TIKKANEN III, ROBERT KENNETH, M.S. Functional Characterization of the *dfn* Polyketide Synthase Gene Cluster in *Bacillus velezensis* ATCC 39374 (2020)
Directed by Dr. Jason J. Reddick. 74 pp.

The polyketide macrolactone, difficidin, is a secondary metabolite of *B. velezensis* ATCC 39374 and 39320 which has been found to exhibit broad-spectrum antibacterial activity against both Gram-positive and Gram-negative bacteria. While the majority of the difficidin structure follows canonical polyketide biosynthesis, it contains an unprecedented exomethylene "β-branch" which is found in only a small number of other polyketide natural products. In past work conducted by the Reddick lab, it had been hypothesized that *bae*C, a malonyl-CoA-acyl carrier protein transacylase from *B.* velezensis strain ATCC 39374, loaded a malonyl acyl group from malonyl CoA onto the free thiol group of the phosphopantetheine arm of *holo*-DfnX acyl carrier protein from the *dfn* biosynthetic gene cluster. While data from these initial experiments appeared to support this initial hypothesis, it was revealed that the promiscuous 4'-phosphopantetheinyl transferase encoded by *sfp* from strain OKB105 used in this *in vitro* system was capable of independently loading the entire, malonyl-containing, phosphopantetheine prosthetic group from malonyl-CoA directly to the *apo* form of DfnX. Therefore, the data produced from these experiments were not reflective of the true malonyl acyl transferase activity of *baeC.* The goal of the research presented in this thesis was to redesign an *in vitro* system that does not require the addition of the 4'-phosphopantetheinyl transferase to activate *apo*-dfnX.

The characterization and assignment "β-branching" activity of genes within the *dfn* and *bae* gene clusters in the *B. subtilis* ATCC 39374 strain had been an ongoing

project conducted by Dr. Brittany Kiel formerly of the Reddick Lab. By modeling the biochemical insights gleaned from the biosynthesis of bacillaene by the *pksX* gene cluster from *B. subtilis* 168, previously characterized by Walsh and Kelleher, and the biosynthesis of bacillaene and difficidin produced by the respective *bae* and *dif* gene clusters in *B. velezensis* (previously classified as *B. amyloliquefaciens* FZB42) characterized by Chen, Piel, and Borris, an on-going long-term goal of this project is to assign the activity of the genes responsible for the incorporation of the "β-branches" in the polyketide synthase gene clusters of *B. subtilis* ATCC 39374.

The overall objective of the work presented in this thesis is to characterize the genes within the *dfn* gene cluster which we hypothesize to constitute the biosynthetic pathway for the formation of the exomethylene β-branch in the difficidin structure by the analysis of polyketide intermediates via tandem electrospray ionization mass spectrometry analysis (ESI-MS/MS). The central hypothesis is that the enzyme encoded by *bae* C catalyzed the transfer of a malonyl acyl group from malonyl-CoA to *holo* form of the acyl carrier protein *dfnX*, which is subsequently decarboxylated by the enzyme encoded by *yhdS* to yield acetyl-S-DfnX (Ac-S-DfnX). What has previously not been shown in preliminary data involves the gene product of *dfnL* hypothesized to catalyze the reaction of acetoacetyl-S-DfnJ-T2 (Acac-S-DfnJ-T2) and Ac-S-DfnX to yield a 3-hydroxymethylglutaryl-S-DfnJ-T2-like (HMG-S-DfnJ-T2) intermediate. The subsequent dehydration and the decarboxylation of HMG-S-DfnJ-T2 is catalyzed by the enzymes encoded by *baeH* and *dfnM* respectively, which is hypothesized to produce an external olefin containing isopentenyl derivative resulting in a difficidin-like "β-branch".

Whereas, the catalysis of this reaction by the genes *baeH* and *baeI* is hypothesized to produce an internal olefin containing isopentenyl derivative bacillaene-like "β-branch". The scope of this project only encompasses the assignment of activity to the genes responsible for the formation of this internal or external olefinic isopentenyl product, however further structural characterization of the gene-specific isopentenyl derivatives beyond analysis of MS^3 fragmentation patterns are a future goal for the lab. The rationale that underlies the research proposed here is that the characterization of the activity of the *dfn* gene cluster in *B. subtilis* ATCC 39374 would give a much deeper and strain specific understanding of the formation of the polyketide "β-branch" in difficidin, rather than a strict reliance on the sequencing data and proposed order of PKS modules within the *dif* cluster of *B. velezensis.*

FUNCTIONAL CHARACTERIZATION OF THE *dfn* POLYKETIDE SYNTHASE GENE CLUSTER IN *BACILLUS VELEZENSIS* ATCC 39374

by

Robert Kenneth Tikkanen III

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2020

Approved by

_____
Committee Chair

APPROVAL PAGE

This thesis, written by Robert Kenneth Tikkanen III, has been approved by the following committee of the Faculty of The Graduate School at the University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

I. A. Background

Difficidin was first discovered in 1987 by the scientists of the Merck Sharp and Dohme Research Laboratories while screening of the fermentation broths of each of two different strains of *B. subtilis* (ATCC 39320 and ATCC 39374) for new antibiotics.[1] This novel macrocyclic polyene lactone phosphate ester demonstrated a broad spectrum potent antibiotic activity. It had been proposed that the *pksX* gene cluster was responsible for the production of difficidin in *B. subtilis* 168, however further studies revealed catalytic biosynthetic machinery within the cluster which produced structural elements not present within the difficidin structure, ultimately disproving this initial hypothesis.[2] The secondary metabolite biosynthesized by the *pksX* gene cluster was later found to be bacillaene, a "β-branch" containing polyketide.[3] While difficidin was isolated from *B. velezensis* ATCC 39374 the genomic DNA of this strain was never sequenced, and has remained largely unstudied since the 1987 publication. Rather, the DNA sequencing information used to model the biosynthetic gene cluster responsible for the formation of difficidin comes from the *pks3* (*dif*) gene cluster from *B. velezensis*, previously classified as *B. amyloliquefaciens* FZB42.[4,5] The *Bacillus velezensis* strain FZB42, is a Gram-positive, plant root-colonizing bacterial strain which is closely related to the model organism *Bacillus subtilis* 168. Its ability to stimulate plant growth and capacity to

biologically synthesize a host of secondary metabolites to suppress plant pathogens and other virulent microorganisms has made it a commercially viable biocontrol agent in the agricultural sector.[6] Specifically, *B. velezensis* FZB42 contains eight gene clusters (*srf, bmy, fen, nrs, dhb, mln, bae, and dif) responsible* for the non-ribosomal synthesis of antimicrobial cyclic lipopeptides and polyketides which have exhibited broad spectrum activity against both aerobic and anaerobic bacteria.[1]

I. A. i. Polyketide Synthesis

Difficidin belongs to a subclass of bacterial secondary metabolites known as polyketides. Polyketides constitute a broad and chemically diverse subclass of natural products with particularly high pharmaceutical importance due to their potent bioactivity contributing to these compounds' allure as antibiotics, immunosuppressants, and antitumor therapeutics.[7] Polyketides are differentiated from other natural products by the catalytic machinery utilized in the biosynthesis of these compounds. Large megaenzyme complexes known as modular polyketide synthases (PKSs) are responsible for the piecewise assembly of polyketide natural products.[8] The protein that makes up each PKS module within the larger megaenzyme complex is responsible for the incorporation of a specific monomeric unit in the growing polyketide chain. The conventional biosynthetic paradigm modeled after type 1 *cis*-acyltransferase (*cis*-AT) polyketide synthases, states that the order or sequence of each PKS module within the complex is indicative of the resultant polyketide structure.[9] The elongation of the chain is dependent on three basic enzymatic functions: a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP). Acyl carrier domains contain a conserved serine residue that is post-

translationally modified by a 4'-phosphopantetheinyl transferase (PPTase) which catalyzes the attachment of a prosthetic phosphopantetheine (PPant) arm from coenzyme A (CoA).[10] The addition of this PPant arm "activates" the acyl carrier protein from its inactive *apo*-form to the CoA-derived free thiol containing *holo*-form, this sulfhydryl substituent can then be modified by different acyl transferase domains (AT).

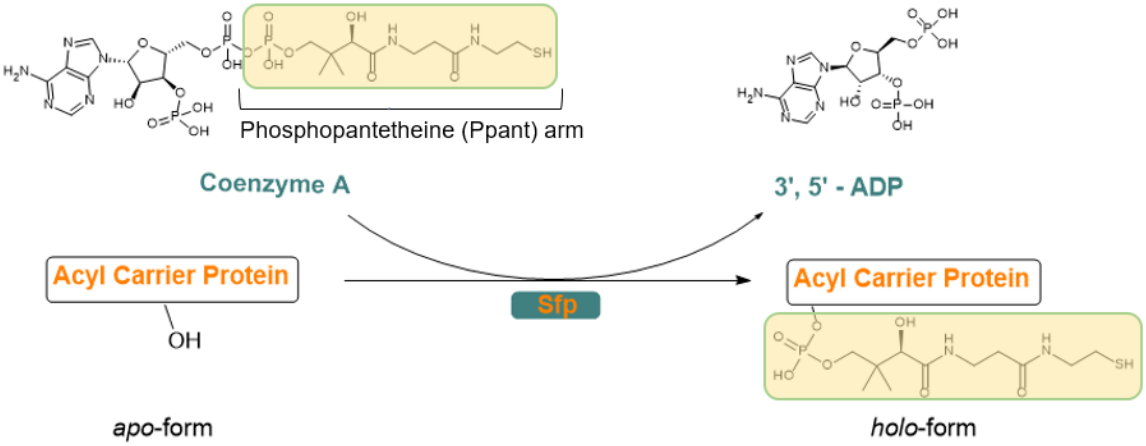
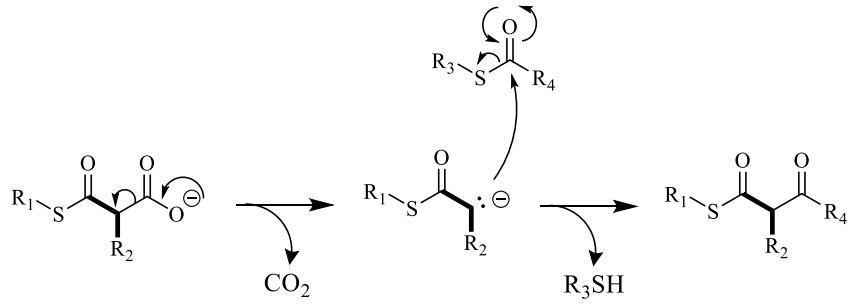


Figure 1. Schematic of the Activation of the *apo*-form of the Acyl Carrier Protein to the *holo*-form by the Transfer of the Phosphopantetheine Prosthetic Group by the 4'-phosphopantheinyl Transferase Encoded by *sfp*.

"Activation", in this case, refers to the ability of the new *holo*-form acyl carrier protein to covalently bind acyl reaction intermediates through the formation of a thioester with the free thiol of the phosphopantetheine arm during chain elongation (Figure 1). These acyl intermediates, or "building blocks" of the multistep assembly of the polyketide include acetyl, aminoacyl, and malonyl units. These AT domains are then able to incorporate acyl CoA building blocks via the transfer the acyl moiety to the ACP PPant arm, which results in the formation of a thioester. Once the ACP has been loaded,

the proteins within the β-ketoacylsynthase domain (KS) facilitate the connection of the polyketide units by catalyzing a decarboxylative, "Claisen-like" condensation, leaving an enzyme-bound β-ketothioester intermediate (Figure 2). Structural diversity within polyketide compounds can come from different accessory domains such as ketoreductase (KR), dehydratase (DH), and enoyl-reductase (ER) or by the incorporation of different extender units.[9]



Type I PKSs:	Type II PKSs:	Type III PKSs:
R_1 = Acyl Carrier Protein (ACP)	R_1 = Acyl Carrier Protein (ACP)	R_1 = Coenzyme A
R_2 = H, CH_3, CH_2CH_3, OH, OCH_3, NH_2	R_2 = H	R_2 = H, CH_3, CH_2CH_3
R_3 = Ketosynthase	R_3 = Ketosynthase	R_3 = PKS
R_4 = CH_2-R	R_4 = CH_2-R	R_4 = CH_2-R

Figure 2. Mechanism for Decarboxylative "Claisen-like" Condensation for the Incorporation of Extender Units for Type I, II, and III PKSs. Figure Adapted from Chan et al.[11]

Within the Type I PKS modular synthases, specifically the *cis*-AT PKS subtype, there exists a strong collinearity between each structural feature within a polyketide natural product and the sequence of each enzymatic PKS module. Thereby, a prediction of the structure of the overall polyketide natural product can be made by the orientation and sequence of the PKS modules encoded within the biosynthetic gene cluster.

I. A. ii. trans-AT PKS

However, certain structural features contained within a smaller class of polyketide natural products seem to defy the convention of *cis*-AT PKS complex polyketide synthesis. While containing ostensibly similar modular architecture, the *trans*-AT PKS modular system distinctly lacks acyltransferase domains and includes modules with unusual domain orders or non-elongating domains (Figure 3). Other unique characteristics include "split-modules" that are divided on two separate protein chains and exhibiting action of domains across different modules.[12] The *bae* gene cluster which is responsible for the production of the polyketide bacillaene contains several of these unconventional features. As a result of these inherent differences as compared to *cis*-AT PKS domains, the canonical collinearity rules do not apply, but opens up the possibility for an extensive array of polyketide structural diversity. In *trans*-AT PKS complexes, the acyl transferase domains are missing from the modules, however this essential activity is provided by a separate malonyl-specific acyltransferase (AT) domain enzymes that add "in *trans"*. These domains are encoded by a single gene in the cluster that interacts with the majority of the other PKS modules in the system. Thereby it follows that the building blocks in trans-AT PKS complexes are often simple malonyl-derived units that are modified either by reduction or methylation of the "building block" post chain elongation.[9]

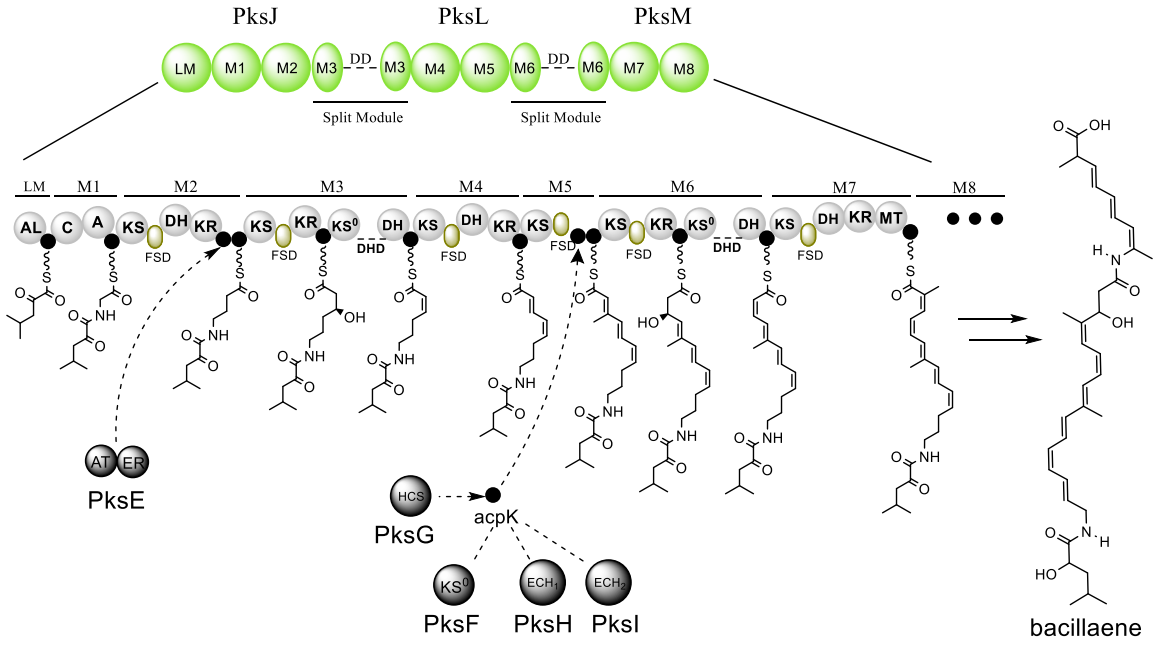


Figure 3. Schematic of a Portion of the *trans*-AT PKS-NRPS Responsible for the Production of the Polyketide Bacillaene. Biosynthetic polyketide intermediates are bound to acyl carrier protein domains (shown in black). The interactions between specified ACP domains and *trans*-acting enzymes are denoted by the dashed arrows. Split modules M3 and M6, another unconventional feature of *trans*-AT domains, are encoded across multiple genes (shown above). The ACP domain abbreviations are as follows: A, adenylation domain; C, condensation domain; AL, acyl-CoA ligase domain; AH, acyl hydrolase; AT, acyl transferase domain; C, condensation domain; DH, dehydratase domain; DHD, dehydratase docking domain; ECH enoyl-CoA hydratase-like enzyme; ER, enoylreductase domain; FSD, flanking subdomain; HCS, HMG-CoA synthase-like enzyme; KS, ketosynthase domain; KS^{Q}, malonyl-ACP decarboxylase domain; KS^{0}, non-elongating ketosynthase domain; KR, ketoreductase domain; MT, methyl transferase domain; Figure adapted from Kosol *et al.*[13]

trans-ATs can exist as a monofunctional enzyme or in certain cases can form a bi- or tri-domain by fusing with enoyl reductase and acyl hydrolase domain proteins which further contributes to the diversity and complexity of polyketide natural products.[9]

Previous analysis of the PKS architecture of both the *bae* and *dif* gene cluster in *B. amyloliquefaciens* FZB42 has revealed several notable deviations from the collinearity rule along with modules lacking domains that would be predicted from the polyketide structure suggesting the presence of *trans*-AT PKS modules within the biosynthetic pathway of difficidin and bacillaene.[14]

I. A. iii. β-branching

The term "β-branching" refers to the polyketide modification which occurs during elongation *in trans,* which entails the conversion of a β-carbonyl function into a carbon branch. There exists some structural diversity within the types of "β-branch" produced in the polyketide natural product based on the enzymatic components present within the *trans*-AT PKS pathway. However, typically the first step within the biosynthesis of a "β-branch" involves the decarboxylation of the malonyl-ACP to yield the acetyl-ACP which is catalyzed by a modified non-elongating keto-synthase enzyme (often denoted as KS*). This modified KS* is missing the necessary amino acid residues required for condensation and is consequently non-elongating. KS* is encoded by the genes *pksF* in *B. subtilis* 168 and *yhdS* in *B. velezensis* ATCC 39374.[15] The enzymes produced by *pksG* and its homologue *dfnL* are proposed to have similar function to the HMGCS, 3-hydroxy-3-methylglutaryl-synthase, enzyme responsible for the primary metabolism in the mevalonate-dependent isoprenoid pathway and in the biosynthesis of terpene natural products.[9]

The HMGCS-like synthase encoded by *pksL* or its *dfn* homologue *dfnJ* catalyze the reaction of the acetyl-ACP and β-keto-ACP. In the proposed mechanism of this

branching enzyme in the case of curacin A biosynthesis, a deprotonation of the acetyl unit on the acetyl-HMGS leads to the formation of an enolate nucleophile which attacks the β-carbonyl of a nearby acyl-ACP by means of an aldol addition. This HMG-ACP intermediate is then hydrolyzed to release HMG-CoA-like polyketide intermediate then undergoes a dehydration and subsequent decarboxylation via an enoyl-CoA-like hydratase enzyme (ECH) to yield an isopentenyl derivative (Figure 4).[16] While this comparison serves a simplified model for the study of polyketide synthesis, HMGCs' yield HMG-CoA, whereas the "β-branching" pathway produces an HMG-CoA-like intermediate where the elongated polyketide chain constructed up until this point within the PKS takes the place of the methyl group in the HMG-CoA structure.

