2014

EPIGENETIC MODIFICATION AND ANALYSIS OF NATURAL PRODUCT GENE CLUSTERS TO ENHANCE DRUG DISCOVERY FROM BACTERIA

Justin D. Schumacher
University of Rhode Island, justin.d.schumacher@gmail.com

EPIGENETIC MODIFICATION AND ANALYSIS OF NATURAL PRODUCT GENE CLUSTERS TO ENHANCE DRUG DISCOVERY FROM BACTERIA

JUSTIN D. SCHUMACHER

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

PHARMACEUTICAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2014

MASTER OF PHARMACEUTICAL SCIENCES THESIS

OF

JUSTIN D. SCHUMACHER

APPROVED:

Thesis Committee:

Major Professor David Rowley

Daniel Udwary

Navindra Seeram

Linda Hufnagel

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2014

ABSTRACT

Natural products, and molecules derived from them, have played an extremely large role in the field of medicine and have contributed as therapies for numerous disease states. Unfortunately, the rate of discovery of medically relevant molecules from natural sources is slowing as secondary metabolite studies from microbes are increasingly yielding the re-isolation of previously known compounds. This study focused on isolating new secondary metabolites through the usage of two different techniques: (1) treatment of fungi and bacteria with epigenetic modifying compounds to induce natural product production, and (2) genome mining/bioinformatic analysis of natural product gene clusters to enhance the isolation of new compounds from previously uninvestigated species. Previous studies in the literature have shown treatment with chemical epigenetic modifiers to be successful in inducing new secondary metabolite production in fungi. At the start of my research, no natural product studies involving epigenetic modifiers had been performed in bacteria, and, to date, only one publication has appeared in the peer-reviewed literature. The approach used during my investigation was to induce new metabolite formation in the bacterium *Bacillus pumilus* strain RI06-95 via epigenetic modification using agents that target several epigenetic regulatory mechanisms. Reproducibility proved a difficulty as treatment of the same organism induced the production of different metabolites in repeat experiments.

The second focus of this study utilized genome mining and bioinformatic analysis of natural product gene clusters in order to isolate new secondary metabolites from organisms previously unstudied for natural product synthesis. The Secondary Metabolite Online Repository (SMOR) is a new tool that rapidly analyzes the well-curated complete

microbial genome sequences located on the National Center for Biotechnology Information Genome Database and flags natural product gene clusters. Many organisms are sequenced for reasons other than determination of their biosynthetic potential, and thus have been largely overlooked for their secondary metabolite production. By using SMOR and other bioinformatic tools, this study sought to isolate natural products, particularly predicted cyclic lipopeptide antibiotics, from the potentially overlooked microorganism *Bradyrhizobium sp. BTAi1*. Extract fractions showed bioactivity against methicillin resistant *Staphylococcus aureus* and structural characterization of a purified bioactive molecule indicated the potential presence of one of the predicted cyclic lipopeptides. Unfortunately due to insufficient material and long culturing times, structural elucidation of the bioactive molecules was not possible.

ACKNOWLEDGMENTS

I would like to greatly acknowledge my major advisor, David Rowley, and previous major advisor, Daniel Udwary, for all their guidance, mentorship, and for sparking my interesting in natural products chemistry and biosynthesis. To Navindra Seeram, thank you for all your teachings and input over the years. Great thanks to Linda Hufnagel who sent me down my research path during my freshman year and for all her mentorship and encouragement since then. I would like to acknowledge all of my lab mates for their help and advice whenever questions arose in the lab. To my family and dear friends, I would not have been able to complete this work without your encouragement and support, thank you. This material is based upon work conducted at a research facility at the University of Rhode Island supported in part by the National Science Foundation EPSCoR Cooperative Agreement #EPS-1004057. Research reported in this thesis was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number 8 P20 GM103430-12.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER 1:
EPIGENETIC MODIFICATION TO ACCESS CRYPTIC NATURAL PRODUCT GENE CLUSTERS

INTRODUCTION

Genome mining uses bioinformatic analyses to identify genetic sequences in an organism responsible for desired functions. This technique has gained use in the field of natural products, and allows researchers to search through an organism's genome to identify gene clusters that may encode enzymes that produce secondary metabolites (Lim, 2012; Medema 2011). These techniques also allow researchers to predict the potential structure of the natural product encoded by the gene cluster. Unfortunately, many compounds that are predicted in the genome are not expressed under normal cultivation conditions and cannot be isolated (Hertweck, 2009). These genes are commonly referred to as cryptic or silent gene clusters and have been found to be prevalent in microbial genomes. The genome for the fungus *Aspergillus niger* contains thirty-one polyketide synthase gene clusters, fifteen non-ribosomal peptide synthase gene clusters, and nine hybrid clusters. Of the identified clusters less than 30% of the corresponding molecules were isolated under normal laboratory conditions while the remaining 70% of potential metabolites remained elusive (Fisch, 2009). In order to gain access to these cryptic gene clusters, researchers are now utilizing a variety of techniques such as co-culturing, ribosomal engineering, alteration of promoter regions and transcription factors, and alteration of growth conditions (Chiang, 2011). Another technique which has shown promising results is the induction of epigenetic modifications through chemical and genetic knockout methods. Upon treatment of *Aspergillus niger* with the epigenetic modifying compound suberohydroxamic acid, forty-eight of the previously mentioned fifty-five gene clusters were found to be up-regulated by qRT-PCR (Fisch, 2009).

Epigenetics is the change in gene expression caused by mechanisms other than the underlying DNA sequence. The epigenome differs from the genome by accounting for gene expression levels under varying conditions. Most of the mechanisms which induce epigenetic change function by altering how DNA is stored by adding additional molecular units to DNA or to histones. Histones contain many positively charged lysine tails which can act as attachment points for molecules such as acetyl groups (Grunstein, 1997), phosphates (Hammet, 2007), ubiquitin (Geng, 2008), and biotin (Hassan, 2006). By altering the histone structure through these modifications, the transcription of the DNA surrounding the histone can be affected. Acetylation of a histone decreases its positive charge density. Therefore, the electrostatic interactions between the positively charged histone and negatively charged DNA are decreased. This affects how tightly the DNA binds to the histone, and relaxes the chromatin structure. The relaxed chromatin structure is more accessible to replicative enzymes allowing for the increased transcription of the previously down-regulated genes (Grunstein, 1997).

By altering the expression or functionality of epigenetic regulatory enzymes through inhibitory compounds or gene knockout, researchers can induce the production of new natural products. Research groups have already demonstrated these techniques for the isolation of novel compounds (Williams, 2008; Asai, 2012; Fisch, 2009; Beau, 2012). A trend can be seen when amongst the types of molecules produced in response to these treatments; most are either antibiotics or toxins. Therefore, using these techniques may lead to the isolation of molecules with clinically important bioactivity. With the diminishing amount of effective antibiotics due to resistance, the discovery of new antimicrobial agents is of extreme importance.

Fungi have been the primary focus for research involving the use of epigenetic modification to access cryptic natural product gene clusters. No identified publications at the start of my research had studied the use of epigenetic modification in bacterial natural product isolation. To date, only one paper was published in which bacteria were treated with epigenetic modifying compounds to elicit metabolite production (Wang, 2013). Bacteria do not have histones so many of the targets and enzymes that researchers alter in fungi to induce epigenetic change do not exist in bacteria. That is not to say that bacteria do not have their own epigenetic regulatory mechanisms. Currently, the best understood method which bacteria use to alter their epigenome is the methylation of DNA (Casadesús, 2006). Different DNA methyltransferases (DNMT) target unique DNA sequences at which they attach methyl groups on cytosine bases. Methylation near gene promoter regions causes variations in gene expression (Suzuki, 2008). It may therefore be possible to access silent natural product gene clusters in bacteria through treatment with drugs such as 5-azacytidine and procaine which inhibit DNMT. With DNMT inhibition reducing overall methylation of DNA, regulatory proteins may interact better with gene promoter regions in a natural product cluster and increase gene expression. By expanding the use of these techniques to bacteria, it may be possible to gain access to an untapped pool of compounds encoded for by silent bacterial natural product gene clusters. Therefore, this study explored the treatment of bacteria as well as fungi with epigenetic modifying compounds with the overarching aim of isolating new natural products.

The bacteria used in this study was a strain of *Bacillus pumilus* isolated from a sponge found in the Pettaquamscutt River in Narragansett, RI. This strain, known as

RI06-95, was previously studied by our lab at which time its genome was sequenced. RI06-95 was chosen to be treated with epigenetic modifying compounds as any metabolites isolated in response to the treatment could then ideally be associated back to their respective gene clusters. Several fungi were also chosen to be treated during this study. These fungi were isolated from sediment samples taken from the South Pacific Gyre during expedition 329 of the Integrated Ocean Drilling Program (Expedition 329 Scientists, 2011). The isolated fungi were already under study for the production of secondary metabolites in our lab and were included in this study in order to increase the potential number of bioactive molecules isolated.

METHODS

Selection of Epigenetic Modifying Compounds and MIC Determination

The epigenetic modifiers (EM) used during this study were selected based on previous studies reported in the literature with fungi (Cichewicz, 2010). Selection was also influenced by the mechanism of action of each EM in order to test the alteration of the epigenome by a wide array of mechanisms. The EMs 5-azacytidine, levofloxacin, prednisone, procaine, suberohydroxamic acid, and trimethoprim were chosen (See Table 1 for list of EMs used and their respective mechanisms of action).

Epigenetic Modifying Compound	Mechanism of Action
5-Azacytidine	DNA Methyltransferase inhibitor, Histone Methyltransferase inhibitor
Levofloxacin	DNA Gyrase and Topoisomerase IV inhbitor
Prednisone	Increases Histone Deacetylation activity
Procaine	DNA Methyltransferase inhibitor
Suberohydroxamic Acid	Histone Deacetylase Inhibitor
Trimethoprim	Dihydrofolate reductase inhibitor

Table 1: EMs utilized and their respective mechanisms of action.

The minimum inhibitory concentration (MIC) of each EM to each organism was determined prior to the treatment of cultures. MICs were identified via a 10x broth dilution series with the highest concentration at 10mM. The broth dilution series were performed in 96 well plates containing YPM media (0.2% yeast extract, 0.2% peptone, 0.4% mannitol, 3.5% Instant Ocean® by Marineland) for fungi and YP media (0.2% yeast extract, 0.2% peptone, 3.5% Instant Ocean® by Marineland) for RI06-95. Organisms and EM were added on the same day and incubated at 30°C for 4 to 7 days while rotating at 200 RPM. The MIC of each organism to each EM was determined by visual inspection. As two EM stock solutions (trimethoprim and prednisone) could only dissolve in DMSO, a 10x broth dilution series was also performed using DMSO with the highest concentration of DMSO at 10% of total volume.

Culturing with Epigenetic Modifying Compounds

RI06-95 was plated on 2% YP agar petri dish. One sterile colony-forming unit was scraped with a sterile pipet tip and dropped into 10 mL of YP media. After 4 days, 25 µL of bacterial culture was transferred into a centrifuge tube containing 25 mL YP media. Cultures were then grown at 30°C at 200 RPM for 18 hours at which time each EM was added into separate tubes (See Table 2 for the list of concentrations used for each EM). Cultures were grown in the same conditions for 4 more days and then were frozen. Four 1 liter cultures of RI06-95 were also created in YP medium at 30°C shaking at 150 RPM for 5 days. Trimethoprim was added on day 2 of culturing. Fungal cultures were created by directly adding 50 µL of thawed spores to 25 mL YPM in centrifuge tubes and grown at 30°C shaking at 200 RPM. After 24 hours each EM was added to separate tubes and cultures were continued for 6 more days and then frozen (See Table

2). Concentrations of EM used were chosen based upon MIC data; one concentration was a tenth the MIC and the second was variable depending on morphological changes observed during MIC testing. The concentration of one tenth the MIC was selected based on studies with fungi previously reported in the literature (Cichewicz, 2010). If morphological changes were seen at higher EM concentrations in some organisms during MIC testing then a second EM concentration up to a half of the MIC would be used during treatment in hopes that morphological changes would correlate to changes in natural product chemistry. Untreated cells were also cultured as a control.

Epigenetic Modifying Compounds	**Treatment Concentrations (Fungi)**	**Treatment Concentrations (RI06-95)**
5-azacytidine	100 µM and 500 µM	100 µM and 500 µM
Levofloxacin	500 µM and 1000 µM	1 µM and 5 µM
Prednisone	500 µM and 1000 µM	---
Procaine	200 µM and 400 µM	500 µM and 1000 µM
Suberohydroxamic Acid	350 µM and 700 µM	250 µM and 500 µM
Trimethoprim	500 µM and 1000 µM	0.1 µM, 1 µM, and 10 µM

Table 2: Treatment concentrations of each EM tested.

Extraction and Detection of Changes in Metabolite Formation

Frozen cells were thawed and then centrifuged to pellet cellular debris. A funnel containing 20 micron filters were placed airtight on top of C18 sep-packs (Maxi-Clean™ 300 mg C18; Alltech Associates Inc., Deerfield, IL). The C18 sep-packs were then placed on a vacuum manifold. Centrifuge supernatant was vacuumed through the sep-packs, rinsed with deionized water to remove salts, and eluted with methanol. The liter cultures of RI06-95 were extracted at different time points: 1 day, 2 days, and two liters 5 days post addition of trimethoprim. Two extraction methods were utilized for each liter

of culture. 500 mL was extracted by liquid/liquid partitioning with ethyl acetate and 500 mL was extracted using XAD resin (Amberlite® XAD16, Sigma Aldrich).

