MUKHERJEE, SRIPARNA, M.S. Characterization of PksD and PksG in *Bacillus subtilis*. (2006)
Directed by Dr. Jason J. Reddick. 72 pp.

Polyketides are a diverse class of natural products that have a wide variety of pharmacological applications. Polyketides are biosynthesized by modular polyketide synthases in a very similar fashion to that of fatty acids. The distinct features of each module govern the degree of reduction in the two carbons added during polyketide synthesis. The highly modular constitution of polyketide synthases makes them adjustable to synthesize unnatural medicinally important polyketides.

Difficidin, an antibiotic polyketide, isolated from *Bacillus subtilis* strain 39320 and 39374, contains an unusual subunit that can not be explained by literature precedent for understood polyketide synthesis. Polyketide synthases have been proposed to biosynthesize difficidin. PksD, being homologous to acyltransferases, is expected to load malonyl CoA to the acyl carrier protein domains found in PksJLMNR, and AcpK. PksG is homologous to hydroxymethylglutaryl-CoA synthases and expected to be involved in the synthesis of the unusual subunit in difficidin.

PksD, from *B. subtilis* strain 168, was successfully cloned, overexpressed, and purified. Attempts to characterize pure PksD by MALDI-MS were not successful. PksG was successfully cloned from *B. subtilis* strain 39374. Attempts at generating a knockout mutant of *pksG* in *B. subtilis* 39374 were not successful. The function of PksG in difficidin production could not be confirmed. However, on the basis of sequence data, we discovered that *B. subtilis* strain 39374 is likely *B. amyloliquefaciens*.

CHARACTERIZATION OF PksD AND PksG IN *BACILLUS SUBTILIS*

by

Sriparna Mukherjee

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

Greensboro

2006

Approved by

_____
Committee Chair

This work is dedicated to my loving husband.

I could not have done this without his love and support.

APPROVAL PAGE

This thesis 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 wish to express my gratitude to my supervisor, Dr. Jason Reddick, whose expertise, understanding, and planning, made this a wonderful graduate experience. I would like to thank the other members of my committee, Dr. Gregory Raner, and Dr. Alice Haddy for their assistance. Finally, I would like to acknowledge the department of chemistry and biochemistry at UNC-Greensboro for the financial support.

I thank my family for the support they provided me through my entire life and in particular, I must acknowledge my roommate Swetha Chakravarthi for her inputs to make our apartment a home. Last but not the least, I thank my labmate Jayme Williams, my friends, especially Ridwana Khan and Debasish Chaudhuri for all their help.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

I.A Polyketide natural products

Polyketides, many of which are produced by environmental bacteria, are a diverse class of natural products that have a wide variety of pharmacological applications.