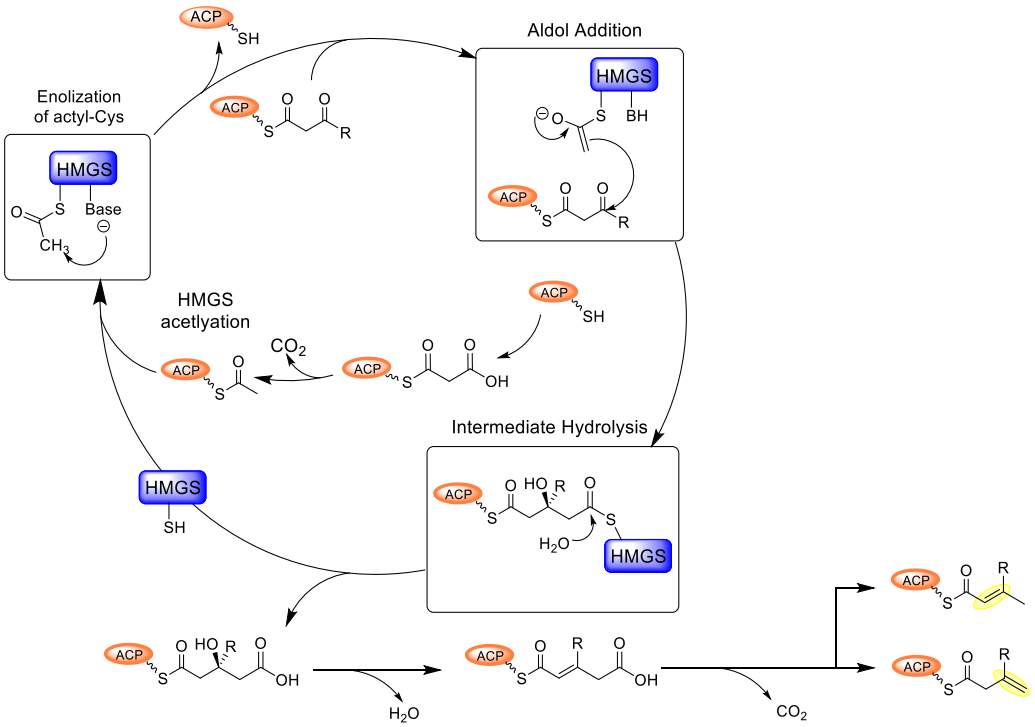


Figure 4. Proposed β-branching Mechanism via the 3-hydroxy-3-methylglutaryl-like Synthase Cassette. Figure adapted from Maloney *et al.*[7]

The subcluster thought to be responsible for the production of the "β-branch" within *trans*-AT PKS pathways encodes the acyl carrier protein (ACP), the modified KS*, and the HMGSs like enzymes along with one or several Enoyl-CoA dehydratases.[9] This subcluster is often referred to as the "β-branching", or "HMG" cassette.

"β -branching", while a relatively new discovery, has been discovered to be encoded in the gene clusters of a host of different polyketide producing bacteria. A few notable examples of polyketide natural products that contain this "β-branch" cassette are bacillaene, pederin, antibiotic TA, mupirocin, difficidin, and the jamaicamide and curacin families of polyketide secondary metabolites.[4] BLAST search analysis of the DNA of the different bacterial strains that produce "β-branch" containing polyketide secondary

metabolites, exhibit a high degree of sequence similarity among their respective polyketide biosynthetic gene clusters (Figure 5).[17,18] However, for the scope of this introduction will be centered around the analysis of the biosynthetic gene clusters responsible for the biosynthesis of bacillaene, pederin, and difficidin. In each case, large *trans*-AT PKS systems were predicted to be involved in the biosynthesis of these natural products based on structural analysis of the multienzyme biomodules produced by the biosynthetic gene clusters and structural elements within the resultant polyketide metabolites which deviated from the collinearity rules of the canonical type-1 *cis*-AT PKS model.[9]

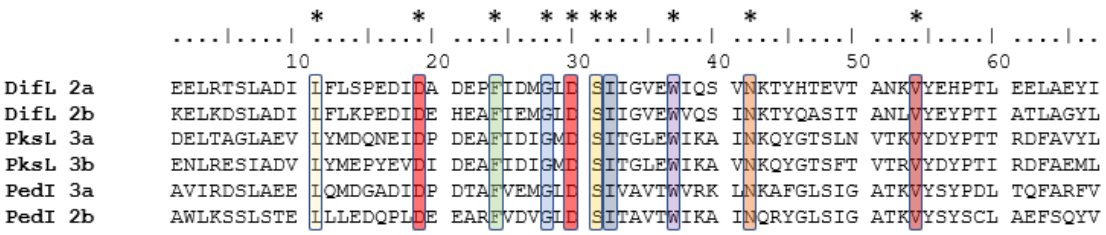


Figure 5. Alignment of the Core Amino Acid Sequences of Acyl Carrier Proteins from Three Type I PKS Clusters Associated with β-branching Activity. The origin of each PKS gene cluster alignment is denoted with the prefix: DifL, difficidin; PksL, bacillaene; PedI, pederin. The 'a' and 'b' correspond to each of the thiolation domains.

I. A. iv. Pederin

Pederin is a polyketide natural product isolated from the *Paederus* rove beetle and a bacterial symbiont of the marine sponge *Theonella swinhoei*.[19,20] The structure of pederin (Figure 6) contains a C4 exomethylene "β-branch" similar to that found in difficidin. The PKS *ped* gene cluster, encoding the biosynthesis of pederin, was found to contain several aberrant features. While a large *trans*-AT PKS system is implicated in the biosynthetic production of pederin, an unusual characteristic of the *ped* cluster architecture is the distribution of genes within the cluster across three distinct genomic regions.[9] While this presented challenges in the overall identification and isolation of the genes involved in the biosynthetic gene cluster, domain analysis of the joined PKS exhibited outstanding architectural agreement with the structure of pederin. However, the highest degree of deviation from the collinearity rule exists within the last six PKS modules with the PedH gene, which were found to be superfluous and could not be attributed to any polyketide counterpart within the pederin molecule.

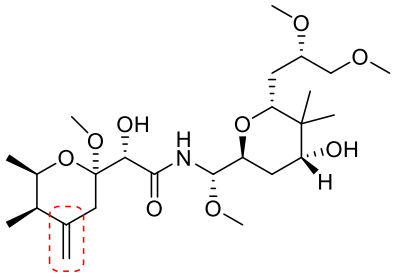


Figure 6. Structure of Pederin. The exomethylene β-branch is indicated by the red dashed line.

The genes thought to be responsible for the "β-branching" within the *ped* cluster are: PedI, F, L, M, N, and P. PedI encodes a homologue of a GCN5-related N-acetyltransferase (GNAT) commonly involved in histone acetylation.[21] It is hypothesized that the GNAT homologue catalyzes the incorporation of an acetyl starter unit similar to that exhibited in biosynthesis of curacin A.[9] PedI also encodes two ECH domains which catalyze dehydration and decarboxylation resulting in the formation of the "β-branch". It is proposed that the ECH domains work in collaboration with a modified KS* found within PedM, an HMGS within PedP, an additional ECH domain from PedL, an acyl carrier protein from PedN, and an unusual occurrence of two DH domains from PedF to form the exomethylene "β-branch".[9]

I. A. v. Bacillaene

Bacillaene is an antibiotic polyketide secondary metabolite initially discovered from its isolation from *Bacillus subtilis*, however was later found to also be a secondary metabolite of *Bacillus velezensis* (previously classified as *B. amyloliquefaciens* FZB42). The structure of bacillaene is a linear polyene molecule with two amide bonds, its structure contains a "β-branch" on the C18 carbon. The PKS gene cluster encoding the

polyketide bacillaene in *Bacillus subtilis* 168 was termed *pksX*. However, speculation arose over the function of the *pskX* pathway when genomic sequencing data of the strain *Bacillus subtilis* 168 was released. Along with many errors found within the sequencing data, a mutation was discovered within the 168 strain, which revealed a loss of function in the gene product of *sfp*, the 4'-phosphopantetheinyl transferase.[17] Without a functioning *sfp* gene, the bacteria is unable to produce the majority of its secondary metabolites including polyketides. Further complicated by bacillaene's non-canonical polyketide biosynthesis made it difficult to characterize its structure. However, the seminal work of the Clardy research group developed methodologies which were able to piece together the structural components of bacillaene, which was expounded upon by the Walsh and Kelleher groups, who then characterized the functionality of the genes within the *pksX* gene cluster by means of enzymatic analysis through the use of Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICRMS).[22] This method analyzed the polyketide monomeric precursors and covalently bound polyketide intermediates bound to the PPant arm of the ACP produced by different enzymes and catalytic domains present within the *pksX* gene cluster.[22] The main findings from these experiments proposed that a *trans*-AT domain, a product of the *PksC* gene, exclusively loaded acyl carrier proteins with malonyl-CoA rather than acetyl- or methylmalonyl-CoA. Through the development of this new analytical proteomic technique, Walsh and Kelleher were the first to observe the conversion of β-keto function to a carbon branch and coined the term "β-branching".

In an earlier 2006 paper published by Calderone, Walsh, Kelleher, and Pieter Dorrestein, detailed the assignment of functional roles to the *pksX* genes responsible for formation of "β-branch" in the previously unsolved bacillaene structure.[17] They proposed that the "β-branching" genes were *AcpK*, *PksC* which encodes a malonyl acyltransferase (AT), *PksL*, which encodes a large multi-modular synthase, in which a smaller sub-component was cloned and expressed to serve as an acyl carrier protein containing a double thiolation domain, *PksF* that encodes a ketosynthase (KS*), *PksG* which encodes a HMG-CoA synthase homologue (HMGS), *PksH* and *PksI* which both code for different enoyl-CoA hydratase homologues (ECH1, ECH2). The characterization and assignment of gene activity of the *pksX* gene cluster of *Bacillus subtilis* 168 was accomplished by phosphopantetheine ejection assay and analysis of the polyketide intermediates by tandem FT-ICRMS outlined above.[22]

I. A. vi. Difficidin

The macrolide structure of difficidin is a 27-carbon phosphorylated macrolide containing a "β-branch" on carbon 3. The polyketide synthase gene cluster responsible for its production is the *dif* cluster from *B. velezensis* which has been previously characterized and has aided in our endeavors to characterize the *dfn* gene cluster of *B. velezensis* ATCC 39374. Crystal structures of the *trans*-AT domain involved in the production of difficidin have been obtained and reveal that the architecture is similar to that of discrete type II PKSs rather than the canonical type I PKSs.[8] We propose that the genes involved in the "β-branching" within the *dfn* cluster are: *dfnX, baeC, yhdS, dfnJ, dfnL, baeH, and dfnM*. We hypothesize that the enzyme encoded by *bae*C (*PksC*)

catalyzes the transfer of a malonyl acyl group from malonyl-CoA to *holo* form of the acyl carrier protein *dfnX (acpK)*, which is subsequently decarboxylated by the enzyme encoded by *yhdS* to yield acetyl-S-DfnX (Ac-S-DfnX).

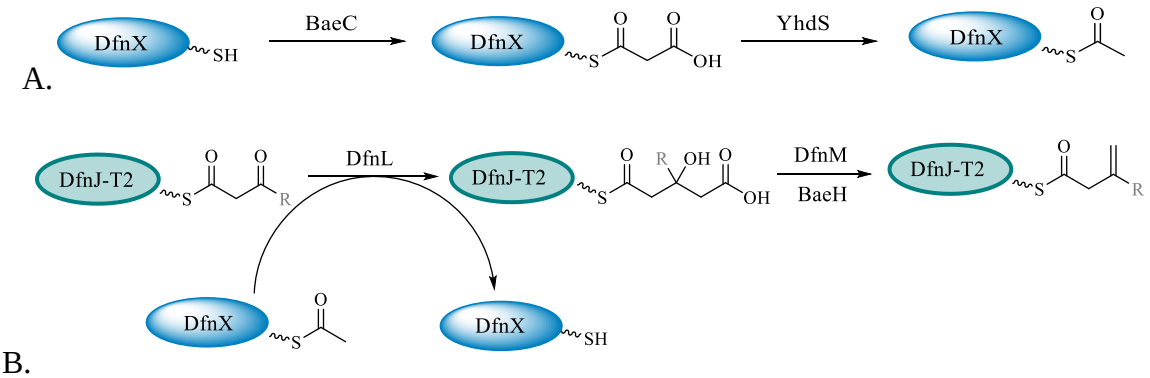


Figure 7. Scheme of the Proposed Biosynthetic Pathway Encoded by the *dfn* Polyketide Synthase Gene Cluster Responsible for the Incorporation of the β-Branch in Difficidin.

What has previously not been shown in preliminary data involves the gene product of *dfnL (PksG)* hypothesized to catalyze the reaction of the beta-keto acyl polyketide intermediate, which was modeled experimentally as Acac-S-DfnJ-T2 (Acac-S-PksL-T2) and acetyl-S-DfnX (Ac-S-AcpK) to yield a HMG -S-DfnJ-T2-like (HMG-S-PksL-T2) intermediate. The subsequent dehydration and the decarboxylation of HMG-S-DfnJ-T2 is catalyzed by the enzymes encoded by *baeH* and *dfnM* (baeI) respectively, which produce an external olefin containing isopentenyl derivative resulting in the "β-branch" that will be incorporated into the difficidin structure (Figure 7). The *pksX* and *bae* homologs for each *dfn* gene have been provided in parentheses to denote the parallels between these biosynthetic pathways, Figure 1 displays the full gene clusters with the

genes hypothesized to be responsible for the incorporation of the "β-branch" are indicated in colors corresponding to each gene's homologue in the respective gene cluster (Figure 8).

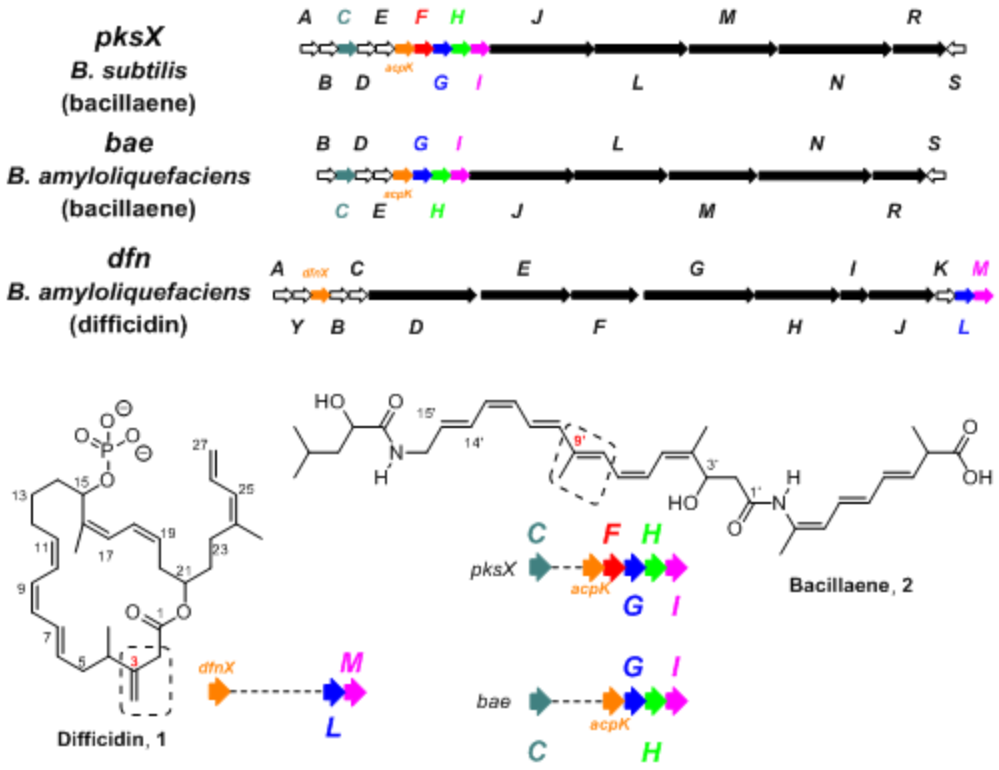


Figure 8. A Depiction of the *pksX*, *bae*, and *dfn* Gene Clusters in which the Homologous Genes Hypothesized to be Responsible for the Incorporation of the "β-branch", Across these Gene Clusters have been Color-coded.

