KIEL, BRITTANY E., Ph.D. Investigating the Biosynthesis of the β-Branch Found in the Polyketide Difficidin Isolated from *Bacillus amyloliquefaciens* (2019)
Directed by Dr. Jason J. Reddick. 74 pp.

Difficidin is a broad-spectrum antibiotic polyketide product from *Bacillus amyloliquefaciens* FZB42. Although most of the difficidin structure is made by conventional polyketide synthase (PKS) chemistry, it also contains a small structural feature that is not fully understood. This novel "β-branch" structure is unusual because it is a disubstituted exocyclic olefin that is likely thermodynamically unstable compared to the tri-substituted β-branching groups found in other polyketides, such as bacillaene. The hypothesis is that the *B. amyloliquefaciens* difficidin biosynthetic pathway utilizes a mechanistic route that avoids the thermodynamically stable conjugated trisubstituted double bond while forming the less stable disubstituted β-branch olefin. Using mass spectrometry and recombinant proteins, the biochemical steps for the biosynthesis of the difficidin β-branch were investigated and analyzed. The reconstitution of the early steps of the difficidin β-branch biosynthetic pathway involving the construction of the acetyl-DfnX substrate that provides the extra carbon of the β-branch were completed first. Then, using a hydroxymethylglutaryl (HMG) group tethered to the acyl carrier di-domain of the modular polyketide synthase DfnJ, a dehydration by BaeH followed by decarboxylation by DfnM successfully generated the β-branch. With this pathway completely reconstituted, we are able to study how the difficidin system constructs the unusual di-substituted β-branch.

INVESTIGATING THE BIOSYNTHESIS OF THE β-BRANCH FOUND IN THE POLYKETIDE DIFFICIDIN ISOLATED FROM *BACILLUS AMYLOLIQUEFACIENS*

by

Brittany E. Kiel

A Dissertation 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
Doctor of Philosophy

Greensboro
2019

Approved by

_____
Committee Chair

APPROVAL PAGE

This dissertation written by Brittany E. Kiel 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

ACKNOWLEDGMENTS

I would like to dedicate this work to my mother for her unwavering love and support for me throughout this journey. All that I am or ever hope to be, I owe to her.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

1.1 Polyketides

Polyketides are a large family of natural product secondary metabolites and are complex structures that are biosynthesized from acyl-CoA precursors through polyketide synthases (PKSs). These natural metabolites are known to possess various antimicrobial activity (erythromycin and tetracycline), antifungal activity (rapamycin), anti-cholesteremic agents (lovastatin), anticancer activity (daunorubicin), parasiticides, and immunomodulators. These polyketides can be characterized into three classes (type I, type II, and type III) and are found in fungi, bacteria, and plants. Not only do these structurally and functionally diverse family of bioactive natural products provide a rich source of pharmaceutical and agrochemical lead compounds, they allow for reengineering or "programming" the modules to generate functionally optimized "non-natural" natural products. With this said, polyketides have valuable pharmaceutical properties and exhibit a high degree of structural diversity which is why the sales of pharmaceuticals derived from polyketides reaches roughly 20 billion dollars annually[1,2].

The three types of PKSs all perform different functions and can form small polyketides consisting of only 6-8 carbons chain, larger cyclized macrolides, to even a 164 carbon chain (maitotoxin) molecule[3]. In general, PKSs consists of multiple sequential modules each with individual domains responsible for selecting and

processing the building blocks needed for chain extension and subsequent β-keto processing[1].

The first type of PKSs (type I) are the most well-known type and also the focus of this research. Type I PKSs are multi-modular enzymes that operate through a non-iterative pathway. This means that domains within each module act as an assembly line, carrying out only one cycle of the polyketide chain elongation (Figure 1A) before being passed to the next acyl carrier protein (ACP). The reaction begins with a loading (or activation) module acyl transferase (AT) that binds the initial precursor through a thioester bond to an ACP. From there, ketosynthase (KS) can catalyze the decarboxylative 'Claisen-like' decarboxylation on the extending unit so that it can form a carbon-carbon bond with the ACP-bound acyl chain. This is completed through the α-carbon of the extender unit attacking the thioester carbonyl of the ACP-bound acyl chain which generates the extension of the growing polyketide by two carbons. Other domains within the PKSs modules carry out various reductive modifications and essentially "tailor" the β-keto group on the ACP before the next round of chain extension. These domains are optional and can be included in the synthesis or not depending on the how each module is designed to modify the β-keto group. The first domain is called ketoreductase (KR) and it is responsible for reducing the β-ketone to a β-hydroxyl. The next domain is called dehydratase (DH) and it is responsible for eliminating the β-hydroxyl down to an internal enoyl double bond. The next domain is called enoyl reductase (ER) and its job is to reduce the β-enoyl double bond to a single acyl bond[4,5].

Once these domains have finished tailoring the β-keto group, the growing chain is passed to the next ACP where KS extends the chain by two more carbons, followed by subsequent tailoring by these 'optional' domains. Once the chain is fully synthesized, the final step of the reaction is thioesterase (TE) which releases the ACP-bound polyketide molecules from the termination module (sometimes creating macrolides).

The second type of PKSs (type II) are composed of three or more mono-functional or bi-functional enzymes that create modular enzymes through iterative PKSs which means that the modules are recycled and reused within synthesis (Figure 1B). Type II PKSs are known as "minimal PKSs" because they consist of only three enzymes; KS_{α}, KS_{β}, and an acyl carrier protein (ACP). Chain initiation begins with the precursor binding to the ACP through a thioester bond which is catalyzed by KS_{α}. Much like PKSs type I, the first chain is constructed through a Claisen condensation reaction which generates the β-keto group that is subsequently reduced by various additional enzymatic units (ketoreductases, cyclases, aromatases) while the chain is anchored to the ACP. This Claisen condensation reaction is repeated numerous times before the release of the polyketide from the ACP where post-modifications occur to the molecule such as; transferases, hydrolases, and oxidases[1,2].

The last type of PKSs (type III) are only composed of one multifunctional active site, the KS domain. This one KS domain catalyzes repetitive condensation of free acyl-CoA through Claisen condensation which typically produces mono-cyclic or bi-cyclic aromatic compounds independent of ACP. Not only does each chain extension occur on

the same KS, but intramolecular condensation and aromatization all occur within the same PKSs active site (Figure 1C). The ability to tailor this mechanism is outstanding due to the adaptability of the KS domain. Depending on the choice of acyl-CoA and the number of elongation steps, there is a high diversity of potential products. Once the chain is released, downstream enzymes can cause additional branching which generates even more variability in its products[1,2].

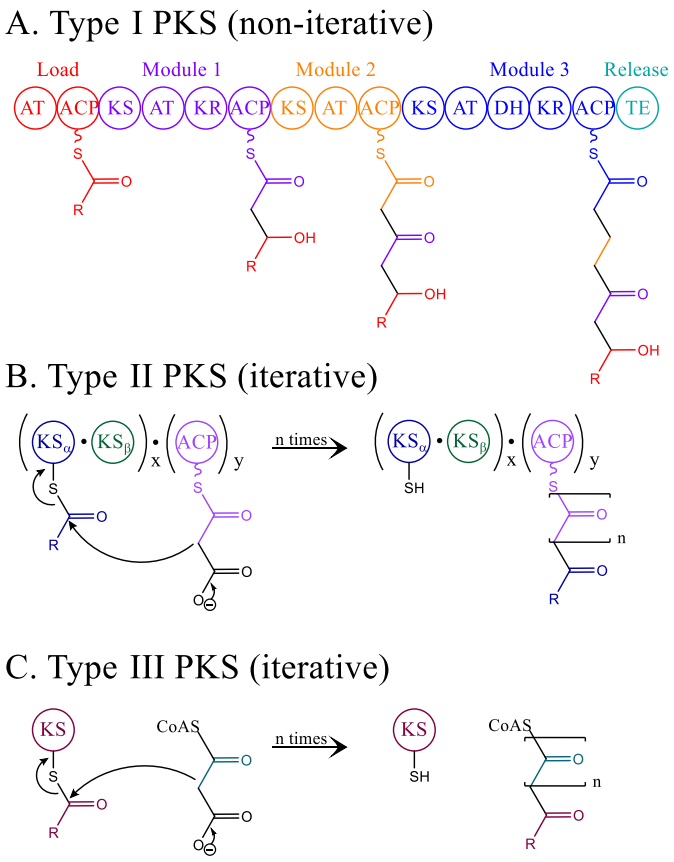


Figure 1. Schematic of Three Types of PKSs Catalytic Reactions and their Different Dynamics.[2]

Erythromycin is a perfect example of the diversity you can achieve using a PKS pathway. Its synthesis follows typical PKSs type I synthesis in a non-iterative fashion and

is a widely used antibiotic that was first isolated in 1952 from the bacteria *Saccharopolyspora erythraea* by McGuire[6]. Since its discovery, erythromycin and its synthesis has been studied extensively and was found to have a 14-membered macrolactone ring known as 6dEB (deoxyerythronolide B) with two deoxysugar moieties attached[7].

In 1991, Leonard Katz was able to analyze and understand the organization of the enzymatic domains present in the multifunctional polypeptides. Erythromycin is formed through the synthesis of 6dEB followed by post-polyketide modifications which include; the hydroxylation of 6dEB at the C-6 and C-12 position, the addition of sugars at the C-3 and C-5 hydroxyls and the O-methylation occurs on the C-3 carbon sugar[7]. The three *eryA* genes involved include six modules with fatty acid synthesis 'FAS-like' domains that contain multi-modular enzymes which form 6dEB precursor in erythromycin. Each of the domains in the six modules contain active site motifs that follow 'normal' PKSs systems and each "DEBS" system contains 2 modules[7,8] which can be seen in Figure 2.

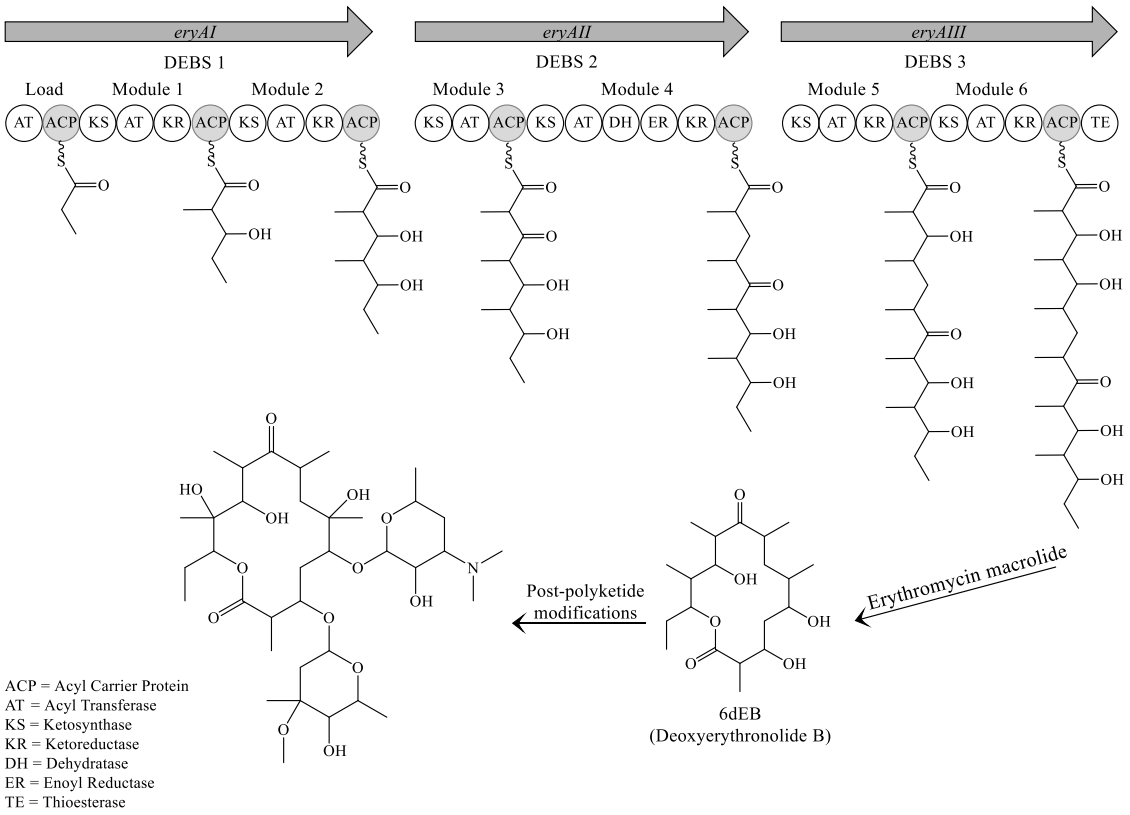


Figure 2. Erythromycin Synthesis by Type I Modular Polyketide Synthases. This Macrolide is Created Using Domains DEBS 1, DEBS 2, and DEBS 3 Found in the *eryA* genes[7,8].

Since erythromycin follows the noniterative type I PKSs, each module performs an addition of a two-carbon chain extension which forms a new C-C bond at each step. The rest of the enzyme domains within the module are then able to complete the subsequent β-keto carbon processing. Once the growing chain is finalized, TE (thioesterase) performs hydroxylation of the C-12 carbon which cyclizes the linear chain into the 6dEB macrolide.

Erythromycin is not only a good example of 'typical PKSs' but it also can easily be modified by re-engineering the modules and the enzymes they possess. In 1992, Leonard Katz was also able to show erythromycin's ability to be tailored. He isolated the

first module in 6dEB synthesis (*eryA1*: DEBS 1) and added the thioesterase (TE) from the last module (*eryAIII*: DEBS 3) to the end of module 1 (*eryA1*: DEBS 1). By doing this, he was able to essentially stop the 6dEB synthesis at module 1 and create a new 6 membered ring (Figure 3) called 3,5-dihydroxy-2,4-dimethyl-n-heptanoic acid-δ-lactone. Katz was able to alter the polyketide structure simply by moving the precise localization of the enzymatic domains and amino acid residues essential for activity[8].

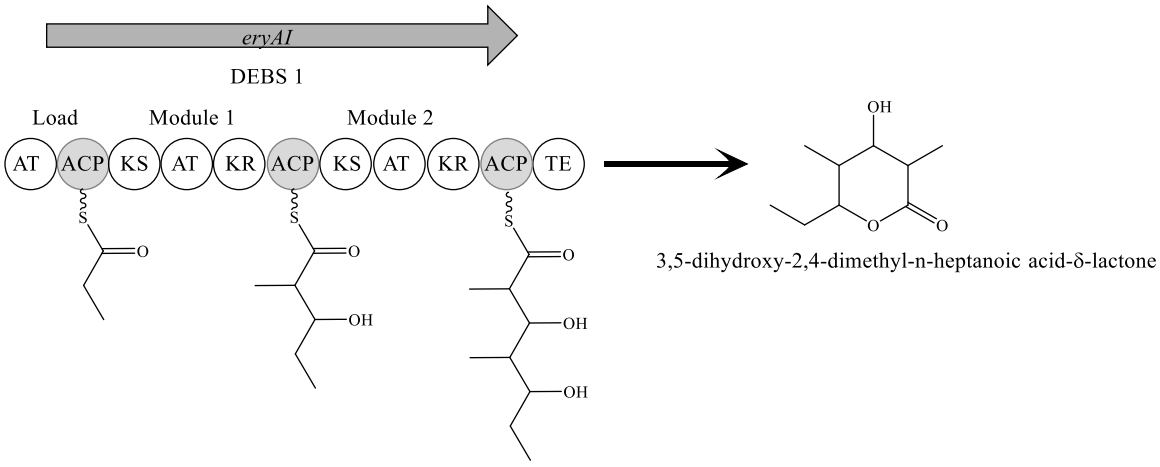


Figure 3. Truncated Version of the Erythromycin Pathway. This Representing *eryAI*:DEBS 1 Using Modules 1 and 2 with the Addition of the Chain-Terminating Thioesterase of DEBS 3. This Generates 3,5-dihydroxy-2,4-dimethyl-n-heptanoic acid-δ-lactone[8].

Not only can polyketide synthesis be tailored by human intervention, but there are various post-polyketide modifications that can occur to alter the predicted PKSs structure. Three have been mentioned previously in the formation of erythromycin. The first is the hydroxylation of 6dEB by cytochrome P450 enzymes at the C-6 and C-12 position generating alcohols on the already methylated carbon. The second is the addition of mycarose sugar at the C-3 hydroxyl and desosamine at C-5 hydroxyl. The last post-

modification step in erythromycin synthesis is the O-methylation that occurs on the C-3 carbon sugar[7].

It is important to note that PKSs can be *cis* AT or *trans* AT to the assembly line. *Cis* AT is when all of the AT domains are in each of the modules in the assembly line. *Trans* AT synthases have modules that use "outside" AT enzymes to select extension units at each step, meaning not all necessary domains are present in the module and therefore things outside of the assembly line are acting upon the growing chain. Since these PKSs modules vary by which domains each module possesses and uses as well as the degree to which the β-keto carbon is reduced, PKSs can generate a vast diversity of complexity of polyketide products.

Another variant that is not seen in erythromycin biosynthesis includes a hybrid of PKSs and NRPSs (non-ribosomal peptide synthetases) synthesis. NRPS domains are synthases that work on peptide biosynthetic pathways independent of ribosomal machinery. The biosynthesis of non-ribosomal peptides occurs through a thiotemplate mechanism reminiscent of PKSs mechanism that incorporates one amino acid residue per module into the growing peptide backbone. There are three key components (domains) of each NRPS module: peptidyl carrier protein (PCP) also known as the thiolation domain (T), the adenylation (A) domain, and the condensation (C) domain. The (A) domain is used to load the amino acid substrate on to the (PCP)/(T) domain and the (C) domain is a condensation domain that catalyzes peptide bond formation[9,10].

Nature uses this hybrid in bacterial modular systems to incorporate acetate building blocks that are then modified in accordance to their domain architecture. One

archetype example of PKSs/NRPSs is BaeJ which is one of the enzymes found in the biosynthesis of the antibiotic bacillaene isolated from *Bacillus amyloliquefaciens* (Figure 4). BaeJ begins with an unusual starter module followed by an NRPS module that is responsible for incorporating glycine. Following this NRPS module is two PKS modules (PKS module 1 and PKS module 2)[11]. The PKSs modules follow the non-iterative type I PKSs that tailors the β-keto carbon according to the domains present.

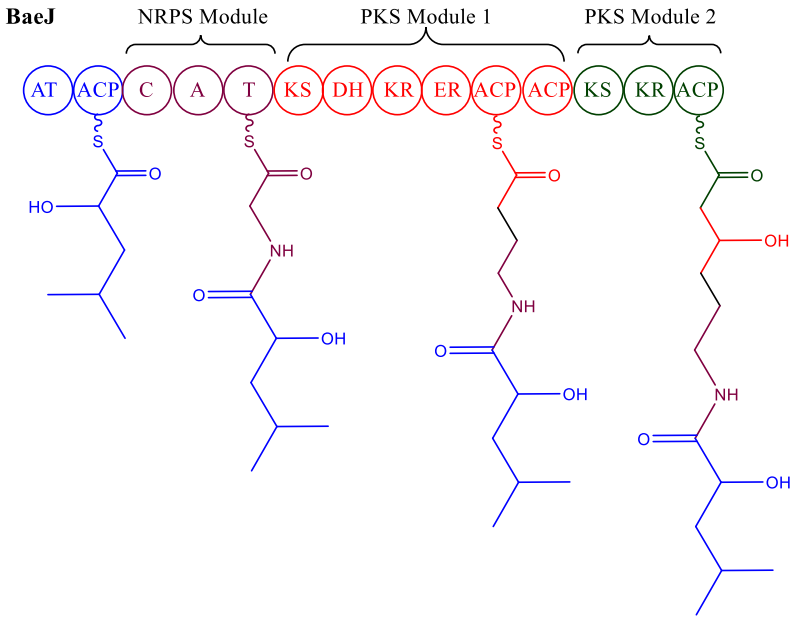


Figure 4. The Domain Architecture and Biosynthesis Products of BaeJ. This is an Example of a Bacterial Modular Non-Iterative Hybrid PKSs/NRPSs.

With there being so many variables in PKSs and PKSs/NRPSs production, there are thousands of possible molecules that can be created. Besides human intervention (knock ins, knock outs) nature does a lot of arranging itself to create important medications and secondary metabolites that are used every day.

1.2 Discovery and Characterization of Difficidin

In 1987**,** two novel macrocyclic polyene lactone phosphate esters (difficidin and oxydifficidin) were discovered in fermentation broths of each of two strains of *Bacillus subtilis*: ATCC 39320 and ATCC 39374. These novel compounds (seen in Figure 5) exhibited activities in vitro against aerobic and anaerobic bacteria, many of which were human pathogens[12].