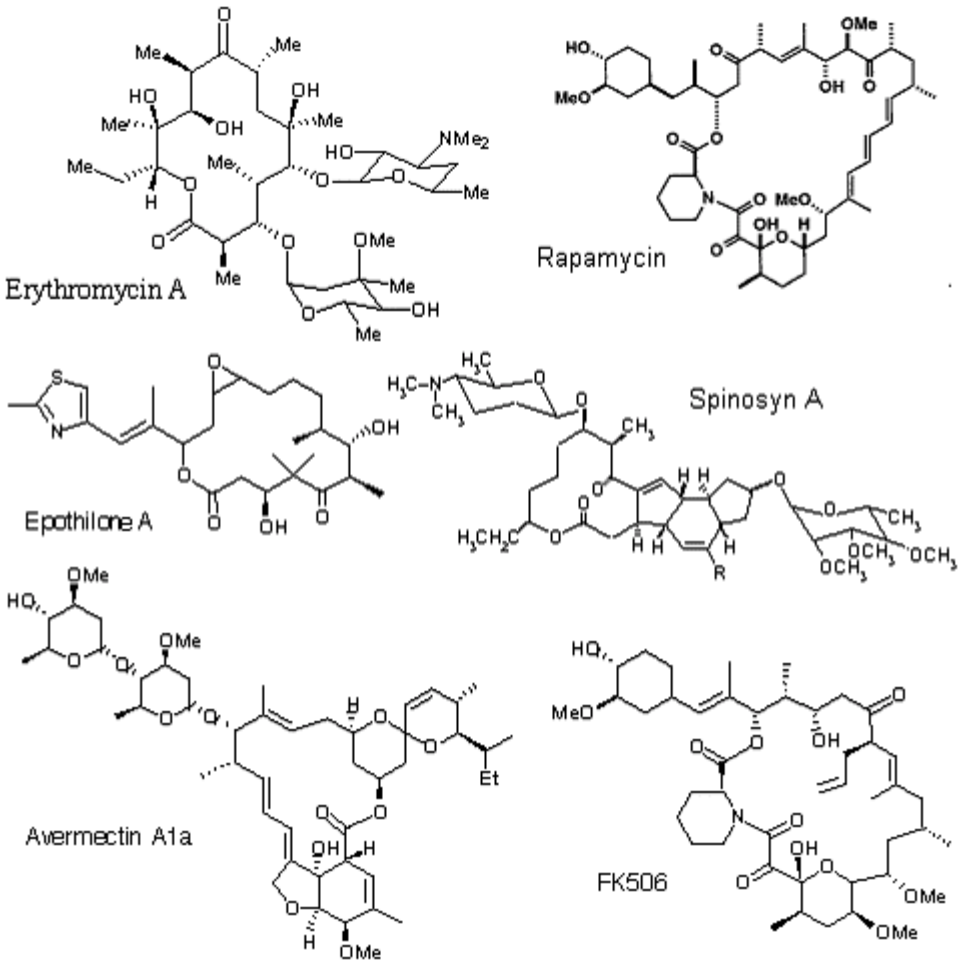

Figure I-1: Examples of some polyketides having significant therapeutic value.

Some of the polyketides (Figure I-1) have extraordinary medicinal value, such as erythromycin A and rifamycin S as antibiotics[I-1], rapamycin and FK506 as immunosuppressants, amphotericin B as antifungals, avermectin as antiparasitics, spinosyn A as insecticidals, and epothilone, a promising anticancer drug[I-2].

The biosynthesis of polyketides occurs by a series of condensation and subsequent reduction of carbon units in a similar fashion to fatty acid synthesis. Long chain fatty acids are produced by fatty acid synthases (FAS) through a process in which a starter unit, generally acetate and an extender unit, generally malonate, are condensed through a decarboxylative Claisen condensation. The resulting β-keto group is fully reduced by β-ketoreduction, dehydration and enoyl reduction, and enters into another cycle through condensation with a new extender malonate unit, resulting in extension of the fatty acid chain. The biosynthesis of polyketides is very similar to that of fatty acids. Polyketides are synthesized by a class of one or more enzymes called polyketide synthases (PKSs). The structural variety of the polyketides can be attributed to the use of different extender units and in the extent of reduction of the β-keto group using β-ketoreduction, dehydration, and enoyl reduction. The PKS genes for a certain polyketide are usually organized in one operon in bacteria and in gene clusters in eukaryotes. Type I PKSs are large proteins structured as modules[I-3], while type II PKSs are aggregates of monofunctional proteins [I-3]. Modular PKSs are found in both Gram-positive and Gram-negative bacteria, although they are particularly abundant in the actinomycetes, which are Gram-positive[I-4].

I.B Polyketide biosynthesis

I.B.1 Erythromycin

Erythromycin A (Figure I-2) was isolated from a Gram-positive bacterium *Saccharopolyspora erythraea*[I-5]. Erythromycin, a polyketide antibiotic, is used as a drug against infections caused by Gram-positive bacteria. In the polyketide biosynthesis field, erythromycin was the first to be well understood in terms of the genetics, activity, and chemistry of its biosynthesis.

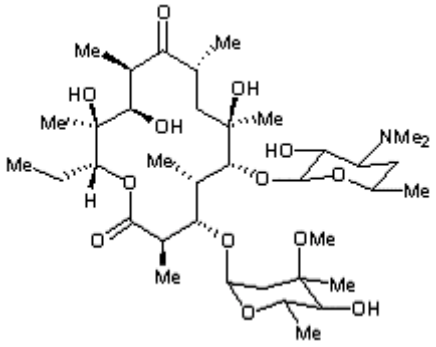

Figure I-2: Erythromycin A

The polyketide portion of erythromycin is synthesized through the condensation of seven propionates via methyl malonyl CoA. The polyketide precursor of erythromycin is 6-deoxyerythronolide B (Figure I-3). 6-Deoxyerythronolide B synthase (DEBS I, II, III) is a large enzyme that catalyses the biosynthesis of 6-Deoxyerythronolide B (6-DEB). DEBS is encoded by three large genes, *eryAI*, *eryAII* and *eryAIII*[I-6]. The three proteins are organized into modules. Each module performs its enzymatic functions for the elongation steps required for 6-DEB formation.