As previously stated, the gene activity of the bacillaene-producing *pksX* and *bae* gene clusters present within *Bacillus subtilis* 168 and *B. velezensis* FZB42 (previously classified as *B. amyloliquefaciens* FZB42) respectively, and the difficidin-producing *dif* gene cluster in *B. amyloliquefaciens* FZB42 have been extensively characterized. However, the long-term goal of the Reddick research lab has centered on the

characterization and assignment of function of the homologous polyketide producing gene clusters in *B. velezensis* strains ATCC 39374 and 39320. Presently, these particular strains' genomic DNA have not yet been sequenced but due to their taxonomic classification and reported secondary metabolite profiles it had been hypothesized that these strains contain genes encoding catalytic machinery analogous to that found within polyketide synthase gene clusters of strain *B. subtilis* 168. However, research initiated by Sriparna Mukherjee, Amy Quattlebaum, and Brittany Kiel, former researchers in the Reddick lab, not only revealed a high incidence of mismatches in the DNA sequence alignment of the *pksG* homologue in ATCC 39374 but a 100% match in sequence alignment with the pksG homologue found within *B. amyloliquefaciens* FZB42. This called into question the validity of the initial taxonomic classification of ATCC 39374 as a strain of *B. subtilis*, and if the strain should be reclassified as a strain of *B. amyloliquefaciens.*

The DNA sequencing data we have obtained from the *B. velezensis* strains ATCC 39374 exhibit almost negligible differences as compared to the DNA sequence of *B. velezensis* FZB42. Furthermore, what few discrepancies of base pairs that do exist between their genomic DNA sequences encode silent mutations that ultimately produce the same amino acid sequences. Since this discovery, our investigations into the *bae* and *dfn* biosynthetic gene clusters, responsible for the production of difficidin and bacillaene in ATCC 39374, have been modeled after the sequence data and previously characterized gene activity of gene homologues found within the *bae* and *dif* PKS gene clusters in *B. velezensis.*

I. A. vii. Introduction to Molecular DNA Cloning with Plasmid Vectors

The crux of analyzing the intricate protein-protein interactions that occur within the biosynthesis of polyketide secondary metabolites is overcoming the inherent complexity of the bacterial cell. The ability to isolate, amplify, and express genes of interest allows for the systematic study of specific proteins and cellular components down to molecular scale. Molecular cloning with recombinant DNA technology has overcome this obstacle allowing virtually any gene to be purified, its DNA sequence studied, and its functionality probed by means of making site directed alterations.[23]

Molecular cloning is a procedure by which a gene of interest is isolated, amplified, and "chemically inserted" into plasmid vector to ultimately create a recombinant DNA molecule from which a multitude of identical copies can be generated by the non-native host's cells. DNA cloning with plasmid vectors enables the mass production of identical DNA molecules by linking the DNA encoding a gene of interest to a vector DNA molecule through a standard 3' → 5' phosphodiester bond. The most widely used vectors are *E. coli* plasmids, circular double-stranded DNA molecules have been engineered into cloning vectors due to their relatively small size, their ability to self-replicate their plasmid DNA and gene of interest into each daughter cell, and variable antibiotic resistance genes.

I. A. viii. Polymerase Chain Reaction

The gene of interest is isolated from its native genomic DNA through the technique of polymerase chain reaction (PCR). This procedure allows DNA from a predetermined region of the genome, typically associated with a gene, to be amplified a

billion fold, effectively isolating this segment of DNA from the residual genomic DNA.[23] An "upstream" and "downstream" set of DNA oligonucleotide "primers", are designed to flank the nucleotide sequence of the gene of interest. These "primers" fulfill their titular role by priming DNA synthesis of the flanked region on the dissociated single strands of the heated genomic DNA. The reaction that produces the newly synthesized DNA is catalyzed by a purified DNA polymerase added to the reaction mixture. The polymerase chain reaction amplifies the gene of interest by turning each newly synthesized double-stranded DNA fragment into two new single stranded templates for the DNA synthesis reaction to occur by way of successive rounds of heating, a thermophilic DNA polymerase, and an abundance of dNTPs present within the PCR Master Mix. After 20-30 cycles of this polymerase chain reaction roughly 2^ (20-30) copies of the DNA that encodes the gene of interest are produced and constitute the most abundant species in the reaction. Recombinant DNA plasmids can be formed from these engineered *E. coli* vectors and amplified PCR fragments of the gene of interest using two types of enzymes: restriction enzymes and a DNA ligase.

I. A. ix. Restriction Digestion

Restriction enzymes, or restriction endonucleases, cut DNA at a specific enzyme dependent 4 to 8 base pair recognition site, leaving a small segment of single stranded DNA with complimentary base pair overhangs (sticky ends). Digestion of the plasmid vector with restriction enzymes will result in the plasmid being cut at the enzyme dependent restriction recognition site present within the vector, thereby linearizing the circular plasmid DNA and leaving the appropriate single stranded complimentary

overhangs. If the purified PCR fragments of the gene of interest are also subjected to digestion with the same restriction enzymes, both the plasmid and the DNA fragments from the gene of interest will contain complimentary overhangs (sticky ends). The addition of DNA ligase to a reaction mixture containing both digested products will covalently bind, or "ligate", the complimentary overhangs of the linearized plasmid to the sticky ends of the digested purified PCR fragments to catalyze the formation of 3' -> 5' phosphodiester bonds resulting in the "chemical insertion" of the gene of interest into the newly reformed circular plasmid vector.

This ligated product is then "transformed" into chemically competent *E. coli* cells and colonies are grown on nutrient rich LB plates containing the antibiotic corresponding to the vector's antibiotic resistant gene. This allows for the exclusive propagation of *E. coli* bacterial cells that contain the vector. These colonies are then selectively grown up and screened for their ability to replicate the plasmid containing the gene of interest by means of PCR and DNA sequencing.

I. A. x. pET-Duet System

The coexpression of multiple genes of interest in *E. coli* is optimal for the analysis of protein complexes. pETDuet-1 is a commercially available T7 promoter expression *E. coli* vector specifically designed for cloning and the coexpression of two target genes. The pETDuet-1 vector contains two multiple cloning site (MCS) regions each with their own set of T7*lac* promoters and a single T7 terminator, which are targeted by the T7 RNA polymerase. The plasmid also contains the *lac I* gene responsible for controlling basal expression rates by upregulating the amount of *lac I* repressor which then binds to

this operator sequence which allows the binding of isopropyl β- d-1-thiogalactopyranoside (IPTG) to induce the expression of the genes within each multiple cloning site.[24] The vector offers a wide array of restriction recognition sites for easy clone transfer (*Bam*H I, *Eco*R I, *Sac* I, *Sal* I, *Hin*d III, and *Not* I for MCS1 and *Nde* I, *Bgl* II, *Mun* I, and *Xho* I for the second MCS2).[24] pETDuet-1 encodes an N-terminal 6-amino acid His•Tag® for purification by nickel nitriloacetic acid (Ni^{2+}-NTA) affinity chromatography (Figure 9).

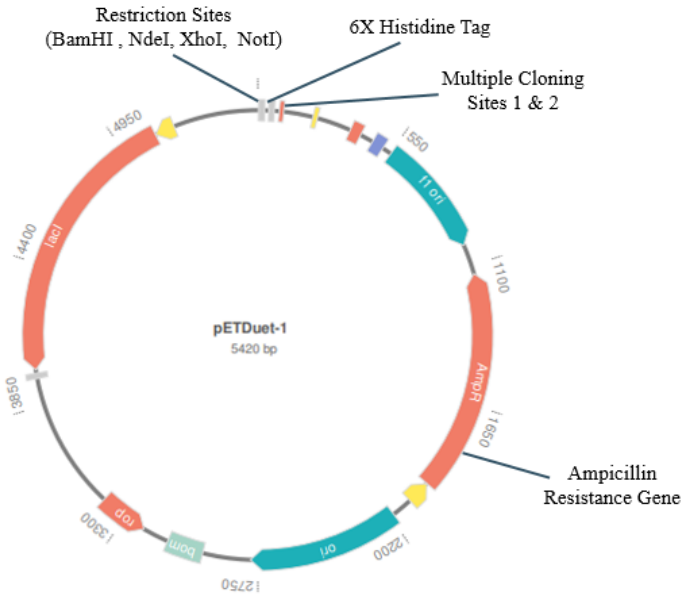


Figure 9. pETDuet-1 Vector Map. Features of the pETDuet-1 vector are as follows: ampicillin resistance gene, multiple cloning sites 1 and 2, the polyhistidine affinity His•Tag®, and restriction site domains.

I. A. xi. Protein Purification by Ni^{2+}-NTA Affinity Chromatography

The purification of all proteins containing the His•Tag® in this work were purified by Ni^{2+}-NTA metal affinity chromatography. This chromatographic method utilizes the tetradentate chelating character of nickel nitriloacetic acid and two histidine

amino acid side chains from the polyhistidine affinity tag as a ligand matrix to coordinate Ni^{+2} in an octahedral coordination environment (Figure 10).[25] The His•Tag® allows for the expressed proteins to bind to the Ni^{2+}-NTA column with significantly higher binding affinity than non-affinity tagged proteins. The column can then be washed to remove non-specific bound proteins by the addition of increasing concentrations of imidazole. Imidazole competitively binds to the ligand matrix eluting the protein from the transition metal containing resin. The protein of interest can also be eluted in this manner by increasing the concentration of imidazole to 100mM or higher.

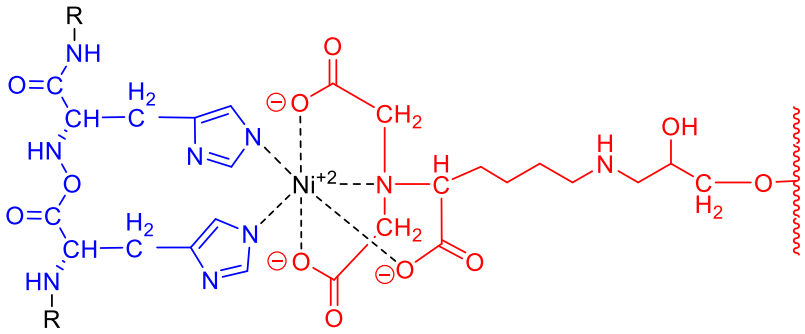


Figure 10. Model of the Interactions between Ni^{2+}-NTA Immobilized Metal Affinity Matrix and the Polyhistidine Affinity His•Tag®.[25] The Ni^{2+} metal center exhibits octahedral coordination by the fixed tetradentate chelating nitriloacetic acid and two histidine side chain residues from the His•Tag® containing protein.

I. B. Overview and Objectives

The polyketide macrolactone, difficidin, is a secondary metabolite of *B. velezensis* ATCC 39374 and 39320 which has been found to exhibit broad-spectrum antibacterial activity against both Gram-positive and Gram-negative bacteria.[1] While the majority of the difficidin structure follows canonical polyketide biosynthesis, it contains an

unprecedented exomethylene "β-branch" which is found in only a small number of other polyketide natural products. In past work conducted by the Reddick lab, it had been hypothesized that *bae*C, a malonyl-CoA-acyl carrier protein transacylase from *B. velezensis* strain ATCC 39374, loaded a malonyl acyl group from malonyl CoA onto the free thiol group of the phosphopantetheine arm of *holo*-DfnX acyl carrier protein from the *dfn* biosynthetic gene cluster. While data from these initial experiments appeared to support this initial hypothesis, it was revealed that the promiscuous 4'-phosphopantetheinyl transferase encoded by *sfp* from strain OKB105 used in this *in vitro* system was capable of independently loading the entire, malonyl-containing, phosphopantetheine prosthetic group from malonyl-CoA directly to the *apo* form of DfnX. Therefore, the data produced from these experiments were not reflective of the true malonyl acyl transferase activity of *baeC*. The ***goal*** of the research presented in this thesis was to redesign an *in vitro* system that does not require the addition of the 4'-phosphopantetheinyl transferase to activate *apo*-dfnX.

The characterization and assignment "β-branching" activity of genes within the *dfn* and *bae* gene clusters in the *B. subtilis* ATCC 39374 strain has been an ongoing project conducted by Brittany Kiel of the Reddick Lab. By modeling the biochemical insights gleaned from the biosynthesis of bacillaene by the *pksX* gene cluster from *B. subtilis* 168, previously characterized by Walsh and Kelleher, and the biosynthesis of bacillaene and difficidin produced by the respective *bae* and *dif* gene clusters in *B. velezensis* (previously classified as *B. amyloliquefaciens* FZB42) characterized by Chen, Piel, and Borris, an on-going long-term goal of this project is to assign the activity of the

genes responsible for the incorporation of the "β-branches" in the polyketide synthase gene clusters of *B. subtilis* ATCC 39374.[26]

The overall objective of the work presented in this thesis is to characterize the genes within the *dfn* gene cluster which we hypothesize to constitute the biosynthetic pathway for the formation of the exomethylene β-branch in the difficidin structure by the analysis of polyketide intermediates via tandem electrospray ionization mass spectrometry analysis (ESI-MS/MS). The central hypothesis is that the enzyme encoded by *bae* C catalyzed the transfer of a malonyl acyl group from malonyl-CoA to *holo* form of the acyl carrier protein *dfnX*, which is subsequently decarboxylated by the enzyme encoded by *yhdS* to yield acetyl-S-DfnX (Ac-S-DfnX). What has previously not been shown in preliminary data involves the gene product of *dfnL* hypothesized to catalyze the reaction of acetoacetyl-S-DfnJ-T2 (Acac-S-DfnJ-T2) and Ac-S-DfnX to yield a 3-hydroxymethylglutaryl-S-DfnJ-T2-like (HMG-S-DfnJ-T2) intermediate. The subsequent dehydration and the decarboxylation of HMG-S-DfnJ-T2 is catalyzed by the enzymes encoded by *baeH* and *dfnM* respectively, which is hypothesized to produce an external olefin containing isopentenyl derivative resulting in a difficidin-like "β-branch". Whereas, the catalysis of this reaction by the genes *baeH* and *baeI* is hypothesized to produce an internal olefin containing isopentenyl derivative bacillaene-like "β-branch".

The scope of this project only encompasses the assignment of activity to the genes responsible for the formation of this internal or external olefinic isopentenyl product, however further structural characterization of the gene-specific isopentenyl derivatives beyond analysis of MS^3 fragmentation patterns are a future goal for the lab. The rationale

that underlies the research proposed here is that the characterization of the activity of the *dfn* gene cluster in *B. subtilis* ATCC 39374 would give a much deeper and strain specific understanding of the formation of the polyketide "β-branch" in difficidin, rather than a strict reliance on the sequencing data and proposed order of PKS modules within the *dif* cluster of *B. velezensis.* Progress toward the long-term goal of the Reddick laboratory and the accomplishment of this proposed master's project's overall objective will be realized through the following ***specific objectives***:

1. **Develop a new gene expression construct that can produce active holo-DfnX and use it for mass spectrometry studies of the malonyl transacylase BaeC.** The *working hypothesis* that will be tested in this objective is that cloning genes *dfnX* and *sfp* into the plasmid pETDuet-1 will allow for the co-expression of these proteins and the exclusive isolation of the his-tagged *holo*-DfnX by Ni-NTA affinity chromatography. The new co-expression system of holo-DfnX will allow studies of the BaeC-dependent malonyl acyl transferase activity, without the artificially increased activity caused by the presence of the Sfp enzyme.
2. **Characterization and assignment of *dfnL* activity in the *dfn* polyketide synthase gene cluster of ATCC 39374.** The *working hypothesis* that will be tested within this objective is that the enzyme encoded by *dfnL* will catalyze the addition of Ac-S-DfnX and Acac-S-DfnJ-T2 to form an HMG-like polyketide intermediate. Assignment of this activity will be performed via HPLC ESI-MS/MS analysis.
3. **Analysis of MS/MS fragmentation patterns of the *baeI* and *dfnM* derived isopentenyl polyketide intermediates.** The *working hypothesis* that will be tested

within this objective is that the location of the double bond within the isopentenyl polyketide intermediate formed will be dependent on whether the reaction is catalyzed by *baeI* or *dfnM*. Preliminary data has shown that both enzymes form an isopentenyl intermediate on DfnJ's double thiolation domain of the same exact mass, the MS^3 fragmentation patterns of the phosphopantetheine ejection species may provide valuable structural insights that could differentiate the two intermediates. Optimization of an intact HPLC ESI- MS^3 method for fragmentation pattern analysis is underway to test this objective.

