoserine lactone contributes to virulence and induces inflammation in vivo. *J. Bacteriol.* **184,** 1132–1139 (2002).
17. Smith, R. S. *et al.* IL-8 production in human lung fibroblasts and epithelial cells activated by the Pseudomonas autoinducer N-3-oxododecanoyl homoserine lactone

is transcriptionally regulated by NF-kappa B and activator protein-2. *J. Immunol.* **167,** 366–374 (2001).

18. Telford, G. *et al.* The Pseudomonas aeruginosa quorum-sensing signal molecule N-(3-oxododecanoyl)-L-homoserine lactone has immunomodulatory activity. *Infection and Immunity* **66,** 36–42 (1998).
19. Ritchie, A. J., Yam, A. O., Tanabe, K. M., Rice, S. A. & Cooley, M. A. Modification of in vivo and in vitro T-and B-cell-mediated immune responses by the Pseudomonas aeruginosa quorum-sensing molecule N-(3-oxododecanoyl)-L-homoserine lactone. *Infection and Immunity* **71,** 4421–4431 (2003).
20. Chhabra, S. R. *et al.* Synthetic analogues of the bacterial signal (quorum sensing) molecule N-(3-oxododecanoyl)-L-homoserine lactone as immune modulators. *J. Med. Chem.* **46,** 97–104 (2003).
21. Cooley, M., Chhabra, S. R. & Williams, P. N-Acylhomoserine lactone-mediated quorum sensing: a twist in the tail and a blow for host immunity. *Chemistry & Biology* **15,** 1141–1147 (2008).
22. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An Immunomodulatory Molecule of Symbiotic Bacteria Directs Maturation of the Host Immune System. *Cell* **122,** 107–118 (2005).
23. Mazmanian, S. K. & Kasper, D. L. The love–hate relationship between bacterial polysaccharides and the host immune system. *Nat Rev Immunol* **6,** 849–858 (2006).
24. Grangette, C. *et al.* Enhanced antiinflammatory capacity of a Lactobacillus plantarum mutant synthesizing modified teichoic acids. *Proc. Natl. Acad. Sci.*

U.S.A. **102,** 10321–10326 (2005).

25. Zaleznik, D. F., Finberg, R. W., Shapiro, M. E., Onderdonk, A. B. & Kasper, D. L. A soluble suppressor T cell factor protects against experimental intraabdominal abscesses. *Journal of Clinical Investigation* **75,** 1023 (1985).
26. Ruiz-Perez, B. *et al.* Modulation of surgical fibrosis by microbial zwitterionic polysaccharides. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 16753–16758 (2005).
27. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–625 (2008).
28. Gevers, D., Pop, M., Schloss, P. D. & Huttenhower, C. Bioinformatics for the Human Microbiome Project. *PLoS Comput Biol* **8,** e1002779 (2012).
29. Metzker, M. L. Sequencing technologies - the next generation. *Nature Publishing Group* **11,** 31–46 (2010).
30. Starks, C. M., Zhou, Y., Liu, F. & Licari, P. J. Isolation and characterization of new epothilone analogues from recombinant Myxococcus xanthus fermentations. *J. Nat. Prod.* **66,** 1313–1317 (2003).
31. Martins, H. P. R., da Silva, M. C., Paiva, L. C. F., Svidzinski, T. I. E. & Consolaro, M. E. L. Efficacy of fluconazole and nystatin in the treatment of vaginal Candida species. *Acta Derm. Venereol.* **92,** 78–82 (2012).
32. Shoemaker, E. H. & Yow, E. M. Clinical evaluation of erythromycin. *AMA archives of internal medicine* **93,** 397–406 (1954).
33. Vézina, C., Kudelski, A. & Sehgal, S. N. Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of

the active principle. *The Journal of Antibiotics* **28,** 721–726 (1975).