The dried extracts were then analyzed by analytical high pressure liquid chromatography (HPLC). HPLC was performed on a Hitachi Elite LaChrom system consisting of a 2130 pump, L-2200 autosampler and a L-2455 diode array detector and equipped with a C18 column (Sunfire C18 5µm, 2.1 mm x 150 mm column). HPLC grade methanol (Pharmaco-AAPER; Brookfield, CT) and deionized water both containing 0.1% trifluoroacetic acid (TFA) were used for the mobile phase (2 minutes isocratic flow 5% MeOH, 16 minutes linear gradient to 100% MeOH, 6 minutes isocratic 100% MeOH; flow rate = 0.5 mL/min). Blanks of each EM were also analyzed using the same gradient to allow for the identification of the EM peak in treated cell extracts. Chromatograms of extracts from untreated cells were then compared to that of treated cells and differences were noted.

Isolation and Structural Characterization of Induced Metabolites From RI-0695

An induced compound produced by RI06-95 in multiple small scale cultures was isolated by semi-preparative HPLC using a C18 column (Phenomenex Luna C18 5 µm, 250 mm x 10 mm column). Methanol and deionized water both containing 0.1% TFA were used for the mobile phase (3 minutes isocratic 40% MeOH, 14 minutes linear gradient to 70% MeOH; flow rate = 2 mL/min). The isolated molecule was then dried *in vacuo*. Mass spectrometry data was obtained by direct infusion on a Q-Star Elite (Applied Biosciences MDS) mass spectrometer equipped with a Turbo Ionspray source. The isolated molecule was dissolved in DMSO and analyzed by H-NMR on an Agilent 500MHz spectrometer (Agilent; Santa Clara, CA).

RESULTS

South Pacific Gyre Fungi Treated with EM

The MIC was determined for each fungus to each of the EMs (See Appendix 1 for list of MICs). Following suit of other previous studies in the literature, concentrations of EM used during treatments were a tenth of the MIC. However, not all fungi had a clear MIC with each EM. Many displayed morphological changes at lower concentrations. With this observation in mind, concentrations up to half the MIC were used during the treatment of fungi as it was hypothesized that altered morphology would lead to the discovery of new chemistry.

Extracts from fungi treated with an EM were compared to both untreated controls and blank runs containing only the EM used. Blank EM HPLC runs made it possible to determine if a new peak found in an extract was a new metabolite rather than the EM or breakdown product of the EM. Chromatograms from treated cells did not contain many new molecules; however, several treatments did stimulate the production of some minor peaks. Suberohydroxamic acid stimulated F53, F65, and F70 to produce new minor peaks. Several molecules in treated cells of F17, F53 and F65 were decreased in relative quantity or removed. More common than the creation of new molecules was the loss or decreased abundance of some molecules found in treated cell extracts. Nearly all fungi treated with the procaine displayed a decreased abundance of certain compounds and in some cases the near complete absence of compounds in response to EM treatment.

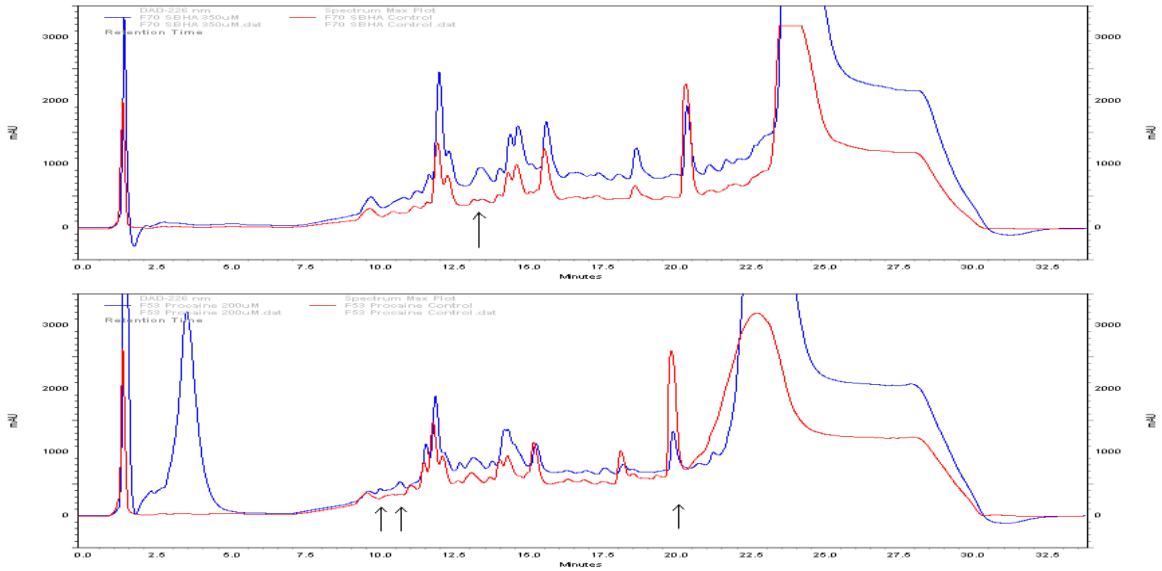


Figure 1:Chromatographs from EM treated and untreated fungi. *(Top) New compound formation in response to SBHA treatment. Blue chromatogram represents F70 treated with 350 µM of SBHA while the red is the untreated control. Note the new peak at 13.5 minutes. (Bottom) Removal of peak and two new peaks: Blue chromatogram represents F53 treated with 200 µM of procaine while the red is the untreated control. Note the decreased intensity of the peak at 20 minutes and the two new small peaks at 10 and 11 minutes. The large peak at 3.5 minutes represents the added procaine modifier.*

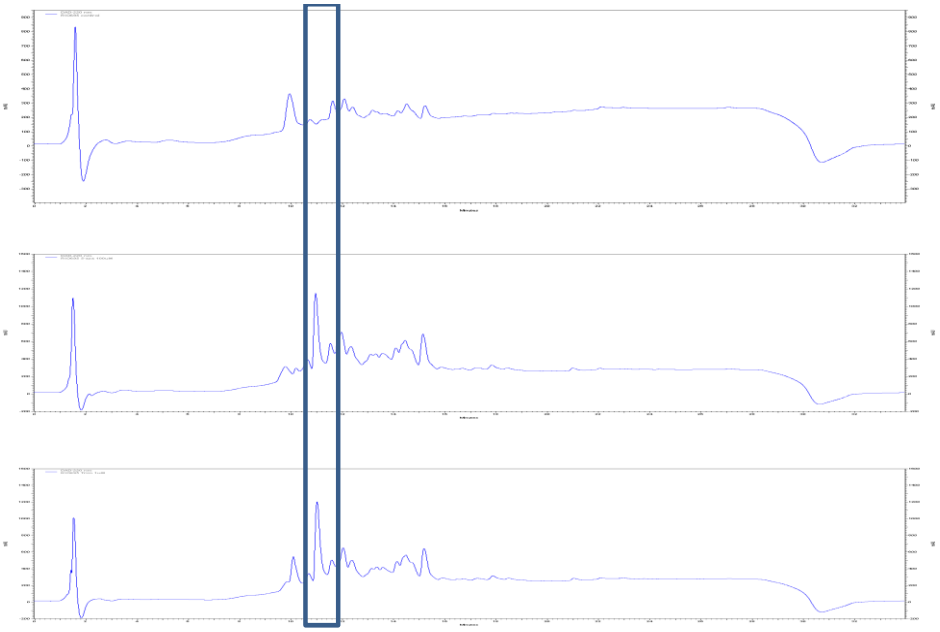
New Molecules Induced by EM in RI06-95

The MIC for RI06-95 was determined for each EM used (See Appendix 1). RI06-95 was then treated with each EM based off a tenth of the observed MIC. The MIC to trimethoprim was unclear and the cells displayed morphological differences at all concentrations despite no observed inhibition of growth. Therefore three different concentrations of trimethoprim were tested: 10 µM, 1 µM, and 0.1 µM.

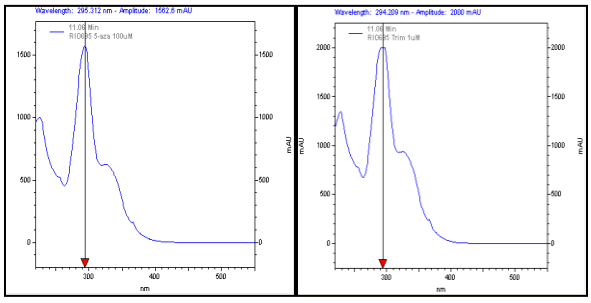
During several small scale treatments, chromatograms from cells treated with trimethoprim and 5-azacytidine displayed the production of an identical new peak compared to untreated controls (See Figure 2). The creation of this peak was dependent on the concentration of EM used. The concentrations of 10 µM and 1 µM trimethoprim induced the production of the new metabolite; however, the 0.1 µM concentration

showed no effect on inducing the compound of interest. Also, the 100 µM 5-azacytidine stimulated compound formation but not the 500 µM concentration. This molecule has three UV maxima: 230 nm, 295 nm, and 330 nm. Due to the low concentration of trimethoprim used, its respective peak was not observed in the HPLC chromatogram of treated cells. 5-Azacytidine degrades in water and after five days of culturing. Thus, a peak corresponding to 5-azacytidine is not observed in the HPLC chromatograms of treated cells.

A)



B)



<u>Figure 2: Treatment of RI06-95 with trimethoprim and 5-azacytidine induces the production of a new compound with the same UV spectrum and HPLC retention time.</u>
A) Chromatograms of untreated controls, 5-azacytidine 100 µM, and trimethoprim 1 µM treated cells (top, middle, and bottom respectively). B) UV spectra of the induced metabolite from both 5-azacytidine (left) and trimethoprim (right).

The molecule stimulated in response to trimethoprim and 5-azacytidine was isolated via semi-preparative HPLC from several small scale extracts yielding 0.9 mg of pure material. During semi-preparative HPLC, several compounds were observed with similar UV spectra to that of the induced molecule indicating the presence of potential analogs. Due to limited material, these potential analogs were not isolated, and only the major peak was collected. Time of flight mass spectrometry (Q-Star Elite in positive ion mode) revealed a single ion correspond to M+H at 362.1771 (See Appendix 2). Fragmentation of the 362.1771 peak yielded fragments of 220.9899, 260.9843, and 317.9950 in positive ion mode. Analysis ^{1}H NMR data was inconclusive due to insufficient material, leading to acquisition of poor spectra, and a potential contaminant (See Appendix 3).

One liter-scale cultures treated with 1 μM trimethoprim did not result in production of the previously mentioned new metabolite. However, the cultures did produce an amicoumacin metabolite which was not present in the small scale cultures. Amicoumacin A is a metabolite already known to be produced by RI06-95 and was dereplicated by retention time and UV spectra (Li, 2012; Socha, 2008). In several repeats of the small scale RI06-95 cultures, the treated cells responded similarly to that of the 1 L upscale cultures. Control cells lost the ability to produce amicoumacin; however, amicoumacin production was restored when treated with either trimethoprim or 5-azacytidine.

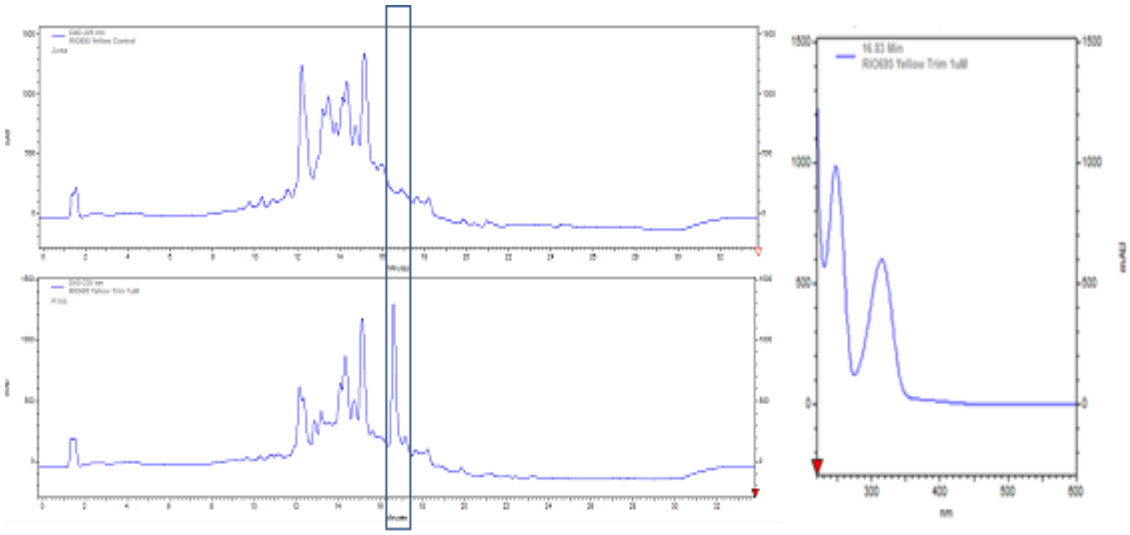


Figure 3: EM induced amicoumacin from 25 mL small scale culture. *(Left top) Chromatogram of untreated control cells and (left bottom) trimethoprim 1 μM treatment cells. (Right)UV spectra of the induced molecule with maxima at 247 nm and 314 nm.*

DISCUSSION

New metabolite formation was observed in several fungi treated in this study. However, the majority of peaks formed in response to EM in this study were relatively minor compared to others in the extract. Of all the EM tested, suberohydroxamic acid appeared to alter the epigenome of the most fungi treated. This finding correlates to previous studies which also found that suberohydroxamic acid is one of the most active compounds in altering the epigenome of fungi (Fisch, 2009; Cichewicz, 2010). Histones play an extremely important role in the storage of DNA in fungi, and therefore it is understandable that a HDAC inhibitor like suberohydroxamic acid was best at inducing epigenetic changes.