The expected significance of the research project proposed here is advancing the knowledge of the genetic sub clusters responsible for the formation of rare structural moieties within polyketide natural products. Thereby the characterization of genes involved in the formation of "β- branching" structural motifs can lead to engineered diversification and production of novel polyketide natural products by reaction of these enzymes with non-native substrates. Another potential future application could be the utilization of PKS enzymes in a semi-synthetic capacity produce complex structural scaffolds for pharmaceutical drug leads as a means to circumvent long yield-diminishing difficult synthetic routes.

CHAPTER II

OVEREXPRESSION AND PURIFICATION OF PETDUET-1-*DFNX-SFP*

II. A. Overview and Objectives

The objective of the work described in this chapter is the development of a new gene expression construct that produces active holo-DfnX for mass spectrometry studies of the malonyl transacylase BaeC. The *working hypothesis* that was tested in this objective is that cloning genes *dfnX* and *sfp* into the plasmid pETDuet-1 will allow for the co-expression of these proteins and the exclusive isolation of the his-tagged *holo*-DfnX by Ni-NTA affinity chromatography (Figure 11). The new co-expression system of holo-DfnX will allow studies of the BaeC-dependent malonyl acyl transferase activity, without the artificially increased activity caused by the presence of the Sfp enzyme. The new *in vitro* construct allows for genes *dfnX* and *sfp* to be coexpressed upon induction, the 4'-phosphopantethiene transferase encoded by *sfp* will catalyze the transfer of the phosphopantetheine prosthetic group from endogenous coenzyme A produced by *E. coli* to the conserved serine of the phosphopantetheine binding site on the DfnX acyl carrier protein. By cloning *dfnX* into multiple cloning site 1 into pETDuet-1 the post-translationally modified *holo*-DfnX product can be exclusively isolated by Ni^{2+}-affinity chromatography

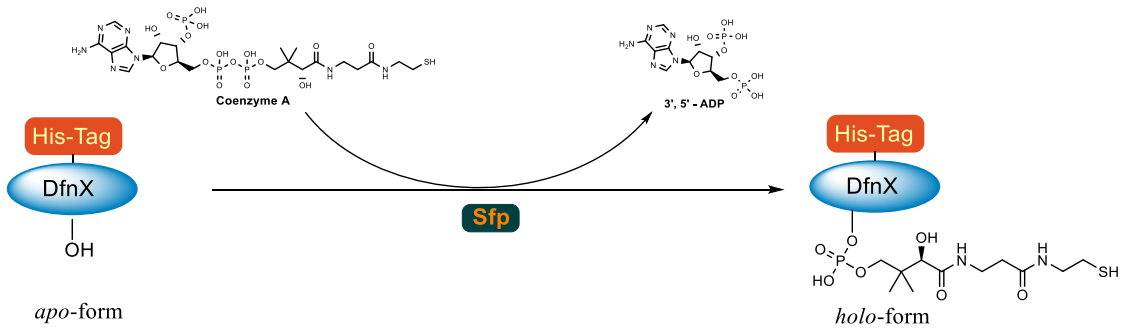


Figure 11. Reaction Scheme of pETDuet-1-*dfnX-sfp in vitro* Construct. The 4'-phosphopantetheinyl transferase encoded by *sfp* is hypothesized to catalyze the addition of a phosphopantetheine prosthetic group onto DfnX from the endogenous Coenzyme A found in *E. coli*. The polyhistidine affinity His•Tag® should allow for the exclusive isolation of the *holo*-form of the DfnX protein.

II. B. Experimental

II. B. i. Preparation of PCR Reactions

Forward and Reverse primers containing restriction sites were designed to amplify these respective genes of interest. 25 µL of Phusion® High-Fidelity PCR Master Mix with HF Buffer, 2 µL of template DNA from their respective sources, 17µL of ddH2O, and 3 µL of each forward upstream and downstream primers made up the 50µL PCR reaction mixture.

The gene *dfnX* was PCR amplified with the oligonucleotides 5'-TGGAGCTGGGATCCGATGGAACAAACCAAG-3' and 5'-ATCGCGGCCGCTTTCTGTTTTCATGATTGAG 3' from the genomic DNA from *B. subtilis* ATCC 39374 isolated using a Promega Wizard® Genomic DNA purification kit. The genomic DNA obtained served as the template DNA for the polymerase chain reaction experiments involving DfnX.

The gene *sfp* encoding the 4'- phosphopantetheinyl transferase was amplified by PCR with oligonucleotides 5'-GATCTCATATGAAGATTTACGGAATTTATATG-3' and 5'-GTTGATGCTCGAGTTATAAAAGCTCTTCGTAC-3' from a pUC8 plasmid containing the intact *sfp* gene from the strain OKB105 of *B. subtilis*, graciously received from Peter Zuber, served as the template DNA for the polymerase chain reactions involving *sfp*.[27]

The thermocycler program utilized consisted of a 30 second initial denaturation at 98 °C, followed by the 30X cycle primer extension steps involving denaturation at 98 °C for 5 seconds, annealing at 60 or 55 °C for 10 seconds, an extension at 72 °C for 1 minute. The final extension performed at 72 °C for 10 minutes and the temperature was held at 4 °C. The annealing temperature varied and was contingent upon the lowest Tm among the upstream and downstream primers. The annealing temperature was 60 °C for the DfnX and 55 °C for the Sfp polymerase chain reactions.

The PCR products were purified using Qiagen QIAquick® PCR Purification kit. The efficacy of the PCR reactions was analyzed by 1% Agarose gel electrophoresis with ethidium bromide. Each gene was inserted into the pETDuet-1 vector individually. A restriction digest was performed by adding the restriction enzymes to both the PCR products and the target plasmid. The desired fragments were then extracted from the agarose gel via QIAquick® Gel Extraction Kit. The digested linearized pETDuet vector and digested DNA fragments were then ligated using T4 DNA ligase, into the multiple cloning site of the pETDuet vector. The ligation reaction was incubated at 37°C for 1 hour, then held at 16 °C overnight in a thermocycler.

II. B. ii. Cloning the DfnX and Sfp Genes and the Construction of petDuet-1-*dfnX*

The gene *dfnX* was PCR amplified with the oligonucleotides 5'-TGGAGCTGGGATCCGATGGAACAAACCAAG-3' and 5'-ATCGCGGCCGCTTTCTGTTTTCATGATTGAG 3' from the genomic DNA from *B. subtilis* ATCC 39374 isolated using a Promega Wizard® Genomic DNA purification kit. The genomic DNA obtained served as the template DNA for the polymerase chain reaction experiments involving DfnX. The amplified DfnX fragment was digested with BamHI and NotI and ligated into the BamHI and NotI restriction sites of double digested linearized pETDuet-1 vector to give the expression plasmid pETDuet-1-*dfnX*.

II. B. iii. Construction of petDuet-1-*dfnX-sfp*

The gene *sfp* encoding the 4'- phosphopantetheinyl transferase was amplified by PCR with oligonucleotides 5'-GATCTCATATGAAGATTTACGGAATTTATATG-3' and 5'-GTTGATGCTCGAGTTATAAAAGCTCTTCGTAC-3' from a pUC8 plasmid containing the intact *sfp* gene from the strain OKB105 of *B. subtilis*, graciously received from Peter Zuber, served as the template DNA for the polymerase chain reactions involving *sfp*.[27] The rationale behind isolating *sfp* from *B. subtilis* OKB105 (*pheA sfp*$^{+}$) is that this strain not only contains a functioning *sfp* gene, but the encoded 4'-phosphopantetheinyl transferase proved effective in the production of all polyketide secondary metabolites in the *pksX* and *dif* gene clusters. The amplified *sfp* PCR product was digested with NdeI and XhoI restriction enzymes and ligated into the NdeI and XhoI sites on the double digested linearized petDuet-1-*dfnX* vector to give the expression plasmid pETDuet-1-*dfnX-sfp.*

II. B. iv. Analysis of pETDuet-1 Restriction Digests and Cloning Reactions

Analysis of the 1% Agarose gel electrophoresis indicated incomplete digestion of the pETDuet-1 vector. The band migration of the pETDuet-1 SalI digested product (Lane 3) was similar to that of the undigested pETDuet-1 vector control (Lane 4). The BamHI restriction site clearly indicated full digestion of the vector as noted by the slowed band migration through the agarose gel as a result of the linearization of the plasmid, whereas the migration pattern of the band produced by the SalI control was similar to that of the undigested circular plasmid. It was concluded that pETDuet-1 system's SalI restriction enzyme recognition site was unable to be accessed or positively identified by the SalI restriction enzyme which hindered the digestion of the vector. Therefore, an additional experiment varying the concentration of SalI restriction enzyme necessary to digest the pETDuet-1 vector was performed (Figure). Despite a five-fold increase in concentration of SalI restriction enzyme, linearization of the pETDuet-1 vector was not observed. As a result of these experiments the downstream *dfnX* primer containing the SalI recognition site was redesigned with the NotI recognition site.

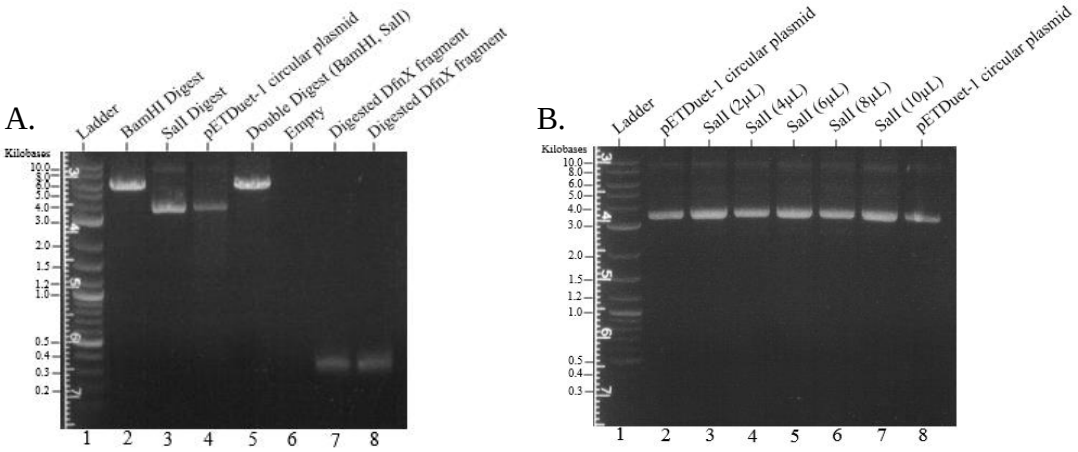


Figure 12. (A) Restriction Digest of pETDuet-1 and *dfnX* PCR Product (B) Analysis of pETDuet-1 Digestion Efficacy using Varied Concentrations of Sal I Restriction Enzyme.

The same restriction digest control experiment was reproduced using the redesigned *dfnX* (NotI) downstream primer (Figure 13) and complete digestion and linearization of the pETDuet-1 plasmid was observed.

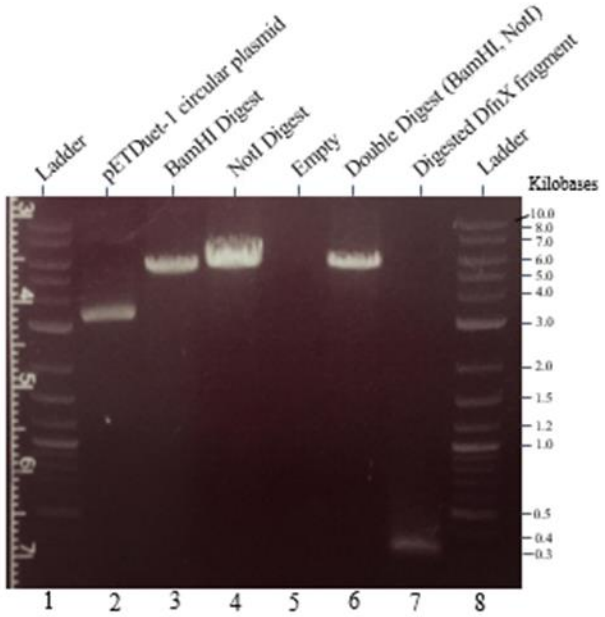


Figure 13. Restriction Digest Analysis of pETDuet-1 and *dfnX* PCR Product using Redesigned *dfnX* (NotI) Downstream Primer.

The gene *dfnX* was cloned into the pETDuet-1 vector. The ligation mixture was transformed into One Shot TOP10 chemically competent *E.coli* cells and plated on LB agar plates containing ampicillin (100 mg/mL). Individual colonies were selected and grown in LB medium containing ampicillin. Plasmids were isolated from these cultures using the QIAprep Spin Miniprep kit and tested for *dfnX* insertion by PCR analysis using the initial cloning primers and visualized via 1 % Agarose gel electrophoresis (Figure 14). Positive pETDuet-1-*dfnX* TOP10 transformants were then concentrated and sequenced by the dye-terminated Sanger method (by Eurofins). The same procedure outlined above, but the primers designed for *sfp* containing the restriction enzyme

recognition sites, NdeI and XhoI, were used to clone the *sfp* gene into the pETDuet-1-*dfnx* vector.

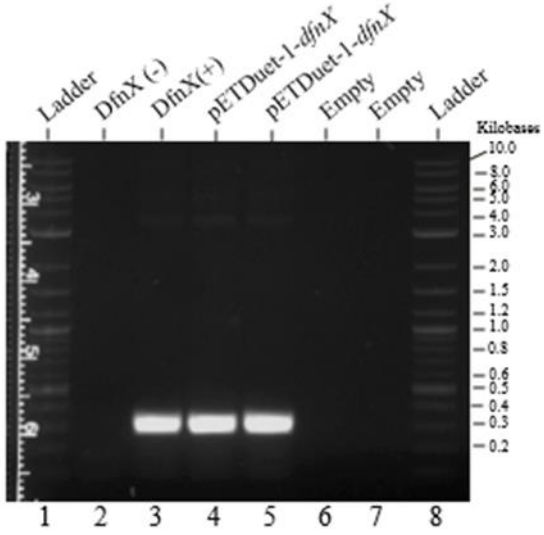


Figure 14. PCR Confirmation Analysis of *dfnX* Gene Insertion into pETDuet-1 to form pETDuet-1-*dfnx*

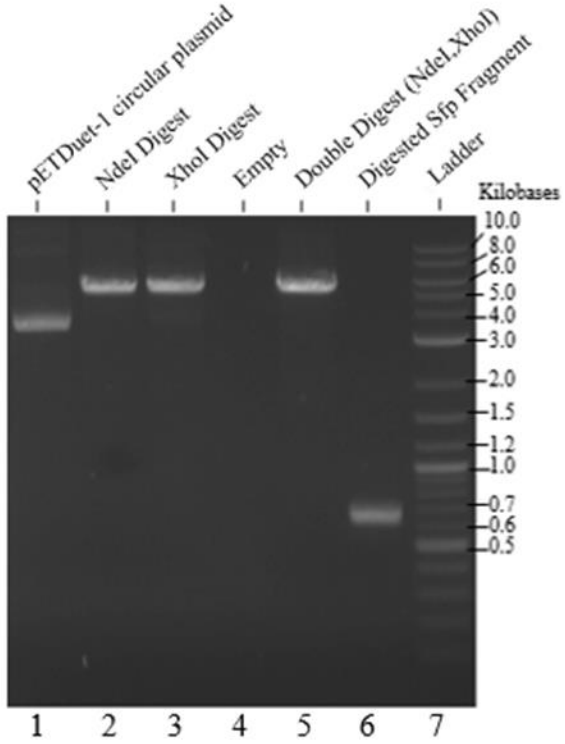


Figure 15. Restriction Digest Analysis of pETDuet-1-*dfnX* and *sfp* PCR Product.

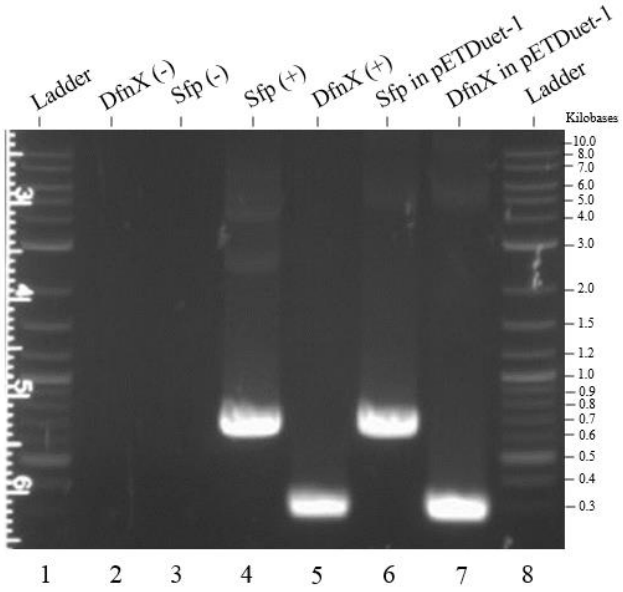


Figure 16. PCR Confirmation Analysis of *sfp* Gene Insertion into pETDuet-1-*dfnX*.

II. C. Experimental

II. C. i. Overexpression and Purification of pETDuet-1-*dfnX-sfp*, BaeC, BaeH, BaeI, DfnJ, DfnL, DfnM, YhdS, and Sfp

The completed pETDuet-1-*dfnX-sfp* plasmid was transformed into BL21 Star (DE3) chemically competent *E. coli* and plated on an LB agar plate containing ampicillin (50µg/mL). An isolated colony was selected and inoculated a starter culture containing 5 mL LB broth containing ampicillin (50µg/mL). A 10% glycerol stock was produced from this culture for long-term storage at -80°C, and the remainder of this starter culture was used in the formation of a large culture. The same protocol was carried out when growing up BaeC, BaeH, BaeI, DfnJ, DfnM, YhdS, and Sfp with the exception that these proteins were cloned into the pET-200 vector which encodes kanamycin resistance, thus kanamycin was substituted for ampicillin in the same concentrations.