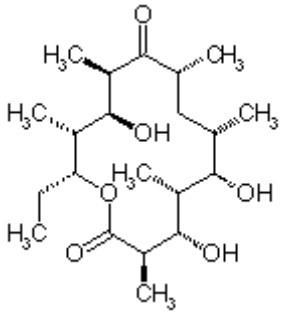

Figure I-3: 6-deoxyerythronolide B, the polyketide precursor of erythromycin.

These include the β-ketoacyl-ACP synthase domain (KS), the acyltransferase domain (AT), the acyl carrier protein domain (ACP), β-ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains. After each condensation step the newly formed β-carbonyl group may undergo different types of reduction according to the different reducing domains present in the module (see Figure I-4).

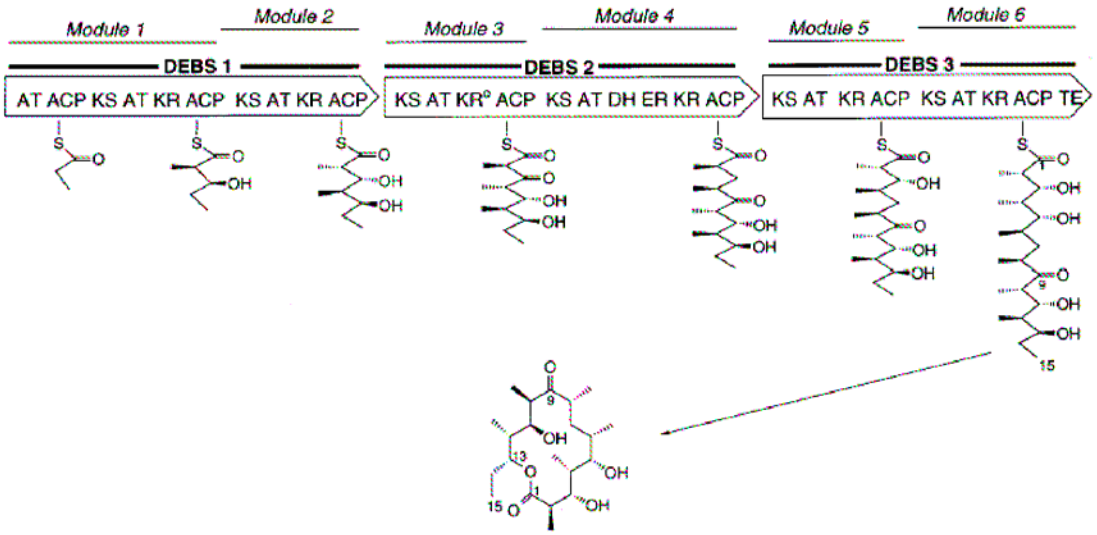

Figure I-4: Six polyketide synthase modules in DEBS.[I-9]

The biosynthesis of erythromycin is known to happen in two major phases (see Figure I-5). In the first phase 6-DEB is produced through the series of condensation of

one propionyl CoA and six methylmalonyl CoAs in succession with the catalytic presence of PKSs. In the second phase, 6-DEB is elaborated by the addition of two sugar moieties, and two cytochrome P450-dependent hydroxylations to complete the formation of erythromycin. PKSs are grouped into modules that are responsible for different enzymatic activities, and for the extension of the polyketide chain by two carbons, followed by different degrees of reduction depending on the presence and sequence of domains in the module.