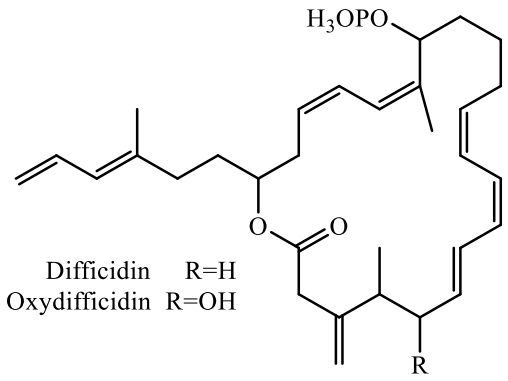


Figure 5. The Structures of Difficidin and Oxydifficidin Isolated from Strains ATCC 39320 and 39374.

In 1993, Claudio Scotti under Alessandra M. Albertini worked with fellow colleagues in Italy who were striving to understand how the life cycle of microorganisms is affected by the functions executed by secondary metabolites. This prompted preliminary molecular and genetic analyses of *Bacillus subtilis* strain 168, which led to the discovery of the "*outG*" locus, which is now called the *pksX* cluster (seen in Figure 6). Genes from the *pksX* cluster showed potential ability to synthesize a secondary metabolite that exhibited polyketidic qualities. This was a major finding because it was the first polyketide synthase analog locus ever discovered in all *Bacilli*[13]. These

researchers opened the door for the unappreciated biosynthetic potential of this well-known bacterial species.

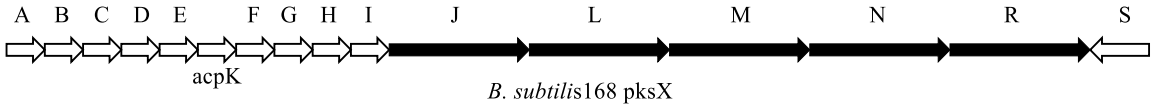


Figure 6. The Organization of the "*outG*" Locus Now Known as the *PksX* Cluster which was Isolated from *Bacillus subtilis* Strain 168.

In 1994, Pramathesh Patel and colleagues from different Bristol-Myers Squibb Pharmaceutical Research Institute departments were on a path to discover new antibiotics that inhibit protein synthesis. They identified a particular antibiotic called bacillaene from fermentation broths of a strain of *Bacillus subtilis*. While screening novel antibacterial agents, this promising molecule with a molecular weight of 580g/mol and empirical formula of $C_{35}H_{48}O_7$ showed a strong potency against hyper-permeable strains of *Escherichia coli*. After the antibiotic was extracted, column chromatography fractions that contained bacillaene were pooled and subjected to further purification[14], but the structure was not elucidated.

In 2006, another research team decided to come up with their own way of isolating the antibiotic (bacillaene). Rebecca Butcher's research team led by Christopher T. Walsh, and Jon Clardy found that bacillaene had a rapid decomposition when exposed to light or room temperature which made it impossible to characterize it in earlier studies. Due to this instability, they adopted a purification strategy based on minimal chromatography coupled with detailed NMR spectroscopic analysis and were finally able to characterize the structure of this unusual secondary metabolite known as bacillaene.

It was determined to be a linear structure that contains two amide bonds which both play a role in its synthesis. The first links the hexaene-containing carboxylic acid to an ω-amino carboxylic acid containing a conjugated triene. The second links an α-hydroxy carboxylic acid to an ω-amino carboxylic acid containing a conjugated hexaene[15]. The structure of bacillaene is unusual due to it being acyclic and having a high degree of conjugation with several *cis* double bonds, coupled with the fact that it incorporates a β-branch on the 18' methyl group which is a feature seen in only a few other polyketides.

After the original discovery of the novel antibiotic bacillaene, Xiao-Hua Chen headed a research team under Rainer Borriss that set out to explore the molecular biology of polyketide biosynthesis in *Bacillis*. They knew that this largely unexplored $Gram^+$ spore forming family of bacterial species was one of the richest bacterial sources of bioactive natural products. Their study characterized three giant polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB42 both structurally and functionally (Figure 7). Of these three giant modular 'trans-acyltransferase type' PKS systems, researchers were able to prove that there was an ortholog to the *pksX* cluster (found in *B. subtilis*) called *pks1* (later named *bae*) that was also able to produce the antibiotic bacillaene. These findings demonstrated that both the *bae* and *pksX* gene clusters were responsible for directing the synthesis of the bacillaene polyketide. Identifying the genes encoding the biosynthesis of this polyene antibiotic played a key role in helping future researchers understand its complicated biosynthesis. Up until this point, *pks2* and *pks3* were still unexplored[16].

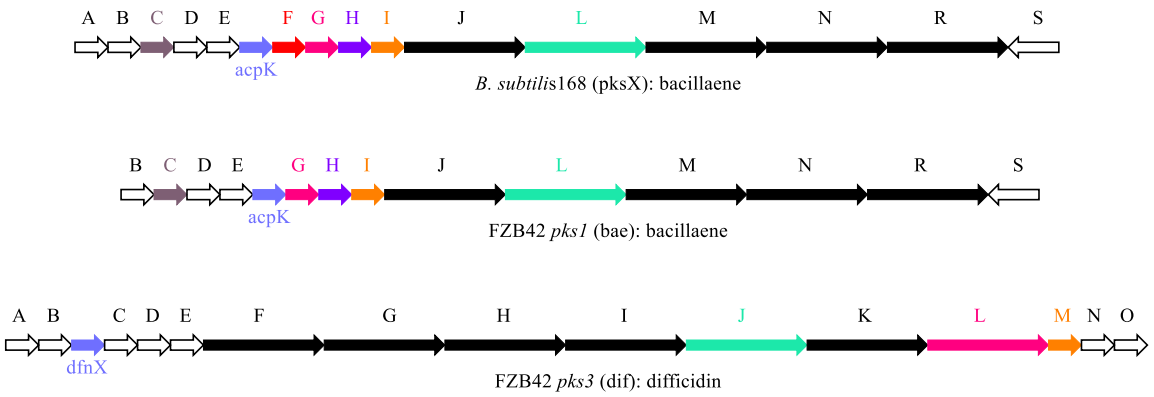


Figure 7. Evaluation of Similarities of the Gene Clusters Involved in Polyketide Synthesis. The Three Operons Represent *Bacillus subtilis* 168 *pksX* (Bacillaene), *Bacillus amyloliquefaciens* FZB42 *pks1* (Bacillaene) and *Bacillus amyloliquefaciens* FZB42 *pks3* (Difficidin). The Colors Highlight the Homologs in the Various Clusters.

Xiao-Hua Chen's research provided insight in to the FZB42 strain and he was able to show that it produces both bacillaene (*pks1*) and difficidin (*pks3*). Although it was originally thought that difficidin was likely synthesized from the *pksX* cluster ("*outG*" locus), the data produced by Chen's research gave insight into the original bacterial strains (ATCC 39320 and ATCC 39374) and the Reddick Lab was able to determine they were not *B. subtilis* but rather unique strains of *Bacillus amyloliquefaciens*. This was discovered when the Reddick Lab acquired sequencing data that showed the ATCC strains were very similar to the FZB42 strain found in *B. amyloliquefaciens*, not *B. subtilis* (data not yet published).

1.3 β-Branch Production in Bacillaene

A study that set out to analyze naturally occurring small molecules that help regulate interactions between the common soil genera *Bacillus* and *Streptomyces* ended up revealing an unprecedented finding in bacterial subcellular organization. Paul Straight, Christopher Walsh, David Rudner, and Robert Kolter were all part of the team that

wanted to identify the small molecule produced by *pksX* synthase. The ≈80kb *pksX* gene cluster that occupies around 2% of the entire *Bacillus subtilis* genome encodes for a 2.5MDa hybrid of non-ribosomal peptide synthases and polypeptide synthases[18]. *PksX* is comprised of 16 genes that start PKSs assembly from the *pksJ* multi-modular enzyme which sequentially adds biosynthetic monomers until bacillaene is released as a linear-free carboxylate at *pksR*. These five genes (*pksJ*, *pksL*, *pksM*, *pksN*, and *pksR*) encode the multi-modular synthetase, while the other ten (*pksB-pksI* and *pksS*) encode individual enzymes that operate *in trans* to the assembly line as the molecule is being built[19].

Knowing that the *pksX* gene cluster encodes a hybrid of NRPSs/PKSs, the molecular assembly can be investigated. PKSs and NRPSs are large modular enzymes that regularly couple malonyl derivatives and amino acids, respectively[15]. They are known for producing a diverse array of secondary metabolites using varied building blocks and tailoring enzymes. Since these are modular, the genes that normally encode them are often co-linear with their corresponding natural product structures. Normally, the molecular assembly allows for prediction of the product's structure[20]. However, bacillaene does not follow a 'normal assembly'.

This uncertainty of assembly is because a typical NRPS module consists of three domains which are responsible for activation. The first is an adenylation (A) domain that is responsible for amino acid activation. The second is a thiolation (T) domain, also called a peptidyl carrier protein (PCP), which is covalently tethered to the activated amino acid. The last is condensation (C) domain that catalyzes the peptide bond formation[10,21]. Similarly, three domains are essential as the basic equipment of a PKS

elongation module as well. The first is an acyltransferase (AT) domain which is responsible for extender unit selection and transfer. The second is an acyl carrier protein (ACP) that acts as an extender loading unit and the third is a ketoacyl synthase (KS) domain for decarboxylative condensation of the extender unit with an acyl thioester[22]. Each module is responsible for one individual chain elongation step and the specific order of these modules is what defines the sequence of the incorporated amino acid (NRPSs) or carboxylic acid (PKSs) building blocks.

With that being said, the *pksX* synthase has unusual features that act much different than the standard assembly-line mode characteristic of polyketide and non-ribosomal peptide biosynthesis which generates a methyl β-branch instead of the typical β-carbon options (ketone, alcohol, double bond, or single bond).

Through combining biochemical and mass spectrometric techniques, a team of researchers led by Christopher Calderone (in the laboratory of Christopher Walsh at Harvard) assigned functional roles to the proteins *pksC*, *acpK*, *pksF*, *pksG*, *pksH*, *pksI*, and *pksL* all found in the *pksX* cluster. They ultimately discovered that the proteins' purpose was to incorporate an acetate derived β-methyl branch on a β-ketoacetyl-S-carrier protein which ends up generating a Δ^2-isoprenyl unit tethered to a PKS carrier protein domain via the unusual thioester linkage[23] (Figure 8). The resulting acetate derived β-methyl branch lies at the β-position in the middle of the hexaene moiety, specifically on carbon 18. This β-alkylation requires the action of five enzymes and one external thiolation domain to function. This particular pathway represents a non-

canonical route of the construction of prenyl units that imitates isoprenoid and polyketide biosynthesis.

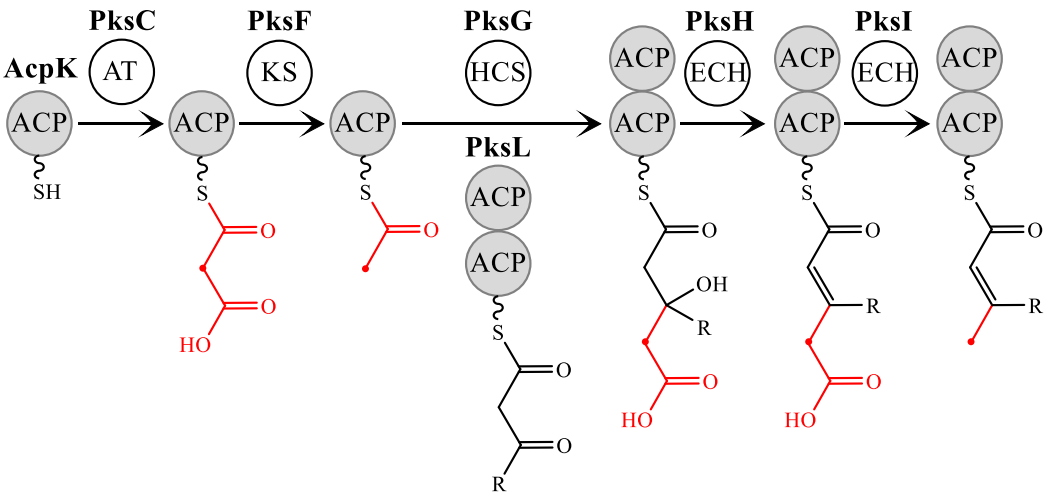


Figure 8. The Insertion of the β-Methyl Branch on PksL by the AcpK/PksC/PksF/PksG/PksH/PksI Subcluster. (HCS)= HMG-CoA Synthase; (ECH)= Enoyl-CoA Hydratase/Isomerase

Although Reddick Lab's previous research (not yet published) has discovered that the original difficidin-producing strains ATCC 39320 and ATCC 39374 are in fact strains of *Bacillus amyloliquefaciens* rather than *Bacillus subtilis* as was previously thought, not much has been done to show how difficidin is actually produced. However, analyzing the structure of difficidin provides some insight into its construction. Much like bacillaene's β-branch on the 18' methyl group, difficidin also exhibits a β-branch. The only difference between the two is that bacillaene's β-branch is an isoprenoid-like unit with an internal trisubstituted double bond and external methyl branch whereas difficidin's β-branch is an external disubstituted olefin (seen in Figure 9).

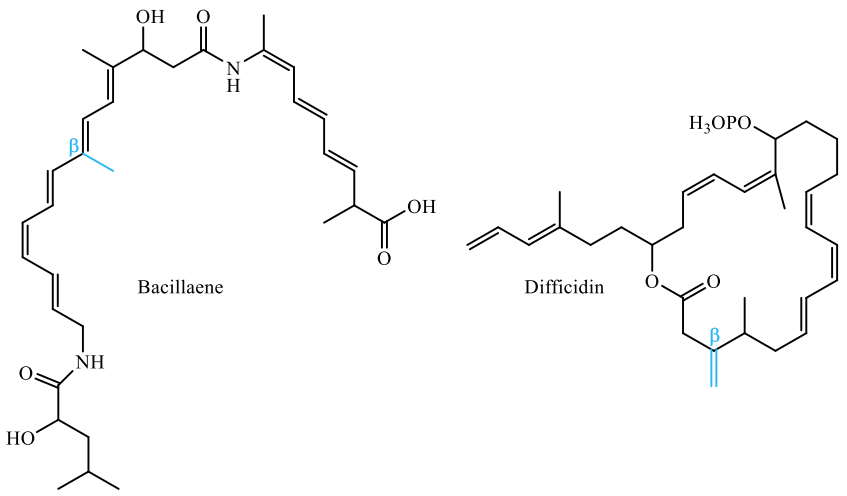


Figure 9. Structures of Bacillaene and Difficidin with their β-Branches Highlighted for Comparison.

Although the biosynthetic gene cluster of difficidin contains some homologs of the β-branching biosynthetic genes found in the bacillaene gene cluster, the different β-branch structures in these two compounds may warrant some differences in enzymatic chemistry that are thus far poorly understood. Since there are thousands of small molecule natural products that display interesting biological activities and exhibit NRPSs, PKSs, or a hybrid of the two, the two antibiotics were compared. Much like the production of bacillaene, difficidin follows the same irregular chemistry demonstrated in the *pksX* cluster. Previous research has shown the that the difficidin biosynthetic gene cluster is not the only polyketide that has homologs to the genes responsible for the β-branch construction in bacillaene. It was discovered that each of the secondary metabolites that have homologous genes to *pksX*, such as; curacin, jamaicamide, mupirocin, and pederin (seen in Figure 10) all display some form of the β-branching chemistry found in the structure of bacillaene[23] and difficidin. For example, jamaicamide

and pederin exhibit the external double bond β-branch exhibited in difficidin whereas mupirocin displays the same internal β-branch found in bacillaene.

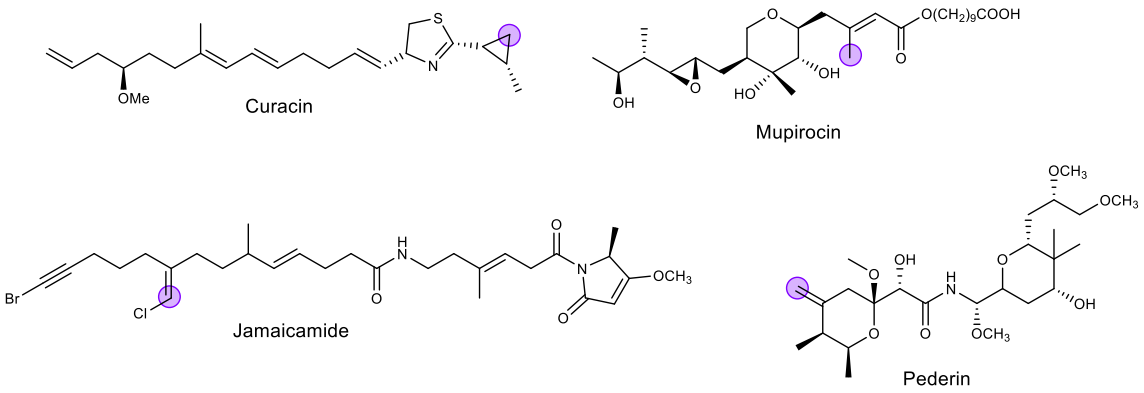


Figure 10. Comparison of Secondary Metabolites that Share Homologs of the β-Branching Genes.

After analyzing their β-branch gene homologs, it appears as though (for pederin and difficidin at least) that both biosynthetic gene clusters of secondary metabolites only have one gene that is a part of the "crotonase superfamily" and therefore potentially carry out different chemistry than the metabolites with two genes in the superfamily (curacin, mupirocin, and bacillaene). Since the *dif* (*dfn)* gene cluster contains homologs of the genes known to be involved in the synthesis of the β-branch methyl group on the 18' carbon in bacillaene, it's likely that the β-branch found in difficidin is constructed in a similar manner as the one found in bacillaene.

1.4 Comparing Pathways

Since difficidin's biosynthetic genes share homology with the β-branch genes associated with bacillaene and other β-branching polyketides, the hypothesis is that *Bacillus amyloliquefaciens* genes *(dfnX, baeC, bamF, dfnJ, dfnL, baeH,* and *dfnM*)

catalyze β-branching synthesis with steps similar to *pksC*, *acpK*, *pksF*, *pksG*, *pksH*, *pksI*, and *pksL* all found in the *pksX* cluster in *Bacillus subtilis* (seen in Figure 11) by utilizing a different pathway to avoid the thermodynamically stable quaternary double bond and form the less stable β-branch olefin.

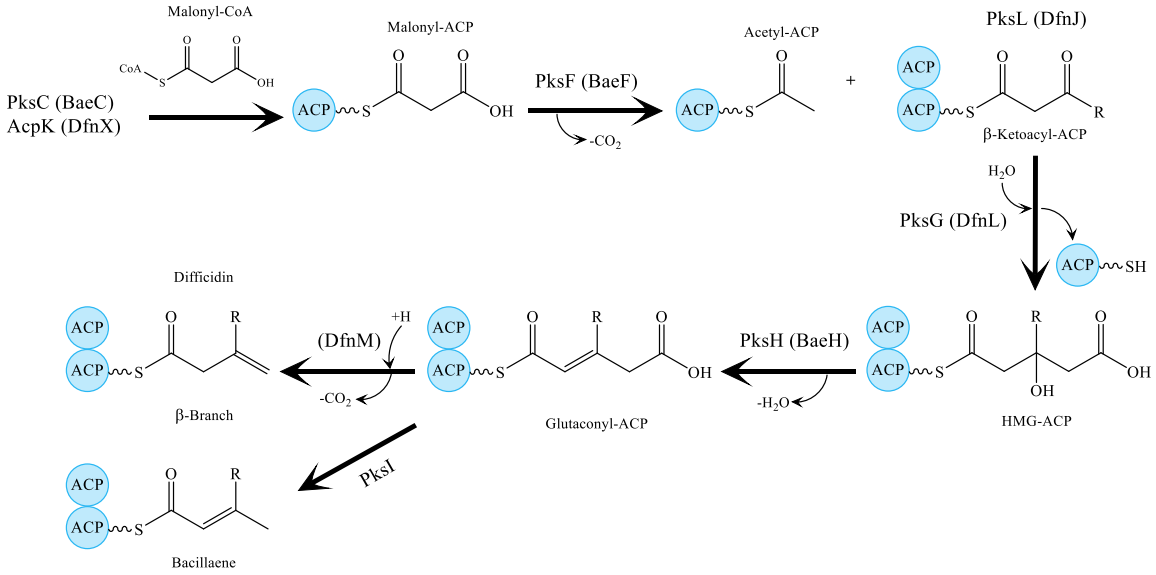


Figure 11. The Synthesis of the β-Branch in Bacillaene and Difficidin. *B. subtilis* Bacillaene Genes are Next to the *B. amyloliqufaciens* Difficidin *dif (dfn)* Homologs in Parenthesis.

There are 7 genes responsible for the synthesis of the β-branch found in bacillaene and each have homologs that appear in the *dif (dfn)* cluster used in the production of difficidin (as seen in Figure 11). The main genes responsible for the two different β-branches (*pksH* and *pksI* in *B. subtilis/ baeH* and *dfnM* in *B. amyloliquefaciens*) - as well as similar homologs for other compounds - all belong to the large family of enzymes called the "crotonase superfamily." The crotonase superfamily (or enoyl-CoA hydratase) consists of enzymes with homologous sequences that catalyze a wide range of metabolic reactions but all share a common structural solution to a mechanistic problem.