Two milliliters of the starter culture were used to inoculate a large culture comprised of 1 liter of LB broth and 1 mL of ampicillin (50 mg/mL) the large culture was incubated with shaking at 205 rpms at 37 °C for 3-4 hours. The optical density of the large culture was monitored by taking 1 mL aliquots which were then analyzed at 595 nm. When the absorbance readings of the large culture aliquots were within the range of 0.5-0.6, the cells were induced with isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. For proteins BaeC, BaeI, DfnM, YhdS, and Sfp the large cultures were incubated at 37 °C overnight post induction. For BaeH, DfnJ, and DfnL the large cultures were incubated at 18 °C overnight post induction. The post induction growth conditions for pETDuet-1-*dfnX-sfp* included a drop in incubation temperature from 37 °C to 18 °C, however the large culture was only incubated for three and a half hours to mitigate degradation of the DfnX protein.

The following day the cells from the large culture were pelleted by centrifugation at 6,500 rpm for 30 minutes at 4 °C. The pellet was resuspended in 15 mL of binding buffer (4 M sodium chloride, 160 mM Tris, 40 mM imidazole, pH=7.9). The resuspended cells were then transferred to a 50 mL centrifuge tube and sonicated on ice at 20 Watts in 30 second bursts for 3 minutes. The lysate was centrifuged at 11,500 rpm for 30 minutes at 4 °C. The supernatant was filtered using a Corning® 0.45 micron, 26 mm syringe filter. The filtered solution was then passed through a Ni-NTA column. A series of binding, wash and elution buffers comprised of: 0.5 M NaCl, 20 mM Tris, 5 mM imidazole; 0.5 M NaCl, 20mM Tris, 60 mM Imidazole; 0.5 M NaCl, 20 mM Tris, 200mM Imidazole respectively. Twenty milliliters of binding buffer, 12 mL of wash

buffer, and 12 mL of elution buffer were passed through to wash the column of undesired proteins and to selectively elute the His-tagged protein of interest.

Eluent was collected in fractions of 1.5 mL aliquots, and these aliquots were tested for the presence of protein by mixing 33 µL of the fraction with 1 mL of Bradford Reagent in a separate microcentrifuge tube. The fractions containing protein were then pooled together in snakeskin dialysis tubing (22mm x 7cm, MWCO = 7,000). Dialysis was performed overnight at 4 °C in 4 L of 25 mM Tris-HCl buffer at pH 8.0. Sterile glycerol was mixed with purified and dialyzed protein to a final concentration of 10% by volume to stabilize the protein for -80 °C storage. The glycerol containing protein was then divided into 200 µL aliquots and stored at -80 °C.

II. C. ii. Determination of Protein Concentration

The concentration of both holo-form DfnX and BaeC were determined spectrophotometrically using the Bradford method, utilizing Bovine Serum Albumin (BSA) as a standard. A five-tier serial dilution of bovine serum albumin with a stock concentration of 2.0 mg/mL was used to produce a standard curve using the following dilutions: 1.0, 0.75, 0.5, 0.25, and 0.125 mg/mL. One milliliter of Bradford reagent was added to each 33 µL aliquot and absorbance was measured spectrophotometrically at 595 nm for each tier of the dilution. The absorbance of each aliquot was plotted as a function of its known concentration to establish a correlation between absorbance and concentration based on the equation of Beer's law. The linear relationship of absorbance to protein concentration was graphically determined as **$y = 0.7287x + 0.0163$ (R^2**

=0.9969) and was used to quantify the unknown concentrations of the overexpressed and isolated proteins of interest.

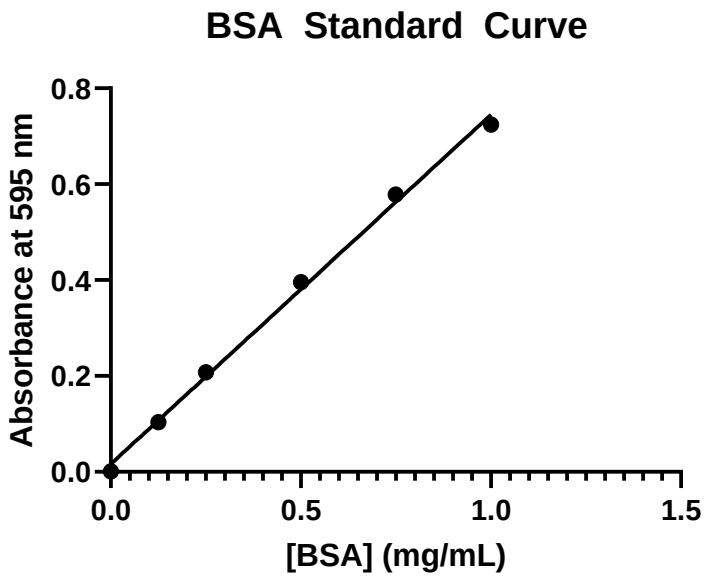


Figure 17. BSA Standard Curve. The concentrations of each overexpressed protein were determined by establishing a linear relationship (**y= 0.7287x + 0.0163) (R^2 =0.9969)** between absorbance and protein concentration by producing of a Bradford standard curve.

II. C. iii. Characterization of BaeC, BaeH, BaeI, DfnJ, DfnM, holo-DfnX, Sfp, YhdS by SDS PAGE

A 10% SDS-PAGE analysis (Figure 18) was performed to qualitatively assess the efficacy of the purification and isolation processes of each of the overexpressed proteins. The bands produced by each protein were compared against a broad-range protein ladder standard to determine if molecular weight of the isolated protein was comparable to the expected molecular weight of the protein (**Table 1**).

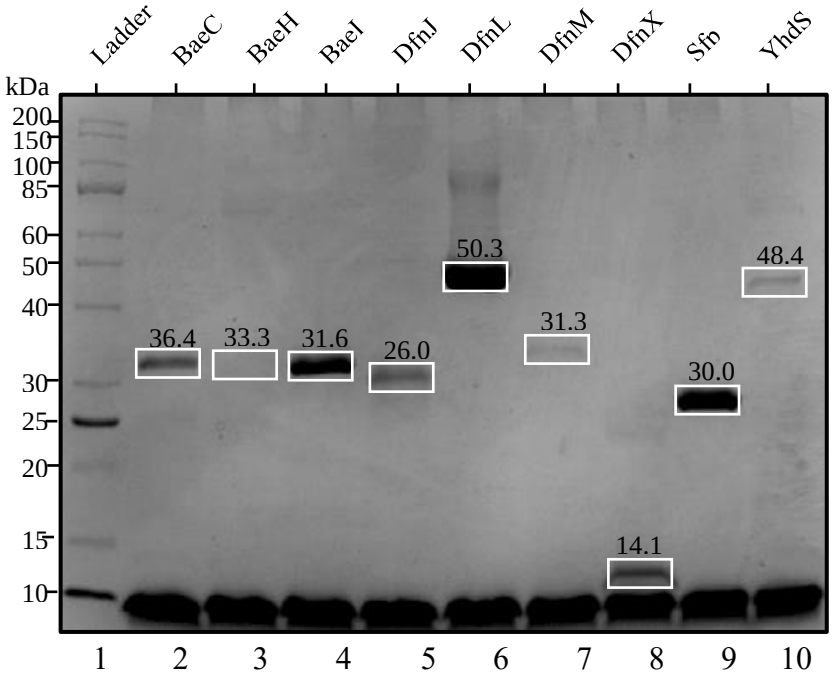


Figure 18. 10% SDS-PAGE Analysis of Molecular Weight of Isolated Proteins. Lane 1: Unstained Protein Ladder, Broad Range (10-200kDa), Lane 2: 20 µL BaeC, Lane 3: 20 µL BaeH, Lane 4: 20 µL BaeI, Lane 5: 20 µL DfnJ, Lane \16: 20 µL DfnL, Lane 7: 20 µL DfnM, Lane 8: 20 µL holo-DfnX, Lane 9: 20 µL Sfp, Lane 10: 20 µL YhdS. Sample composition: Protein loading dye to protein (4:1)

Table 1. The Molecular Weights (kDa) of each of the Purified and Isolated Enzymes and their Yield per Liter of Culture as Determined by the Bradford Assay

Enzyme	**BaeC**	**BaeH**	**BaeI**	**DfnJ**	**DfnM**	***holo*-DfnX**	**Sfp**	**YdhS**	**DfnL**
Protein MW (kDa)	36.47	33.36	31.64	26.04	31.36	14.18	30.03	48.48	50.35
Protein Yield (mg)	3.08	1.35	4.1	2.45	1.23	3.26	5.72	2.55	13.4

The 10% SDS PAGE analysis of the isolated and purified products yielded results that were congruent with the expected molecular weight for each protein. The purified proteins and tryptic peptides were then subjected to mass spectrophotometric analysis to determine exact mass (**Table 2**) and their polyketide intermediates (**Figure 19)** as a function of their mass to charge ratio and to characterize the function of each enzyme. The spectral data acquired during these experiments provide further supporting evidence of the identity of the isolated proteins. The molecular weight of the tryptic peptides produced by the trypsin digestion match the predicted values, and the pantetheine ejection fragments obtained through the enzymology reaction experiments agree with the hypothesized biosynthetic pathway.

Table 2. Expected m/z Values of Intact Protein (Post-N-terminus Methionine Excision) and Tryptic Peptides with Respective PPT or acyl-PPT Group

Protein	**Charge State**	**Expected m/z**
Intact Proteins		
holo-DfnX	+10, +9, +8	1185.66, 1317.40, 1481.94
malonyl-S-DfnX	+10, +9, +8	1194.36, 1327.06, 1492.57
Tryptic Peptides		
holo-DfnX (V48-K96)	+4	1402.41
malonyl-S-DfnX (V48-K96)	+4	1424.41
holo-DfnJ-T2 (E167-K205)	+4	1197.09
acetyl-S-DfnJ-T2 (E167-K205)	+4	1413.41
acetoacetyl-S-DfnJ-T2 (E167-K205)	+4	1434.41
HMG-S-DfnJ-T2 (I63-K101)	+4	1233.10
glutaconyl-S-DfnJ-T2 (I63-K101)	+4	1228.59
isopentenyl-S-DfnJ-T2 (internal) (I63-K101)	+4	1217.85
isopentenyl-S-DfnJ-T2 (external) (I63-K101)	+4	1217.85

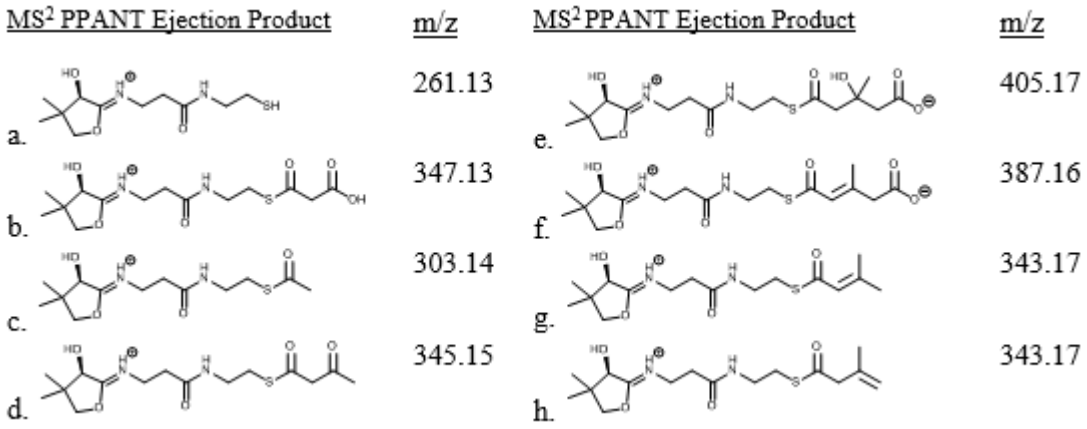


Figure 19. Expected m/z Values of Hypothesized MS/MS Pantetheine Ejection Products

Figure 19 contains a table of the hypothesized pantetheine ejection products and their respective theoretical m/z values. The pantetheine fragments included in the table are as follows. a. Pantetheine, b. Malonyl-loaded pantetheine, c. Acetyl-loaded pantetheine, d. Acetoacetyl-loaded pantetheine, e. HMG-loaded pantetheine, f. Dehydrated HMG-loaded pantetheine, g. Isopentenyl-loaded pantetheine containing internal olefin, h. Isopentenyl-loaded pantetheine containing external olefin.

II. D. Experimental

II. D. i. Confirmation of Intact *holo*-DfnX expressed from pETDuet-1-*dfnX*-*sfp* by High-performance Liquid Chromatography-electrospray Ionization (HPLC-ESI) MS

HPLC-ESI MS was performed on a ThermoFisher Scientific LTQ Orbitrap XL coupled with a Water's Acquity Ultra-High Performance Liquid Chromatography system. Purified protein samples were fractionated by reversed-phase HPLC on a Phenomenex 00B-4167-E0 Jupiter 5µm C4 300Å, LC column (50x 4.60mm) at a flow-rate of 1 mL/min. The composition of the polar solvent A1, was H_2O and 0.1% formic acid, and

the organic solvent, B1, was composed of acetonitrile and 0.1% formic acid. The HPLC purification method of the intact proteins was written as follows: 0-2 min 70% A1 and 30% B1, from 2-14.5 min 40% A1 and 60% B1, from 14.5-25min 10% A1 and 90% B1, from 25-26 min 70% A1 and 30% B1. A subsequent series of 3 wash steps in which the circulating solvents were alternated in quick 1 minute succession between 10% A1 and 90% B1 to 70% A1 and 30% B1 was used to elute any residual protein still present on the column prior to the next run to reduce any instance of carryover. Finally, minutes 30-33 were held at 90% A1 and 10% B1. The total run time for this method was 33 minutes. The eluent from the HPLC was held at a flow rate of 1 mL/min but was diffused with a T-splitter to a flow rate of 350 µL/min into the heated electrospray ionization (HESI) source. Every sample was measured in positive mode. For MS^2 fragmentation, ions were subjected to collision induced dissociation with a normalized collision energy of 30 eV. Source fragmentation, with a collision energy of 30 eV, was turned on for this experiment resulting in the ejection of PPant and Pant species and the presence of their respective peaks in the full MS scan.

II. D. ii. MS/MS Characterization of Intact *holo*-DfnX Expressed from pETDuet-1-*dfnX-sfp*

The observed parent mass of the *holo*-DfnX protein obtained from the spectra was found to be 11,848.5 Daltons, which was 131.6 Daltons lower than the anticipated mass of the protein which we had calculated to be 11,979.1 Daltons. However, we attribute this loss of 131 Daltons to the loss of an N-terminus methionine hypothetically caused by N-terminal methionine excision by the endogenous Methionine aminopeptidase found within *Escherichia coli*. The sequence data obtained for the *dfnX* gene cloned within

pETDuet-1-*dfnX*-*sfp* differed from the amino acid sequence within the *dfnX* gene used by Brittany Kiel. The newly designed pETDuet-1-*dfnX*-*sfp* construct had been found to encode a glycine amino acid residue following the methionine, whereas Brittany's gene encoded an arginine residue in the same location. Studies have found that the statistical probability of the occurrence of N-terminal methionine excision is up to 97% more likely when the amino acid sequence contains a glycine residue followed by an alanine residue in the P1' position direction following the N terminus methionine. The underlying principle dictating this increased probability of this excision is that diminutive size of these two amino acid residues allows for the P1' position to fit within the binding pocket of the methionine aminopeptidase.[28] While the molecular weight of methionine in its native form is 149 Daltons, 18 Daltons contributed by water are not present within the peptide form resulting in mass congruent with the corresponding 131 Dalton loss. The first scan of the *holo*-DfnX full MS spectra yielded a protein envelope within the range of 1100-1500 m/z consisting of the +8, +9, +10 charge states at 1185.66, 1317.40, 1481.94 m/z respectively. In subsequent scan events the three most abundant charge states of the intact protein peaks were fragmented via CID at a collision energy of 30 eV, but PPant and Pant ejection fragment ions were not observed in the MS^2. The phosphopantetheine ejection fragment ion (261.13) was detected in the full MS scan of *holo*-DfnX from the source collision induced dissociation (CID) normalized at 30 eV.

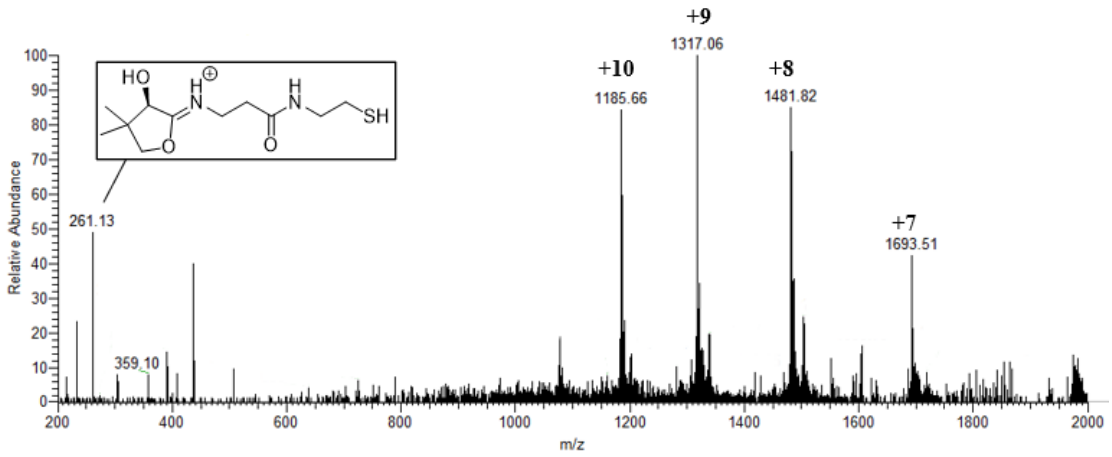


Figure 20. Formation of Intact *holo*-DfnX Protein. Full mass scan of intact *holo*-DfnX protein with most abundant charge states (+7, +8, +9, +10), the pantetheine and phosphopantetheine ejection fragments (261.13 and 359.10 m/z) were generated via a source CID fragmentation energy of 30 eV.