Upon validating our technique for the induction of new peak formation in fungi, the bacterium RI06-95 was tested. Unlike fungi, bacteria lack histones and rely upon other means to regulate gene transcription, one of the most important being the methylation of DNA (Casadesús,2006). Therefore, it was hypothesized that if any EM

would induce the production of new compounds in bacteria through epigenetic changes it would be one of the DNA methyltransferase inhibitors. The results observed during this study confirmed this hypothesis. 5-Azacytidine, an EM which stimulated the production of the new metabolite, inhibits DNA methyltransferase. Although trimethoprim does not directly inhibit DNA methylation, it does prevent the formation of folic acid, which is a cofactor used by enzymes as a methyl donor. Suberohydroxamic acid, an HDAC inhibitor which influenced the epigenome of the fungi used in this study the greatest, displayed little affect on RI06-95.

The stimulated production of the same molecule by both trimethoprim and 5-azacytidine poses questions as to the cause of the up-regulation and function of the induced molecule. Was the molecule formed due to epigenetic regulatory mechanism changes or was the molecule perhaps produced in response to a perceived microbial threat? As trimethoprim is an antibiotic, RI06-95 may have sensed that it was under attack by another organism and therefore produced its own secondary metabolites in response. The restoration of the production of the antimicrobial amicoumacin by RI06-95 in response to trimethoprim aligns to this theory. Interestingly though, a trend can be seen in the literature as to the types of molecules produced in response to EM. The majority of epigenetically induced compounds are either antimicrobials or toxins (Asai, 2012; Wang,2010; Wang, 2013; Williams, 2008; Zutz, 2013). Therefore, it may be have been a combination of these two mechanisms at work in RI06-95.

The results from this study provide preliminary evidence that treatment of bacteria with EM may be a useful technique in isolating additional bioactive molecules beyond those produced under normal laboratory conditions. A recent study in the

literature has investigated the plausibility of increasing secondary metabolite production in bacteria through epigenetic modification (Wang, 2013). This study however was inconclusive in regards to whether the observed changes in metabolite formation were due to epigenetic modifying compound treatments or co-culturing. Extracts from neither EM treated cells or co-cultured cells showed increased bioactivity in antimicrobial assays, yet co-cultures treated with EMs were more bioactivity. Also, in mono-cultures treated with EMs the researchers were also unable to distinguish whether the observed changes in the LC-MS fingerprint were due to new compound formation or from breakdown products of the EMs (Wang, 2013). Reproducibility in our study posed a difficulty as different cultures of the same strain of bacteria induced the production of different compounds. With potential refinement of the methods and techniques used, induction of specific metabolites may become more reproducible. Bacteria treated with EM in both this study and the study performed by the Hamann group increased the production of antimicrobials; restoration of amicoumacin production by RI06-95 and increased production of antimicrobial quinolones and rhamnolipids from *Pseudomonas sp.* co-cultures. With the potential to discover new bioactive molecules, especially antibiotics, the usage of EM should be expanded beyond their current focus on fungi.

CHAPTER 2:
GENOME MINING AND BIOINFORMATIC ANALYSIS OF NATURAL PRODUCT GENE CLUSTERS TO GUIDE COMPOUND ISOLATION FROM THE ORGANISM *BRADYRHIZOBIUM SP. BTAI1*

INTRODUCTION

Natural products are formed by organisms through a variety of molecular pathways such as non-ribosomal peptide synthesis (Challis, 2004), polyketide synthesis (PKS; Staunton, 2001), shikimate pathway (Maeda, 2012), and terpene synthesis (Withers, 2007). Each of these pathways consists of many enzymes that function together to create the final compound. In microbes, the genes encoding for all the enzymes responsible for the formation of a particular natural product are often clustered together in the genome. Each enzyme in a synthetic pathway functions as part of an assembly line, adding a new moiety to the molecule or catalyzing one or several reactions before passing the molecule off to the next enzyme. The next enzyme may then add another moiety or act as a catalyst for other reactions before passing off the molecule again. Assembly continues in this fashion until the natural product is completely formed. The order in which the enzymes function in the assembly line often matches the order in which they are encoded in the genome, a concept known as colinearity (Fischbach, 2006). By analyzing the DNA sequence and predicted amino acid sequence of each enzyme in a natural product gene cluster, one can predict the functionality of each enzyme, and therefore also make predictions of the overall structure of the secondary metabolite that the gene cluster is responsible for generating(Bachmann, 2005; Li, 2009; Lautru, 2005).

With recent advances in gene sequencing technology, numerous organisms have had their genomes sequenced. The sequences of many of these organisms are posted on the National Center for Biotechnology Information (NCBI) webpage allowing for the creation of a database of publically available genomes. This database represents a pool of secondary metabolite data that can be analyzed for natural product gene clusters,

thereby leading to the identification of secondary metabolite producing organisms. Many organisms are sequenced for reasons other than determination of their biosynthetic potential, and thus have been largely overlooked in regards to their secondary metabolite production. The Secondary Metabolite Online Repository (SMOR; www.secondarymetabolism.com) is a new tool that rapidly analyzes the well-curated complete microbial genome sequences located on the NCBI Genome Database and flags natural product gene clusters. Upon analysis by SMOR, many organisms not previously studied for the synthesis of natural products have been found to contain numerous natural product gene clusters. These organisms represent an untapped pool of potential novel compounds.

Bacteria belonging to the genus *Bradyrhizobium* are photosynthetic Gram-negative bacilli possessing a singular flagellum. They are soil bacteria that are known form symbiotic relationships with legumes (Willems, 2006). *Bradyrhizobium sp*. nodulate around root hairs and fix nitrogen in the soil for the plant, and in return, the plant provides the bacterium with nutrients. To date, it is this symbiotic relationship which researchers have focused their interest the most regarding *Bradyrhizobium* due to potential agricultural implications (Zahran, 2001; Torres, 2011). Four species of *Bradyrhizobium* have now been sequenced in order to gain insight into the symbiotic relationships these bacteria form with plants (Giraud, 2007; Kaneko, 2002; Okubo, 2012). The sequenced genomes are publicly available on NCBI. In this project, these genomes were analyzed by SMOR for the presence of natural product gene clusters. The species *Bradyrhizobium sp. BTAi1* (*BTAi1*) was found to contain five natural product gene clusters. Two of the five clusters were identified for the hypothesized production of

potentially important bioactive molecules, two predicted lipopeptide antibiotics. As no literature could be found regarding secondary metabolite isolation from *BTAi1*, this organism attracted our attention as a prime candidate for the production of previously overlooked secondary metabolites.

METHODS

Bioinformatic Analysis of Natural Product Gene Clusters:

Organisms with numerous natural product gene clusters were identified using SMOR. A background literature search was performed for each organism identified this way, and those that had not undergone secondary metabolite isolation studies were noted. Gene clusters were then analyzed using SMOR, Pfam (Punta, 2012), BLAST, ClustalW2 (Larkin, 2007), and NRPSpredictor2 (Rottig, 2011) in order to predict the structure of the metabolite formed by the cluster. The SMOR interface allows for the rough determination of the order in which enzymatic reactions take place. It also provides the predicted primary amino acid sequence for every gene in the cluster. These sequences can then be put in several analytical programs in order to determine the functionality of each gene product. Pfam allows for the identification of protein motif regions that may give clues to the functional motifs of the gene (Punta, 2012). NRPSpredictor2 analyzes NRPS gene motifs and predicts which amino acid will be incorporated into the secondary metabolite (Rottig, 2011). ClustalW allows for the comparison of a gene within the cluster to genes with known functionality using sequence alignment algorithms and phylogenetic tree formation (Larkin, 2007). By doing so, it is possible to determine if a gene is functionally similar to a comparator. After analyzing the gene cluster of several organisms using the previously mentioned bioinformatics tools and performing

background literature searches, the organism *BTAi1* was selected for this study and was purchased from the American Type Culture Collection (ATCC).

Culturing and Extraction:

Dry spores of *BTAi1* were received from the ATCC and were transferred to 5 mL of a medium similar to the recipe used by the ATCC (2.5g/L glutamate, 2 g/L yeast extract, 0.66 g/L K_2HPO_4, 0.05 g/L NaCl, 0.1 g/L $MgSO_4*7H_2O$, 0.04 g/L $CaCl_2$, 0.004 g/L $FeCl_2.4H_2O$) and grown at 30 °C shaking at 150 RPM for 5 days. The cell culture was then further expanded by transferring 2 mL into 50 mL of sterile medium. The remaining broth from the 5 mL culture was then used to create cryovials: 0.5 mL culture medium was added to 1 mL of 50% sterile filtered glycerol and stored at -80°C. The 50 mL culture was then grown for two weeks shaking at 150 RPM at 30°C.

As *BTAi1* is a photosynthetic organism, 2 mL of the 50 mL culture mentioned above was placed into two flasks containing 50 mL of fresh media; one covered in foil and one exposed to light. These cultures were grown for one month at 30 °C while shaking at 100 RPM under a growth light. The cultures were centrifuged at 2000 RPM for 4 minutes. Supernatant was passed through a 20 micron filter and then a C18 sep-pack (Maxi-Clean^{TM} 300 mg C18; Alltech Associates Inc., Deerfield, IL) via vacuum. The sep-pack was rinsed with 2 mL of deionized water to remove salts. The cell pellet was resuspended in 5 mL methanol, vortexed, and then centrifuged again at 2000 RPM for 4 minutes. The methanol was then passed through the sep-pack, collected, and dried *in vacuo*. Each extract was brought to 10 mg/mL in methanol and analyzed by high pressure liquid chromatography (HPLC). HPLC was performed on a Hitachi Elite LaChrom system consisting of a 2130 pump, L-2200 autosampler and a L-2455 diode

array detector and equipped with a C18 column (Sunfire C18 5µm, 2.1 mm x 150 mm column). Methanol and deionized water both containing 0.1% TFA were used for the mobile phase (Isocratic 5% MeOH for 2 minutes, linear gradient to 100% MeOH over 16 minutes, isocratic 100% MeOH for 6 minutes).

A 50 mL culture was created and grown for two weeks shaking at 150 RPM at 30°C. Liter cultures were then started by transferring 2 mL from the 50 mL culture into five flasks containing 1 L of sterile medium. Cells were grown for over a month at 25 °C while shaking at 150 RPM. The cultures were extracted using ethyl acetate partitioning and dried *in vacuo*.

Compound Isolation:

Crude extract was adsorbed to C18 resin and fractionated by CombiFlash medium pressure liquid chromatography (MPLC) equipped with a C18 column (RediSep SF High Performance Gold 50 g C18 column; Teledyne, Lincoln, NE). A mobile phase of water and methanol was used (Isocratic 5% MeOH for 5 minutes, linear gradient to 100% MeOH over 30 minutes, isocratic 100% MeOH for 10 minutes; flow rate = 40 mL/min).

Compound isolation was guided by disk diffusion assays against methicillin-resistant *Staphylococcus aureus* (ATCC 25923; MRSA). MRSA cultures were grown overnight at 37 °C on tryptic soy agar (Difco™ Tryptic Soy Agar; Becton, Dickinson and Co, Sparks, MD). Select colonies were picked, inoculated into 10 mL of tryptone soy broth (Bacto™ Tryptic Soy Broth, Soybean-Casein Digest Medium; Becton, Dickinson and Co, Sparks, MD) and grown for 24 h at 37 °C. Crude extract (5 mg/disk and 0.5 mg/disk) and fractions (0.150 mg/disk) were dissolved in methanol, loaded onto paper disks, and then dried at ambient temperature. MRSA was swabbed from the broth

cultures onto tryptic soy agar plates and disks were then arranged onto the plate. For each disk diffusion assay, methanol blank and ciprofloxacin disks were used as negative and positive controls. A control disk prepared by loading extract from sterile media (2.5 mg/disk) was also tested. Initially, the ciprofloxacin controls contained 0.625 mg/disk. Due to large zones of inhibition, later experiments used 0.150 mg/disk of ciprofloxacin.

The bioactive fraction JSc1-95-2 was found to contain two major peaks. These peaks molecules were isolated via semi-preparative HPLC. A C18 column (Phenomenex C18 5 μm, 10mm x 250 mm) was used with a mobile phase of methanol and water both containing 0.1% TFA (Isocratic 5% MeOH for 2 min, linear gradient to 70% MeOH over 11 min, linear gradient to 100% MeOH over 3 min, isocratic 100% MeOH for 2 min; flow rate of 5 mL/min).

Structural Characterization of Isolated Fractions:

Mass spectrometry data was obtained by direct infusion on a Q-Star Elite (Applied Biosciences MDS) mass spectrometer equipped with a Turbo Ionspray source. Both JSc1-95-2-2 and JSc1-95-2-3 were dissolved in DMSO-d_6 in a nitrogen atmosphere and analyzed by ^{1}H NMR on an Agilent 500 MHz spectrometer (Agilent; Santa Clara, CA).

RESULTS AND DISCUSSION

Natural Product Gene Cluster Analysis

Review of the SMOR database and background literature search led to the selection of *BTAi1* for this study. *BTAi1* was found to have five natural product gene clusters: a potentially broken cluster, a polyketide synthesized by an iterative type III PKS, a tetrapeptide, and two potential lipopeptide antibiotics (See Figures 4-8). *BTAi1*

was originally sequenced by researchers investigating its ability to fix soil nitrogen. No biosynthetic studies had been previously done with *BTAi1* in pure culture. However, several studies previously examined the effect of *BTAi1* on the production of phytoestrogens by yeast expressing soybean chalcone isomerases (Ralston, 2005). The lack of biosynthetic research, ability to easily acquire the organism from the American Type Culture Collection (ATCC), and presence of the two potential lipopeptide antibiotics clusters made *BTAi1* a potentially ideal candidate for this study.