There are several types of reactions that target various β-substituted thioesters such as; dehydratases, dehalogenase, hydratase, isomerase activities, etc. Overall, the object of this superfamily is to stabilize an enolate anion intermediate derived from an acyl-CoA substrate using an oxyanion hole[24].

Although *pksH* and *pksI* align well with each other (identity: 17.2%, similarity: 36.6%), they play two different roles in the production of the β-branch in bacillaene. A BLAST search revealed that the *dfn* cluster only has one homolog from the crotonase superfamily (*dfnM*). With that being said, the original hypothesis was that *dfnM* did the work of both *pksH* and *pksI* in one concerted step instead of two (Figure 12). Since *dfnM* is the sole crotonase homolog in the *dif (dfn)* cluster, it was possible that it solely created the external β-branch olefin found in difficidin and may therefore use a different mechanism than the *pksH/pksI* pair found in *B. subtilis*.

However, when homologs were aligned together, the amino acid sequences align at a higher percentage between *dfnM* and *pksI* (identity: 55.8%, similarity: 72.7%) compared to *dfnM* and *pksH* (identity: 17.8%, similarity: 38.3%). This information generated a new hypothesis in which *dfn* cluster utilizes *baeH* from *pks1* (bacillaene) found in *Bacillus amyloliquefaciens* to complete its pathway. *BaeH* comes in to dehydrate HMG-ACP and then *dfnM* decarboxylates glutaconyl-ACP. The most likely difference is that *pksI* (*B. subtilis*) protonates the enolate at the end of the chain creating the internal β-branch double bond instead of *dfnM* protonating difficidin at the α-carbon right after the electrons drop down off the enolate creating the external β-branch olefin (seen in Figure 12).

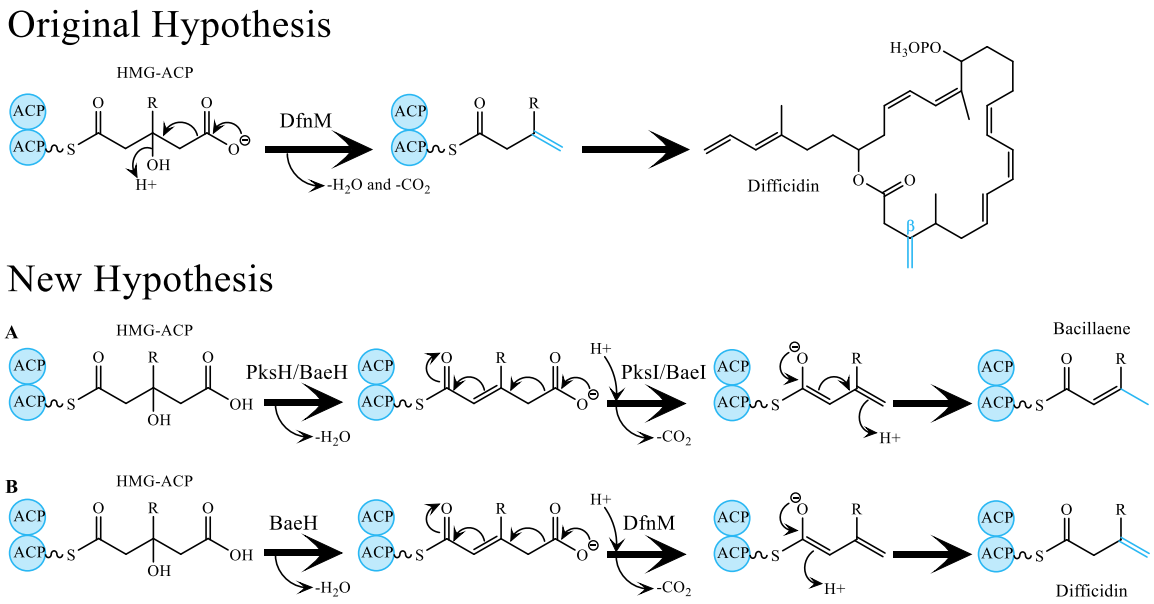


Figure 12. Original Hypothesis vs. New Hypothesis. This is the proposed difference in function between *dfnM* and its homolog *pksI*. This also shows the novel hypothesized chemistry occurring at *dfnM*. A) Bacillaene β-branch pathway in *B. subtilis*. B) Difficidin β-branch pathway in *B. amyloliquefaciens.*

The goal of this research is to discover new chemistries utilized in the biosynthesis of polyketide natural products and understand how an external β-branching double bond is installed during the biosynthesis of the polyketide difficidin. The central hypothesis of this paper is that the biosynthesis of difficidin utilizes a shortened variant of the isoprene-like assembly pathway that was observed for the internal β-branching double bond observed in bacillaene.

CHAPTER II

EXPERIMENTAL PROCEDURES

2.1 Isolating Genomic DNA

Isolation of genomic DNA started with streaking a lysogeny broth (LB) plate with *Bacillus amyloliquefaciens* FZB42 bacteria strains ATCC 39374 and ATCC 39320. From there, a 5 mL culture of the gram-positive bacteria was grown overnight at 37°C. The DNA was then purified using the Wizard® Genomic DNA Purification Kit protocol. These two strains of DNA are the template DNA that each of the difficidin genes were isolated from.

2.2 Primers

With the genomic DNA isolated, primers for each of the genes were then constructed (Figure 13) and they are listed below. These primers are short, single stranded DNA often referred to as oligonucleotides that 'bracket' target DNA for PCR. There is one upstream primer and one downstream primer required for each gene. Each primer binds to one strand of the double stranded target DNA using their complementary sequence and sequester the targeted region of DNA. Using these primers, the genes of interest (*dfnX, baeC, bamF, dfnJ, dfnL, baeH, dfnM*) were 'bracketed' by the primers and isolated from strains ATCC 39374 and ATCC 39320. CACC additions were required for directional cloning into the commercially prepared pET-200 vector using topoisomerase-based cloning.

Genes	Forward "Up" Primers
dfnX	CACCATGGAACAAACCAAGGAAATGAAGC
baeC	CACCATGATCACATACTTATTCCCCGGAC
bamF	CACCATGTCGAGGAGTAATAATGGCACG
dfnJ	CACCTTGGATTGTACTGTTTCTGTCCAGCAGCCAGAAAAAGC
dfnL	CACCATGAATGTAGGAATCGAAGCAATGAAC
baeH	CACCGTGACCTATGAGACGATAAACGTC
dfnM	CACCATGAATCCAATGATTGATGTACAGGAAAC

Genes	Reverse "Down" Primers
dfnX	TCATGATTGAGCACTGTCCATATATTTTTG
baeC	CTACATGGCGGGTGAAGCTTCAG
bamF	TTATCTGCTTCGGAAATGATTTTCTAAACAAAGCGC
dfnJ	TTAGACAGAACCGGTCAGATACCCG
dfnL	CTAGGTCGTCCATTCATACTTTCTGTGAAAC
baeH	CTAACTGTTCTCCGTTTCCCAAGGAAGC
dfnM	TTAACTGCCGTATAAGCGGCTGATTC

Figure 13. A List of all Fourteen Primers (Both Forward and Reverse) Used to Isolate the Seven Genes of Interest. CACC Additions to the 5' End of Each Gene are Underlined.

2.3 Polymerase Chain Reaction

Once each of the seven genes listed above were isolated using their "up" and "down" primers, they were amplified using polymerase chain reaction (PCR). This technique is vital to laboratories because it exponentially amplifies DNA sequences by generating copious amounts (up to millions of copies) of the targeted DNA fragment. This is done though a number of denaturing, annealing, and elongation steps which exposes reactants to different temperatures needed for their temperature-dependent reactions. Each PCR reaction requires four main components; template DNA, primers, DNA polymerase, and free dNTPs (deoxyribonucleotide triphosphates). The first step is denaturing which splits or 'melts' the double stranded DNA template through high thermal temperatures. Once the two strands have separated, the temperature is lowered so

the primers can bind to their complementary sequence. This then becomes the generator of replication for DNA polymerase to enzymatically assemble new strands of DNA using free dNTPs.

Each PCR mixture contained 1 µL of purified genomic DNA, 3 µL each of "up" and "down" primers, 25 µL of 'master mix" (1µL of deoxyribonucleotides (dNTPs), 10 µL of 5X Phusion high fidelity buffer, 0.5 µL of Phusion polymerase, and 17 µL of sterile distilled deionized water (ddH_2O) for a total reaction volume of 50 µL. Negative controls were also performed replacing the genomic DNA with ddH_2O to maintain the final volume of 50 µL. The PCR thermocycle program starts with its initial denaturation at 98°C for 30 minutes. Then, the program repeats the denaturation/anneal/extend 30 times. These cycles begin with another 5 minutes of denaturation at 98°C, then an annealing temperature of 55°C for 10 minutes, and then extension at 72°C for 1 minute. Lastly, there was a final extension at 72°C for 10 minutes before the reaction was stored at 4°C until used.

2.4 DNA Sequencing

To test the success of the DNA amplification, the genes were verified using a 1% agarose gel. These gels can be read using UV light and their bands correspond to a certain mass (in this case, the number of base pairs in the gene of interest) which can be verified using a standard ladder. Larger masses will migrate towards the top of the gel, whereas the smaller fragments can pass through the agarose pores easier and will be found lower down the gel.

Gels were prepared using 30 mL of 1x TAE (tris-acetate-EDTA) buffer, 0.3g of agarose, and 2 µL of ethidium bromide (10 mg/mL) which allowed them to be examined under UV light. For the DNA, the gels were loaded with 16 µL of PCR product and 4 µL of 6x loading dye. For the 2-log DNA ladder, the gels were loaded with 1 µL of ladder, 1 µL of 6x loading dye, and 6 µL of ddH_2O. The gels were then put in 1X TAE buffer in the electrophoresis chamber and ran at 120V for approximately one hour. After getting positive results for each of the seven genes, each were cloned into pET-200 as described in the procedure below.

2.5 Vector Cloning

After confirming the success of PCRs for each gene through sequencing, the amplified fragment was then purified using the Qiagen® QIAquick PCR Purification Kit using the quick start protocol. This is to remove any leftover primers or excess DNA from PCR and to essentially clean it up. The purified PCR product was then cloned into the pET200 vector using the Champion™ pET Directional TOPO® Expression Kit according to the user manual. This process inserted the purified PCR product into the TOPO cloning site vector (indicated in the Figure 14) which incorporates the gene and it becomes part of the plasmid. The cloning reaction contained 4 µL of purified PCR product, 1 µL of salt solution (provided), and 1 µL of the TOPO® vector. This reaction was incubated at room temperature for 5 minutes before being used to transform One Shot® TOP10 cells.

This pET200 vector also contains a 6 consecutive histidine tag (6xHis) that is crucial for protein purification performed later in the experiment. By having a 6xHis tag, the protein is able to be purified through nickel affinity chromatography.

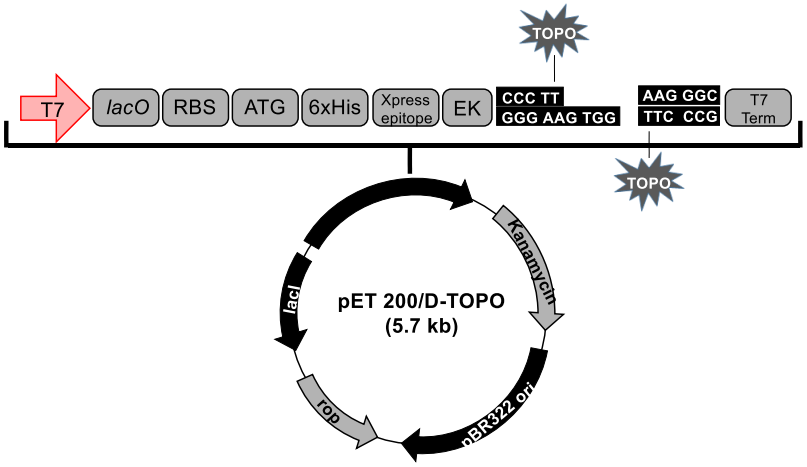


Figure 14. The pET200 Vector and TOPO Cloning Site Used to Incorporate the Target Genes. (Champion™ pET Directional TOPO® Expression Kit user manual)

2.6 Cell Transformation

Each plasmid containing the target genes was then transformed following the OneShot® TOP10 Chemical Transformation protocol found in the Champion™ pET Directional TOPO® Expression Kit user manual. These TOP10 chemically competent *E. coli* cells are ideal for high efficiency cloning and plasmid reproduction because of their tremendous transformation ability. Once the cells were transformed, they were plated on Luria-Bertani (LB) agar plates that contained 50 µg/mL kanamycin and incubated overnight at 37°C. Each LB plate had multiple colonies, six were picked for each gene and added to separate 5 mL LB cultures containing 50 µg/mL kanamycin. These cultures incubated at 37°C overnight were then centrifuged at 13,000 rpm for 5 minutes and the

supernatant was removed. Plasmid DNA from each chosen colony was then isolated using the Qiagen QIAprep® Miniprep kit. These plasmid products were then analyzed on gel electrophoresis using the protocol previously mentioned. The reason behind picking more than one is because not every colony will have the cloned insert and until performing PCR and analysis with an agarose gel, it will not be evident which ones are positive transformants.

Each purified plasmid was then used to transform BL21 Star™ Chemically Competent *E. coli* cells following the Transforming BL21 Star™ (DE3) One Shot® Cells protocol. These BL21 Star™ have a truncated RNase E that lacks the C-terminal domain that is required for mRNA degradation which in turn generates a large amount of protein. After transformation, 100μL of the reaction was plated on LB agar plates containing 50 μg/mL kanamycin and incubated overnight at 37°C. The colonies that had successful growth were put in 5 mL culture with 50 μg/mL kanamycin at 37°C overnight and then the transformed strain was stored at -80°C in 10% glycerol.

2.7 Protein Overexpression

After each plasmid was transformed into BL21 Star™, a 1-liter LB culture (10g Bacto Tryptone, 10g NaCl, and 5g yeast extract at pH 7.5) of cells containing 50 mg/mL kanamycin was grown at 37°C for roughly 4 hours at 220 rpm until the OD_{595} reached an absorbance between 0.5 and 0.6. Once the growth phase of the media was at an optimal absorbance, 0.2383g (1mM) isopropyl β-D-1-thiogalactopyranoside (IPTG) was added. IPTG is a reagent that triggers transcription of the lac operon so it is often used to induce protein expression when the gene is under the control of the lac operon operator. After

induction, the culture was shaken overnight at 18°C. The lower temperatures allowed the proteins to grow with more stability and to help prevent the common problem of inclusion body interference. After overnight growth occurred, the culture was then centrifuged at 6,500 rpm for 30 minutes at 4°C. The supernatant was discarded and cell pellets were stored at -80°C. This procedure worked for all of the proteins except for DfnX.

Due to DfnX being so small, allowing it to grow for more than 4 hours after induction proved to degrade the protein. Since there was such poor yield, this particular protein was grown at 37°C until the OD_{595} showed an absorbance between 0.5 and 0.6, and then it was induced with 1mM IPTG and grown for another 4 hours before the cells were centrifuged at 6,500 rpm for 30 minutes at 4°C. The supernatant was discarded and cell pellets were stored at -80°C.

2.8 Protein Purification

As mentioned before, the proteins' ability to be purified comes from the 6xHis tag provided by the pET200 vector. This His tag allows the proteins to bind to the nickel in the column and stay bound until the final elution step.

Each of the proteins were purified at 4°C using the following conditions. The bacterial pellet from 1 liter of culture was resuspended in 20 mL of binding buffer which contained 0.5M NaCl, 20 mM Tris, 5 mM imidazole, and 10% glycerol (pH 7.9). After 3 minutes of sonicating the cells on ice, they were lysed and the debris centrifuged at 11,500 rpm for 30 minutes. The supernatant was added to a pre-equilibrated 2 mL nickelnitrilotriacetic acid (Ni-NTA) column after the supernatant was passed through a

0.45 μm syringe filter to remove any particulates. After being loaded, another 20 mL of binding buffer was passed through followed by the 12 mL wash buffer which contained 0.5M NaCl, 20 mM Tris, 60 mM imidazole, and 10% glycerol (pH 7.9). The elution buffer containing 0.25 M NaCl, 10 mM Tris, 25 mM imidazole, and 10% glycerol (pH 7.9) was passed through until the column fractions had no more protein in them. This was determined using Bradford analysis.

2.9 Dialysis/FPLC

Once the proteins had been eluted from the Ni-NTA column, they were combined into a Snakeskin dialysis tubing (7000 molecular weight cutoff) and put into a 4L dialysis buffer of 25mM Tris (pH 9) and left in 4°C overnight. All of the proteins survived this treatment except for DfnL, which slightly precipitated coming off the column and then completely precipitated overnight in dialysis. Since DfnL exhibited post-column insolubility, this generated the need for the FPLC (fast protein liquid chromatography) with a slightly more acidic buffer of 25 mM Tris (pH 7.5) to get the excess imidazole away from the protein as fast as possibly to prevent it from precipitating. After DfnL eluted off the de-salting column, it was then put through a Vivaspin concentrator and centrifuged at 8,000xg for 30 minutes which re-concentrated the protein. Each protein had 10% glycerol added before putting in the -80°C for storage.

2.10 Trypsin Reaction

Using Trypsin Gold, the proteins were cut at every lysine and arginine which gave fragments of different masses. This method helped analyze mass spectrometry more

accurately. The trypsin method is as follows, the proteins were first dissolved in 8 M, 50 mM Tris-HCl (pH 8) (volume dependent on protein concentration) and heated for 45-60 minutes in 37°C to stop their reactions. After the reactions had been stopped, 50 mM Tris-HCl, 1 mM $CaCl_2$ (pH 8) was added until the urea concentration was less than 1 M and then trypsin gold was added at a protein ratio of 1:20 (w/w). The reaction was then incubated in 37°C overnight.

The overnight incubation worked until the DfnJ, Sfp, AcAcCoA reaction which required only 1-2 hours of incubation before it started to degrade and was no longer viable. Due to this shorter trypsin digestion period being successful, the other reactions were also able to be shortened to roughly 3-4 hours and still work correctly. Not much difference was observed between the 3 hour trypsin digestion vs 18 hours.

2.11 Sfp in Vitro

Sfp is a 4' phosphopantetheine transferase that comes from *Bacillus subtilis*. It is widely used in biochemical laboratories because of its ability to transfer a phosphopantetheine (Ppant) arm from various coenzyme A's on to a non-active apo protein or peptide and make it active or holo. This allows for the development of standards that can be compared to targeted molecules for analysis. Sfp was added at a 126.1 μM concentration to almost every reaction to aid in the transfer of CoA's (1 mM) Ppant arm to each of the ACPs: DfnX (35.97 μM) and DfnJ (61.4 μM).

2.12 ESI–MS with Jupiter Column

To prepare the samples for mass spectrometry, 50 μL of DfnX (35.97 μM) was incubated with 20 μL Sfp (126.1 μM), 10 μL of CoA (10 mM), and 20 μL reaction buffer of 50 mM Tris-HCl, 12.5 mM $MgCl_2$, 50 mM NaCl (pH 8). This mixture was incubated at room temperature for 1 hour before being de-salted in a spin column microcentrifuge tube and put into a sample vial. A Jupiter 5m C4 300E column was used for these undigested, intact proteins.

Solvent A contained H_20 and 0.1% formic acid while solvent B was acetonitrile and 0.1% formic acid. The method used for this purification was 0-10 min 90% A and 10% B, 10-12.5 min 70% A and 30% B, 12.5-25 min 40% A and 60% B, then 25-30 min 10% A and 90% B. This was then followed by a series of washes which repeatedly changed from 10% A and 90% B back to 70% A and 30% B to avoid carryover from any of the previous samples with a total run time of 43 minutes. The flow rate for this column was 0.3 mL/min after using a T-splitter to divert excess material.

2.13 ESI–MS with BEH Peptide Column

A peptide BEH C18 1.7 μM column was used for trypsin digested proteins. This column is good for assaying samples for proteins characterization, proteomics, and peptide synthesis. The method used for this purification had an initial column equilibration of 90% A and 10% B. Solvent A contained H_20 and 0.1% formic acid while solvent B was acetonitrile and 0.1% formic acid. The instrument method used was 0-17 min 40% A and 60% B, 17-21 min 0% A 100% B, 21-25 min 90% A and 10% B. This was then followed by a 10 minutes shutdown of 0% A and 100% B to avoid carryover

from the any of the previous samples with a total run time of 35 minutes. The flow rate was 0.15 mL/min.

2.14 Sample Preparation for Sfp Dependent Standards

The reaction buffer used for these reactions was 50 mM Tris-HCl, 12.5 mM $MgCl_2$, 50 mM NaCl (pH 8). These Sfp standards (excluding propionyl CoA and butenoyl CoA) all mimic the intermediates found in the bacillaene/difficidin pathway and can therefore be compared to the in vitro intermediates generated.