II. E. Experimental

II. E. i. Formation of Malonyl-S-DfnX

The reaction of malonyl-S-DfnX was prepared by the addition of 50µL of purified and dialyzed *holo*-DfnX (0.88 mg/mL), 50 µL of BaeC (0.89 mg/mL), 15 µL of malonyl-CoA (10mM), and 35 µL of reaction buffer (50mM Tris-HCL, 12.5 mM $MgCl_2$, pH=8). The reaction was allowed to proceed at room temperature overnight. The reaction was purified by HPLC before being analyzed by mass spectrometry.

II. E. ii. MS/MS Characterization of *holo*-DfnX, BaeC, and Malonyl-CoA Intact Protein (Top Down) Method

The full scan MS spectra of the purified reaction of *holo*-DfnX with BaeC and malonyl-CoA yielded a protein envelope within the range of 1100-1500 m/z consisting of the 8+, 9+, 10+ charge states at 1194.36, 1327.06, 1492.57 respectively, which corresponded to a 87 Dalton increase over the same charge states produced by *holo*-DfnX

suggesting the BaeC-dependent addition of a malonyl substituent. The m/z values of the three most abundant peaks were congruent with the intact malonyl loaded PPant containing protein envelope and were subjected to MS^2 fragmentation via CID at a normalized collision energy of 30 eV in later scan events. The MS^2 fragmentation of these peaks did not yield any of the expected Pant or PPant ejection products. However the Pant and PPant ejection fragments produced by CID (30 eV) source fragmentation indicated a noticeable decrease in abundance of the 261.13 m/z ion and the appearance of a new PPant ejection fragment ion at 347.13 within the full MS scan event indicative of the malonyl loaded pantetheine fragment. After extensive time and sample being devoted to method development with no notable progress being made, optimization of the intact method was tabled to allow the utilization of a method of involving a trypsin digest of the protein and analysis of the resultant tryptic peptides.

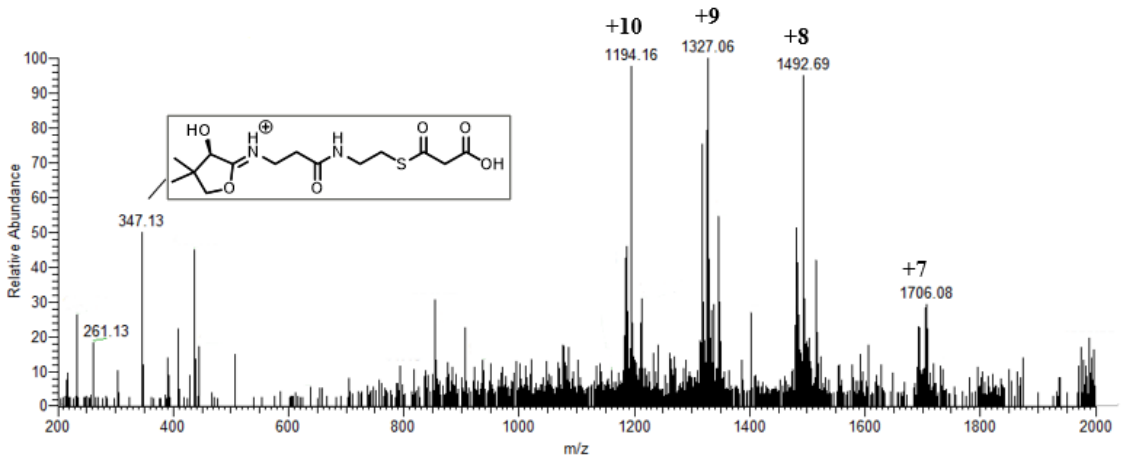


Figure 21. Formation of Intact Malonyl-S-DfnX Protein. Full mass scan of intact malonyl-S-DfnX protein with most abundant charge states (+7, +8, +9, +10), the pantetheine ejection fragment (347.13 m/z) was generated via a normalized source CID fragmentation energy of 30 eV.

II. F. Trypsin Digest Method

Trypsin catalyzes the hydrolysis of peptide bonds to a high degree of specificity only at sites on the protein containing lysine or arginine amino acid residues. A generic trypsin digest protocol from the manufacturers of Trypsin Gold was used to systematically digest the whole purified and dialyzed proteins into tryptic peptides prior to analysis. The rationale behind trypsin digestion is that the decreased size of the tryptic peptides facilitated quicker transition into the gas phase and a higher degree of analytical reproducibility as compared with ionization of the intact protein.

All reactions were quenched with 8 M Urea, 50mM Tris-HCl (pH=8) and incubated at 37°C for 1 hour. Post incubation an addition of 50 µL of 50 mM Tris-HCl, 1mM $CaCl_2$ (pH=8) in order to decrease the concentration of urea to less than 1 M in the solution prior to the addition of trypsin which was added at a 1:20 (w/w) protein

ratio. This reaction was then incubated at 37°C for a minimum of 3 hours prior to analysis of the tryptic peptides.

II. F. i. Analysis of Trypsin Digested Proteins Ultra High-Performance Liquid Chromatography-electrospray Ionization (UPLC-ESI) MS/MS

UPLC-ESI MS/MS was performed on a ThermoFisher Scientific Q Exactive Plus coupled with a Water's Acquity Ultra-High Performance Liquid Chromatography system. Purified protein samples were subjected to a trypsin digest outlined above and the tryptic peptides were fractionated by reversed-phase UPLC on a Acquity UPLC Peptide BEH C18 300Å 1.7µm (1x100mm) Column. The composition of polar solvent, A1, was H_2O and 0.1% formic acid and the organic solvent, B1 consisted of acetonitrile and 0.1% formic acid in this reversed phase UPLC method. The UPLC purification method of the tryptic peptides was written as such: 0-10 min 90% A1 and 10% B1, 10-12.5 min 70% A1 and 30% B1, 12.5-25 min 40% A1 and 60% B1, 25-30 min 10% A1 and 90% B1. A subsequent series of 3 wash steps in which the circulating solvents were alternated in quick succession between 10% A1 and 90% B1 to 70% A1 and 30% B1 was used to elute any residual protein still present on the column prior to the next run to reduce any instance of carryover. The total run time of this method was 43 minutes.

II. F. ii. MS/MS Analysis of pETDuet-1-*dfnX-sfp* Tryptic Peptides

The protein expressed by the pETDuet-1-*dfnX-sfp* produced a tryptic peptide peak with a +4 charge at 1402.41 m/z which corresponds to the hypothesized phosphopantetheine containing tryptic peptide of *holo*-DfnX. MS^2 analysis was performed by fragmenting the tryptic peptide at 1402.41 m/z, which resulted in the ejection of the expected pantetheine ejection product at 261.13 m/z.

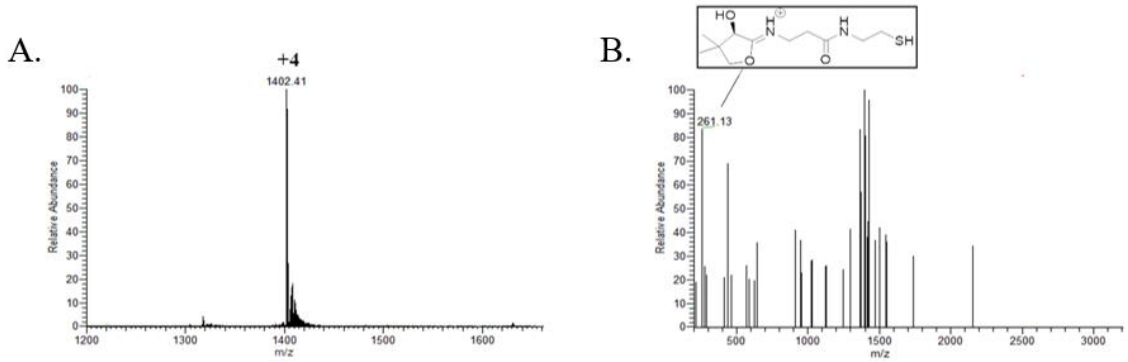


Figure 22. Full Mass Scan of *holo*-DfnX Tryptic Peptide with Most Abundant Charge State (+4). The pantetheine ejection fragment (261.13 m/z) were generated via CID fragmentation of the tryptic peptide peak at 1402.41 m/z.

II. F. iii. Formation of Malonyl-S-DfnX

The reaction of malonyl-S-DfnX was prepared by the addition of 50µL of purified and dialyzed *holo*-DfnX (0.88 mg/mL), 50 µL of BaeC (0.89 mg/mL), 15 µL of malonyl-CoA (10mM), and 35 µL of reaction buffer (50mM Tris-HCL, 12.5 mM $MgCl_2$, pH=8). The reaction was allowed to proceed at room temperature overnight. The reaction was subjected to trypsin digestion and the tryptic fragments were purified by UPLC before they were analyzed by mass spectrometry.

II. F. iv. MS/MS Analysis of Malonyl-S-DfnX Tryptic Peptides

The reaction of the *holo*-DfnX protein with BaeC yielded a new tryptic peptide peak in the +4 charge state at 1424.41 m/z, which corresponds to a BaeC dependent 88 Dalton increase over the *holo*-DfnX tryptic peptide peak in the +4 charge state at 1402.41 m/z. We hypothesize that this 88 Dalton increase is a result of the addition of a malonyl acyl unit, catalyzed by the malonyl acyl transferase encoded by *baeC* MS^2 fragmentation of the +4 charge tryptic peptide at 1424.41 m/z yielded the malonylated pantetheine

ejection fragment at 347.13 m/z, further supporting the hypothesized addition of the malonyl acyl unit to the free thiol of the phosphopantetheine arm of *holo*-DfnX.

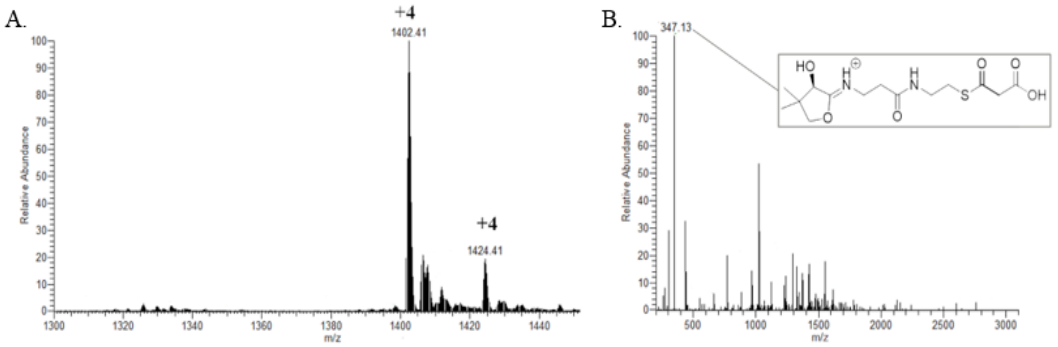


Figure 23. Full Mass Scan of Malonyl-S-DfnX Tryptic Peptide with Most Abundant Charge State (+4). The pantetheine ejection fragment (347.13 m/z) were generated via CID fragmentation of the tryptic peptide peak at 1424.41 m/z.

II. F. v. Formation of Ac-S-DfnX

The formation of Ac-S-DfnX was prepared by the addition of 50 µL of purified and dialyzed *holo*-DfnX (0.88 mg/mL), 50 µL of BaeC (0.88 mg/mL), 15 µL of malonyl-CoA (10mM), and 35 µL of reaction buffer (50mM Tris-HCL, 12.5 mM $MgCl_2$, pH=8). The initial reaction was allowed to proceed for 2.5 hours at room temperature, followed by the addition of 50 µL of YhdS (0.63 mg/mL) to the reaction which was subsequently incubated for an additional hour. An additional reaction involving all of the aforementioned reactants at their respective concentrations and volumes were mixed simultaneously rather than the subsequent addition described above. The rationale behind the simultaneous addition involved a desire to test if yields would hypothetically increase from the promiscuous Sfp PPTase loading both malonyl-CoA starting material and

acetyl-CoA produced if malonyl-CoA itself were a non-native substrate of YhdS and a decarboxylation of malonyl-CoA starting material was catalyzed prior to being loaded onto DfnJ. All reactions were quenched with 20 µL of 8M urea then digested with trypsin. The tryptic peptides were purified by HPLC before they were analyzed by mass spectrometry.

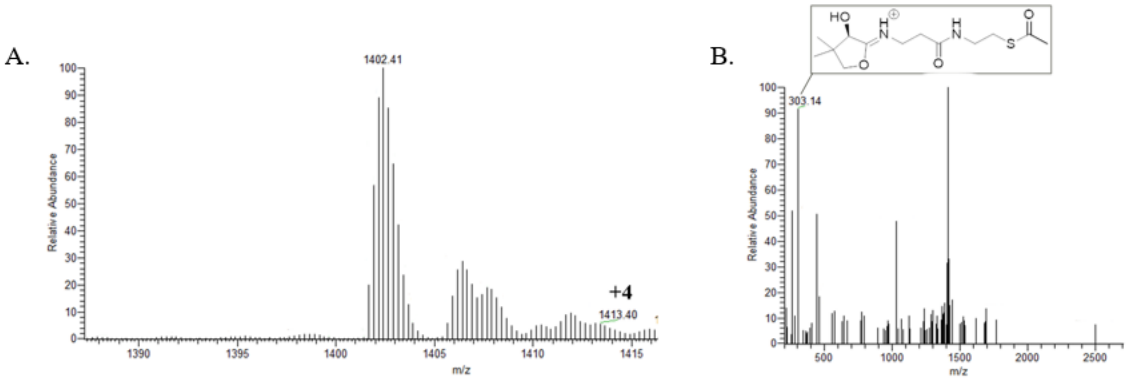


Figure 24. Full Mass Scan of Acetyl-S-DfnX Tryptic Peptide with Most Abundant Charge State (+4). The pantetheine ejection fragment (303.14 m/z) were generated via CID fragmentation of the tryptic peptide peak at 1413.41 m/z.

II. F. vi. Characterization of MS/MS of Ac-S-DfnX

The reaction of YhdS with malonyl-S-DfnX yielded a new +4 charged peak at 1413.41 m/z, is indicative of a 44 Dalton decrease corresponding to a loss of CO_2. We hypothesized that enzyme encoded by *yhdS* catalyzed the decarboxylation of the malonyl substituent on malonyl-S-DfnX to yield acetyl-S-DfnX. Further evidenced by the MS^2 fragmentation of the +4 charge state tryptic peptide at 1413.41 m/z which exhibited the ejection of the acetylated pantetheine fragment (303.14 m/z).

II. G. Conclusions

The newly designed *in vitro* construct pETDuet-1-*dfnX*-*sfp* was found to successfully coexpress *dfnX* and *sfp*. The dual-expression vector yielded the hypothesized polyhistidine affinity-tagged *holo*-DfnX protein product which was isolated by means of Ni^{2+}-NTA affinity chromatography. The presence of the 4'-phosphopantethienyl transferase product encoded by *sfp* was neither observed in the 10% SDS-PAGE analysis, nor subsequent ESI-MS/MS experiments.

The malonyl-transacylase activity of the enzyme encoded by *baeC* was observed via the reaction of holo-DfnX, BaeC, and Malonyl-CoA, without the presence of *sfp.* This was evidenced by the 87 Dalton increase, corresponding to the addition of a malonyl substituent, in both the intact protein and tryptic peptide masses, and further confirmed by the presence of the hypothesized 347.13 m/z Ppant ejection species.

The decarboxylase activity of the enzyme encoded by *YhdS* was also observed utilizing the malonyl-S-DfnX, produced by the reaction of *holo*-DfnX, malonyl-CoA, and BaeC. This was exhibited by the 44 Dalton loss, corresponding to the loss of CO_2, in both the intact protein and tryptic peptide masses in the ESI-MS spectra. This was further supported by the presence of the MS^2 Ppant ejection species at 303.17 m/z, indicating the presence of an acetyl-containing pantetheine ejection product.

CHAPTER III

CHARACTERIZATION OF *DFNL*

III. A. Overview and Objectives

We proposed that the HMGS-like synthase encoded by *dfnL* was responsible for catalyzing the adol addition of Ac-S-DfnX and Acac-DfnJ to form HMG-S-DfnJ-T2 (Figure 25**)**. While Brittany Kiel, a former PhD in the Reddick lab was able to produce a few promising spectra which evidenced the successful formation of the HMG-S-DfnJ-T2 by this proposed pathway. However, based on the sequence of experiments run on the particular day these spectra were produced, we are hesitant to rule out contamination or residual remnants of reaction products that produce similar spectral peaks left on the column from previous experiments, Dr. Kiel was unable to reconstitute this reaction due to a shortage of commercial acetoacetyl-CoA. Past lots of commercial acetoacetyl-CoA received by our lab were found to have apparent quality control issues. This was first evidenced through the formation of only the acetyl-DfnJ-T2 within the spectra acquired from these reactions and later confirmed by direct MS/MS analysis of the defunct acetoacetyl-CoA lot. The order was put in for more acetoacetyl-CoA, however issues with commercial supply led to this coenzyme being on back order for several months without any clarification from the supplier, which ultimately stalled the characterization of this biosynthetic step as we were unable to form the necessary polyketide reaction intermediate Acac-DfnJ-T2.