The first natural product gene cluster in *BTAi1* contains only one gene that may produce secondary metabolites. This gene is a Type III Polyketide Synthase (PKS) that is 1052 bases long. Numerous forms of Type III PKS exist which utilize different substrates as building blocks. In order to predict the structure of the potential metabolite, a phylogenetic tree was created from the primary protein sequences of numerous forms of Type III Polyketide Synthases via ClustalW (BLOSOM alignment, gap extension penalty = 0.1, gap opening penalty = 10). Included in the tree were two alkylresorcinol synthases (srsA *Streptomyces fulvissimus*, AGK81780.1; srsA *Streptomyces griseus*, YP001821984.1), three dihydroxyphenylacetylglycine synthases (dpa *Amycolatopsis orientalis*, CCD33159.1; dpgA *Streptomyces toyocaensis*, AAM80548.1; dpgA *Actinoplanes teichomyceticus*, CAE53371.1), three alkylpyrone synthases (gcs *Streptomyces coelicolor*, 3V7I_A; PKS18 *Mycobacterium tuberculosis*, Q7D8I1.1; arsC *Azotobacter vinelandii strain DJ*, ACO79120.1), three tetrahydroxynaphthlene synthases (rppA *Streptomyces griseus*, YP001828132.1; rppA *Streptomyces lividans*, BAB91445.1; rppA *Streptomyces antibioticus*, BAB91443.1), two octaketide synthases (octaketide synthase *Aloe arborescens*, AAT48709.1; octaketide synthase *Hypericum*

perforatum, ACF37207.1), two valerophenone synthases (valerophenone synthase *Humulus lupulus*, ACM17227.1; valerophenone synthase *Psilotum nudum*, BAA87923.1), two acridone synthases (acridone synthase *Ruta graveolens*, CAC14058.1; acridone synthase *Huperzia serrata*, ABI94386.1), one benzalacetone synthase (benzalacetone synthase *Rheum palmatum*, AAK82824.1), and two chalcone synthases (chalcone synthase *Medicago sativa*, AAB41559.1; chalcone synthase *Glycine max*, AAB01004.1). As expected, the forms of Type III PKS found in plants (octaketide synthases, valerophenone synthases, acridone synthases, benzalacetone synthases, chalcone synthases) grouped together. Also, the multiple sequences used for each form of Type III PKS grouped amongst themselves except for the alkylpyrone PKS18 sequence from *Mycobacterium tuberculosis*. The identified Type III PKS in BTAi1 showed the most homology to the alkylresorcinol synthase protein sequences. Other nearby genes found in the gene cluster encode for a DNA lysase, an isoprenylcysteine carboxyl methyltransferase domain, and a putative multidrug resistance protein. Therefore, this molecule is predicted to be a relatively small polyketide molecule, possibly an alkylresorcinol derivative, containing potential methylation. Due to the proximity of the multidrug resistance protein to the cluster, it is possible this compound is secreted by the cell into the surrounding environment.

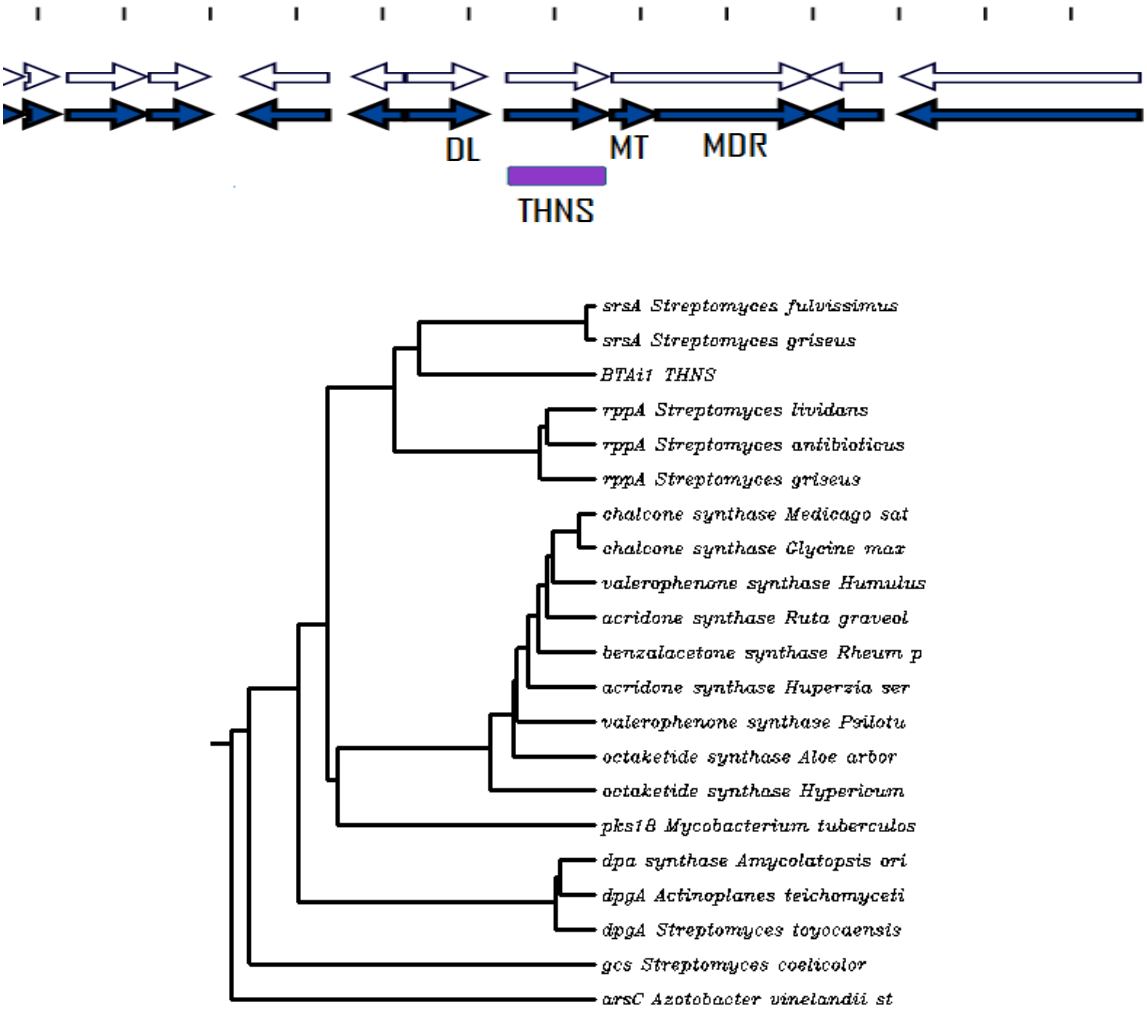


Figure 4: (Top) BTAi1 cluster 1 layout and (bottom) Type III PKS phylogenetic tree. *(Top) DL = Predicted DNA Lysase, THNS = Predicted Type III PKS, MT = Predicted Methyltransferase, MDR = Predicted Multi-Drug Resistance Transporter. (Bottom) Phylogenetic tree created from protein sequences of different forms of Type III PKS and the identified Type III PKS in BTAi1.*

The second natural product gene cluster is predicted by SMOR to encode for a secondary metabolite consisting of nine amino acids. Upon further analysis of the genes in the cluster, the adenylation domain predicted by SMOR in the loading unit was instead found to be a fatty acyl-AMP ligase (e-value = 0+00) by the conserved domain database. The NRPS Predictor 2 program predicted the sequence of the eight amino acids as Asp-Gly-Phe-Gly-Leu-Glu-Gly-Leu. This cluster includes three epimerization domains which

may be responsible for altering the stereochemistry of the aspartic acid, central leucine and glutamic acid residues. All amino acid predictions are strong except for both leucines. Nearby genes include ornithine oxidases and therefore the leucines may actually be oxidized ornithines. This natural product appears to belong to the class of molecules known as cyclic lipopeptides; a class of antimicrobials known to be active against MRSA. Another gene cluster in *BTAi1* lipopeptide is annotated for its similarity to arthrofactin synthetase. Arthrofactin is a partially cyclic lipoptide with aspartic acid and leucine as the amino acid residues adjacent to the fatty acyl chain (Morikawa, 1993), just like the metabolite for this second gene cluster. The location of the neighboring leucine and aspartic acid residues in the predicted cluster two metabolite are reversed compared to arthrofactin; leucine on the C-terminal and aspartic acid on the N-terminal of the peptide chain . Therefore this metabolite is predicted to cyclize in the same way as arthrofactin. The genes bordering the cluster are all predicted as hypothetical proteins except for one which is predicted to be a pyoverdine ABC transporter permease/ATP-binding protein. As both pyoverdine and the predicted metabolite are cyclized NRPS molecules it is probably that this compound is secreted outside the cell as would be expected for an antibiotic. Over 13,000 bases away there is also a glycosyltransferase which may indicate, though unlikely due to the distance away from the cluster, that the metabolite may be glycosylated.

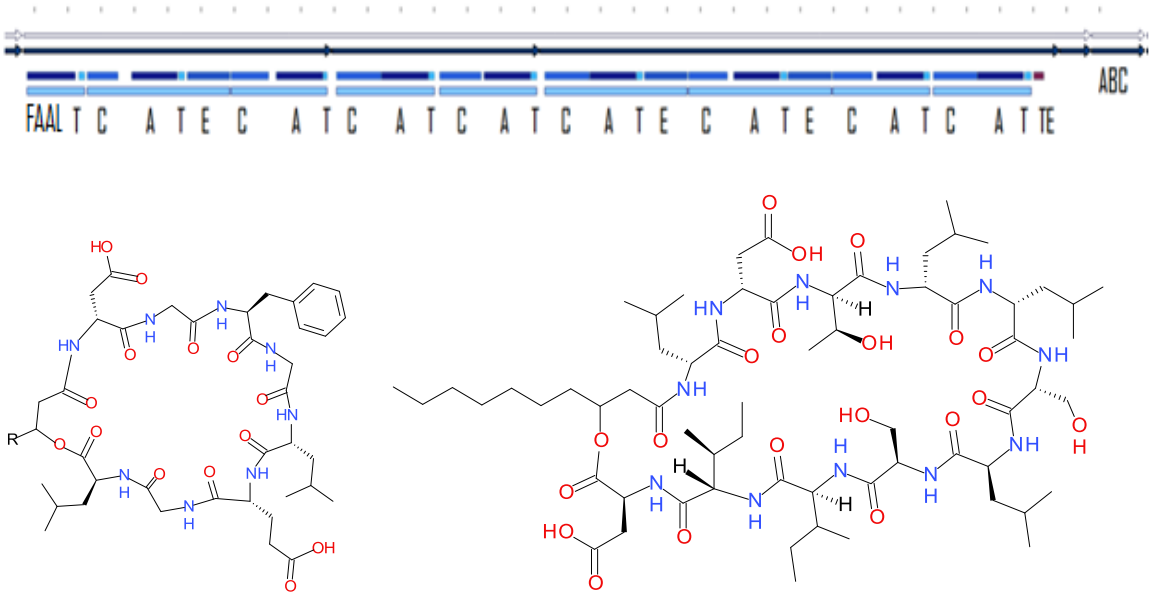


Figure 5: BTAi1 cluster 2 layout and structures of the predicted metabolite and arthrofactin. *(Top) Layout of the second BTAi1 natural product gene cluster. A = Adenylation domain, T = Carrier protein, C = Condensation domain, E = Epimerization domain, TE = Thiol Esterase, ABC = Predicted ABC transporter, FAAL = Fatty Acyl-AMP. (Bottom Left) Predicted structure of the cluster two metabolite. R-group is predicted to be a fatty acid chain. (Bottom right) Structure of arthrofactin. The metabolite from the second natural product gene cluster is predicted to cyclize in the same fashion as arthrofactin.*

The third natural product gene cluster in *BTAi1* appears to be responsible for the synthesis of a tetrapeptide connected to a formyl unit. This formyl unit may serve as an attachment point for other molecules or to cyclize the metabolite. The predicted amino acid sequence for the natural product is Ser-Orn-Orn-β-Ala. The β-Ala was a peak prediction and could potentially be Asp, Asn, Glu, Gln, or Aad (aminoadipic acid) instead. The second ornithine is predicted to be epimerized and contain stereochemistry different from the previous ornithine.

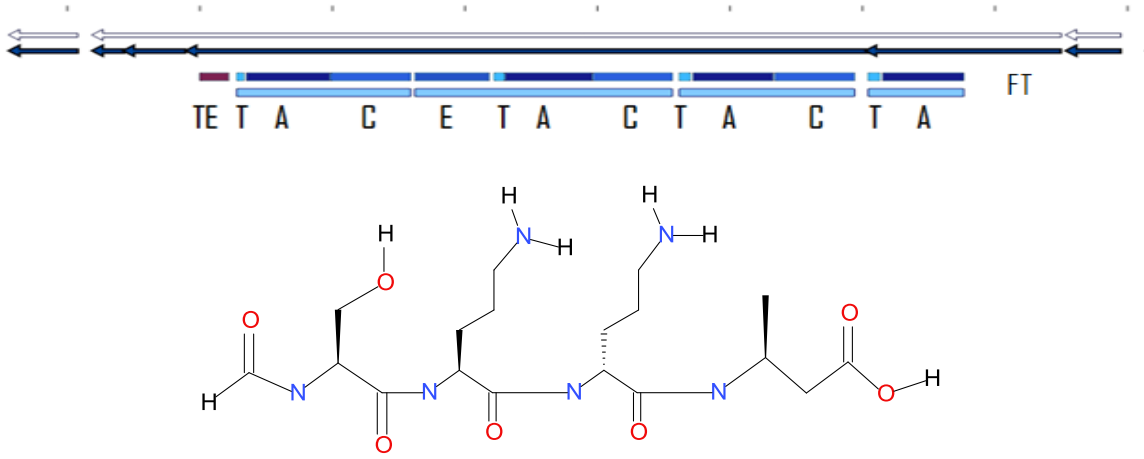


Figure 6: BTAi1 cluster 3 layout and structure of the predicted metabolite. *(Top) Layout of the third BTAi1 natural product gene cluster. A = Adenylation domain, T = Carrier protein, C = Condensation domain, E = Epimerization domain, TE = Thiol Esterase, FT = Formyltransferase. (Bottom) Predicted structure of the cluster three metabolite.*

The fourth natural product gene cluster identified in *BTAi1* has several interesting properties. First of all, it has an extra condensation domain situated at the end of the cluster without an adenylation domain following it. Secondly, this gene cluster lacks a thiolesterase domain. A putative alkanesulfonate monooxygenase gene neighbors the cluster and may be responsible for removing the secondary metabolite from the final carrier protein. However, the alkanesulfonate monooxygenase is at the beginning of the gene cluster and not at the end where thiolesterase domains normally reside. Where the TE domain should reside is an acyl carrier protein phosphodiesterase (ACPP). The ACPP may also be responsible for the release of the molecule. The first adenylation domain predicted by SMOR was later identified instead as a fatty acyl-AMP ligase. It was identified as being similar to the fatty acid CoA ligase called Fum10p (e-value = 3.01×10^{-137}). Fum10p has been determined to act like an adenylation domain responsible for the ester side groups found in the class of compounds known as fumonisins (Huffman, 2010). The second adenylation domain was weakly predicted to incorporate a

β-Ala. Other potential amino acids that may be used as substrates instead include asp, asn, glu, gln, and aad. Based on the above analysis, the metabolite produced by this gene cluster appears to be either used as a building block for other molecules or the gene cluster is broken and does not produce any metabolites.