1) **Holo-DfnX (DfnX + Sfp + CoA):** 50 μL of DfnX (35.97 μM) was incubated with 20 μL Sfp (126.10 μM), 1 μL of CoA (100 mM), and 29 μL reaction buffer. After incubating at room temperature overnight, it was analyzed on the Jupiter column.
2) **Acetyl-DfnX (DfnX + Sfp + Acetyl CoA):** 50 μL of DfnX (35.97 μM) was incubated with 20 μL Sfp (126.10 μM), 10 μL of acetyl CoA (10 mM), and 19 μL reaction buffer. After sitting at room temperature overnight, the trypsin procedure was performed and then analyzed on the BEH peptide column.
3) **Malonyl-DfnX (DfnX + Sfp + Malonyl CoA):** 50 μL of DfnX (35.97 μM) was incubated with 20 μL Sfp (126.10 μM), 10 μL of malonyl-CoA (10 mM), and 19 μL reaction buffer. After sitting at room temperature overnight, the trypsin procedure was performed and then analyzed on the BEH peptide column.
4) **Propionyl-DfnJ (DfnJ + Sfp + Propionyl CoA):** 50 μL of DfnJ (61.4 μM) was incubated with 20 μL Sfp (126.10 μM), 10 μL of propionyl-CoA (10 mM), and 19 μL reaction buffer. After sitting at room temperature overnight, the trypsin procedure was performed and then analyzed on the BEH peptide column.

5) **Butenoyl-DfnJ (DfnJ + Sfp + Butenoyl CoA):** 50 μL of DfnJ (61.4 μM) was incubated with 20 μL Sfp (126.10 μM), 10 μL of butenoyl-CoA (10 mM), and 19 μL reaction buffer. After sitting at room temperature overnight, the trypsin procedure was performed and then analyzed on the BEH peptide column.
6) **HMG-DfnJ (DfnJ + Sfp + HMG CoA):** 50 μL of DfnJ (61.4 μM) was incubated with 20 μL Sfp (126.10μM), 10 μL of HMG-CoA (10 mM), and 19 μL reaction buffer. After sitting at room temperature overnight, the trypsin procedure was performed and then analyzed on the BEH peptide column.

2.15 Sample Preparation for Hypothesized Substrates

To prepare the samples for mass spectrometry, the reaction buffer used was 50 mM Tris-HCl, 12.5 mM $MgCl_2$, 50 mM NaCl (pH 8). Each of the following proteins is an intermediate in the biosynthetic pathway of difficidin.

1) **Malonyl-DfnX:** 50 μL of DfnX (35.97 μM) was incubated with 20 μL Sfp (126.10 μM), 1 μL of CoA (100 mM), and 14 μL reaction buffer. After sitting at room temperature for 1 hour, 15 μL of malonyl-CoA (10 mM) was added with 50 μL of BaeC (38.94 μM) and let react for another hour before starting the trypsin procedure.
2) **Acetyl-DfnX:** 50 μL of DfnX (35.97 μM) was incubated with 20 μL Sfp (126.10 μM), 1 μL of CoA (100 mM), and 14 μL reaction buffer. After sitting at room temperature for 1 hour, 15 μL of malonyl-CoA (10 mM) was added with 50 μL of

BaeC (38.94 μM) and let react for another hour before adding 50 μL BamF (99.27 μM). After reaction for another hour, the trypsin procedure was started.

3) **β-Ketoacyl-DfnJ:** 50 μL of DfnJ (61.4 μM) was incubated with 20 μL Sfp (126.10 μM), 1 μL of CoA (100 mM), and 19 μL reaction buffer. This mixture was incubated at room temperature for 1 hour before adding 10 μL acetoacetyl-CoA (AcAc-CoA) (10 mM) for 15 minutes. Then the mixture was put through a trypsin reaction and analyzed on the BEH peptide column.
4) **HMG-DfnJ:** 50 μL of DfnJ (61.4 μM) was incubated with 20 μL Sfp (126.10 μM), 1 μL of CoA (100 mM), and 19 μL reaction buffer. This mixture was incubated at room temperature for 1 hour before adding 10 μL AcAc-CoA (10 mM) for 15 minutes. In another reaction, 50 μL DfnX (35.97 μM) was incubated with 20 μL Sfp (126.10 μM), 10 μL of acetyl-CoA (10 mM), and 20 μL reaction buffer and incubated at room temperature for 1 hour. The two reactions were then mixed and 50 μL DfnL (76.24 μM) was added and incubated at room temperature for another hour before starting the trypsin procedure.
5) **Glutaconyl-DfnJ:** 50μL of DfnJ (61.4 μM) was incubated with 20μL Sfp (126.10 μM), 10 μL of CoA (100 mM), and 19μL reaction buffer. This mixture was incubated at room temperature for 1 hour before adding 10 μL AcAc-CoA (10 mM) for 15 minutes. In another reaction, 50 μL DfnX (35.97 μM) was incubated with 20 μL Sfp (126.10 μM), 10 μL of acetyl-CoA (10 mM), and 20 μL reaction buffer and incubated at room temperature for 1 hour. The two reactions were then

mixed and 50 μL DfnL (76.24 μM) and 50 μL BaeH (52.56 μM) were added and incubated for another hour before performing the trypsin procedure.

6) **β-Branch:** 50 μL of DfnJ (61.4 μM) was incubated with 20 μL Sfp (126.10 μM), 1 μL of CoA (100 mM), and 19 μL reaction buffer. This mixture was incubated at room temperature for 1 hour before adding 10 μL AcAcCoA (10 mM) for 15 minutes. In another reaction, 50 μL DfnX (35.97 μM) was incubated with 20 μL Sfp (126.10 μM), 10μL of acetyl-CoA (10mM), and 20 μL reaction buffer and incubated at room temperature for 1 hour. The two reactions were then mixed and 50 μL DfnL (76.24 μM), 50 μL BaeH (52.56 μM), and 50 μL DfnM (12.46 μM) was added and incubated for another hour before performing the trypsin procedure.

CHAPTER III

GENE AND PROTEIN ISOLATION

3.1 Target Genes

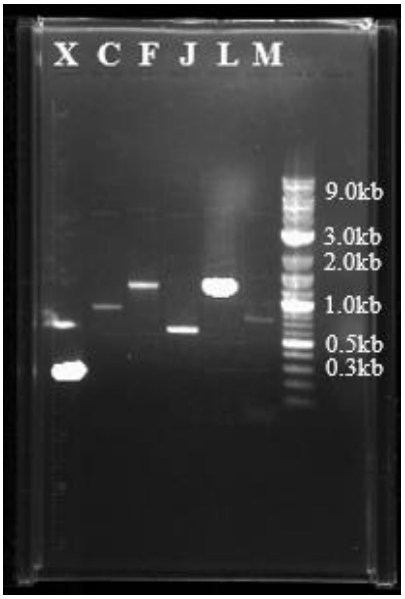
PCR targeting the seven genes of interest (*dfnX, baeC, bamF, dfnJ, dfnL, baeH, dfnM*) plasmids were visualized on agarose gels to see if they displayed the expected bands. After analyzing the intensity of the bands from gel electrophoresis, it was determined that the cloning efficiency was minimal and not showing the desired product. This was because there was too much DNA being added to the pET200 which was causing the cells to not incorporate the gene into the plasmid properly. Since the efficiency of TOPO cloning decreases significantly if the ratio of PCR product to vector is out of the range of a 0.5:1 to 2:1 molar ratio, the optimal ratio of 1:1 was used during transformation which resulted in considerably more colonies that exhibited a correct size band after PCR and gel electrophoresis.

A problem that kept occurring was the gel electrophoresis gave false positive data which was caused from contamination DNA left from the PCR reaction. After several attempts to correct the contamination, DNA sequencing kept coming back with inconclusive data. Some of the DNA was bracketed by the primers and some were littered with fungal DNA. This meant that some of the plasmids had no PCR insert or had completely foreign sequence. After changing several factors in hopes of containing the problem, nothing worked. This prompted re-starting everything from scratch starting with

isolating new genomic DNA and re-making all new solutions and buffers to clone the rest of the genes. To ensure there were no false positives, the 'downstream' primers of the target gene were used along with a primer that targeted with the T7 'upstream' promoter in the pET200 plasmid. Doing this eliminated any false/positive due to contamination by confirming that the gene was in fact present and incorporated into the plasmid.

The only plasmid that required additional work was *bamF* because it would continually show false positives (confirmed through gel electrophoresis and DNA sequencing) even after using fresh materials. After multiple attempts, another strain of *E. coli* was used called "Copy Cutter EPI400" and after only a few tries, the targeted gene (*bamF*) was successfully isolate. This strain is optimal for cloning toxic or unstable DNA because it clones at a reduced plasmid copy number and then induces a high copy number for DNA recovery.

After successfully cloning the targeted genes into pET-200, they were analyzed on gel electrophoresis (Figure 15). The gene *dfnX* consists of 273 bp (base pairs) and the lowest band is showing between 200-300 base pairs. *BaeC* consists of 870 bp and its position is slightly under the 1000 base pair marker. *BamF* consists of 1245 bp and its band is showing up between 1200-1500 base pairs. *DfnJ* consists of 592 bp and its band's position is right under the 600 base pair mark. *DfnL* consists of 1248 bp and its band is showing up in between 1200-1500 base pairs. *DfnM* consists of 747 bp and its position is between 700-800 base pairs. This gel demonstrates each gene was successfully isolated and correctly matches their expected size.



Gene	Gene Size (base pair)
dfnX	273 bp
baeC	870 bp
bamF	1245 bp
dfnJ	592 bp
dfnL	1248 bp
dfnM	747 bp

Figure 15. Agarose Gel Showing Six Genes and their Correlating Sizes Isolated from *Bacillus amyloliquefaciens.* A Standard 2-log DNA ladder was for Analysis.

3.2 Target Proteins

After isolating each of the targeted genes, the purified plasmids were then transformed into BL21 Star™ (DE3) One Shot® Cells. This particular strain of *E. coli* has a truncated RNase E that lacks the C-terminal domain required for mRNA degradation. This is important because it allows for a large amount of mRNA to convey genetic information to ribosomes where they amplify the targeted protein based of the amino acid sequence that it's given. This generates a large quantity of proteins known as 'overexpression'.

A 12% Tris Glycine sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) run at 150 V for 30 minutes validated that the protein bands present are the correct size (seen in Figure 16) and exhibit significant intensity which correlates to a sizable amount of purified protein. DfnX has a mass of 14.13 kDa and its band is

showing up right under 15 kDa. BaeC is 36.47 kDa and its position is slightly under the 40 kDa marker. BamF is 48.48 kDa and its band is showing up slightly above the 40 kDa marker. DfnJ has a mass of 26.04 kDa and its position in the gel is above the 25 kDa maker. DfnL is 50.37 kDa and its band is between 40 kDa and 80 kDa close to the 60 kDa marker. DfnM has a mass of 31.36 kDa and its position is between the 25 kDa and 40 kDa marker. Each of these proteins correlate with their expected size and these SDS-PAGE gels show they were all accurately isolated.

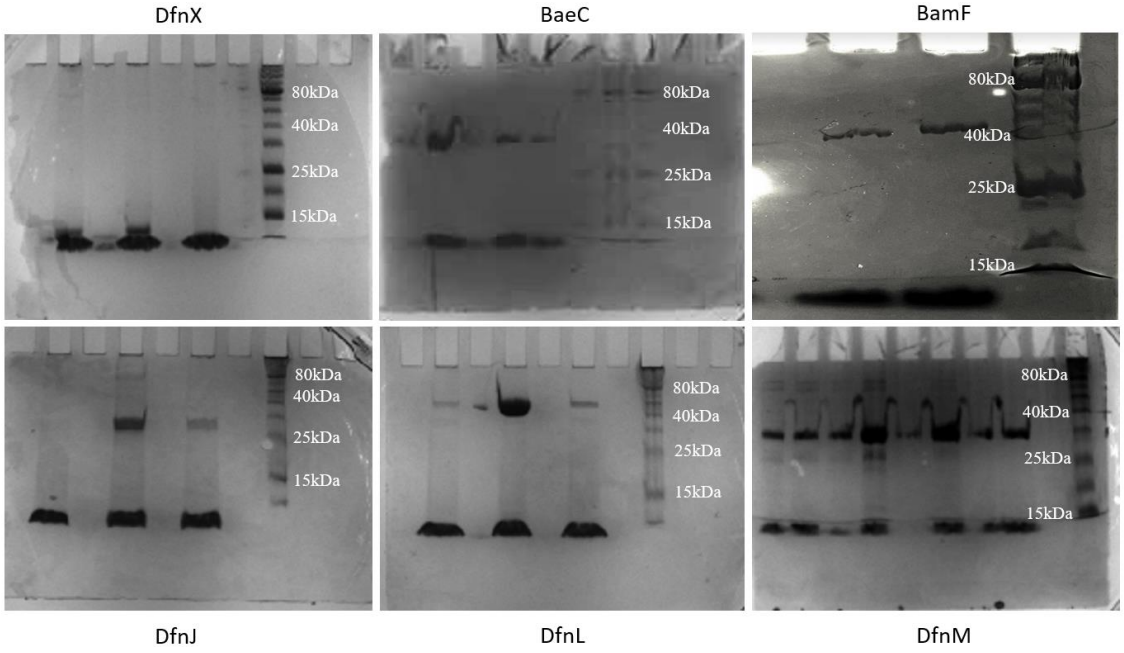


Figure 16. SDS-PAGE Gels Showing Six Overexpressed Proteins from *Bacillus amyloliquefaciens.*
From Left to Right: (Top) DfnX (14.13 kDa), BaeC (36.47 kDa), and BamF (48.48 kDa)
From Left to Right: (Bottom) DfnJ (26.04 kDa), DfnL (50.37 kDa) and DfnM (31.36 kDa)
*The 'Bands' on the Bottom of the Gels are the Dye Front

There are multiple lanes that contain bands, this is because they are the fractions being tested as they are eluting straight off the nickel (Ni-NTA) column (*Note: the 'bands' across the bottom of each gel are the dye front of the loading dye used to load the

proteins on to the gel). Some fractions contain more protein than the others due to the time it takes to elute all of the proteins off the column. The broader the band means the greater the yield and the more overexpression has occurred.

3.3 Bradford Analysis

By changing several factors to help increase protein yield and purity, the targeted proteins were successfully overexpressed and purified. The standard Braford method of 33 μL of protein combined with 1000 μL Bradford reagent revealed that the proteins had substantially high yields between 0.7 mg/mL to over 1.9 mg/mL. The yields of the full 1-liter cultures ended up being between 2.1 mg/L to 6.6 mg/L in total mass.

CHAPTER IV

MASS SPECTROMETRY ANALYSIS

4.1 Phosphopantetheine Assay

A particular assay used to characterize the enzymatic steps carried out on PKSs carrier domains is called the Phosphopantetheine (Ppant) ejection assay and was developed by Pieter Dorrestein[25]. This assay essentially observes the 'prosthetic' arm that binds between the acyl carrier protein (ACP) and the target substrate. The Ppant arm is there to extend the bond between the ACP and substrate for access during polyketide biosynthesis (shown in Figure 17). During MS, the arm 'ejects' the Ppant substrate from the acyl carrier protein and preserving the thioester linkage to the substrate. This allows the mass of these substrates to be readily deducted from the known mass of the corresponding Ppant fragments. This method allows for observation of the biosynthetic intermediates generated by the purified *dfn* proteins.

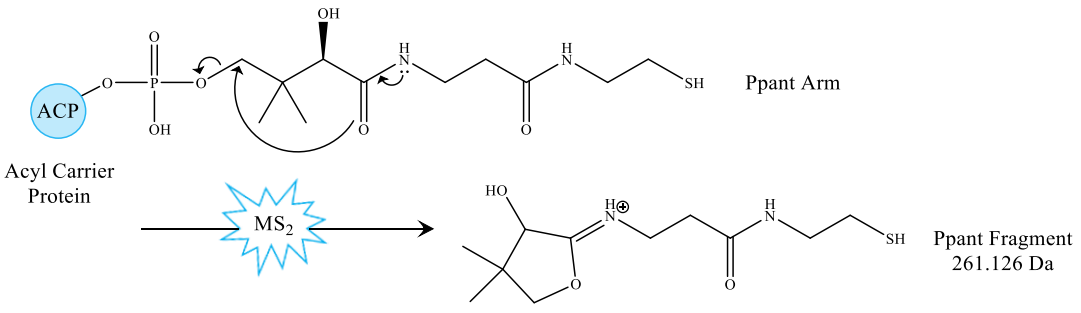


Figure 17. Phosphopantetheine Arm Mechanism. The Phosphopantetheine Arm Attaches to Acyl Carrier Protein and Ejects the MS^2 Ppant Fragment with a Known Mass of 261.126 Da.

4.2 Sfp (4’ Phosphopantetheine Transferase)

Sfp is a 4’ phosphopantetheine transferase (PPTase) and a well-known enzyme that transfers the phosphopantetheine arm from any CoA on to an ACP or PCP (peptidyl carrier protein). Essentially, the CoA [4'-phosphopantetheine] + any apo-acyl (non-active) carrier protein can receive the PPT, which leaves 3',5'-ADP (adenosine diphosphate) and holo-acyl (active) carrier protein (seen in Figure 18). Normal PPTase’s are only activating the ACP/PCP with the phosphopantetheine of CoA-SH whereas Sfp has a high tolerance for activating the ACP/PCP with CoA or any acyl-CoA. This is useful in the lab because Sfp activates the ACPs with an array of CoAs that are used as substrates in the difficin pathway.

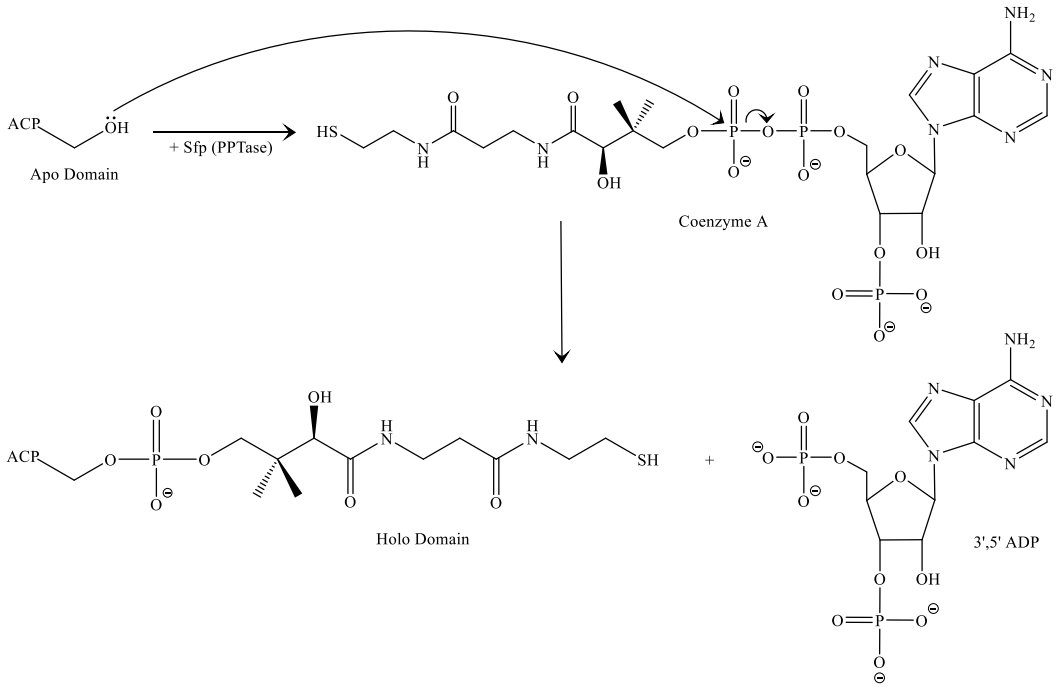


Figure 18. Sfp (4’ Phosphopantetheine Transferase) Mechanism. This Turns Apo (Non-Active) Proteins into their Holo (Active) Form by Adding the PPant Arm to the Acyl Carrier Protein (ACP).

4.3 DfnX Protein

Using the Phosphopantetheine (Ppant) ejection assay, the analysis of the proteins was successful and Dorrestein's method did in fact work for the *dfn* proteins. This was determined by analyzing DfnX with an HPLC Jupiter column on UPLC and analyzing the chromatogram (seen in Figure 19). By adding DfnX, Sfp, and CoA together, holo-DfnX (the active form) was made.

The chromatogram below shows holo-DfnX eluting off the column and providing a nice 'protein envelope' of mass-to-charge data. This data confirmed that the observed mass matched the theoretical mass of 14.16 kDa and holo-DfnX was present in the sample. Not only that, but the MS/MS (MS^2) data from these ions also showed the 261.13 Da Ppant fragment, which indicates that some of the protein Dfnx (ACP) was phosphopantetheinylated by the expression host. This holo-form of DfnX has a thiol group from the phosphopantetheine arm that attacks other CoAs in order to gain the Ppant fragment. This phosphpantetheine arm comes from the 4' phosphopantetheine transferase ability of Sfp. This data corroborates with previous research findings involving the Ppant arm and further supports the LC-MS method for determining the protein's masses.