To combat this limitation an orthogonal approach was pursued through the analysis of *dfnL*'s ability to catalyze the reverse reaction yielding Ac-S-DfnX and Acac-S-DfnJ-T2 from HMG-S-DfnJ-T2 (Figure 26**)**.

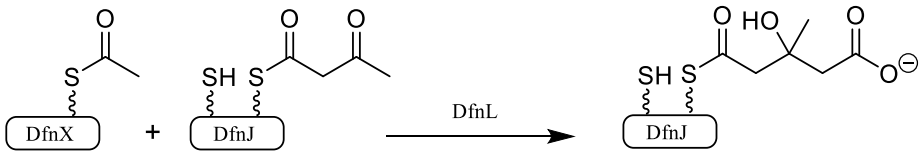


Figure 25. Scheme of the Reaction Catalyzed by DfnL. We hypothesize that the HMGS-like synthase DfnL catalyzes the reaction of Ac-S-DfnX and Acac-S-DfnJ-T2 to form HMG-S-DfnJ-T2.

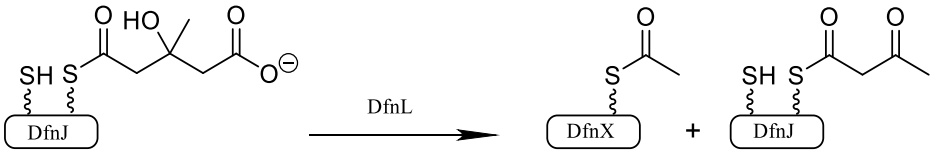


Figure 26. Scheme of the Reverse Reaction Catalyzed by DfnL. We hypothesize that DfnL is capable of catalyzing the reverse reaction of HMG-S-DfnJ-T2 to form Ac-S-DfnX and Acac-S-DfnJ.

III. B. Experimental

III. B. i. Formation of Acac-S-DfnJ-T2

The formation of Acac-S-DfnJ-T2 has yet to be accomplished at this time due to the aforementioned supply issues of Acac-CoA. Preliminary MS/MS data of the analysis of Acac-S-DfnJ-T2 tryptic peptides, suggested there was a high abundance of the acetylated version of holo-DfnJ-T2, as evidenced by the acetylated pantetheine ejection species (303.14 m/z) suggesting either the loss of an acetyl group prior to analysis or defunct Acac-CoA starting material. In conjunction with buffer optimization, an analysis

of the intact form of Acac-S-DfnJ-T2 was pursued as a means to circumvent trypsin digestion for this reaction. The intact (Top Down) methods for the analysis of Acac-DfnJ were unsuccessful as the mass spectra did not yield any peaks congruent with the presence of DfnJ or Acac-S-DfnJ-T2.

III. B. ii. Formation of HMG-S- DfnJ-T2

The preliminary data acquired for the formation of HMG-S-DfnJ-T2 has only been achieved artificially via the loading of HMG-CoA onto the apo-DfnJ-T2 substrate by the 4'-phosphopantetheinyl transferase encoded by *sfp*. This reaction was prepared by the addition of 50 µL of *apo*-DfnJ-T2 (0.79 mg/mL), 20 µL of Sfp (1.8 mg/mL), and 10 µL of HMG-CoA (10mM) and 20 µL of reaction buffer (50mM Tris-HCL, 12.5 mM $MgCl_2$, pH=8). The reaction was incubated at room temperature overnight, the reaction was quenched with 20 µL of 8M urea, a trypsin digest was performed on the newly formed product, and the tryptic peptides were purified by HPLC before where were analyzed by mass spectrometry.

Future work is scheduled for the production of HMG-S-DfnJ-T2 by means of an enzymatic route more in line with the proposed difficidin biosynthetic pathway involving the reaction of Ac-S-DfnX and Acac-S-DfnJ-T2 catalyzed by DfnL to yield the HMG-S-DfnJ-T2 product.

III. B. iii. MS/MS Characterization of HMG-S-DfnJ-T2

The addition of the HMG-CoA containing phosphopantetheine arm to *apo*-DfnJ-T2 by means of Sfp yielded a +4 charged peak at 1233.10 m/z. MS^2 fragmentation of the +4 charge state tryptic peptide at 1233.10 m/z ejected the pantetheine fragment at 405.17

m/z consistent with expected mass of the loading of an HMG-CoA phosphopantetheine prosthetic group.

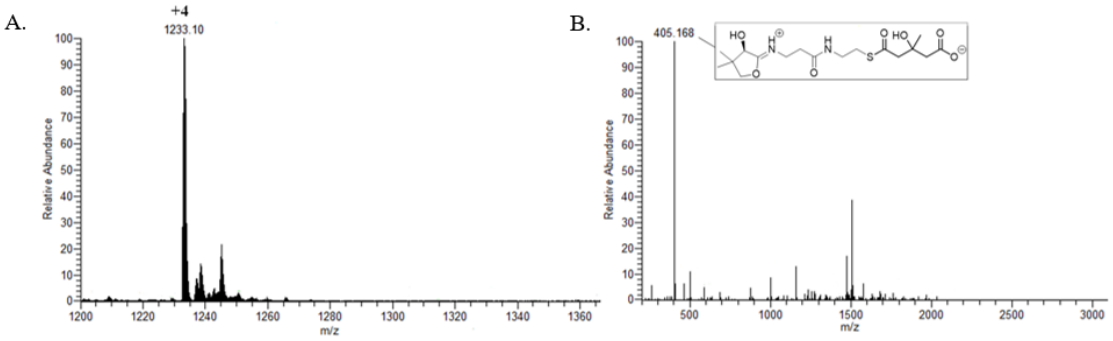


Figure 27. Full Mass Scan of HMG-S-DfnX Tryptic Peptide with Most Abundant Charge State (+4). (A) The pantetheine ejection fragment (405.17 m/z) were generated via CID fragmentation of the tryptic peptide peak at 1233.10 m/z (B).

III. B. iv. Formation of Acetyl-S-DfnX and Acac-S-DfnJ-T2 from HMG-S-DfnJ-T2 and DfnL

Due to issues regarding the accessibility of commercial acetoacetyl-CoA an orthogonal approach was used to characterize the function of the enzyme encoded by *dfnL*. This approach capitalizes on the equilibrium and reversibility of enzymatic reactions, by using the HMG-synthase encoded by *dfnL* to convert HMG-S-DfnJ-T2 to Acetyl-S-DfnX and Acac-S-DfnJ-T2. 3-hydroxy-3-methylglutaryl-CoA synthase intermediate complexes have been previously observed which provided a literature precedent to test the reversibility of this enzymatic reaction.[29] Though the polyketide intermediates of the reverse reaction were observed by means of trapping and crystalizing the substrate-enzyme complex and analyzed via X-ray diffraction. To our knowledge these intermediates have not been observed via mass spectroscopic techniques.

HMG-S-DfnJ-T2 was formed artificially by the loading the prosthetic group of HMG-CoA onto the apo-DfnJ-T2 substrate by the 4'-phosphopantetheinyl transferase encoded by *sfp*. This reaction was prepared by the addition of 50 µL of *apo*-DfnJ-T2 (0.79 mg/mL), 20 µL of Sfp (1.8 mg/mL), and 10 µL of HMG-CoA (10mM) and 20 µL of reaction buffer (50mM Tris-HCL, 12.5 mM $MgCl_2$, pH=8). This reaction was incubated for 1 hour at room temperature. Then 50 µL of *holo*-DfnX produced from the pETDuet-1-*dfnX*-*sfp in vitro* construct described above, 50 µL of DfnL, and an additional 15 µL of reaction buffer (50mM Tris-HCL, 12.5 mM $MgCl_2$, pH=8) were added to the HMG-S-DfnJ-T2 reaction and was incubated for 144 hours (six days) at room temperature. The reaction was quenched with 20 µL of 8M Urea, a trypsin digestion was performed on the newly formed products, and the tryptic peptides were purified by UPLC before analysis via tandem mass spectrometry.

The use of the *holo*-DfnX produced by pETDuet-1-*dfnX*-*sfp* is an important consideration regarding this reverse experiment as the reaction more susceptible to problems if using the apo DfnX with Sfp. The promiscuity of 4'-phosphopantetheinyl transferase encoded by *sfp* is highly likely to indiscriminately load the HMG-CoA phosphopantetheine or regular phosphopantetheine prosthetic groups on either of the *apo*-acyl carrier proteins leading to an abundance of other side products in this enzymatic reaction.

III. B. v. MS/MS Characterization of the Reverse Reaction Catalyzed by DfnL

The reverse reaction of HMG-DfnJ-T2 and DfnL was hypothesized to yield a both Acac-S-DfnJ-T2 and Ac-S-DfnX moieties. The expected tryptic peptides corresponding

to Acac-S-DfnJ-T2 were not observed within the spectra, however a tryptic peptide peak of very low abundance with a +4 charge was observed at 1413.41 corresponding to Ac-S-DfnX. When this tryptic peptide peak was subjected to MS^2fragmentation it yielded a Ppant ejection product at 303.22 m/z. Given the robustness of the Ppant ejection assay and its accuracy the peaks present do not provide sufficient evidence that either the Acac-S-DfnJ-T2 or Ac-S-DfnX products were present in these spectra.

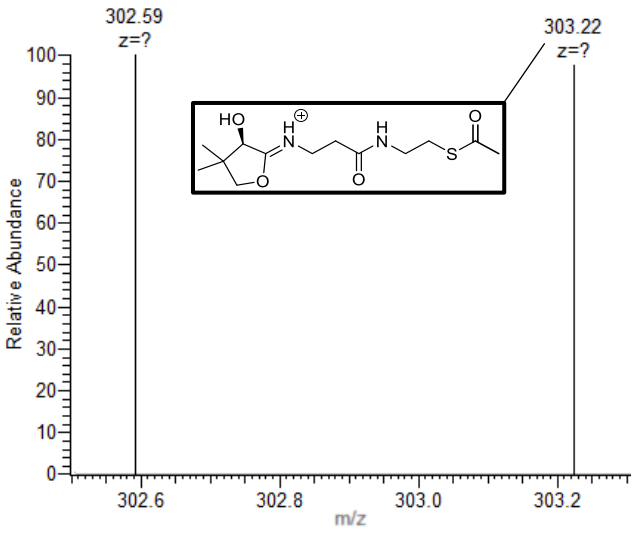


Figure 28. MS^2 Fragmentation the +4 Charged Tryptic Peptide Peak at 1413.41 m/z Yielding a Pantetheine Ejection Product of 303.22 m/z Suggesting the Presence of the Ac-S-DfnX Polyketide Intermediate.

III. C. Conclusions

Giving the benefit of the doubt to the quality control department of the supplier we pursued other hypotheses as to why we were observing acetyl-CoA rather than acetoacetyl-CoA. To control for the potential degradation of the acetoacetyl-CoA buffer conditions were altered to remove $CaCl_2$ from the trypsin reaction buffer, as its role was

ambiguous and not commonly found in trypsin protocols. An intact method for characterization of *dfnL* activity was pursued as a means to circumvent trypsin digestion altogether, to control for the possibility that the trypsin digestion, itself, may be inhibiting the formation of the Acac-S-DfnJ-T2 product.

Multiple intact proteins methods were tested including a method utilizing QToF instrumentation, along with methods similar to the intact (Top Down) methods performed on Orbitrap and QExactive mass spectrometers used for the characterizations of *baeC* and *yhdS* activities. However, peaks were not observed for DfnJ within any of the mass spectra that contained intact DfnJ. Our initial hypotheses were that either the protein was not eluting off of the column either due to precipitation or general column incompatibility, or that the protein was unable to be ionized effectively by the ESI-source within our facility. After performing several iterations of direct injection methods, which bypass liquid chromatography altogether, no peaks were observed for DfnJ. Dr. Brittany Kiel, a former member of the Reddick lab encountered similar difficulties with the ionization of the DfnJ protein. Future work would involve the optimization of a method utilizing nano-ESI to increase the ionization efficiency of this large (26 kDa) protein.[30]

CHAPTER IV

DIFFERENTIATION OF BAEI AND DFNM DERIVED ISOPENTENYL AND EXOMETHEYLENE B-BRANCH INTERMEDIATES

IV. A. Overview and Objectives

In order to complete the objective outlined in the third specific aim of the work presented in this thesis involving the characterization of the enzymes directing the location of the double bond within the isopentenyl polyketide intermediate formed will be dependent on differentiating the MS^3 fragmentation patterns of the reactions catalyzed by *baeI* and those catalyzed by *dfnM*. We hypothesize that enzyme encoded by *baeH* catalyzes a dehydration of the HMG-S-DfnJ-T2 intermediate to yield glutaconyl-S-DfnJ-T2. A subsequent decarboxylation is catalyzed by the decarboxylase *dfnM* to yield an external olefinic pentenyl product, whereas the decarboxylation catalyzed by *baeI* is hypothesized to yield the internal olefinic isopentenyl product (Figure 28). Preliminary data has shown that both enzymes form an isopentenyl intermediate on DfnJ's double thiolation domain of the same exact mass, the MS^3 fragmentation patterns of the phosphopantetheine ejection species may provide valuable structural insights that could differentiate the two intermediates. Due to the inability to detect PPant ejection products in targeted MS^2 fragmentation of other reaction intermediates, pursuit of developing

intact HPLC ESI- MS^3 method on the LTQ Orbitrap was abandoned. However due to the limitation of not being able to perform MS^3 experiments on the ThermoFisher™ Q Exactive Plus, led to the development of a method utilizing source fragmentation to release the phosphopantetheine ejection species and targeted MS^2 fragmentation to study the individual fragmentation patterns of each pentenyl moiety (Figure 32).

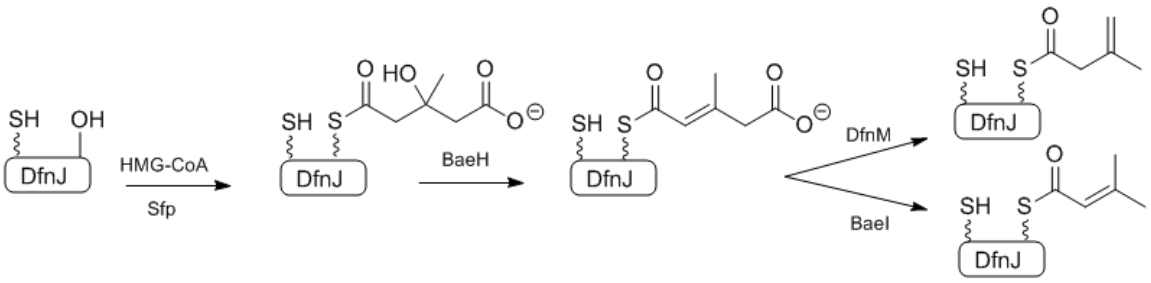


Figure 29. Scheme for the Incorporation of the Two Alternative β-branches Formed by the Decarboxylases DfnM and BaeI from the *dif* and the *bae* Biosynthetic Gene Clusters in *B. velezensis* FZB42. The reaction is denoted as taking place at only a single site on DfnJ for clarity. The proposed "difficidin-like" external olefin product is catalyzed by DfnM whereas a "bacillaene-like" internal olefin product is hypothesized to be catalyzed by BaeI.

These fragmentation patterns were then compared against a S-acyl N-acetylcysteamine thioester (SNAC) specifically synthesized to mimic the structure of the internal olefinic isopentenyl ejection product. This project was done in collaboration with the Peterson research group at the University of North Carolina at Greensboro; Tyler Greenstein was responsible for the synthesis of the internal olefinic isopentenyl SNAC standard (Figure 32 C.).

IV. B. Experimental

IV. B. i. Dehydration of HMG-S-DfnJ-T2 by BaeH to form Glutaconyl-S-DfnJ-T2

The formation of the dehydrated HMG-S-DfnJ-T2 product, Glutaconyl-S-DfnJ-T2 was formed by utilizing the same reaction conditions as stated above for the formation of the HMG-S-DfnJ-T2 intermediate with the addition of 50 µL of BaeH. This reaction was allowed to incubate at room temperature overnight, the reaction was quenched with 20 µL of 8M urea and was subjected to trypsin digestion. The tryptic peptides were purified by UPLC prior to analysis on the mass spectrometer.

IV. B. ii. MS/MS Characterization of Dehydration of HMG-S-DfnJ-T2 by BaeH Tryptic Peptides

The dehydration catalyzed by reaction with BaeH yielded the new tryptic fragment peak 1228.59. MS^2 fragmentation of the +4 charge state tryptic peptide at 1228.59 m/z yielded a pantetheine ejection fragment at 387.16 m/z. The mass of the new PPant ejection fragment ion corresponded to a loss of 18 m/z compared to the HMG pantetheine ejection fragment.