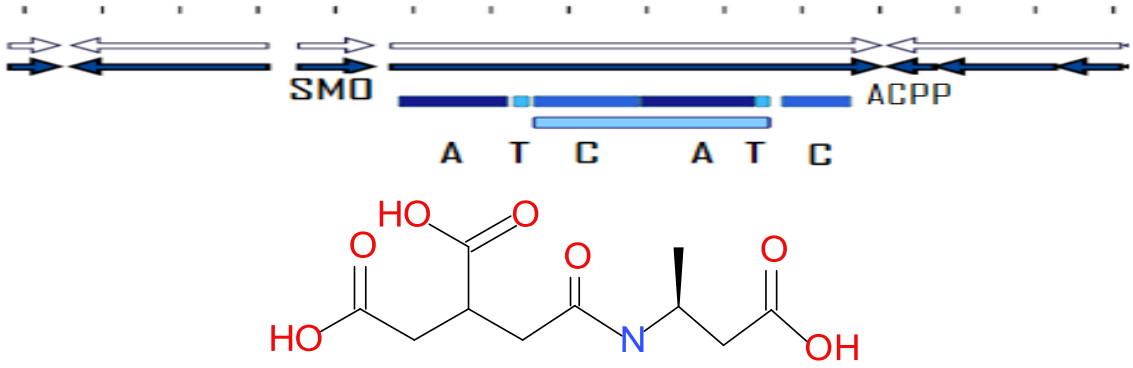


Figure 7: BTAi1 cluster 4 layout and structure of the predicted metabolite. *(Top) Layout of the fourth BTAi1 natural product gene cluster. A = Adenylation domain, T = Carrier protein, C = Condensation domain, ACPP = predicted putative acyl carrier protein phosphodiesterase, SMO = predicted alkanesulfonate monooxygenase. (Bottom) Predicted structure of the cluster four metabolite.*

The final natural product gene cluster identified by SMOR in *BTAi1* consists of three genes encoding fourteen NRPS adenylation domains. Interestingly, there appears to be no loading domain. Large gaps exist between the gene cluster and other adjacent genes. Genes encoding transcription factors are 10kb away from cluster, and it is unclear if they may affect the transcription of the cluster. The predicted sequence of the 14 amino acid metabolite by NRPS Predictor2 is Leu-Ser-Ala-Val-Val-Ala-Ser-Leu- Leu-Thr-Ser-Ser-Val-Leu. All predictions in the metabolite were strong except for both alanines which may vary and be either serines or threonines. All three genes in this cluster were found to be arthrofactin synthetase/syringopeptin synthetase C-related non ribosomal peptide synthetases. Arthrofactin, like this predicted molecule, contains

numerous leucines and amino acids containing hydroxyl groups. This cluster varies from arthrofactin though as it does not contain any genes that appear to incorporate a lipid component, however, the numerous hydroxyl groups may serve as attachment points. The predicted molecule also contains an additional two amino acid residues compared to arthrofactin.

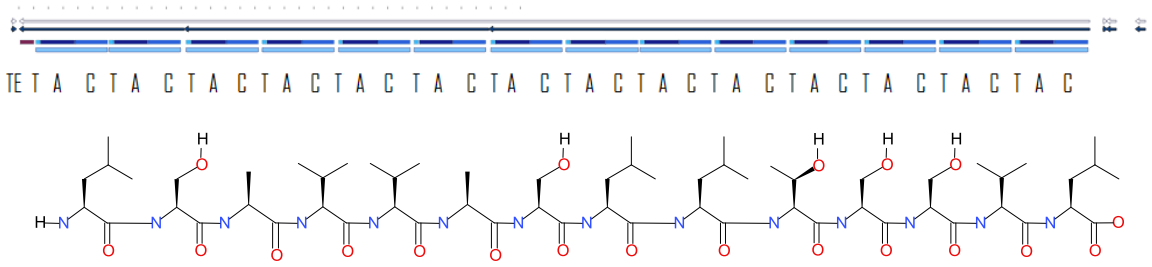


Figure 8: BTAi1 cluster 5 layout and structure of the predicted metabolite. *(Top) Layout of the fourth BTAi1 natural product gene cluster. A = Adenylation domain, T = Carrier protein, C = Condensation domain, TE = Thiol Esterase. (Bottom) Predicted structure of the uncyclized cluster four metabolite.*

Cell Culture and Extraction Yield

BTAi1 is a photosynthetic organism. To determine if light would alter secondary metabolite formation, two small scale cultures of *BTAi1* were grown; one under a growth light and the second wrapped in tin foil. After C18 sep-pack adsorption and elution with methanol, broth extracts from light and dark treatments were compared by analytical HPLC. Overlays of the two chromatographs (See Figure 9) showed only the difference of one extra peak produced in the cells grown in the dark. With this is mind, upscale cultures were grown without growth lights. Cultures maintained under light and darkness grew at the same relative rate. However, the growth curve observed was much slower than previously mentioned in existing literature; cultures took over thirty days to show full growth despite previous reports stating seven days (Wettlaufer, 1992).

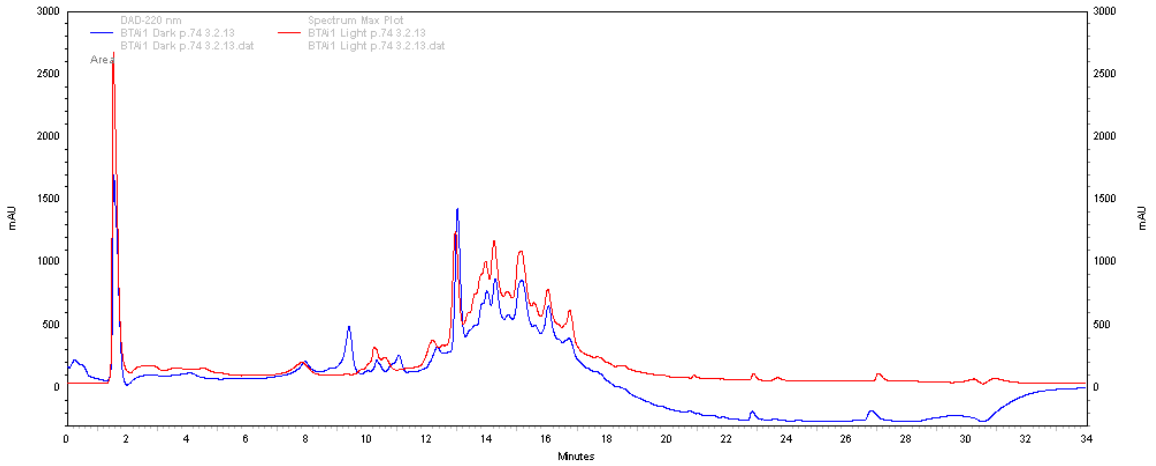


Figure 9: Overlay of the extracts from BTAi1 cultures grown under light and in darkness. *The cells grown in the dark (blue) displayed an additional peak at 9 minutes compared to the cells grown under lights (red).*

When BTAi1 was cultivated in 1 L broth cultures, slow growth was still observed and cultures needed more than a month to show sufficient growth to warrant extraction. The cells did not grow dispersed within the medium but instead in clumps on the flask just above the level of the medium when shaking. This growth pattern was observed in the small scale cultures but not as noticeably as in larger scale. This observation may be due to the fact that in its natural environment, *BTAi1* grows in the soil around root nodules. Extraction of the medium via liquid/liquid partitioning with ethyl acetate yielded a low quantity of material. A total of 172.3 mg of crude extract was obtained.

Isolation Chemistry and Structure Characterization

Disk diffusion assay against MRSA with the crude extract showed a zone of inhibition with the higher concentration disk. Fractionation of the crude extracted by reverse phase MPLC generated nine fractions. In order to attempt to locate the two predicted lipopeptide antibiotics among these fractions, isolation was guided by disk diffusion assay against MRSA. One fraction, JSc1-95-2, was found to have activity against MRSA and another fraction, JSc1-95-4, was found to have partial activity.

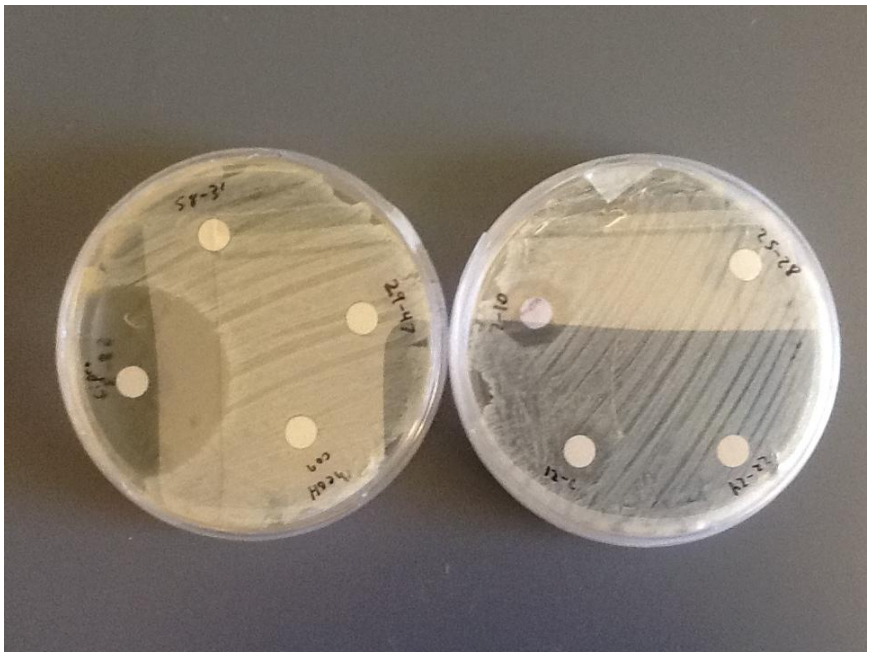


Figure 10: Disk diffusion assay of BTAi1 MPLC fractions against MRSA. *Fraction JSc1-95-2 (disk labeled 7-10 corresponding to MPLC tube collections 7-10) showed a 14 mm zone of inhibition. Fraction JSc1-95- 4 (labeled 17-21; lower left disk on right plate) displayed a weaker, but still evident, zone of inhibition.*

JSc1-95-2 consisted predominantly of two peaks when analyzed by reverse phase C18 analytical HPLC. These two compounds, denoted JSc1-95-2-2 and JSc1-95-2-3, were subsequently isolated by semi-preparative HPLC with final yields of 0.3 mg and 1 mg respectively (See Appendix 4). Unfortunately, isolates were not pure and with minimal material, structure elucidation of the compounds was not possible. Enough material was present however to allow for the initial characterization of these molecules.

Only 0.3 mg of JSc1-95-2-2 was isolated and analysis by mass spectrometry indicated that it was not entirely pure. The major ion of JSc1-95-2-2 was observed best in negative ion mode at 363.4843 (See Appendix 5). The UV spectrum of JSc1-95-2-2 contains a maxima at 328 nm. These pieces of data indicate that JSc1-95-2-2 is a small highly conjugated molecule. The only identified natural product gene cluster in the

BTAi1 genome that could form a molecule of this size is the predicted tetrapeptide. However, the predicted tetrapeptide structure is inconsistent with the observed UV spectra.

Analysis of fraction JSc1-95-2-3 by mass spectrometry did show the presence of ions residing near the masses off the predicted lipopeptides with the largest at 1123.0052 in negative ion mode (See Appendix 6). A mass fragmentation pattern (1032.4199, 1046.3720, 1060.3140, 1074.2662) in negative ion mode indicated the sequential loss of methylenes which could correlate to the alkyl chain of the predicted lipopeptide. MS/MS analysis of the 499.2464 peak in negative ion mode yielded only one product ion of 113.9627 which corresponds to the mass of aspartic acid (See Appendix 7). In the prediction of the cluster two metabolite, the amino acid residue adjacent to the fatty acyl chain was predicted to be an aspartic acid. Therefore, the potential mass of the acyl chain may be 385.2837. The loss of methylene fragmentation pattern potentially indicative of the alkyl chain is also present in the 300 mass range (297.8745, 311.8536, 325.8316, 339.8069) in negative ion mode. Currently the exact mass of the fatty acyl chain remains unclear as does the connectivity of the acyl chain to the peptide. Identification of the peptide sequence by MS/MS fragmentation studies was unclear and challenged by the fact that the peptide is predicted to be cyclized and not linear. The UV spectrum of JSc1-95-2-3 has a maximum at 284 nm which may be resultant of the predicted phenylalanine residue in the cluster two metabolite. Unfortunately due to low quantity of material and presence of impurities, NMR analysis was inconclusive (See Appendix 8 and 9).