4.3.1 Holo-DfnX

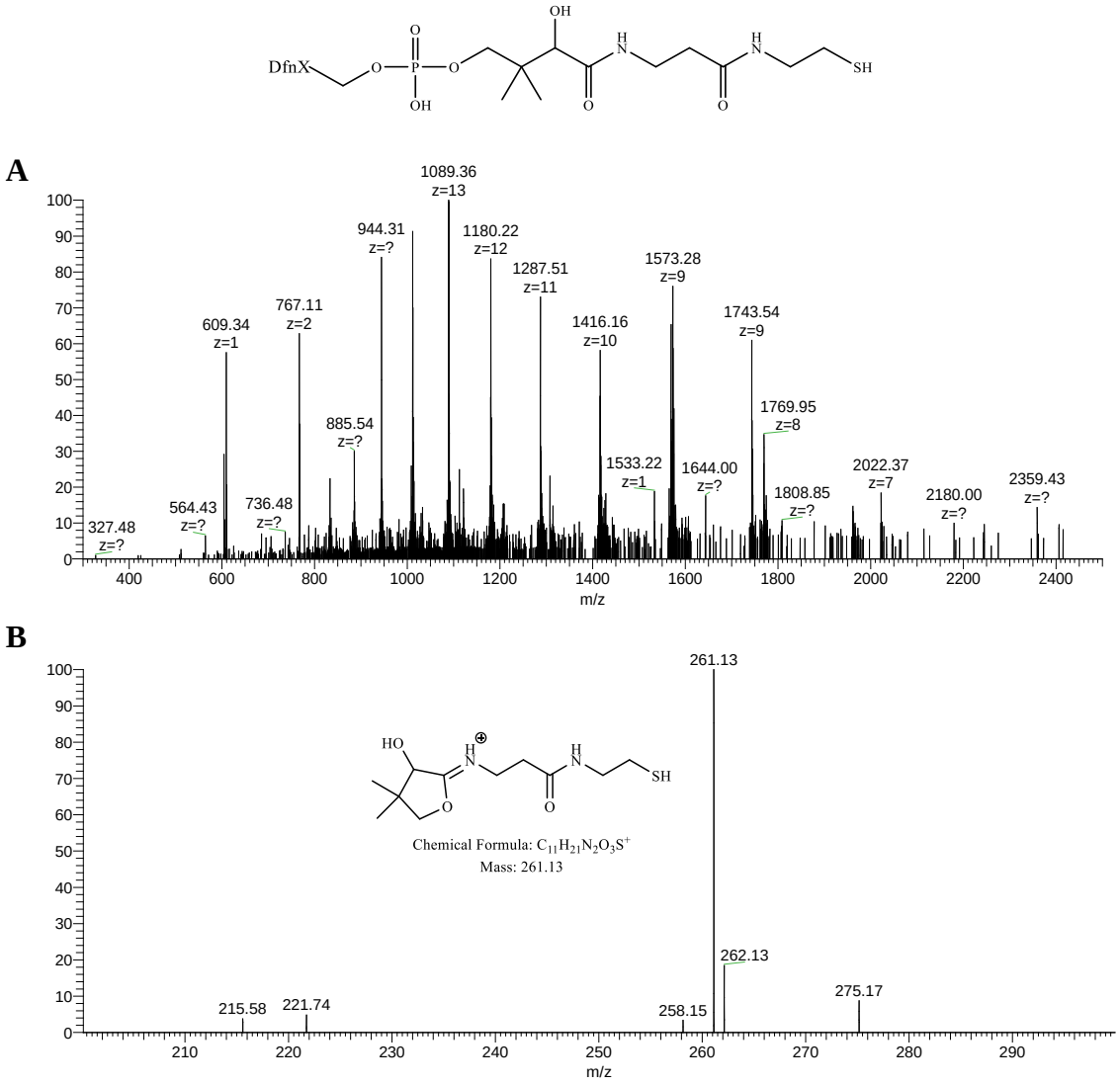


Figure 19. Holo-DfnX Mass Spectrometry Chromatograms

A) Protein Envelope of Holo-DfnX m/z Data. B) MS^2 of the Ppant Fragment of Holo-DfnX.

Holo-DfnX Theoretical Mass: 14.16 kDa

Holo-DfnX Observed Mass: (1089.39*13) = 14.16 kDa

4.4 Sfp Dependent Standards on DfnX

Acyl-ACP generated by Sfp standards are essential because there are no store-generated substrates available to compare in vitro proteins to. By using these standards, there is a baseline analysis of what should be targeted when analyzing chromatograms of the pathway's intermediates.

The Jupiter column was initially used to view the holo-DfnX substrate but after analyzing DfnX's and DfnJ-T2's (double thiolation domain / two acyl carrier proteins) Sfp standards, it was decided that anything larger than the holo-DfnX would be better suited for analysis if the proteins were in smaller fragments (Figure 20). Not only that, but in later steps involving DfnJ there is no choice but to cut the fragments due to the proteins' large size. With that being said, a trypsin digestion was used for every reaction to analyze the targeted masses.

Mass	Peptide Sequence
5265.541	VIEDLGFSSLDIAQLVAQME METGVDPFSQGETISSITTV GSICDIYQK
2074.093	QIIAQIIQDIQEQQSYR
931.4479	AVEAGDDVR
801.3083	YMDSAQS
636.3021	MEQTK
407.1959	EMK

Figure 20. Tryptic DfnX Fragments. DfnX has a Mass of 10,026 Da and the Phosphopantetheine Arm can be Found Attached to a Serine in the 5265.541 Fragment. These Masses are all +1 Charge.

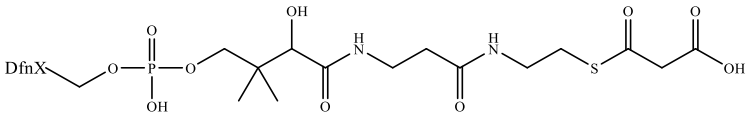
The first standard observed through MS post trypsin digestion was malonyl-DfnX and it mimics the first intermediate in the difficidin pathway. Malonyl-ACP was generated using DfnX, Sfp, and malonyl-CoA. Since the phosphopantetheine (PPT) arm is found on the 5265.54 fragment, the targeted tryptic mass has a +4 m/z charge fragment

consisting of ((5265.54 DfnX) + (427.1 malonyl-CoA PPant arm) + (4 hydrogen)) / (4 charge) giving an expected fragment of 1424.16 and a MS^2 of 347.13 for the malonyl-PPT ejection. The observed tryptic mass of this standard was 1424.42 with a MS^2 phosphopantetheine peak of 347.13 verifying the formation the first standard, malonyl-ACP.

The second standard to be analyzed by trypsin digestion was acetyl-DfnX. This was generated using DfnX, Sfp, and acetyl-CoA and is the second intermediate in the difficidin pathway. The expected tryptic mass with a +4 m/z charge was ((5265.54 DfnX) + (383.1 acetyl-CoA PPant arm) + (4 hydrogen)) / (4 charge) leaving a fragment of 1413.16. The expected MS^2 of the ejected acetyl-ACP phosphopantetheine fragment is 303.14. The tryptic mass observed in vitro was 1413.16 and its MS^2 verified the 303.14 peak conforming the formation the second standard, acetyl-ACP.

Both of these standards are products in the bacillaene/difficidin pathway and demonstrated that this methodology was effective using the UNCG facility.

4.4.1 Malonyl-DfnX Standard



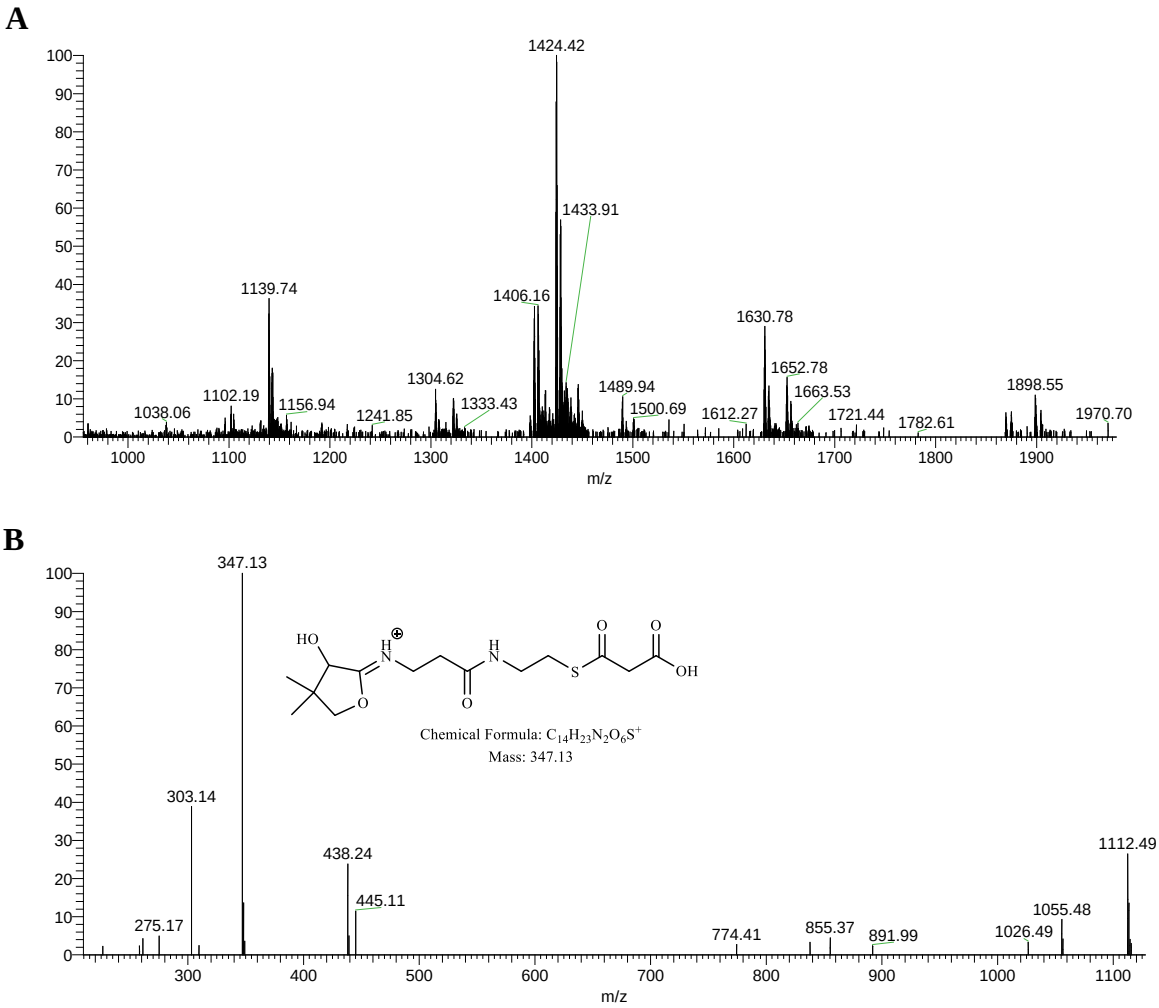


Figure 21. Malonyl-DfnX Standard Mass Spectrometry Chromatograms
A) Tryptic Fragment of Malonyl-DfnX. B) MS^2 of the Ppant Fragment of Malonyl-DfnX.
Tryptic Malonyl-DfnX Theoretical Mass: 1424.16 ((5,692.64+4)/4))
Tryptic Malonyl-DfnX Observed Mass: 1424.42 (+4 m/z)

4.4.2 Acetyl-DfnX Standard

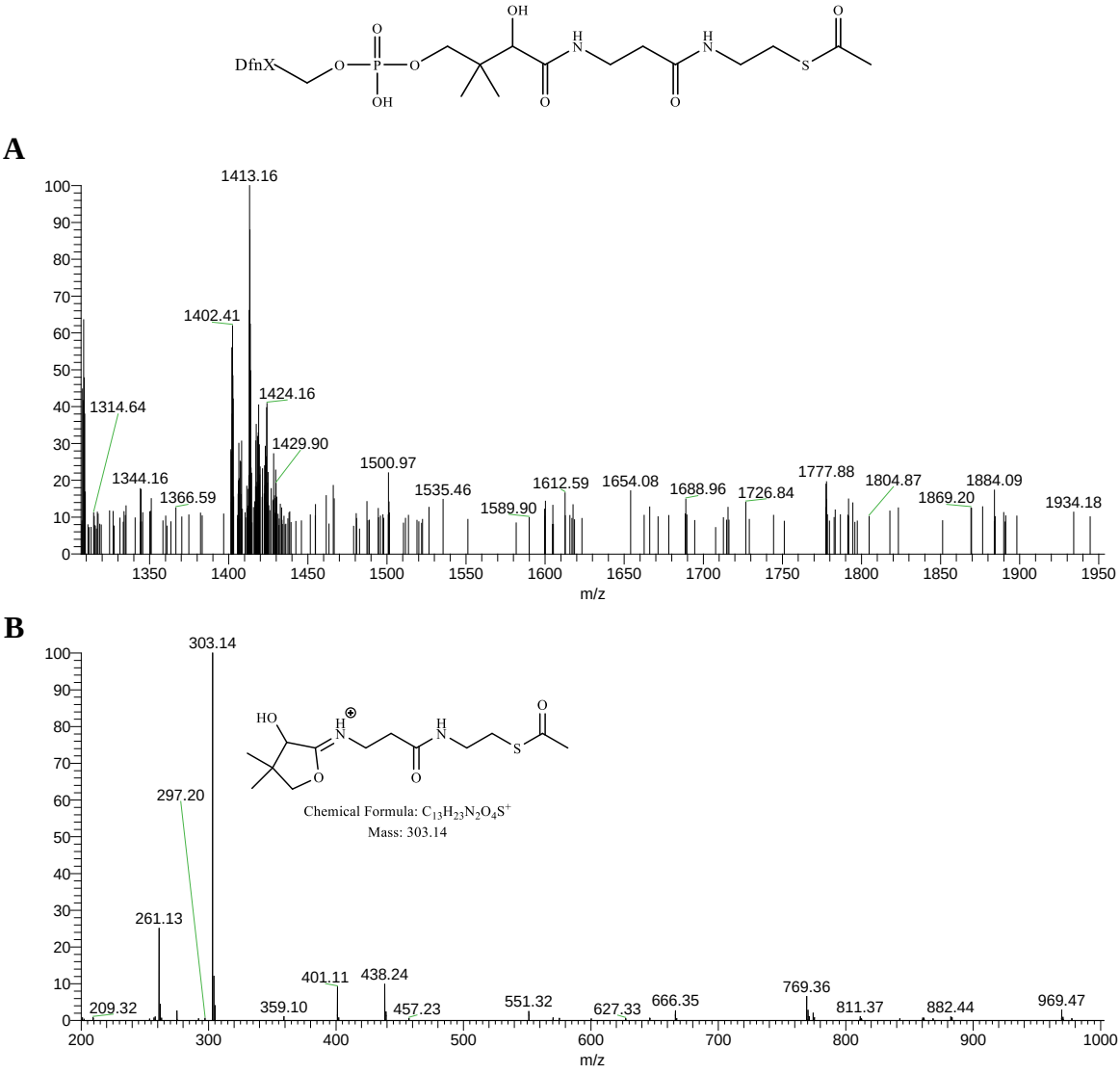


Figure 22. Acetyl-DfnX Standard Mass Spectrometry Chromatogram
A) Tryptic Fragment of Acetyl-DfnX. B) MS^2 of the Ppant Fragment of Acetyl-DfnX.
Tryptic Acetyl-DfnX Theoretical Mass: 1413.16 ((5,648.64+4)/4))
Tryptic Acetyl-DfnX Observed Mass: 1413.42 (+4 m/z)

4.5 Sfp Dependent Standards on DfnJ

The Jupiter column proved to be ineffective in the later steps involving DfnJ-T2 due to the larger size of the DfnJ-T2 di-domain, which is what prompted the trypsin digestion. The trypsin-digested DfnJ-T2 yielded smaller peptide fragments (Figure 23) which were more easily separated on an Acquity UPLC Peptide BEH C18 column. This allowed for the smaller fragments to reach the gas phase necessary for MS detection more readily.

Mass	Peptide Sequence
4443.232	ESLADILFLKPEDIDEHEAF IEMGLD**S**IIGVEWVQSINK
4332.188	ISLADILFLSPEDIDADEPF IDMGLD**S**IIGVEWIQTVNK
3438.739	KPEYAEAEETAVRPQIPEAE TPLAGISDGTLR
2980.519	TYQASITANLVYEYPTIATL AGYLTGSV
2277.967	GSHHHHHHGMASMTGGQQMG R
2203.128	VYEHPTLEELAEYIAGQIK
1943.917	DHPFTLDCTVSVQQPEK
1177.585	TYHTEITANK
998.3949	DLYDDDDK
873.504	TLLEELR
790.3689	AVSEDNR
389.2394	ELK
306.1594	MR
147.1128	K

Figure 23. Tryptic DfnJ-T2 Fragments. DfnJ-T2 has a Mass of 26,070 Da and the Phosphopantetheine Arm can be Found Attached to the Serine Both the 4443.232 Fragment and 4332.188 Fragment. The Conserved Sequence for the Phosphopantetheine Arm is D**S**xxxxxW. These Masses are all +1 Charge.

By investigating the mass of each molecule, the proteins were analyzed to determine if they had carried out their expected function. This method was used for determining DfnJ/Sfp dependent standards for comparison to the expected m/z data. Since DfnJ-T2 is a double ACP (two acyl carrier proteins), there will be two fragments

that carry the phosphopantetheine arm, each with its own mass. This di-domain can be visualized through MS with two different expected peaks with +4 m/z charges attached to the 4443.23 fragment (#1) and the 4332.19 fragment (#2).

The first standard analyzed on the di-domain of DfnJ was propionyl-DfnJ. This is not an intermediate in the pathway, it was just a substrate used to confirm the LC-MS method for DfnJ would work as accurately as it did with the single ACP of DfnX. This standard is generated using DfnJ, Sfp, and propionyl-CoA. Since both the 4443.23 fragment 1 and the 4332.19 fragment 2 carry the Ppant fragment, there are two expected peaks of 1197.34 and 1183.34 with a MS^2 PPT ejection peak of 317.15. These are generated as such; ((4785.36 propionyl-DfnJ fragment 1) + (4 hydrogen)) / (4 charge) gives a mass of 1197.34, and ((4729.36 propionyl-DfnJ fragment 2) + (4 hydrogen)) / (4 charge) gives a mass of 1183.34. The observed fragments were 1197.33 and 1183.34 with a MS2 of 317.15 which confirms both ACPs are carrying acetyl as well as the Ppant fragment.

Butenoyl-DfnJ was the second di-domain standard tested, again just to check the LC-MS on DfnJ. Butenoyl-DfnJ was made using DfnJ, Sfp, and butenoyl-CoA and its two expected peaks are 1214.10 and 1186.33 with a MS^2 peak of 329.15. These are generated as such; ((4852.40 butenoyl-DfnJ fragment 1) + (4 hydrogen)) / (4 charge) gives a mass of 1214.10, and ((4741.32 butenoyl-DfnJ fragment 2) + (4 hydrogen)) / (4 charge) gives a mass of 1186.33. The observed peaks were 1214.34 and 1186.33 with a Ppant fragment of 329.15 which confirms this standard was accurately achieved.