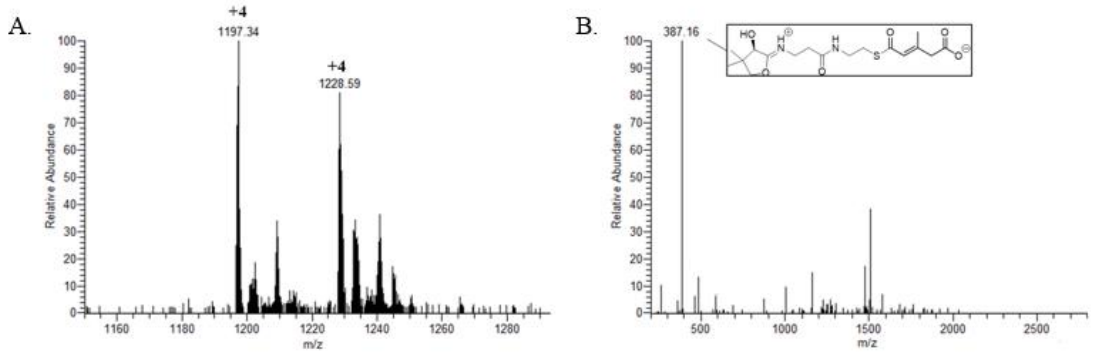


Figure 30. Full Mass Scan of Glutaconyl-S-DfnJ-T2 Tryptic Peptide with Most Abundant Charge State (+4). (A) The pantetheine ejection fragment (387.96 m/z) were generated via CID fragmentation of the tryptic peptide peak at 1228.59 m/z (B).

IV. C. Experimental

IV. C. i. Dehydration and Decarboxylation of HMG-S-DfnJ-T2 by BaeH and DfnM/ BaeI

This reaction was prepared by the addition of 50 µL of *apo*-DfnJ-T2 (0.79 mg/mL), 20 µL of Sfp (1.8 mg/mL), and 10 µL of HMG-CoA (10mM), 35 µL of reaction buffer (50mM Tris-HCL, 12.5 mM $MgCl_2$, pH=8), and 50 µL of BaeH (0.59 mg/mL). The reaction was incubated for 2.5 hours at room temperature and 50 µL of BaeI or DfnM (1.6 mg/mL /0.49 mg/mL) were added to the respective reaction and allowed to incubate for an additional 2 hours before being quenched with 20 µL of 8M urea. A trypsin digest was performed on the newly formed product, and the tryptic peptides were purified by HPLC before where were analyzed by mass spectrometry.

IV. C. ii. MS/MS Characterization of Dehydration and Decarboxylation of HMG-S-DfnJ-T2 by BaeH and DfnM/ BaeI

The decarboxylation catalyzed by both DfnM and BaeI in their respective reactions yielded a similar new tryptic fragment peak at 1217.85 m/z. MS^2 fragmentation of the +4 charge state tryptic peptide at 1217.85 m/z yielded a pantetheine ejection fragment at 343.17 m/z in both cases. The mass of the new tryptic peptide at 1217.85 m/z indicated a loss of 44 m/z corresponding to the loss of CO_2, supporting the catalysis of a decarboxylation by both BaeI and DfnM. PPant ejection fragment ions produced in both cases also mirrored this decarboxylation as compared with the dehydrated HMG pantetheine ejection fragment. No differences were indicated in the exact masses of the tryptic peptides produced between these reactions nor were there any distinguishing features in the MS^2 fragmentation patterns of the internal and external olefinic pentenyl derivatives formed by DfnM or BaeI.

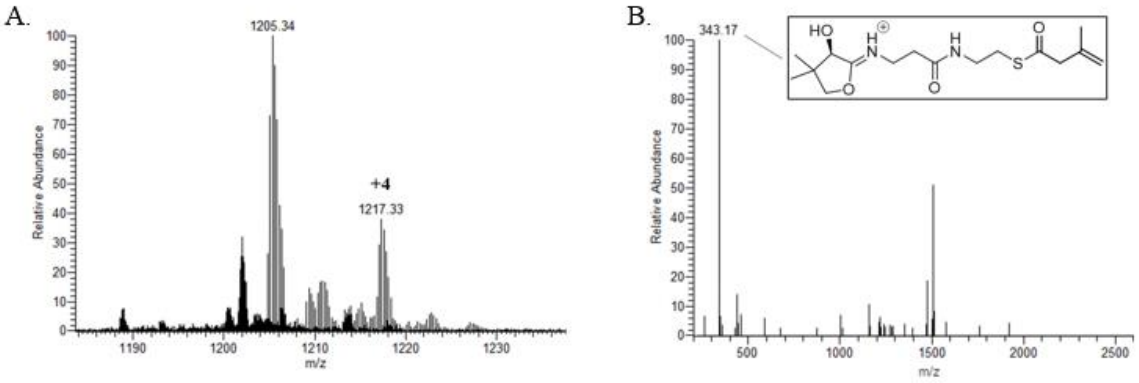


Figure 31. Full Mass Scan of the External Olefinic Pentenyl-S-DfnJ-T2 Tryptic Peptide Generated from DfnM with Most Abundant Charge State (+4). (A) The pantetheine ejection fragment (343.17 m/z) were generated via CID fragmentation of the tryptic peptide peak at 1217.33 m/z (B).

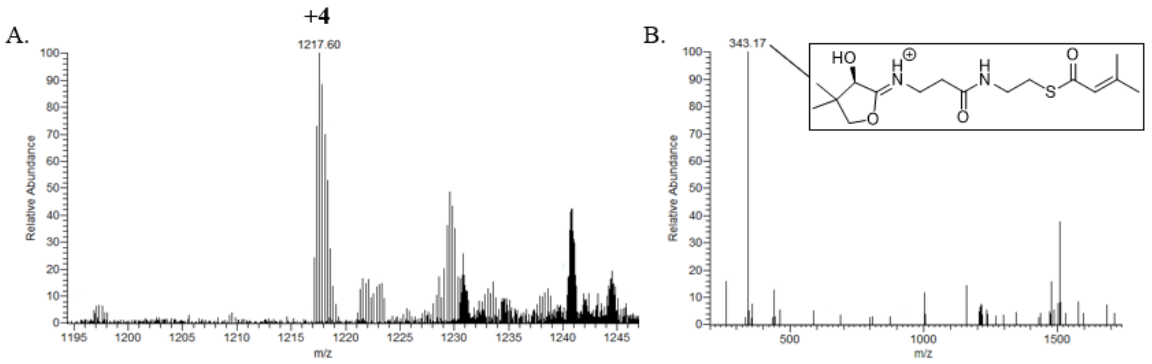


Figure 32. Full Mass Scan of Internal Olefinic Isopentenyl-S-DfnJ-T2 Tryptic Peptide with Most Abundant Charge State (+4) (A), the pantetheine ejection fragment (343.17 m/z) were generated via CID fragmentation of the tryptic peptide peak at 1217.60 m/z (B).

IV. C. iii. Tandem MS Fragmentation Analysis of Internal and External Pentenyl Moieties Compared to a "SNAC" Standard

Provided the two moieties formed from DfnM and BaeI could not be differentiated by mass alone, an analysis of the fragmentation patterns of each species was performed. We hypothesized that the location of the double bond, the internal versus the external olefin, within these two moieties would produce unique fragmentation patterns by which they could be differentiated. These fragmentation patterns could then be compared against a synthetic standard containing the internal olefin to confirm if the decarboxylases, *baeI or dfnM*, dictate the location of the double bond (Figure 32).

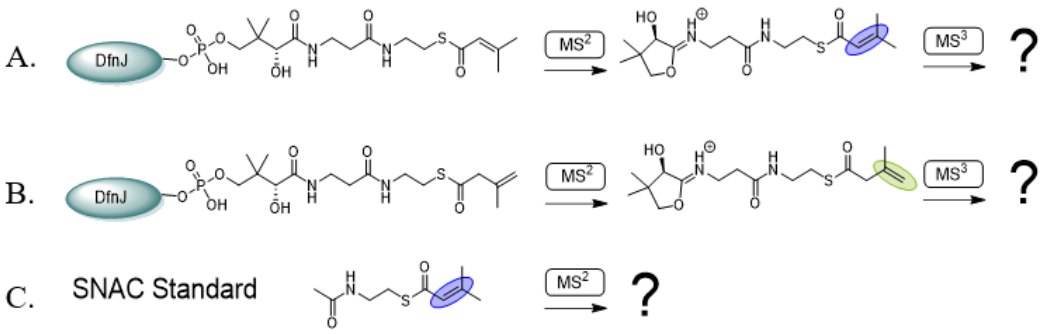


Figure 33. Scheme of Tandem MS Fragmentation Analysis of Isopentenyl and Exomethylene β-branch Containing Pentenyl Products Compared to a SNAC Standard.

The isopentenyl and external olefinic β-branch containing intermediates were formed using the procedure outlined above (**IV. B. i.)** .Source fragmentation with a normalized collision energy of 30 eV was applied to the tryptic peptides from both the *baeI* and *dfnM* containing reactions to produce their respective 343.17 m/z pantetheine ejection products. A targeted MS^2 fragmentation analysis was performed by mass filtering, then exclusively fragmenting the ions at 343.1668 m/z.

The S-acyl N-acetylcysteamine thioester (SNAC) standard was then analyzed on the ThermoFisher QExactive, utilizing similar LCMS instrument methods, and was found to produce a $[M+H]^+$ peak at 202.08864 m/z (Figure 33). This ion was subjected to a normalized collision energy of 30 eV. The fragmentation patterns produced in each spectra were then compared to elucidate any distinguishing features between the two β-branch containing pantetheine ejection species.

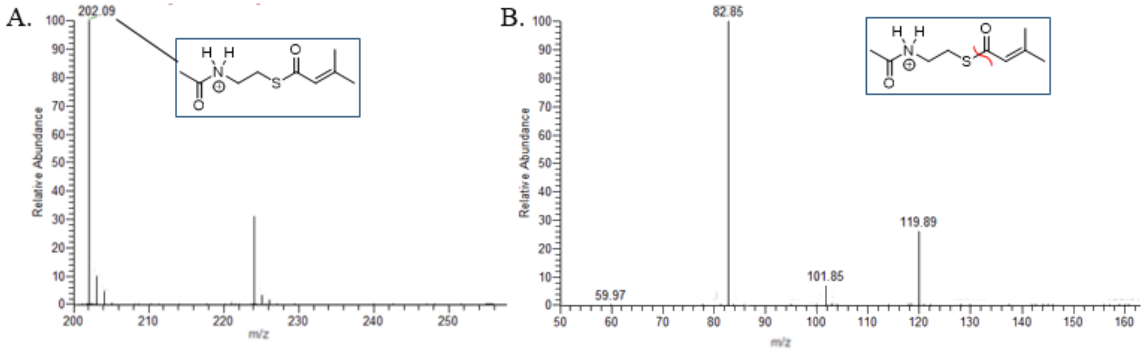


Figure 34. MS/MS Spectra of the SNAC Standard. (A.) Full MS scan reveals a $[M+H]^+$ peak at 202.08 m/z. (B.) MS^2 spectra exhibits peaks at 82.85 and 110.89 m/z.

IV. D. Conclusions

The spectra obtained for the SNAC derivative was congruent with the expected mass based on its molecular formula. The MS^2 fragmentation spectra obtained for this compound exhibits peaks consistent with cleavage at the thioester bond, yielding the fragment peaks at 82.85 and 110.89. Unfortunately, fragmentation at this thioester position would not yield peaks unique to the isopentenyl moiety contained within the molecule as both the *baeI* and *dfnM* catalyzed reactions both contain this feature. Preliminary MS spectral data obtained from the enzymology reactions exhibit similar peaks characteristic of cleavage at the thioester position, which yielded peaks at 83.0497 and 261.1267 m/z consistent with a protonated thiol fragment. Further experimentation and optimization of these methods is yet required to make a definitive determination as to the location of the double bond within these pantetheine ejection products.

One of the main limitations of these experiments was the inability to obtain the required sensitivity of low abundant fragments on an instrument capable of MS^N

fragmentation. The ThermoFisher QExactive mass spectrometer used in these experiments had the appropriate sensitivity, but was unable to perform MS^3 analysis. To get around this limitation, source fragmentation was utilized to simulate the MS^2 fragmentation of the peptide peaks, albeit without the selectivity of targeting specific peptide peaks. While there was plenty of evidence of pantetheine ejection ions at 343.17 m/z, these ions were not obtained during the expected retention time on the chromatogram where the pantetheine containing peptides elute off of the column. Despite the lack of transparency as to which peptides these pantetheine ejection products originated from, there is continuity in both the exact mass and the fragmentation patterns which provide strong evidence to the identity of the β-branch containing polyketide intermediates. Future work includes the development of a nuclear magnetic resonance (NMR) based analysis to identify splitting patterns of the hydrogens associated with the nearby carbons to identify and further differentiate the decarboxylase activities that product the isopentenyl and exomethylene β-branch containing polyketide intermediates.

BIBLIOGRAPHY

1. Zimmerman, SB. *et al. DIFFICIDIN AND OXYDIFFICIDIN: NOVEL BROAD SPECTRUM ANTIBACTERIAL ANTIBIOTICS PRODUCED. J Antibiot (Tokyo).* Dec;40(12):1677-81 (1987)

2. Pieter C. Dorrestein, *et al.* Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *Biochemistry* 45, 42, 12756-12766 (2006).

3. Calderone, C. T. Isoprenoid-like alkylations in polyketide biosynthesis. *Nat. Prod. Rep.* **25**, 845 (2008).

4. Calderone, C. T., Kowtoniuk, W. E., Kelleher, N. L., Walsh, C. T. & Dorrestein, P. C. Convergence of Isoprene and Polyketide Biosynthetic Machinery: Isoprenyl-S-Carrier Proteins in the pksX Pathway of *Bacillus subtilis. Proc. Natl. Acad. Sci. U. S. A.* **103**, 8977–8982 (2006).

5. Dunlap, C. A., Kim, S.-J., Kwon, S.-W., Rooney, A. P. & Christopher Dunlap, C. A. Bacillus velezensis is not a later heterotypic synonym of Bacillus amyloliquefaciens; Bacillus methylotrophicus, Bacillus amyloliquefaciens subsp. plantarum and ‘Bacillus oryzicola’ are later heterotypic synonyms of Bacillus velezensis based on phylogenom. *Int. J. Syst. Evol. Microbiol.* **66**, 1212–1217 (2016).

6. Chen, X. H. *et al.* Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium Bacillus amyloliquefaciens FZB42. *Nat. Biotechnol. Vol.* **25**, (2007).

7. Maloney, F. P., Gerwick, L., Gerwick, W. H., Sherman, D. H. & Smith, J. L. Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10316–21 (2016).

8. Wagner, D. T. *et al.* Structural and Functional Trends in Dehydrating Bimodules from trans-Acyltransferase Polyketide Synthases. *Structure* **25**, 1045-1055.e2 (2017).

9. Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996 (2010).

10. Mootz, H. D., Finking, R. & Marahiel, M. A. 4'-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis. *J. Biol. Chem.* **276**, 37289–98 (2001).

11. Quadri, L. E. N. *et al.* Characterization of Sfp, a Bacillus subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Biochemistry*, 37, 6, 1585-1595 (1998).

12. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231–316 (2016).

13. Kosol, S., Jenner, M., Lewandowski, R. & Challis, G. L. Protein-protein interactions in trans-AT polyketide synthases. *Nat. Prod. Rep.* **35**, 1097-1109 (2018).

14. Chen, X.-H. *et al.* Structural and Functional Characterization of Three Polyketide Synthase Gene Clusters in Bacillus amyloliquefaciens FZB 42. *J. Bacteriol.* **188**, 4024–4036 (2006).

15. Kiel, B. Investigating the Biosynthesis of the β-Branch Found in the Polyketide Difficidin Isolated from Bacillus amyloliquefaciens. (University of North Carolina at Greensboro, 2019).

16. Maloney, F. P. *et al.* Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate *Proc Natl Acad Sci U S A.* 37, 10316-21. (2016)

17. Dorrestein, P. C. *et al.* Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *Biochemistry*, 45, 42, 12756-12766 (2006).

18. Haines, A. S. *et al.* A conserved motif flags acyl carrier proteins for β-branching in polyketide synthesis. *Nat. Chem. Biol.* **9**, 685–692 (2013).

19. Piel, J. *et al.* Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16222–16227 (2004).

20. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14002–14007 (2002).

21. Tucker, A. C. & Escalante-Semerena, J. C. Acetoacetyl-CoA synthetase activity is controlled by a protein acetyltransferase with unique domain organization in *Streptomyces lividans*. *Mol. Microbiol.* **87**, 152–167 (2013).

22. Butcher, R. A. *et al. The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis*. Proc. Natl. Acad. Sci. *U. S. A.* 104 (5) 1506-1509 (2007).

23. Lodish H, Berk A, Zipursky SL, et al. DNA Cloning with Plasmid Vectors - Molecular Cell Biology - NCBI Bookshelf. in *Molecular Cell Biology* (2000).

24. pETDuet^{TM}-1 DNA - Novagen | 71146. Available at: https://www.emdmillipore.com/US/en/product/pETDuet-1-DNA-Novagen,EMD_BIO-71146#documentation. (Accessed: 6th March 2020)

25. Bornhorst, J. A. & Falke, J. J. Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology* **326**, 245–254 (2000).

26. Rabbee, M. F. *et al.* Bacillus velezensis: A Valuable Member of Bioactive Molecules within Plant Microbiomes. *Molecules* (2019).

27. Vollenbroich, D., Mehta, N., Zuber, P., Vater, J. & Kamp, R. M. Analysis of surfactin synthetase subunits in srfA mutants of Bacillus subtilis OKB105. *J. Bacteriol.* **176**, 395–400 (1994).

28. Frottin, F. *et al.* The proteomics of N-terminal methionine cleavage. *Mol. Cell. Proteomics* **5**, 2336–49 (2006).

29. Theisen, M. J. *et al. 3-hydroxy-3-methylglutaryl-CoA synthase intermediate complex observed in ‘‘real-time’’. Proc. Natl. Acad. Sci. U. S. A.* 101(47):16442-7 (2004).

30. Konermann, L. Addressing a Common Misconception: Ammonium Acetate as Neutral pH “Buffer” for Native Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 1827–1835 (2017).