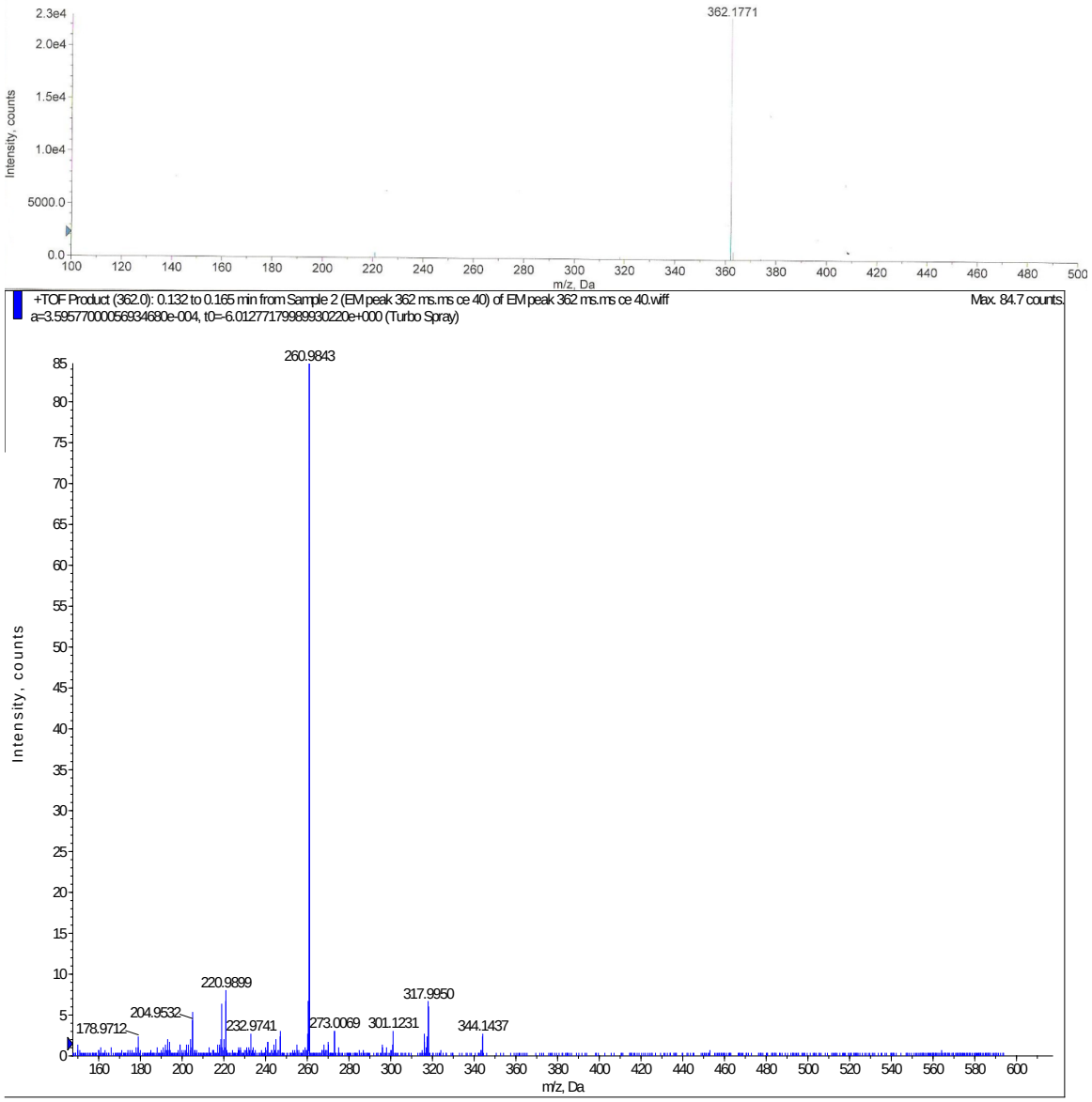
CONCLUSION

Preliminarily, it appears promising that *BTAi1* is creating lipopeptide antibiotics as predicted based on observed antimicrobial bioactivity and mass spectrometry data. Unfortunately due to long culturing times and low extraction yields, purification and structure elucidation of the bioactive molecules was not possible at this time. Culturing times of *BTAi1* was much greater than previously stated in the literature. As BTAi1 is soil microbe, one potential way to optimize growth may be to grow the organism on solid agar surfaces rather than broth in order to best mimic the organism's natural environment. Should liquid cultures be repeated, the inclusion of glass beads should be used to keep the cells in the media rather than clumping to the sides of the flask at the media surface line. Alterations of media constituents could also be performed which mimic those found naturally in the soil. Extraction methodology also requires optimization as the solvent partitioning with ethyl acetate yielded less than 30 mg per liter culture. Extraction using a resin may be preferable.

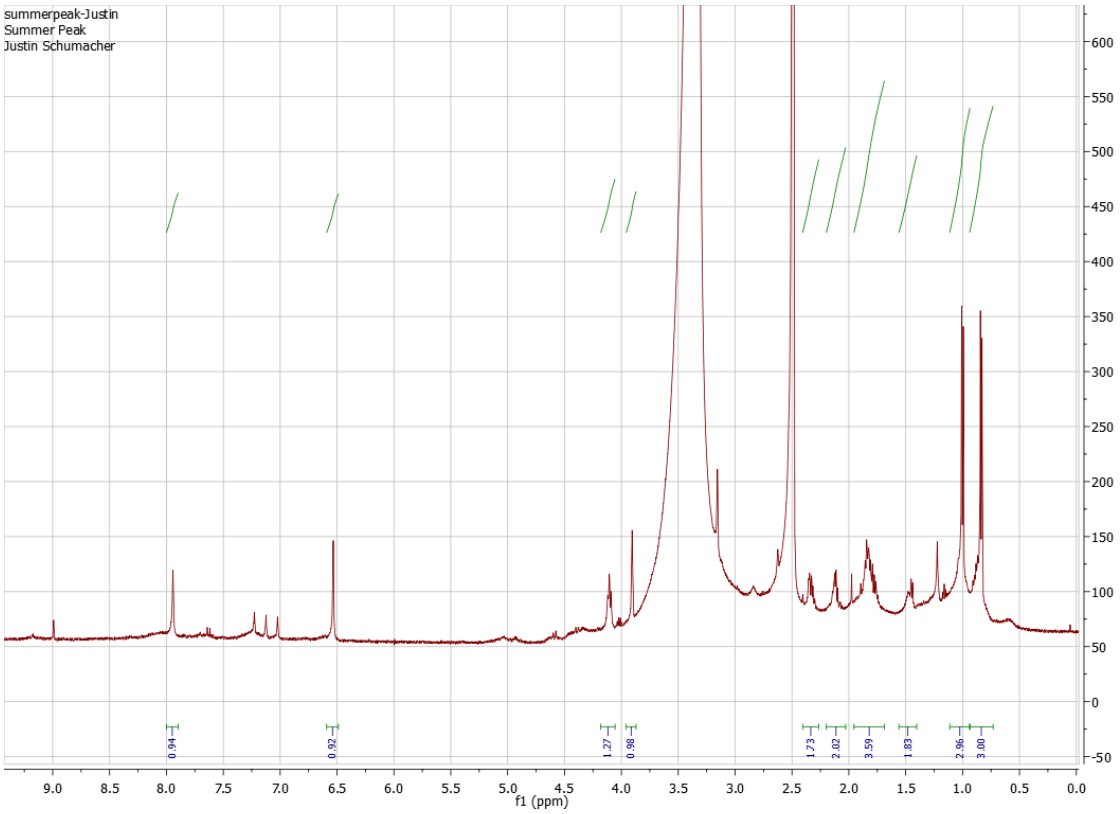
With future optimization of growth conditions and extraction methodology, elucidation of the structure of the responsible antimicrobial compounds may be conceivable. With higher amounts of starting material, isolation of the other predicted molecules may be explored as well. As drug resistant organisms continue to become ever more prevalent, discovery of new antimicrobials is of the utmost importance. Genome mining and identification of organisms overlooked for secondary metabolite production may pose as a practical technique for the isolation of these clinically needed molecules.

Organism	5-azacytidine	DMSO	Levofloxacin	Prednisone	Procaine	SBHA	Trimethoprim
SPG11-F17	10 mM*	10%	10 mM*	10 mM	10 mM*	>10 mM	10 mM/ 1 mM*
SPG11-F34	10 mM*	10%/1%*	>10 mM	10 mM/ 1 µM*	10 mM	10 mM*	10 mM
SPG11-F47	1 mM/ 100µM*	10%*	10 mM*	10 mM/ 1 µM*	10 mM	10 µM*	10 mM/ 1 mM*
SPG11-F53	10 mM	1%*	10 mM*	10 mM*	10 mM*	10 µM*	10 mM
SPG11-F65	1 mM	>10%	>10 mM	10 mM*	10 mM	>10 mM	10 mM
SPG11-F70	1 mM/ 100µM*	10%/1%*	10 µM*	1 mM/ 100 µM*	10 mM	10 µM*	10 mM/ 100 µM*
SPG11-F73	10 mM	>10%	>10 mM	10 mM	>10 mM	1 mM*	10 mM
SPG11-F77	10 mM/ 1 µM*	10%*	10 mM*	10 mM/ 1 µM*	10 mM*	100 µM*	10 mM
RI06-95	10 mM/ 100 µM*	10%*	10 mM/ 10 µM*	10 mM*	10 mM*	>10 mM	1 µM*

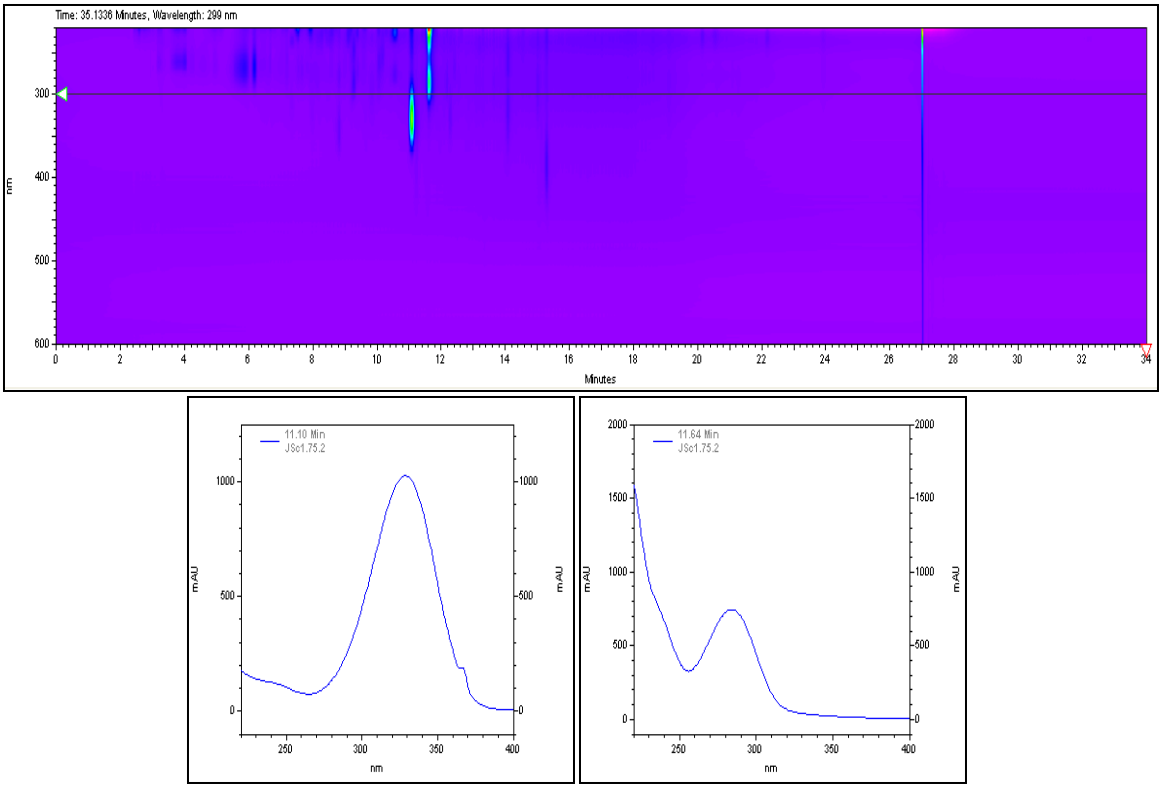
Appendix 1: MIC values of each microorganism tested to each epigenetic modifying compound. Values listed with an asterisk (*) were the highest concentrations that visibly altered cell morphology though did not inhibit growth. DMSO control was test as DMSO was required to fully dissolve both prednisone and trimethoprim. Listed percent values are vol/vol.