4.5.1 Propionyl-DfnJ Standard

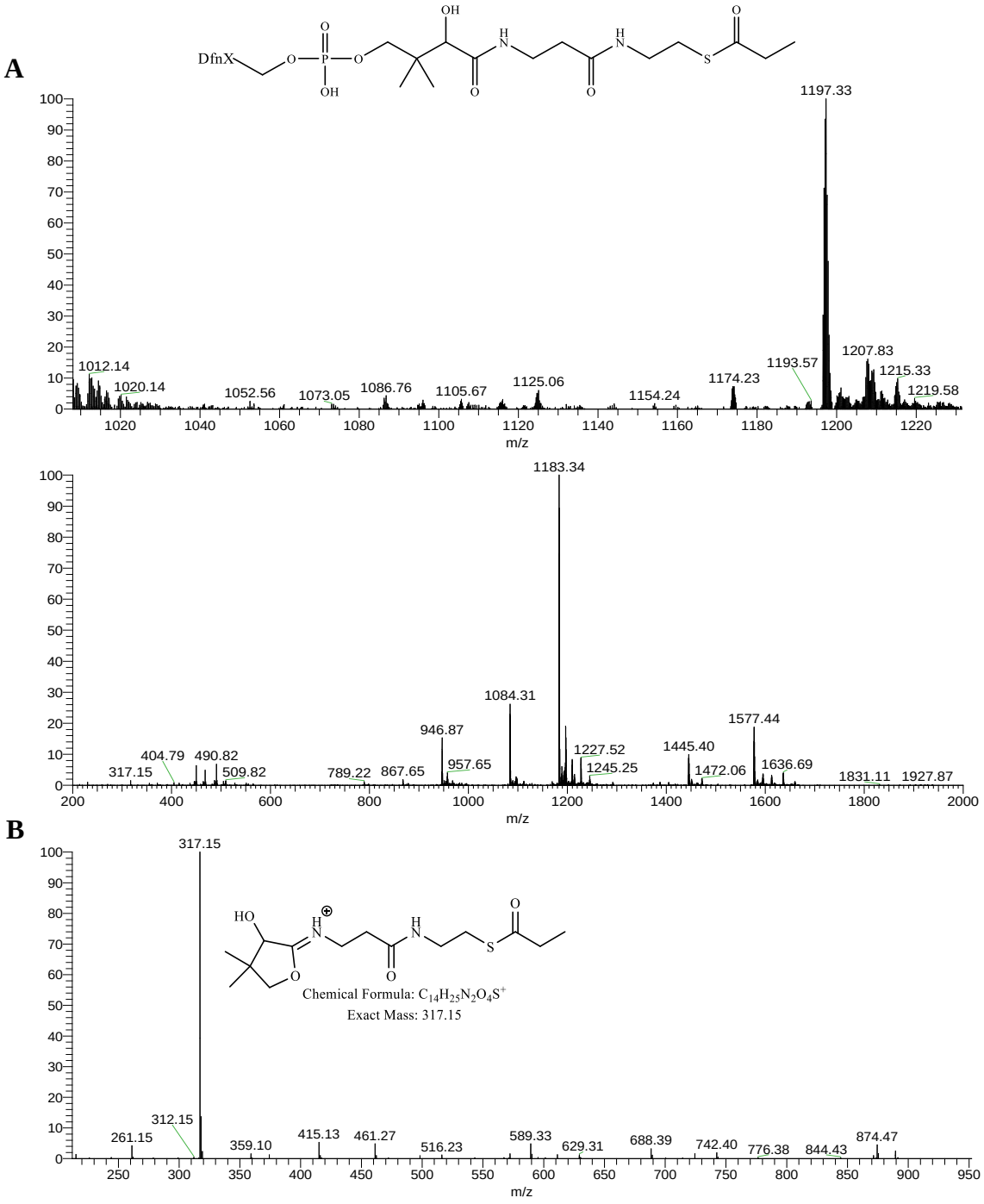


Figure 24. Propionyl-DfnJ Standard Mass Spectrometry Chromatograms
A) Tryptic Fragments of Propionyl-DfnJ. B) MS^2 of the Ppant Fragment of Propionyl-DfnJ.
(+4 m/z) Tryptic Propionyl-DfnJ Fragment 1: Theoretical Mass: 1197.34 ; Observed Mass: 1197.33
(+4 m/z) Tryptic Propionyl-DfnJ Fragment 2: Theoretical Mass: 1183.34 ; Observed Mass: 1183.34

4.5.2 Butenoyl-DfnJ Standard

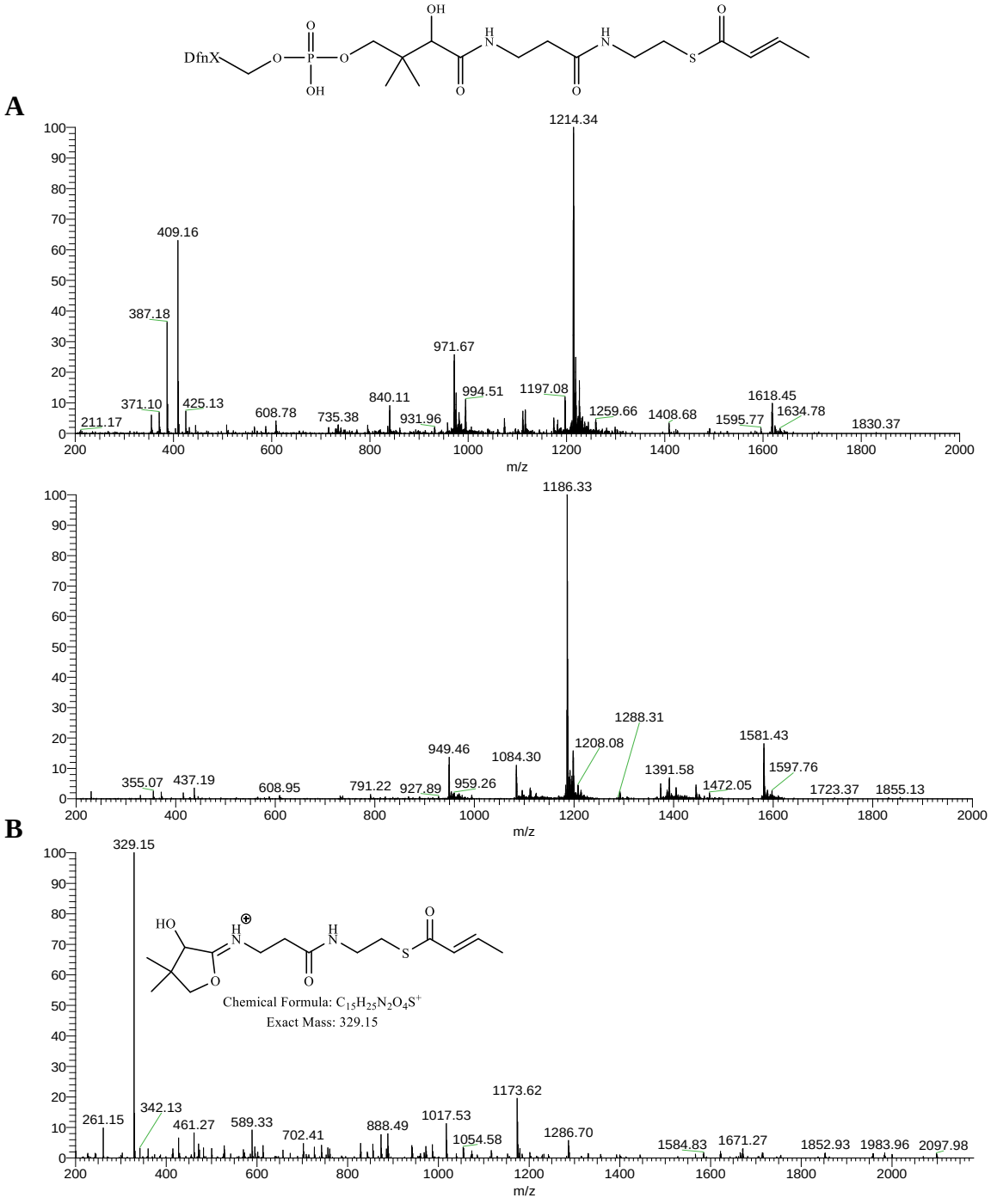


Figure 25. Butenoyl-DfnJ Standard Mass Spectrometry Chromatograms
A) Tryptic Fragments of Butenoyl-DfnJ. B) MS^2 of the Ppant Fragment of Butenoyl-DfnJ.
(+4 m/z) Tryptic Butenoyl-DfnJ Fragment 1: Theoretical Mass: 1214.10 ; Observed Mass: 1214.34
(+4 m/z) Tryptic Butenoyl-DfnJ Fragment 2: Theoretical Mass: 1186.33 ; Observed Mass: 1186.33

The last standard generated was HMG-DfnJ and it is a key intermediate in the β-branch synthesis and is vital to be able to compare to the in vitro synthesized HMG-DfnJ. This standard was created using DfnJ, Sfp, and HMG-CoA and has two expected peaks of 1233.09 and 1205.33 with a MS^2 peak of 405.17. These are generated as such; ((4928.36 HMG-DfnJ fragment 1) + (4 hydrogen)) / (4 charge) gives a mass of 1233.09, and ((4817.32 HMG-DfnJ fragment 2) + (4 hydrogen)) / (4 charge) gives a mass of 1205.33. The observed fragments matched up perfectly confirming the 1233.09 fragment as well as the 1205.33 with both peaks carrying the MS^2 405.17 Ppant arm. This confirms that the artificial HMG-DfnJ standard was created.

4.5.3 HMG-DfnJ Standard

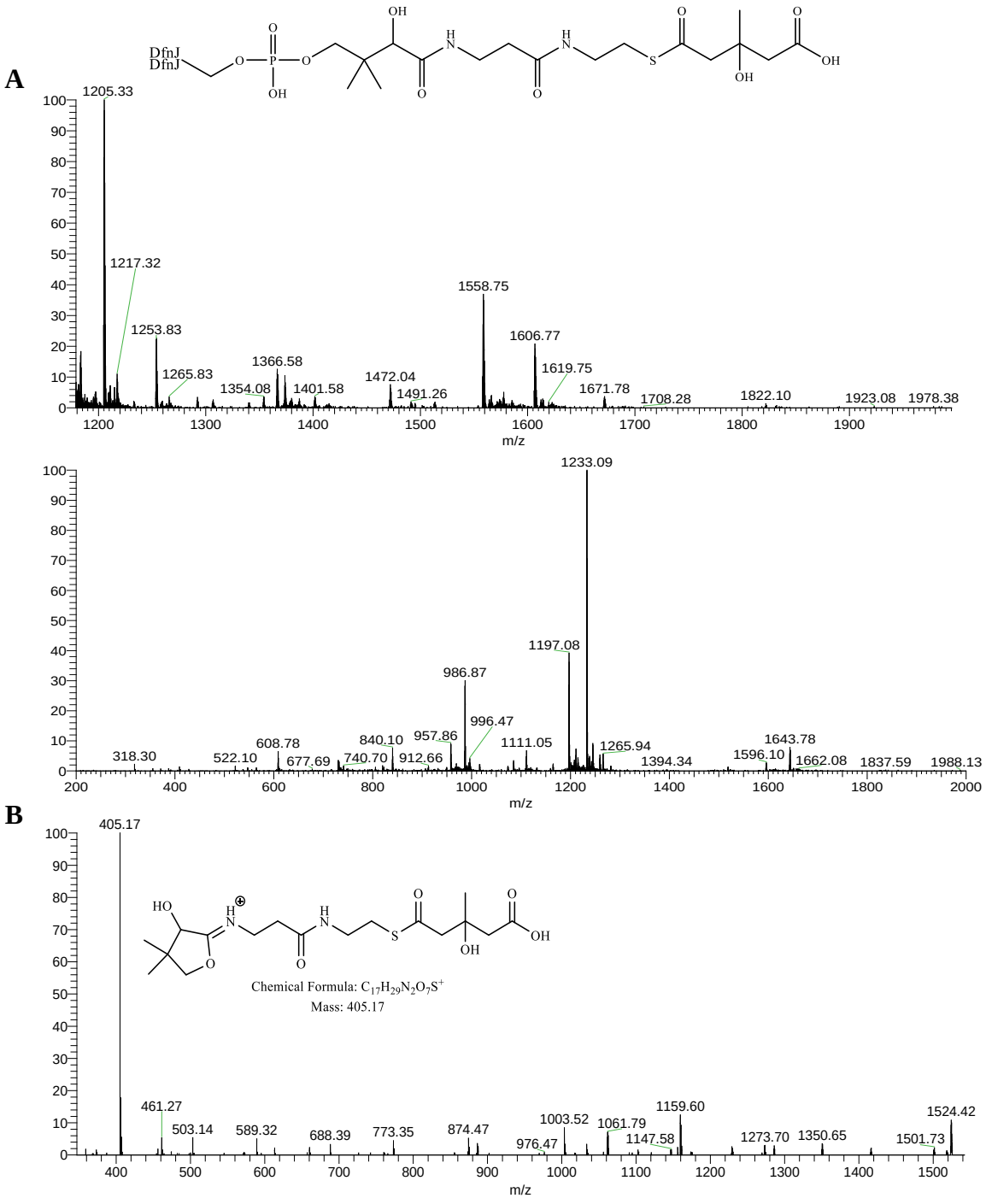


Figure 26. HMG-DfnJ Standard Mass Spectrometry Chromatograms

A) Tryptic Fragments of HMG-DfnJ. B) MS^2 of the Ppant Fragment of HMG-DfnJ.

(+4 m/z) Tryptic HMG-DfnJ Fragment 1: Theoretical Mass: 1205.33 ; Observed Mass: 1205.33

(+4 m/z) Tryptic HMG-DfnJ Fragment 2: Theoretical Mass: 1233.09 ; Observed Mass: 1233.09

CHAPTER V

RESULTS AND DISCUSSION

5.1 Trypsin Reactions of Intermediates in the Difficidin Pathway

In the upcomping chromatograms, **A** will indicate the +4 charged trypsin digested frament(s) coming off of the BEH peptide column while **B** will show the MS^2 of a selected ion where the phosphopantetheine fragment can be seen. Each intermediate of the difficidin pathway (Figure 27) was anaylzed and the chromatagrams indicate they were all successfully constructed.

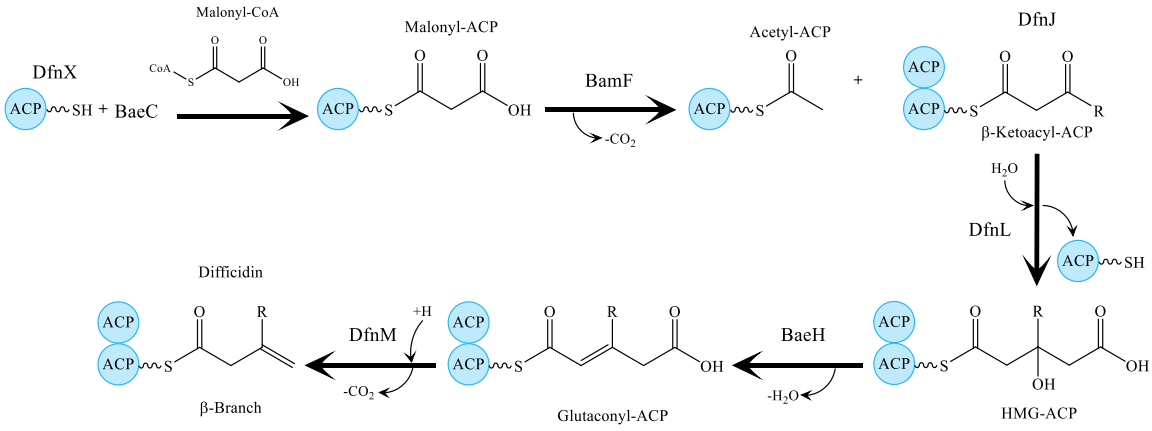


Figure 27. The Biosynthetic Pathway of Each Intermediate Involved in the Formation of Difficidin and its β-Branch.

The first hypothesized intermediate in the difficidin pathway is malonyl-DfnX. This starts with the formation of holo-DfnX via Sfp and CoA, then an addition of BaeC fuses the malonyl group to the thiol group on holo-DfnX leaving CoASH. Since DfnX

only has one acyl group, the Ppant fragment will be attached to only one of its tryptic fragments, specifically the 5265.54 fragment. The expected mass of the malonyl-DfnX intermediate is ((5692.64 malonyl-DfnX) + (4 hydrogen)) / (4 charge) resulting in a 1424.16 fragment and a MS^2 of 347.13. The observed tryptic mass of this intermediate was 1424.42 which matches up with the malonyl-ACP standard with a MS^2 phosphopantetheine peak of 347.13. This confirms the first intermediate in the pathway, malonyl-DfnX, was generated.

The second intermediate thought to be generated is acetyl-DfnX. This molecule which was generated using the previously created malonyl-ACP (holo-DfnX, malonyl-CoA, BaeC) and adding the protein BamF. The hypothesis is that BamF decarboxylates malonyl-ACP and acetyl-ACP is formed. The expected tryptic mass of acetyl-DfnX with a +4 m/z charge is ((5648.64 acetyl-DfnX) + (4 hydrogen)) / (4 charge) leaving a fragment of 1413.16 with a MS^2 of 303.14. The tryptic mass observed was 1413.16 which also matched the previously created standard and its MS^2 verified the 303.14 peak confirming the formation of the second intermediate, acetyl-ACP.

5.1.1 Malonyl-DfnX Intermediate

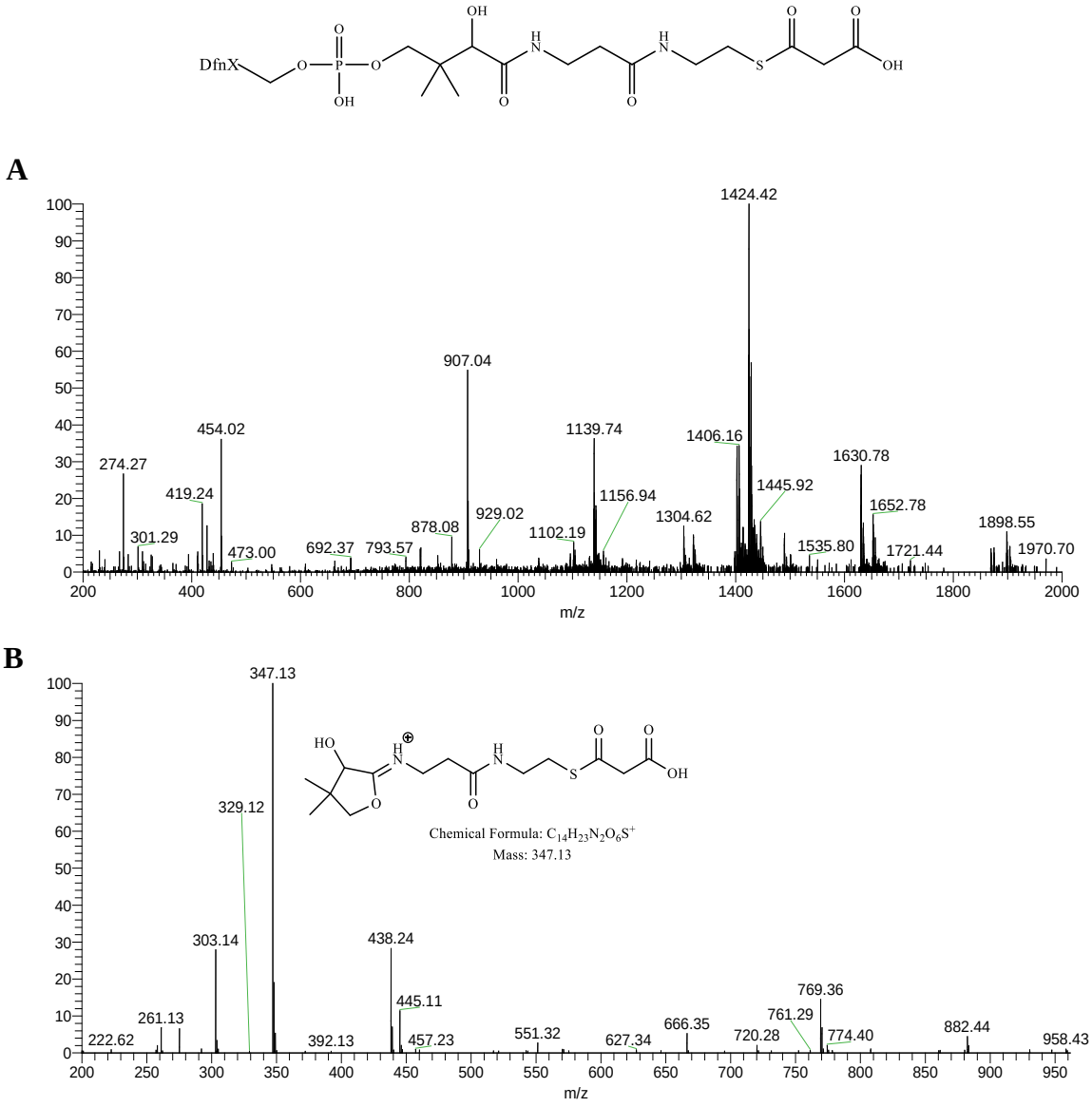


Figure 28. Malonyl-DfnX Intermediate Mass Spectrometry Chromatograms
A) Tryptic Fragment of Malonyl-DfnX. B) MS^2 of the Ppant Fragment of Malonyl-DfnX.
Tryptic Malonyl-DfnX Theoretical Mass: 1424.16
Tryptic Malonyl-DfnX Observed Mass: 1424.42 (+4 m/z)