34. Feuerstein, N., Huang, D. & Prystowsky, M. B. Rapamycin selectively blocks interleukin-2-induced proliferating cell nuclear antigen gene expression in T lymphocyte. Evidence for inhibition of CREB/ATF binding activities. *J. Biol. Chem.* **270,** 9454–9458 (1995).

35. Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British journal of experimental pathology* **10,** 226 (1929).

36. Nagarajan, R. Antibacterial activities and modes of action of vancomycin and related glycopeptides. *Antimicrob. Agents Chemother.* **35,** 605–609 (1991).

37. Pittillo, R. F., Woolley, C. & Rice, L. S. Bleomycin, and antitumor antibiotic: improved microbiological assay and tissue distribution studies in normal mice. *Appl Microbiol* **22,** 564–566 (1971).

38. Shaw, L. M. Advances in cyclosporine pharmacology, measurement, and therapeutic monitoring. *Clin. Chem.* **35,** 1299–1308 (1989).

39. Ferrini, S., Moretta, A., Biassoni, R., Nicolin, A. & Moretta, L. Cyclosporin-A inhibits IL-2 production by all human T-cell clones having this function, independent of the T4/T8 phenotype or the coexpression of cytolytic activity. *Clin. Immunol. Immunopathol.* **38,** 79–84 (1986).

40. Anand, S. *et al.* SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Research* **38,** W487–96 (2010).

41. Khaldi, N. *et al.* SMURF: Genomic mapping of fungal secondary metabolite

clusters. *Fungal Genet. Biol.* **47,** 736–741 (2010).

42. Starcevic, A. *et al.* ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Research* **36,** 6882–6892 (2008).

43. Weber, T. *et al.* CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *Journal of biotechnology* **140,** 13–17 (2009).

44. Röttig, M. *et al.* NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research* **39,** W362–7 (2011).

45. 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,** 185 (2009).

46. 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).

47. 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).

48. Magarvey, N. A., Ehling-Schulz, M. & Walsh, C. T. Characterization of the cereulide NRPS α-hydroxy acid specifying modules: Activation of α-keto acids and chiral reduction on the assembly line. *J. Am. Chem. Soc.* **128,** 10698–10699 (2006).

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

50. 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).

51. Yanes, O. *et al.* Metabolic oxidation regulates embryonic stem cell differentiation. *Nat Chem Biol* **6,** 411–417 (2010).

52. Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research* **37,** W652–W660 (2009).

53. Saccenti, E., Hoefsloot, H. C., Smilde, A. K., Westerhuis, J. A. & Hendriks, M. M. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* 1–14 (2013). doi:10.1007/s11306-013-0598-6

54. Ibrahim, A. *et al.* Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 19196–19201 (2012).

55. Ajdić, D. *et al.* Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 14434–14439 (2002).

56. Shun, C. T. *et al.* Glucosyltransferases of Viridans Streptococci Are Modulins of Interleukin-6 Induction in Infective Endocarditis. *Infection and Immunity* **73,** 3261–3270 (2005).

57. Krzyściak, W., Pluskwa, K. K., Jurczak, A. & Kościelniak, D. The pathogenicity

of the Streptococcus genus. *European Journal of Clinical Microbiology & Infectious Diseases* **32,** 1361–1376 (2013).

58. Wang, X., Du, L., You, J., King, J. B. & Cichewicz, R. H. Fungal biofilm inhibitors from a human oral microbiome-derived bacterium. *Org. Biomol. Chem.* **10,** 2044 (2012).

59. Sudbery, P. E. Growth of Candida albicans hyphae. *Nat. Rev. Microbiol.* **9,** 737–748 (2011).

60. Shareck, J. & Belhumeur, P. Modulation of morphogenesis in Candida albicans by various small molecules. *Eukaryotic Cell* **10,** 1004–1012 (2011).

61. Jimenez, M. Treatment of irritable bowel syndrome with probiotics. An etiopathogenic approach at last. *Rev Esp Enferm Dig (Madrid)* **101,** 553–564 (2009).