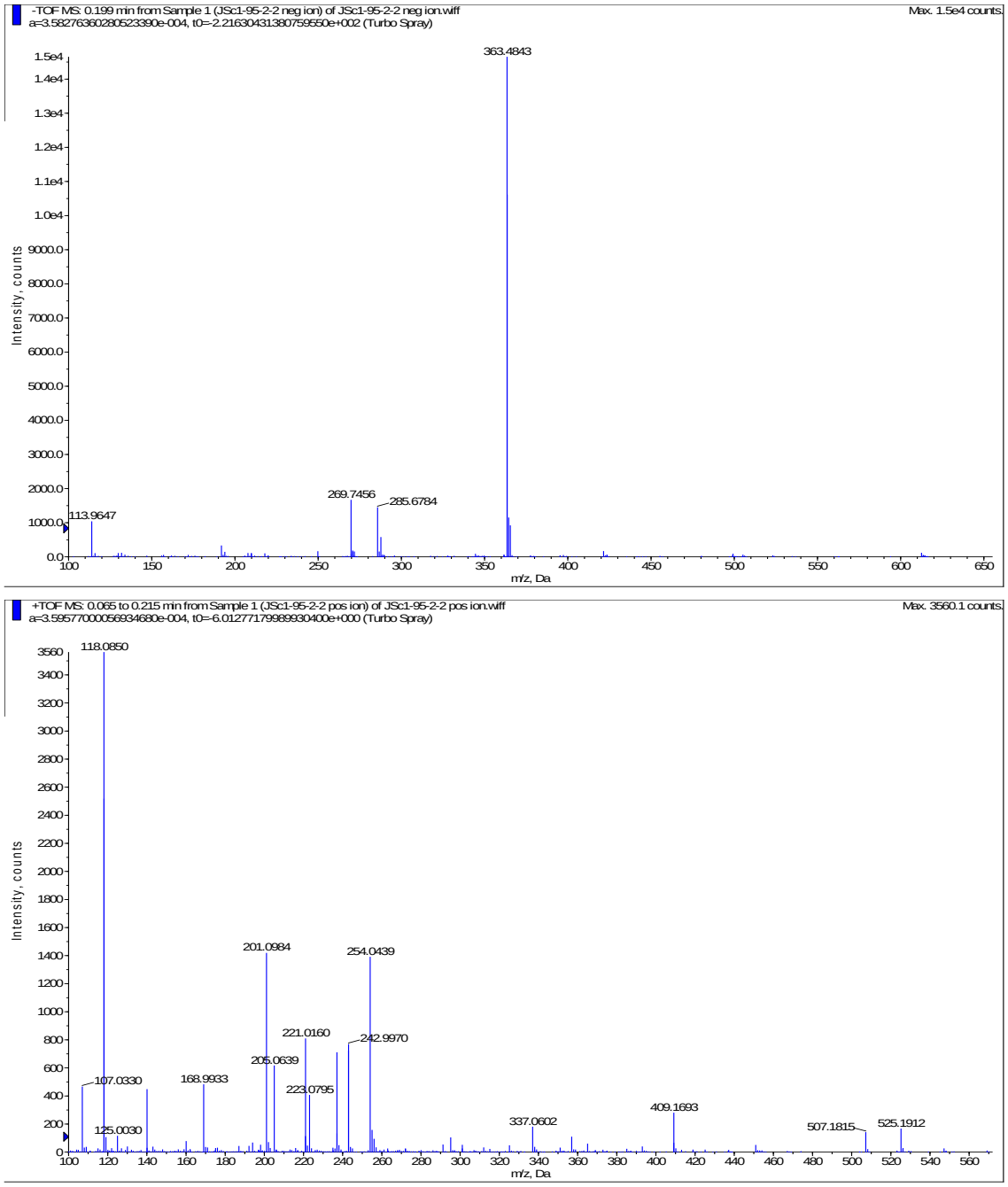
Appendix 2: MS and MS/MS data in positive ion mode for the molecule produced by RI06-95 in response to 5-azacytidine and trimethoprim.



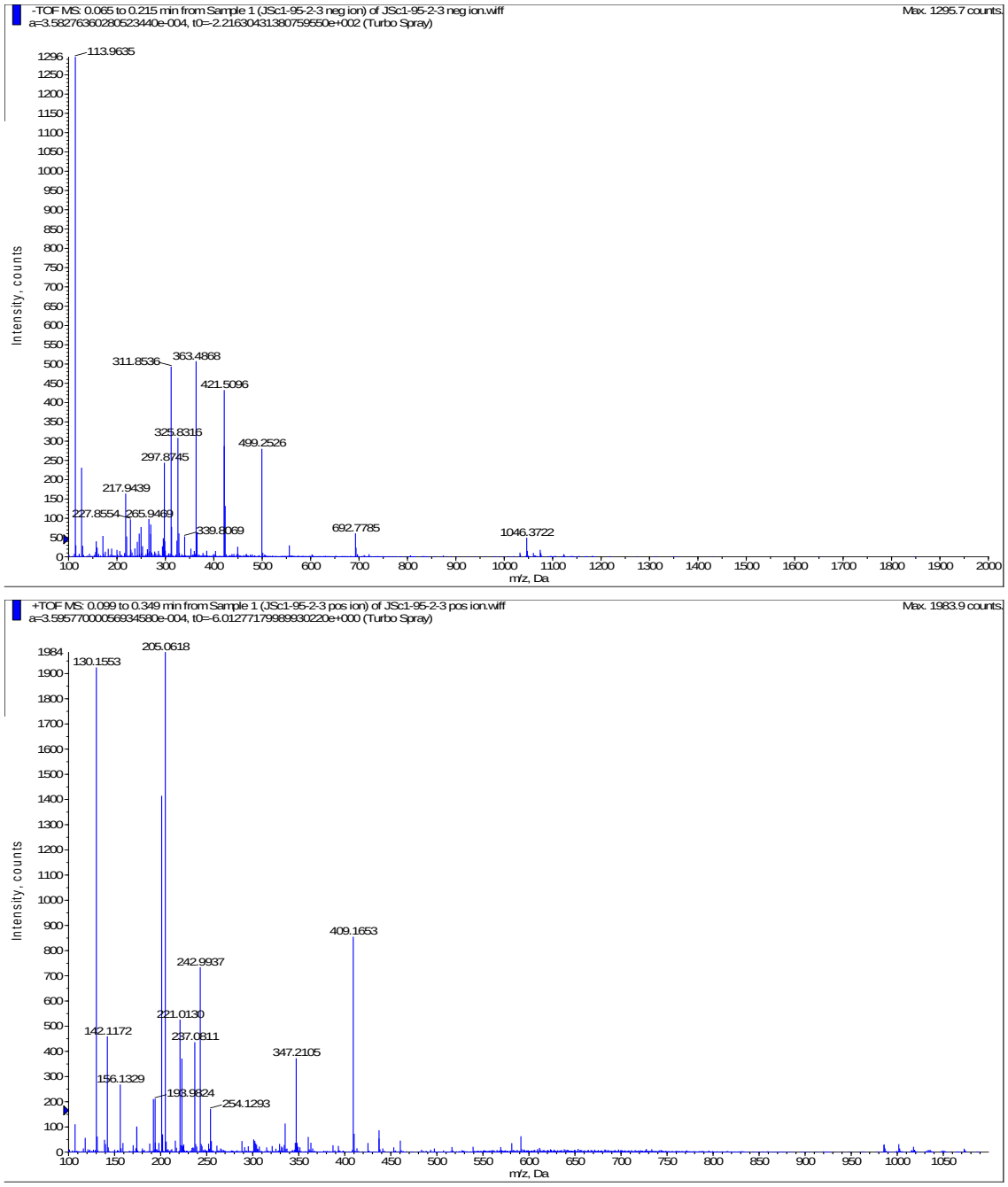
Appendix 3: ^{1}H-NMR spectrum of the molecule produced by RI06-95 in response to 5-azacytidine and trimethoprim. Spectrum obtained in DMSO.



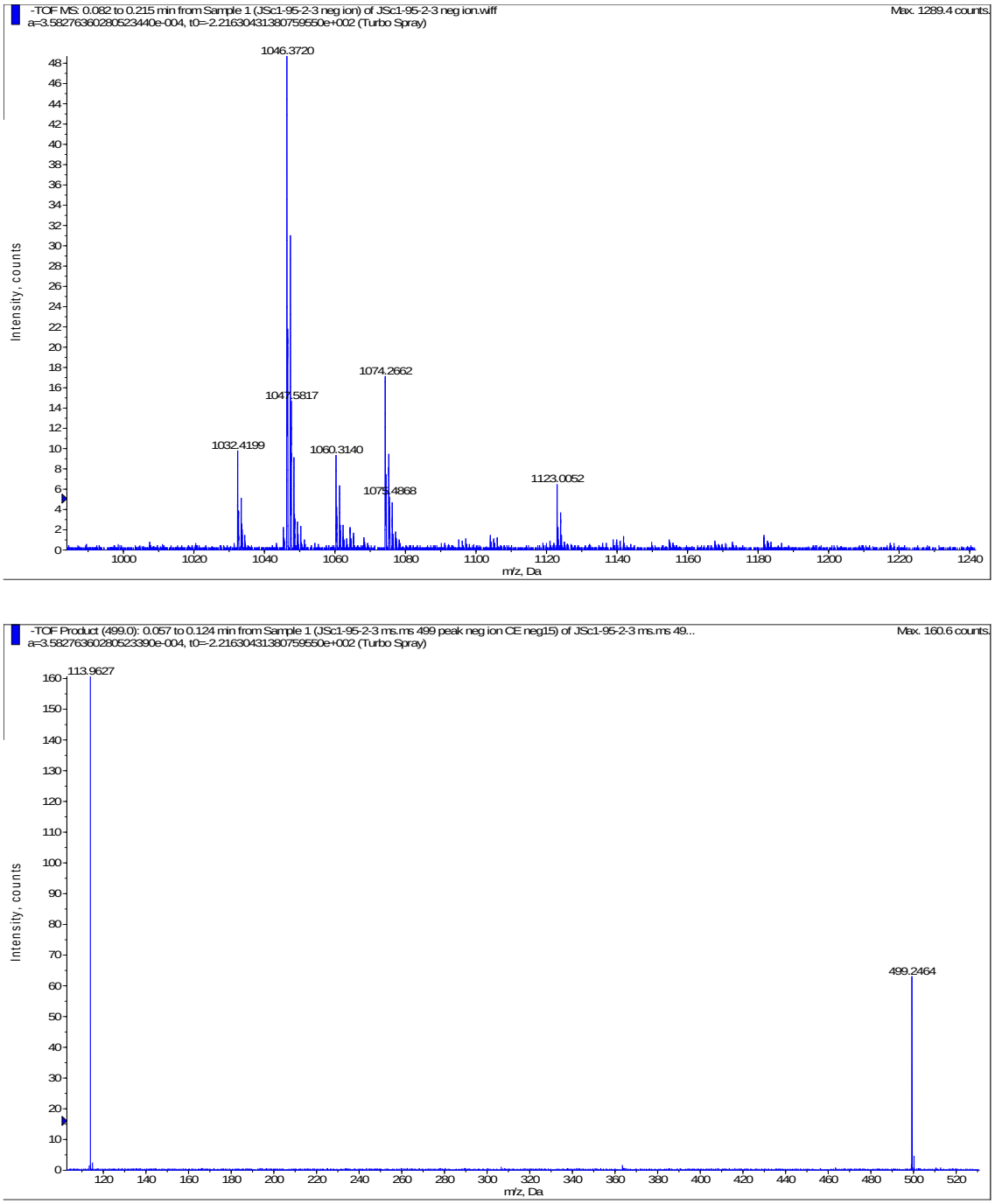
Appendix 4: (Top) HPLC-DAD analysis from the semi-preparative HPLC of JSc1-95-2 yielding 0.3 mg of JSc1-95-2-2 (11 minutes) and 1 mg of JSc1-95-2-3 (12 minutes). (Bottom left) UV spectrum of JSc1-95-2-2 with a maximum at 328 nm. (Bottom right) UV spectrum of Jsc1-95-2-3 with a UV maximum at 284 nm.



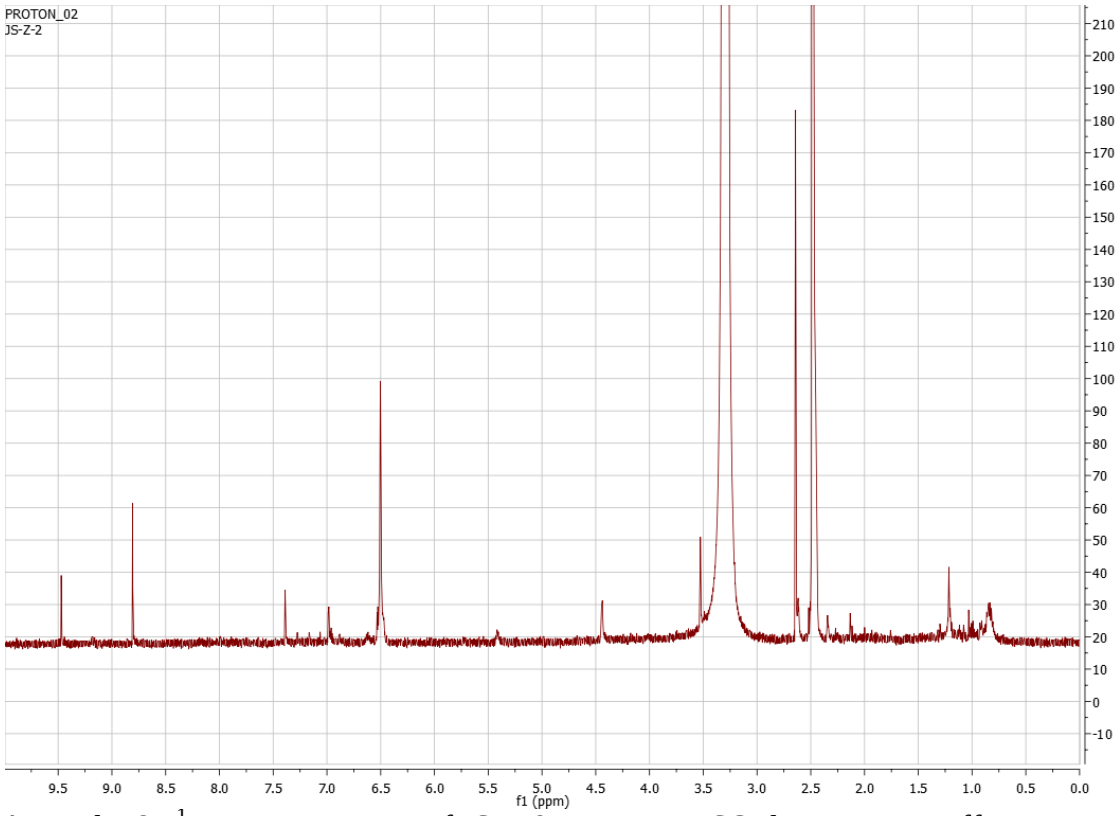
Appendix 5: Mass spectroscopy of JSc1-95-2-2 in negative (top) and positive (bottom) ion modes.