5.1.2 Acetyl-DfnX Intermediate

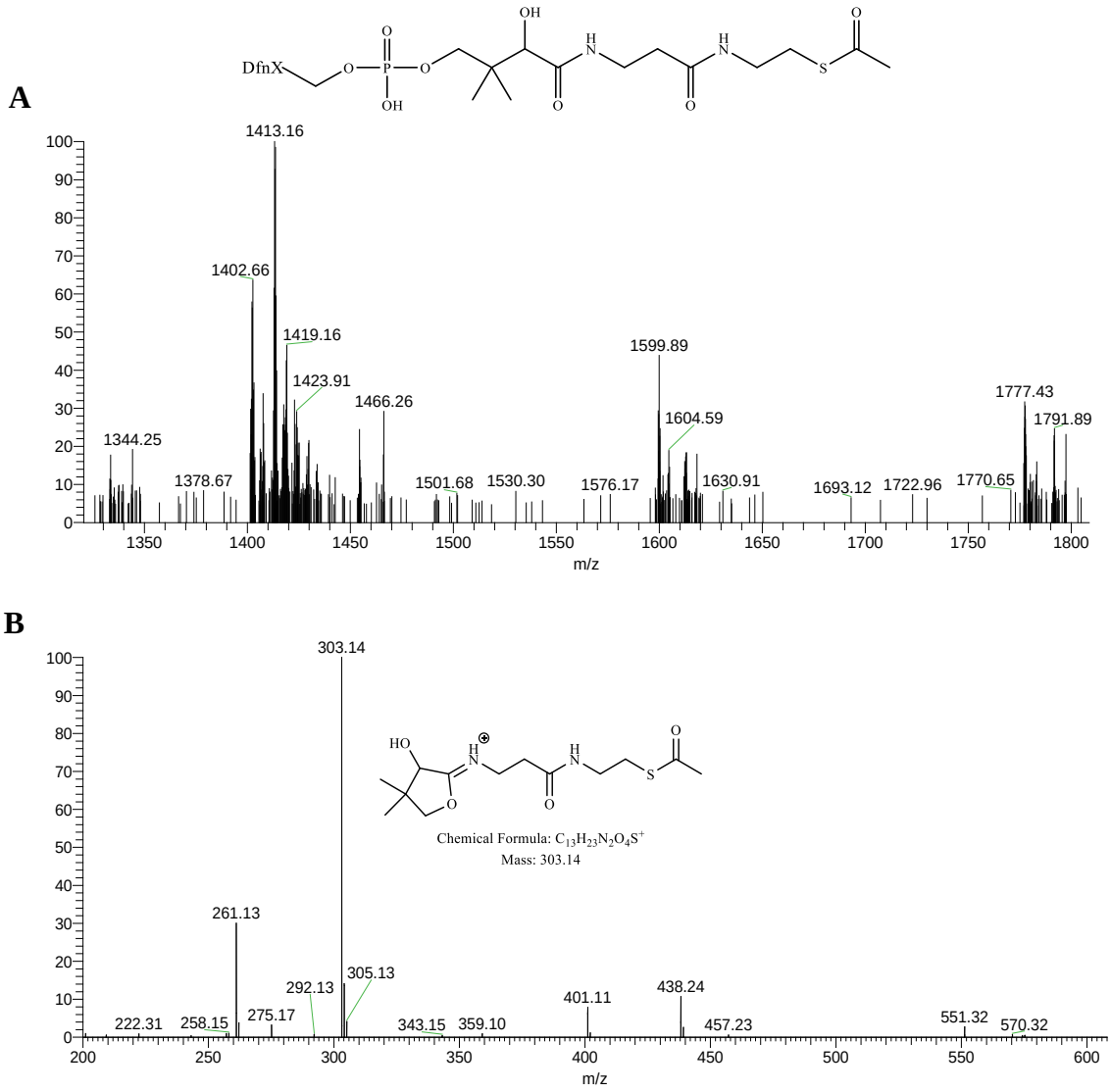


Figure 29. Acetyl-DfnX Intermediate Mass Spectrometry Chromatograms
A) Tryptic Fragment of Acetyl-DfnX. B) MS^2 of the Ppant Fragment of Acetyl-DfnX.
Tryptic Acetyl-DfnX Theoretical Mass: 1413.16
Tryptic Acetyl-DfnX Observed Mass: 1413.42 (+4 m/z)

The β-ketoacyl-DfnJ intermediate is the formation of the AcAc-DfnJ-T2 di-domain. In diffidicin's biosynthesis, the 'R' chain is much longer on this β-ketoacyl-ACP, however the β-carbon chemistry is happening close to the T2 domain. Therefore, a shortened variant, acetoacetyl-CoA (AcAc) has been used in its place to generate the β-ketoacyl-DfnJ which would otherwise be difficult to prepare.

Originally, the thought was that the apo (not-active) form of DfnJ was attacking AcAc-CoA through Sfp and adding on the shortened variant to the protein. After numerous tries on the LC-MS using trypsin, this desired product was not created. Unfortunately, the expected masses for AcAc-DfnJ were not observed. Instead, the observed m/z signals corresponded with acetyl-CoA attaching to DfnJ. These first reactions were done using Tris-HCl buffers. The pH's of these buffers started at a pH of 7.5 and when it did not work, a variety of pH's varying from pH 8 - pH 6.5 were tested. Still yielding no results, a HEPES buffer was used in place of Tris-HCl. Neither buffer yielded the desired β-ketoacyl-DfnJ. No matter what was being tested, the AcAc was being cut into acetyl and attaching to the DfnJ.

This prompted trials using AcAc and DfnX (single acyl carrier protein) to see if possibly AcAc-DfnX could be generated. Still, using both Tris and HEPES buffers with varying pH's, the AcAc-DfnX would not form. Again, the chromatograms were showing acetyl-DfnX indicating the AcAc was still being affected. Holding off on the trypsin to see if maybe that method was generating the acetyl fragment, the AcAc-DfnX mixture was used intact in the Jupiter column as well as directly infused on to the mass spec and the results were still inconclusive.

The next thought was that the proteins were somehow affecting the AcAc-CoA before it had a chance to react. This led to trichloroacetic acid (TCA) trials using each protein and testing it individually with plain AcAc-CoA. After mixing the proteins and AcAc and incubating for an hour, TCA was added to precipitate the proteins. The remaining solution was tested on the LC-MS to determine the protein's detriment to the AcAc-CoA. Surpisingly, the proteins did have some affect on the AcAc causing the amount in the sample to lower, but not enough to impair the reaction.

With nothing working, the next step was to re-create the exact same conditions that Walsh[23] used to generate the *B. subtilis* bacillaene version of β-ketoacyl-ACP. Even after changing the conditions to perfectly match his method, the β-ketoacyl-DfnJ still wasn't being generated.

Searching through literature yielded a paper by Pieter Dorrestein[26] who ran into the same issue described above, so they used a different approach to get the β-ketoacyl-ACP to develop. His research shows the apo (non-active) form of the ACP generating the holo form (active) by Sfp and CoA first before the AcAc-CoA is added. Once the holo form of the ACP was generated, AcAc was added for 15 minutes (anything less or more yielded less product) and then immediately put into the trypsin reaction where it was incubated for 1 hour. With this method, (Figure 30) Dorrestein was able to generate β-ketoacyl-ACP.

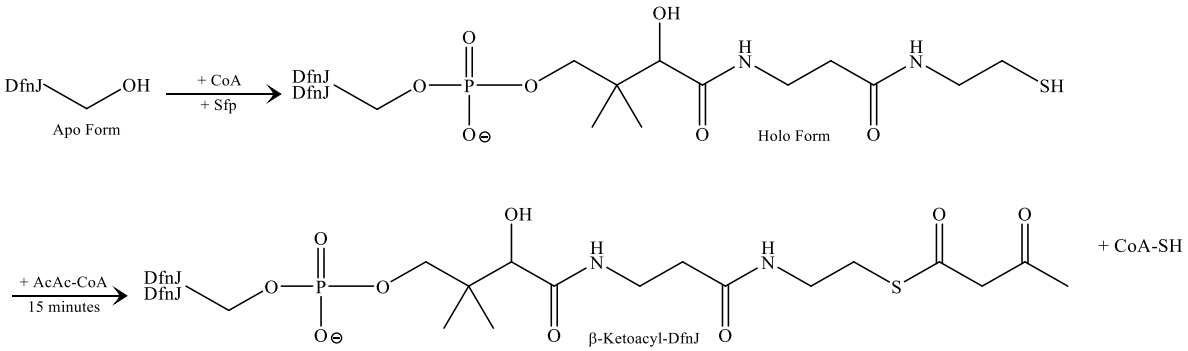


Figure 30. Generating the β-Ketoacyl-DfnJ Intermediate[26].

Following his procedure provided the missing link to generating the β-ketoacyl-DfnJ needed for the difficidin pathway. This β-ketoacyl-DfnJ di-domain can be visualized through MS with two expected peaks of 1190.34 and 1218.10 with a MS^2 peak of 345.17. These two peaks are generated as such; ((4868.39 β-ketoacyl-DfnJ fragment 1) + (4 hydrogen)) / (4 charge) gives a mass of 1218.10, and ((4757.35 β-ketoacyl-DfnJ fragment 2) + (4 hydrogen)) / (4 charge) gives a mass of 1190.34. The observed fragments following the Dorrestein procedure yielded a 1218.09 fragment and a 1190.28 fragment with both peaks carrying the MS^2 345.15 Ppant arm. This verifies the generation of the β-ketoacyl-DfnJ intermediate.

5.1.3 β-Ketoacyl-DfnJ Intermediate

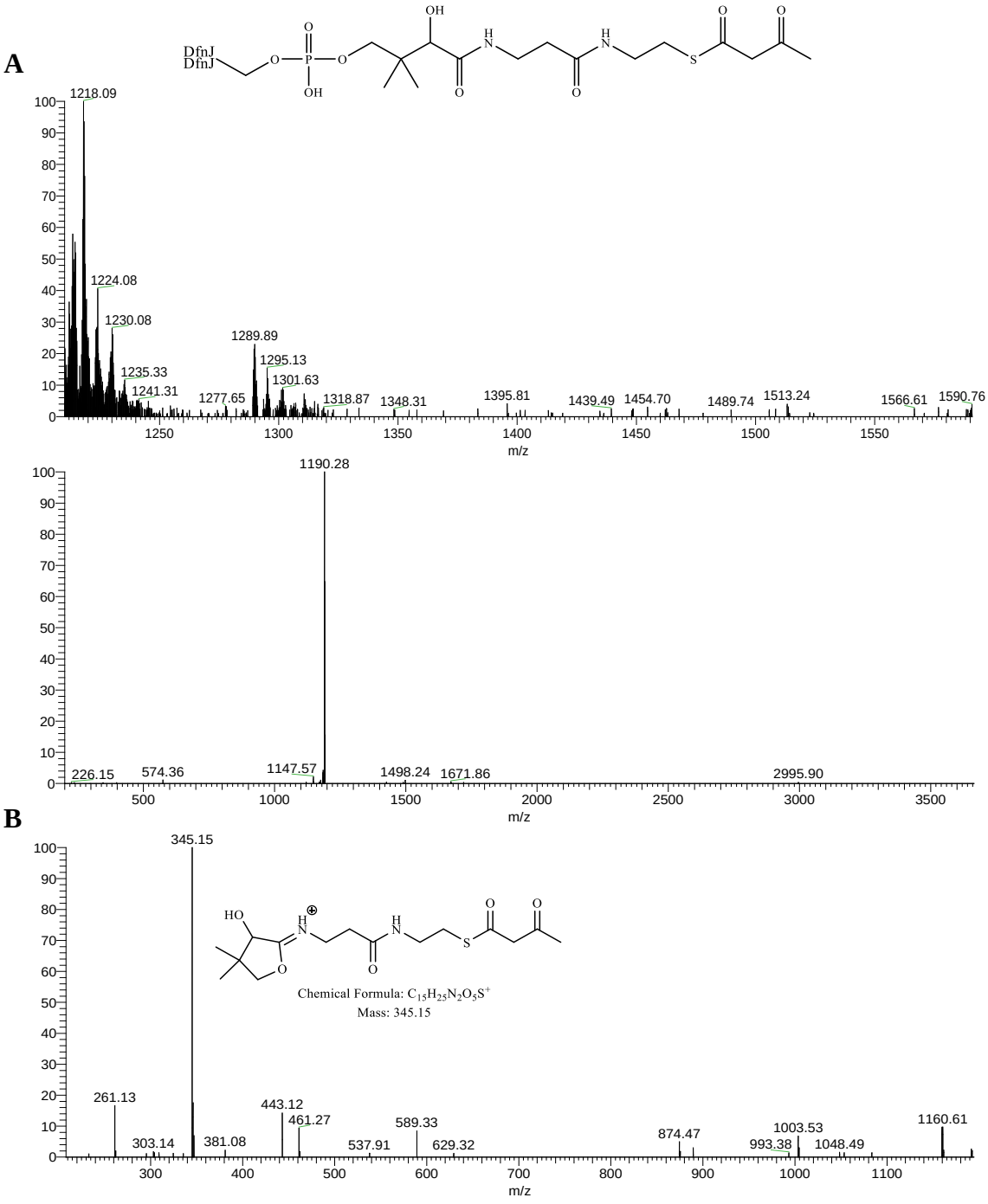


Figure 31. β-Ketoacyl-DfnJ Intermediate Mass Spectrometry Chromatograms

A) Tryptic Fragments of β-Ketoacyl-DfnJ. B) MS^2 of the Ppant Fragment of β-Ketoacyl-DfnJ

(+4 m/z) Tryptic β-Ketoacyl-DfnJ Fragment 1: Theoretical Mass: 1218.10 ; Observed Mass: 1218.09

(+4 m/z) Tryptic β-Ketoacyl-DfnJ Fragment 2: Theoretical Mass: 1190.34 ; Observed Mass: 1190.28

The HMG-DfnJ intermediate is expected to be generated through the combination of DfnJ-T2 di-domain (β-ketoacyl-DfnJ) and acetyl-DfnX through the addition DfnL and releasing the ACP (DfnX). This reaction is responsible for adding the extra carbon needed for the β-branch and the reaction is shown in Figure 32.

The two expected tryptic peaks are of 1233.09 and 1205.09 with a MS^2 peak of 405.17. These are generated as such; ((4928.36 HMG-DfnJ fragment 1) + (4 hydrogen)) / (4 charge) gives a mass of 1233.09, and ((4817.32 HMG-DfnJ fragment 2) + (4 hydrogen)) / (4 charge) gives a mass of 1205.33. The observed fragments matched up perfectly with the expected products and also match the HMG-DfnJ standard made using HMG-CoA and Sfp. The analysis showed a 1233.09 fragment and a 1205.33 fragment with both peaks carrying the MS^2 405.17 Ppant arm confirming the formation of the HMG-DfnJ intermediate.

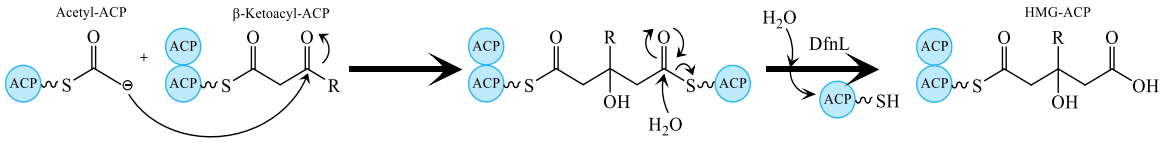


Figure 32. The Formation of HMG-DfnJ.

5.1.4 HMG-DfnJ Intermediate

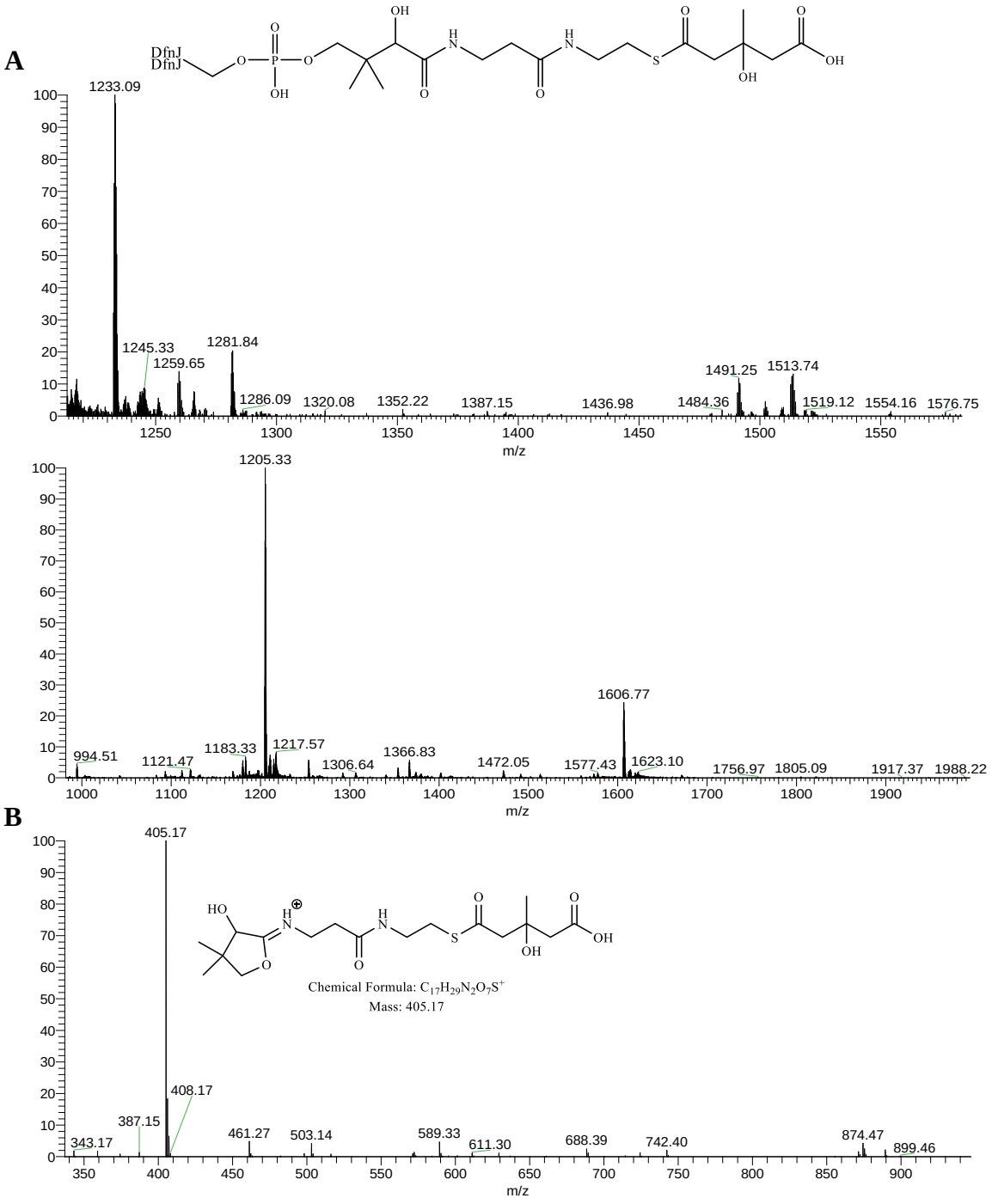


Figure 33. HMG-DfnJ Intermediate Mass Spectrometry Chromatograms
A) Tryptic Fragments of HMG-DfnJ. B) MS^2 of the Ppant Fragment of HMG-DfnJ.
(+4 m/z) Tryptic HMG-DfnJ Fragment 1: Theoretical Mass: 1205.33 ; Observed Mass: 1205.33
(+4 m/z) Tryptic HMG-DfnJ Fragment 2: Theoretical Mass: 1233.09 ; Observed Mass: 1233.09

The next expected intermediate in the pathway is glutaconyl-DfnJ which is formed when BaeH dehydrates the previous intermediate HMG-DfnJ. The original hypothesis was that DfnM (which had no homologs in the PksX cluster) was dehydrating and decarboxylating HMG-DfnJ in one concerted step. After generating HMG-DfnJ, DfnM was added to the mixture to see if the next intermediate could be generated. After several failed attemps at generating glutaconyl-DfnJ, the thought was that the *pks3 (dif)* operon in *B. amyloliquefaciens* was 'borrowing' a dehydrating enzyme from the *pks1 (bae)* operon (operons can be seen in Figure 7). After isolating BaeH from the *pks1 (bae)* operon, it was added to the HMG-DfnJ intermediate mixture and analysis showed the generation of glutaconyl-DfnJ.

Glutaconyl-DfnJ has two expected peptide masses of 1228.60 and 1200.84 with a MS^2 peak of 387.16. These are generated as such; ((4910.40 glutaconyl-DfnJ fragment 1) + (4 hydrogen)) / (4 charge) gives a mass of 1228.60, and ((4799.36 glutaconyl-DfnJ fragment 2) + (4 hydrogen)) / (4 charge) gives a mass of 1200.84. The observed fragments from the enzyme reaction involving the recruited BaeH matched up accurately with the expected values of 1228.60 and 1200.84, as well as the 405.17 Ppant fragment which means that BaeH successfully dehydrated HMG-DfnJ to create glutaconyl-DfnJ intermediate.