62. Wells, J. M. Immunomodulatory mechanisms of lactobacilli. *Microb. Cell Fact.* **10 Suppl 1,** S17 (2011).

63. Anderssen, E. L., Diep, D. B., Nes, I. F., Eijsink, V. G. & Nissen-Meyer, J. Antagonistic activity of Lactobacillus plantarum C11: two new two-peptide bacteriocins, plantaricins EF and JK, and the induction factor plantaricin A. *Applied and Environmental Microbiology* **64,** 2269–2272 (1998).

64. Yasuda, E., Serata, M. & Sako, T. Suppressive Effect on Activation of Macrophages by Lactobacillus casei Strain Shirota Genes Determining the Synthesis of Cell Wall-Associated Polysaccharides. *Applied and Environmental Microbiology* **74,** 4746–4755 (2008).

65. van Baarlen, P. *et al.* Differential NF- B pathways induction by Lactobacillus plantarum in the duodenum of healthy humans correlating with immune tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 2371–2376 (2009).

66. Schroder, K., Hertzog, P. J., Ravasi, T. & Hume, D. A. Interferon-gamma: an overview of signals, mechanisms and functions. *Journal of Leukocyte Biology* **75,** 163–189 (2004).

67. Pritt, B., O'Brien, L. & Winn, W. Mucoid Pseudomonas in cystic fibrosis. *Am. J. Clin. Pathol.* **128,** 32–34 (2007).

68. Rada, B., Lekstrom, K., Damian, S., Dupuy, C. & Leto, T. L. The Pseudomonas toxin pyocyanin inhibits the dual oxidase-based antimicrobial system as it imposes oxidative stress on airway epithelial cells. *The Journal of Immunology* **181,** 4883–4893 (2008).

69. Diggle, S. P. *et al.* The Pseudomonas aeruginosa 4-Quinolone Signal Molecules HHQ and PQS Play Multifunctional Roles in Quorum Sensing and Iron Entrapment. *Chemistry & Biology* **14,** 87–96 (2007).

70. Kim, K. *et al.* HHQ and PQS, two Pseudomonas aeruginosa quorum-sensing molecules, down-regulate the innate immune responses through the nuclear factor-κB pathway. *Immunology* **129,** 578–588 (2009).

71. Déziel, E. *et al.* Analysis of Pseudomonas aeruginosa 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 1339–1344 (2004).

72. Haas, B., Murphy, E. & Castignetti, D. Siderophore synthesis by mucoid

Pseudomonas aeruginosa strains isolated from cystic fibrosis patients. *Can. J. Microbiol.* **37,** 654–657 (1991).

73. Jimenez, P. N. *et al.* Role of PvdQ in Pseudomonas aeruginosa virulence under iron-limiting conditions. *Microbiology* **156,** 49–59 (2009).

74. Lamont, I. L. Identification and characterization of novel pyoverdine synthesis genes in Pseudomonas aeruginosa. *Microbiology* **149,** 833–842 (2003).

75. Serino, L. *et al.* Biosynthesis of pyochelin and dihydroaeruginoic acid requires the iron-regulated pchDCBA operon in Pseudomonas aeruginosa. *J. Bacteriol.* **179,** 248–257 (1997).

76. Blackie, J. A. *et al.* The identification of clinical candidate SB-480848: a potent inhibitor of lipoprotein-associated phospholipase A2. *Bioorganic & Medicinal Chemistry Letters* **13,** 1067–1070 (2003).

77. Johnston, C. W., Zvanych, R., Khyzha, N. & Magarvey, N. A. Nonribosomal assembly of natural lipocyclocarbamate lipoprotein-associated phospholipase inhibitors. *ChemBioChem* **14,** 431–435 (2013).

78. Zvanych, R. *et al.* Small molecule immunomodulins from cultures of the human microbiome member Lactobacillus plantarum. *The Journal of Antibiotics* **67,** 85–88 (2014).