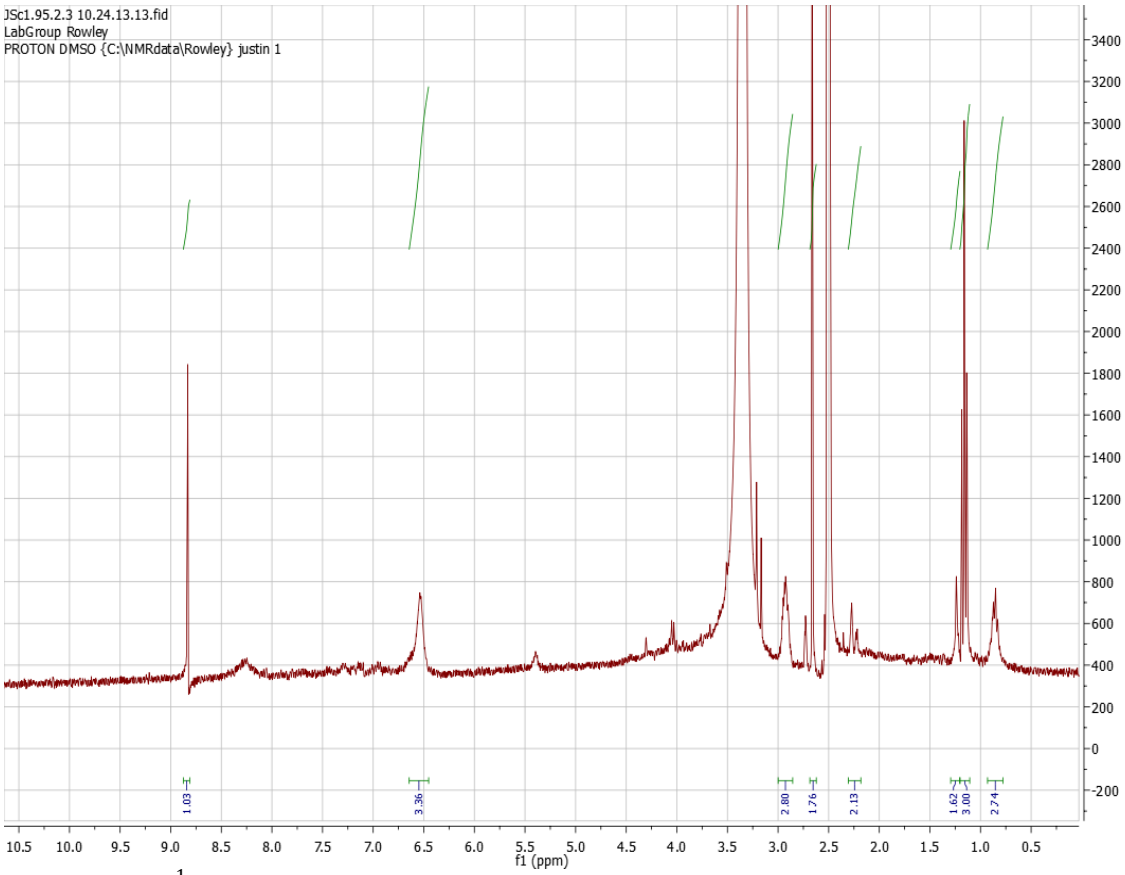
Appendix 6: Mass spectroscopy of JSc1-95-2-3 in negative (top) and positive (bottom) ion modes.



Appendix 7: (Top) Ions observed during TOF-MS analysis in negative ion mode of JSc1-95-2-3 which indicate the loss of methylenes. (Bottom) JSc1-95-2-3 MS/MS of 499.2464 peak in negative ion mode. 113.9627 correlates to the mass of aspartic acid. In the prediction of the cluster two metabolite the amino acid residue adjacent to the fatty acyl chain was predicted to be aspartic acid.



Appendix 8: ^{1}H-NMR spectrum of JSc1-95-2-2 in DMSO-d_6. Due to insufficient material, solvent and water peaks drowned out the signal from the sample.



Appendix 9: ^{1}HNMR spectrum of JSc1-95-2-3 in DMSO. Two molecules can be observed: one abundant and the other near baseline. The more abundant molecule is not the molecule corresponding to the predicted lipopeptide.

BIBLIOGRAPHY

Asai, T., Yamamoto, T., Oshima, Y. (2012) Aromatic polyketide production in *Cordyceps indigotica*, an entomopathogenic fungus, induced by exposure to a histone deacetylase inhibitor. Org Lett, **14**: 2006-2009.

Bachmann, B. (2005) Decoding chemical structures from genomes. Nat Chem Biol, **1**: 244-245.

Beau, J., Mahid, N., Burda, WN., Harrington, L., Shaw, LN., Mutka, T., Kyle, DE., Barisic, B., van Olphen, A., Baker, BJ. (2012) Epigenetic tailoring for the production of anti-infective cytosporones from the marine fungus *Leucostoma persoonii*. Mar Drugs, **10**: 762-74.

Casadesús, J., Low, D. (2006) Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev, **70**: 830-56.

Challis, GL., Naismith, JH. (2004) Structural aspects of non-ribosomal peptide biosynthesis. Current Opin Struct Biol, **14:** 748-756.

Chiang, YM., Chang, SL., Oakley, BR., Wang, CC. (2011) Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. Curr Opin Chem Biol, **15**: 137-43.

Cichewicz, R. (2010). Epigenome manipulation as a pathway to new natural product scaffolds and their congeners. Nat Prod Rep, **27**: 11-22.

Expedition 329 Scientists, 2011. South Pacific Gyre subseafloor life. IODP Prel. Rept., 329.

Fisch, KM., Gillaspy, AF., Gipson, M., Henrikson, JC., Hoover, AR., Jackson, L., Najar, FZ., Wagele, H., Cichewisz, RH. (2009) Chemical induction of silent biosynthetic pathway transcription in *Aspergillus niger*. J Ind Microbiol Biotechnol, **36**: 1199–1213.

Fischbach, MA., Walsh, CT. (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev, **106**: 3468-3496.

Geng, F., Tansey, WP. (2008) Polyubiquitylation of histone H2B. Mol Biol Cell, **19**: 3616-24.

Grunstein, M. (1997) Histone acetylation in chromatin structure and transcription. Nature, **389**: 349-52.

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, JC., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M.,

Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, WL., Chang, WS., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Verméglio, A., Médigue, C., Sadowsky, M. (2007) Legumes Symbioses: Absence of *Nod* Genes in Photosynthetic Bradyrhizobia. Science, **316**: 1307-12.

Hammet, A., Magill, C., Heierhorst, J., Jackson, SP. (2007) Rad9 BRCT domain interaction with phosphorylated H2AX regulates the G1 checkpoint in budding yeast. EMBO Rep, **8**: 851-857.

Hassan, YI,. Zempleni, J. (2006) Epigenetic regulation of chromatin structure and gene function by biotin. J Nutr, **136**: 1763-65.

Hertweck, C. (2009) Hidden biosynthetic treasures brought to light. Nat Chem Biol, **5**: 450-452.

Huffman, J., Gerber, R., Du, L. (2010) Recent advancements in the biosynthetic mechanisms for polyketide-derived mycotoxins. Biopolymers, **93**: 764-76.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M., Tabata, S. (2002) Complete Genomic Sequence of Nitrogen-fixing Symbiotic Bacterium *Bradyrhizobium japonicum* USDA110. DNA Res, **9**:189-97.

Larkin, MA., Blackshields, G., Brown, NP., Chenna, R., McGettigan, PA., McWilliam, H., Valentin, F., Wallace, IM., Wilm, A., Lopez, R., Thompson, JD., Gibson, TJ., Higgins, DG. (2007) Clustal W and Clustal X version 2.0. Bioinformatics, **23**: 2947-2948.

Lautru, S., Deeth, R., Bailey, L., Challis, GL. (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nat Chem Biol. **1**, 265-269.

Li, M., Ung, P., Zajkowski, J., Garneau-Tsodikova, S., Sherman, DH. (2009) Automated genome mining for natural products. BMC Bioinformatics, **10**: 185.

Li, Y., Xu, Y., Liu, L., Han, Z., Lai, PY., Guo, X., Zhang, X., Lin, W., Qian, PY. (2012) Five new amicoumacins isolated from a marine-derived bacterium *Bacillus subtilis*. Mar Drugs, **10**: 319-328.

Lim, FY., Sanchez, JF., Wang, CCC., Keller, NP. (2012) Toward awakening cryptic secondary metabolite gene clusters in filamentous fungi. Method Enzymol, **517**: 303-324.

Maeda, H., Dudareva, N. (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. Annu Rev Plant Biol, **63**: 73-105.

Marchler-Bauer, A. et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res, **39**: 225-9.

Medema, MH., Blin, K., Cimermancic, P., Jager, V., Zakrzewski, P., Fischbach, MA., Weber, T., Takano, E., Breitling, R. (2011) antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res, **39**: W339-W346.

Morikawa, M., Daido, H., Takao, T., Murata, S., Shimonishi, Y., Imanaka, T. (1993) A new lipopeptide biosurfactant produced by *Arthrobacter sp. Strain MIS38*. J Bacteriol, **175**: 6459-66.

Okubo, T., Tsukui, T., Maita, H., Okamoto, S., Oshima, K., Fujisawa, T., Saito, A., Futamata, H., Hattori, R., Shimomura, Y., Haruta, S., Morimoto, S., Wang, Y., Sakai, Y., Hattori, M., Aizawa, S., Nagashima, KV., Masuda, S., Hattori, T., Yamashita, A., Bao, Z., Hayatsu, M., Kajiya-Kanegae, H., Yoshinaga, I., Sakamoto, K., Toyota, K., Nakao, M., Kohara, M., Anda, M., Niwa, R., Jung-Hwan, P., Sameshima-Saito, R., Tokuda, S., Yamamoto, S., Yamamoto, S., Akutsu, T., Nakamura, Y., Nakahira-Yanaka, Y., Takada Hoshino, Y., Hirakawa, H., Mitsui, H., Terasawa, K., Itakura, M., Sato, S., Ikeda-Ohtsubo, W., Sakakura, N., Kaminuma, E., Minamisawa, K. (2012) Complete genome sequence of *Bradyrhizobium sp. S23321*: insights into symbiosis evolution in soil oligotrophs. Microbes Environ, **27**: 306-15.

Punta, M., Coggill, PC., Eberhardt, RY., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, A., Sonnhammer, ELL., Eddy, SR., Bateman, A., Finn, RD. (2012) The Pfam protein families database. Nucleic Acids Res, **40**: D290-D301.

Ralston, L., Subramanian, S., Matsuno, M., Yu, O. (2005) Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. Plant physiol, **137**: 1375-88.

Rottig, M., Medema, MH., Blin, K., Weber, T., Rausch, C., Kohlbacher, O. (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res, **39**: W362-W367.

Secondary Metabolism Online Repository (SMOR). www.secondarymetabolism.com

Socha, AM. (2008) Chemistry of antibiotics from Atlantic Actinomycete and Bacillus bacteria (Doctoral Dissertation). Retrieved from ProQuest Dissertation and Theses, Accession Order No. 3346858.

Staunton, J., Weissman, K. (2001) Polyketide biosynthesis: a millennium review. Nat Prod Rep, **18**: 380-416.

Suzuki, M., Bird, A. (2008) DNA methylation landscapes: Provocative insights from epigenomics. Nat Rev Genet, **9**: 465-476.

Torres, MJ., Bueno, E., Mesa, S., Bedmar, EJ., Delgado, MJ. (2011) Emerging complexity in the denitrification regulatory network of *Bradyrhizobium japonicum*. Biochem Soc Trans, **39**: 284-8.

Wang, B., Waters, AL., Sims, JW., Fullmer, A., Ellison, S., Hamann, MT. (2013) Complex marine natural products as potential epigenetic and production regulators of antibiotics from a marine *Pseudomonas aeruginosa*. Microb Ecol, **65**: 1068-75.

Wang, X., Sena Filho, JG., Hoover, AR., King, JB., Ellis, TK., Powell, DR., Cichewicz, RH. (2010) Chemical epigenetics alters the secondary metabolite composition of guttate excreted by an atlantic-forest-soil-derived *Penicillium citeonigrum*. J Nat. Prod, **73**: 942-8.

Wettlaufer, SH., Hardy, RWF. (1992) Effect of light and organic acids on oxygen uptake by BTAi1, a photosynthetic rhizobium. Appl Environ Microbiol, **58**: 3830-33.

Willems, A. (2006) The taxonomy of rhizobia: an overview. Plant and Soil, **287**: 3-14.

Williams, RB., Henrikson, JC., Hoover, AR., Lee, AE., Cichewicz, RH. (2008) Epigenetic remodeling of the fungal secondary metabolome. Org Biomol Chem, **6**: 1857-2020.

Withers, S., Keasling, J. (2007) Biosynthesis and engineering of isoprenoid small molecules. Appl Microbiol Biotechnol, **73**: 980-990.

Zutz, C., Gacek, A., Sulyok, M., Wagner, M., Strauss, J., Rychli, K. (2013) Small Chemical Chromatin Effectors Alter Secondary Metabolite Production in *Aspergillus clavatus*. Toxins, **5**: 1723-1741.

Zahran, H. (2001) Rhizobia from wild legumes: diversity, taxonomy, ecology, nitrogen fixation and biotechnology. J Biotechnol, **91**: 143-53.