5.1.5. Glutaconyl-DfnJ Intermediate

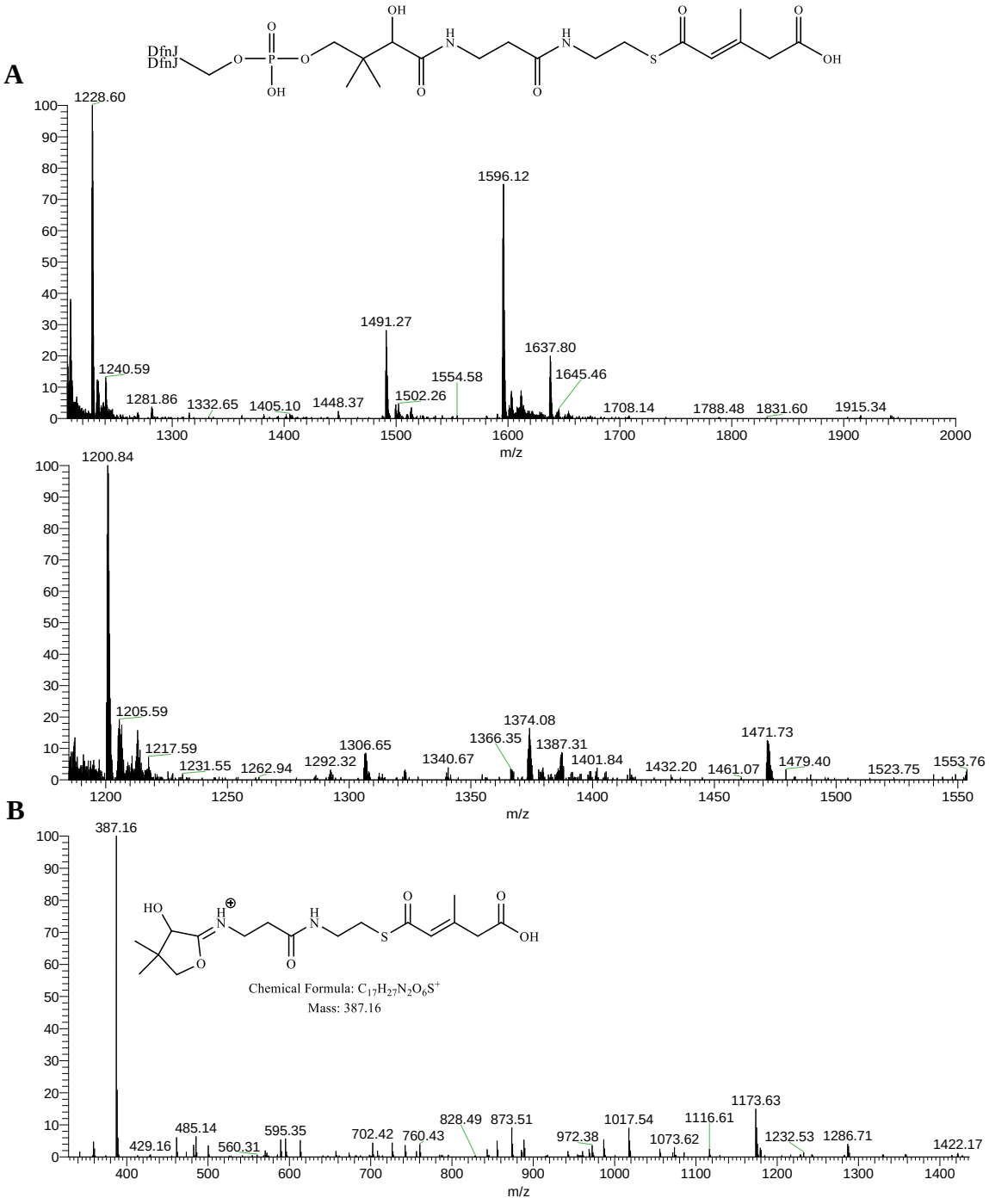


Figure 34. Glutaconyl-DfnJ Intermediate Mass Spectrometry Chromatograms

A) Tryptic Fragments of Glutaconyl-DfnJ. B) MS^2 of the Ppant Fragment of Glutaconyl-DfnJ.

(+4 m/z) Tryptic Glutaconyl-DfnJ Fragment 1: Theoretical Mass: 1228.60 ; Observed Mass: 1228.60

(+4 m/z) Tryptic Glutaconyl-DfnJ Fragment 2: Theoretical Mass: 1200.84 ; Observed Mass: 1200.84

The last step in the diffidicin pathway is the formation of the β-Branch. This is generated through DfnM decarboxylating glutaconyl-DfnJ and giving two expected peaks of 1217.84 and 1190.08 with a MS^2 peak of 343.14. These are generated as such; ((4867.37 β-Branch fragment 1) + (4 hydrogen)) / (4 charge) gives a mass of 1217.84, and ((4756.33 β-Branch fragment 2) + (4 hydrogen)) / (4 charge) gives a mass of 1190.08. The observed fragments of 1217.84 and match up perfectly with the expected values and both peaks carry the 343.14 MS^2 Ppant fragment which means that DfnM successfully decarboxylated glutaconyl-DfnJ to create the elusive β-Branch.

Although the *dfn* genes are homologs to the *B. subtilis* bacillaene *pks* genes, they exhibit different chemistry when producing the β-branch. How the chemistry likely differs between them is shown in Figure 35 below. Both HMG-ACP's are generated and dehydrated by their dehydrating enzymes, but the difference of chemistry is occurring during the decarboxylation step. *B. subtilis'* PksI is protonating the 'dienolate' at the γ-carbon which leaves an internal double bond β-Branch shown in bacillaene. In *B. amyloliquesfaciens* however, HMG-DfnJ is protonated at the α-carbon, which is leaving the external olefin β-Branch found in difficidin.

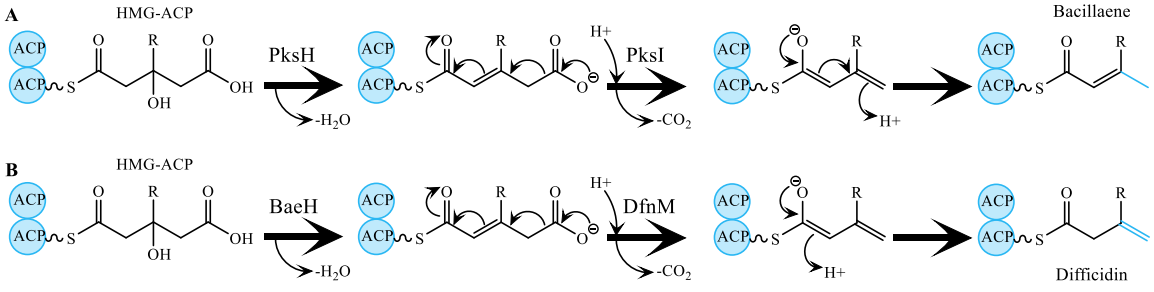


Figure 35. The Difference in Chemistry of the β-Branch's Production. (A) *B. subtilis* - Bacillaene and (B) *B. amyloliquefaciens* - Difficidin.

5.1.6 β-Branch

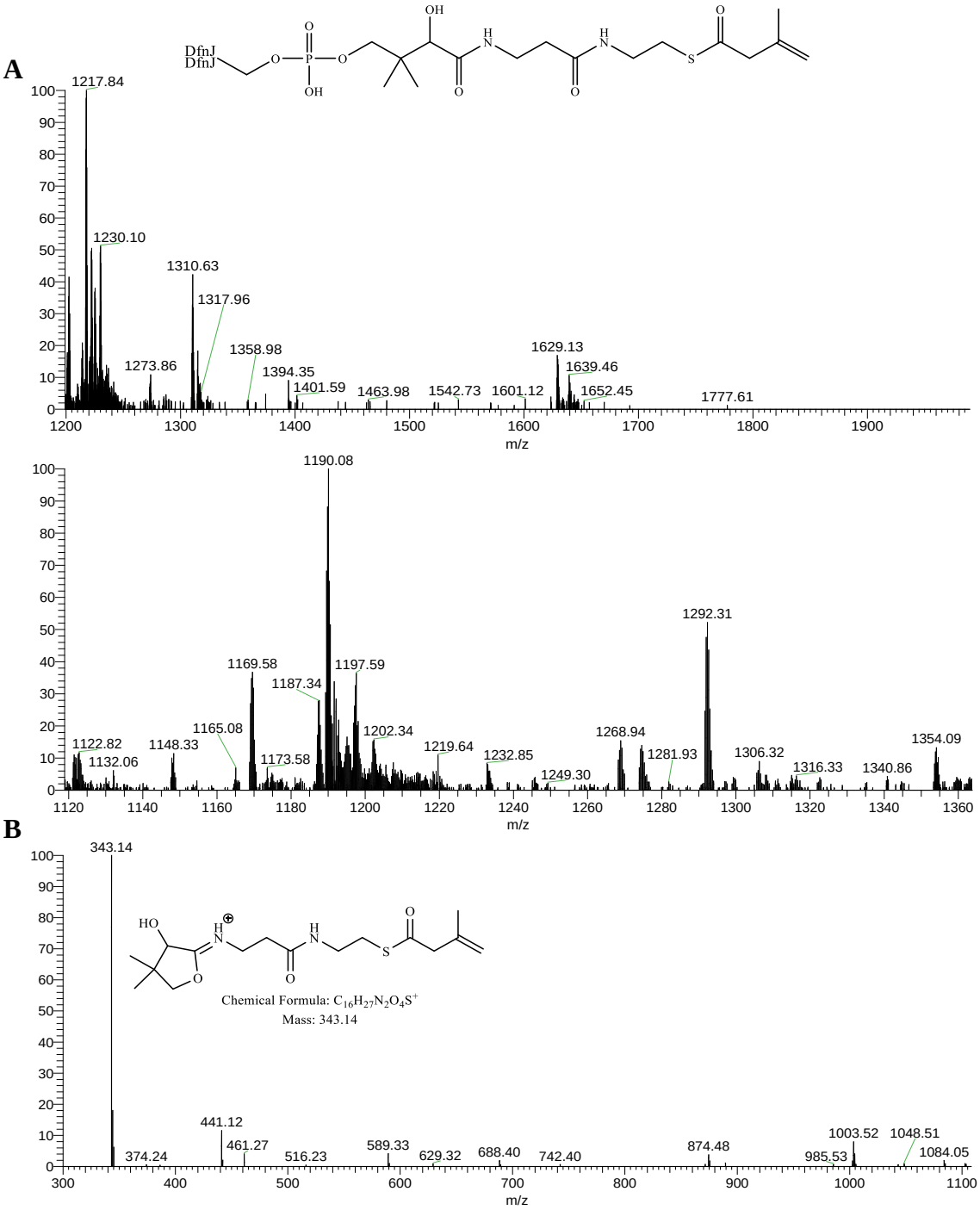


Figure 36. β-Branch Mass Spectrometry Chromatograms

A) Tryptic Fragments of the β-Branch. B) MS^2 of the Ppant Fragment of the β-Branch.

(+4 m/z) Tryptic β-branch Fragment 1: Theoretical Mass: 1217.84 ; Observed Mass: 1217.84

(+4 m/z) Tryptic β-branch Fragment 2: Theoretical Mass: 1190.08 ; Observed Mass: 1190.08

5.2 Hypothesis Proven

The mass spectrometry data collected analyzed difficidin's biosynthetic pathway and each intermediate was verified. Difficidin production does follow a very similar pathway of that found in bacillaene synthesis with a few variants. Originally, the hypothesis was that DfnM was dehydrating and decarboxylating the HMG-DfnJ intermediate all in one concerted step. However, after analyzing reactions that had HMG-DfnJ and DfnM (lacking BaeH), there was no β-branch produced.

The second hypothesis however, was correct. Difficidin is in fact recruiting BaeH from another operon (*pks1:bae*) within the *Bacillus amyloliquefaciens* genome and not acting in one concerted step (Figure 37). After dehydrating HMG-DfnJ with the addition of BaeH, DfnM was then able to decarboxylate the glutaconyl-DfnJ intermediate to generate the β-branch olefin.

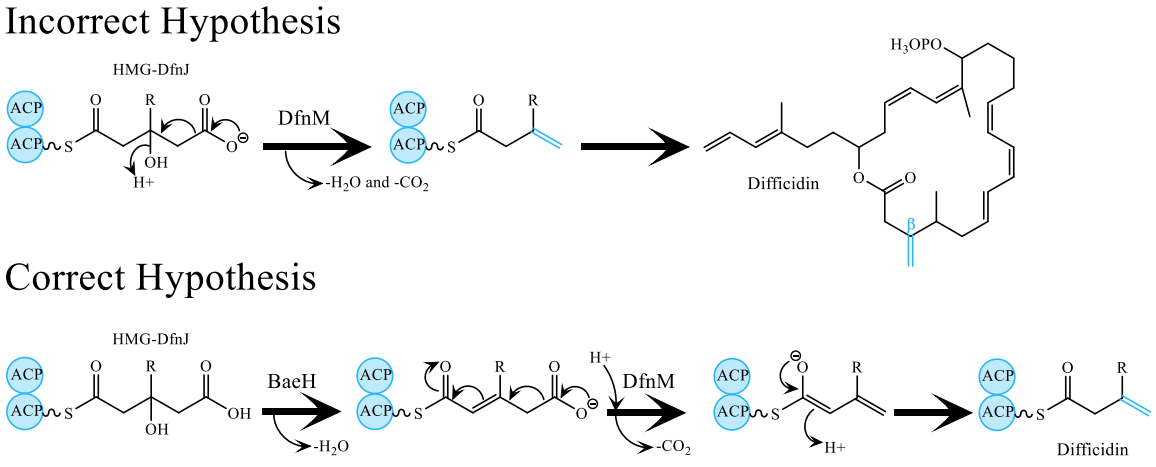


Figure 37. Incorrect Hypothesis vs. Correct Hypothesis. This Shows the Biosynthetic Pathway of Difficidin from the HMG-DfnJ Intermediate Using BaeH From *pks1:bae* and DfnM from the Difficidin Operon (*pks3*).

CHAPTER VI

CONCLUSION

This research shows the complete pathway of how *Bacillus amyloliquefaciens* synthesizes the β-branch found in the macrolide phosphate ester known as difficidin. Not only was every intermediate in the difficidin pathway confirmed through mass spectrometry analysis but the second hypothesis was also validated.

This novel chemistry can open several new doors in the understanding of how β-branches are constructed and how they differ relative to the genes responsible for the synthesis. Not only can it provide insight into other poorly understood secondary metabolites that also exhibit difficidin's external olefin. With this new information, there will be many ways to manipulate and re-design polyketide synthesis and create new products. Not only can this research be applied to other β-branching metabolites, there is a possibility that this can also be incorporated into metabolites that are not known for displaying β-branching, thus opening up a world of new polyketide metabolites.

Future work includes the analysis of the phosphopantetheine fragment containing the β-branch to ensure the external olefin is forming from this pathway and not the internal double bond found in bacillaene. Since both the internal olefin (bacillaene) and external olefin (difficidin) of the $R=CH_3$ model have the same molecular weight, the electrospray mass spectrometry method cannot differentiate these structures. Instead, there are several ideas to test to circumvent this problem.

The first approach is using MS/MS/MS (MS^3) to select the phosphopantetheine fragment (261 Da) and fragment it. Essentially, the precursor 'parent' ion is fragmented in MS^2 to create 'daughter' ions which are then selected and fragmented further into 'granddaughter' ions through MS^3. By exclusively analyzing the fragmentation of the Ppant fragment ensures the observed MS^2 fragments are only coming from the phosphopantetheine appendage. This method of using MS^3 can possibly lead to different fragment patterns depending on where the double bond is located (internal vs. external). However, this fragmentation might be beyond the capabilities of the gentle nature of ESI (electron spray ionization) and bombardment method used in the orbitrap MS.

Another experiment can be done on GC-MS using SNAc (S-nitro-N-acetylcysteine) chemistry since literature indicates there is a specific peak at 72 that is associated with the external olefin that does not form from the internal olefin GC-MS. The reason for GC-MS is the due to its more powerful ionization energy compared to ESI. However, with this method, the PPT appendage needs to be removed from DfnJ (the acyl carrier protein) before analysis on GC-MS because the large protein would cause it not to work.

There is even a possibility that NMR can be used to detect these different bond positions. By creating the β-branch and then separating the external olefin from DfnJ and running the resulting product on NMR, the position of the β-branch could possibly be detected. These will be ideas to test in the future.

Since polyketide β-branching is rare and poorly understood, applying this research could provide insight into how β-branches are created in other secondary

metabolites. Considering many drugs and antibiotics are produced through polyketide biosynthesis, this research could offer different ways to identify and create new products to combat a variety of illnesses. Knowledge of this chemistry will continue to grow over time and possibly encourage new research that has previously been unexplored.

BIBLIOGRAPHY

1. Gomes ES, Schuch V, de Macedo Lemos EG. *Braz J Microbiol*. **2014**. 44(4):1007–1034. DOI:10.1590/s1517-83822013000400002

2. Weissman KJ. Methods in Enzymology. *Academic Press*. **2009**. 459: 3–16.

3. Murata, M., Naoki H., Matsunaga, S., Satake, M., and Yasumoto, T. *J. Am. Chem. Soc*. **1994**. 116: 7098-7107

4. Chen, H. and Du, L. *Appl. Microbiol Biotechnol*. **2016**. 100: 541. DOI:10.1007/s00253-015-7093-0

5. Fischbach M.A.; Walsh C.T. *Chem. Rev.* **2006**. *106*(8): 3468–3496. DOI:10.1021/cr0503097

6. McGuire, J. M., R. L. Bunch, R. C. Anderson, H. E. Boaz, E. H. Flynn, H. M. Powell, and J. W. Smith. *Antibiot. Chemother*. **1952**. 2: 281-283.

7. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., Katz, L. *Science*. **1991.** 252(5006): 675-679.

8. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., Katz, L. *Gene*. **1992.** 15: 97-103.

9. Drake, E.J., Miller B.R., Shi. C., Tarrasch, J.T., Sundlov, J.A., Allen, C.L., Skiniotis, G., Aldrich, C.C., Gulick, A.M. *Nature*. **2016**. 529(7585): 235-238.

10. Martinez-Nunez, M.A. and Lopez y Lopez, V.E. *Sustainable Chemical Processes*. **2016**. 4(13). DOI: 10.1186/s40508-016-0057-6

11. Fisch, K.J. *RSC Advances*. **2013**. 3: 18228-18247. DOI: 10.1039/c3ra42661k

12. Zimmerman, S.B., Stapley, E.O. et al; *The Journal of Antibiotics*. **1986**. XL(12) 1677-1681

13. Scotti, *C.;* Piatti M.; Cuzzoni, *A.;* Perani, P.; Tognoni, A.; Grandi G.; Galizzi, A.; Albertini, *A.M.; Gene.* **1993**. 130(1): 65-71.

14. Patel, P.S.; Huang, S.; Fisher, S.; Pirnik, D.; Aklonis, C.; Dean, L.; Meyers, E.; Fernandes, P.; Mayerl, F. *J. of Antibiot (Tokyo).* **1995**. 48: 997–1003.

15. Butcher, R.A.; Schroeder, F.C.; Fischbach, M.A.; Straight, P.D.; Kolter, R.; Walsh, C.T.; Clardy, *J. PNAS.* **2006**. 104(5): 1506–1509. DOI: 10.1073/pnas.0610503104

16. Chen, *X.H.;* Vater, J.; Piell, J.; Franke, P.; Scholz, R.; Schneider, K.; Koumoutsi, A.; Hitzeroth, G.; Grammel, N.; Strittmatter, A.W.; Gottschalk, G.; Süssmuth, R.D.; Borriss, R. *J. Bacteriol*. **2006**. 188(11): 4024-4036. DOI**:** 10.1128/JB.00052-06

17. Wu, L.; Wu, H.: Chen, L.; Yu, X.; Borris, R.; Gao, X. *Nature.* **2015**. 5: 12975. DOI:10.1038/srep12975

18. Straight P.D.; Fischbach M.A.; Walsh C.T.; Rudner D.Z.; Kolter R. *PNAS.* **2007**. 104: 305–310.

19. Vargas-Bautista, C.; Rahlwes, K.; Straight, P. *J. Bacteriol*. **2014**. 196(4): 717-728, DOI: 10.1128/JB.01022-13\

20. McAlpine J.B.; Bachmann B.O.; Piraee M.; Tremblay S.; Alarco A.M.; Zazopoulos E.; Farnet C.M. *J. Nat. Prod.* **2005**. 68: 493-496

21. Schwarzer, D.; Finking, R.; Marahiel, M.A. *Nat. Prod. Rep.* **2003**. 20: 275–287. DOI: 10.1039/B111145K

22. Cane, D.E. *Chem. Rev*. **1997**. 97(7): 2463–2464. DOI: 10.1021/cr970097g

23. Calderone, C.T.; Kowtoniuk, W.E.; Kelleher, N.L.; Walsh, C.T.; Dorrestein, P.C. *PNAS.* **2006**. 103(24): 8977-8982. DOI:10.1073/pnas.0603148103

24. Holden, H.M..; Benning, M.M.; Haller, T.; Gerlt, J.A. *Acc Chem Res.* **2001**. 34(2):145-57. DOI: 10.1021/ar0000531

25. Meluzzi, D.; Zheng, W.H.; Hensler, M.; Nizet, V.; Dorrestein, P.C.; *Bioorganic & Medicinal Chemistry Letters.* **2008**. 18: 3107–3111

26. Meehan, M.J., Xie, X., Zhao, X., Xu, W., Tang, Y., Dorrestein, P.C. *Biochemistry*. **2011**. 50(2): 287–299. DOI:10.1021/bi1014776