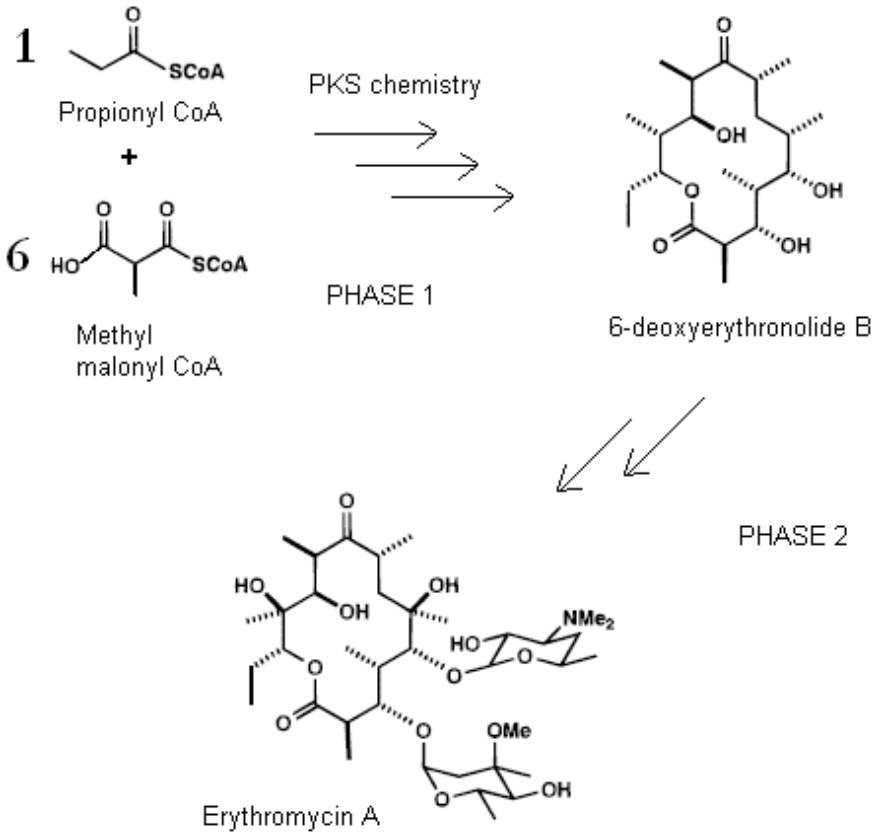

Figure I-5: Scheme of biosynthesis of erythromycin.

In the first phase the condensation reaction between the starter and extender units is a Claisen condensation by decarboxylation in the presence of three domains, ACP, AT, and

KS. The AT domain loads the extender unit onto ACP. The condensation reaction is initiated by KS (see Figure I-6). Each extension of the polyketide backbone is extended by two more carbons.

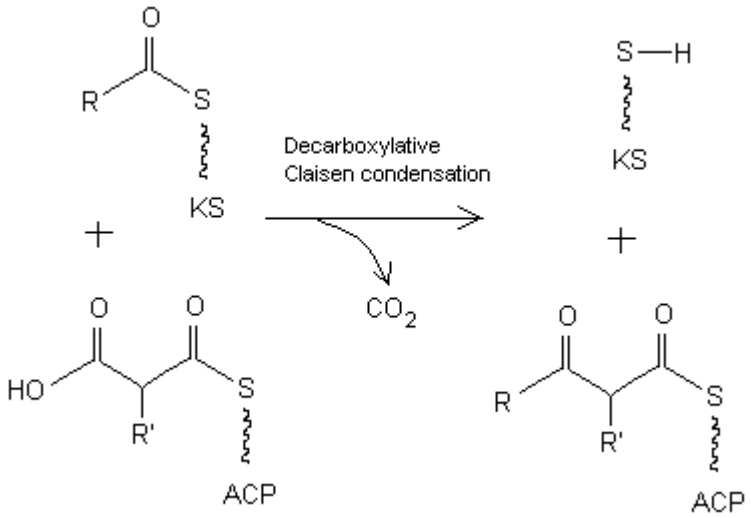

Figure I-6: Claisen-like condensation by decarboxylation.

A series of reductions follows after this, depending on the type of domains present in the module. If no further domains are present in the module, a ketone is produced. The presence of a ketoreductase (KR) reduces the β-carbonyl to an alcohol and the electron provider is NADPH (see Figure I-7). The next possible step is dehydration performed by dehydratase (DH), which produces an α, β double bond. An enoyl reductase (ER) reduces the α, β double bond to another degree to form a fully saturated carbon chain, using NADPH as the electron source. Depending on the structural differences of modular PKSs, the number of cycles of chain extension and subsequent reduction can vary which can explain the structural diversity of known polyketides.

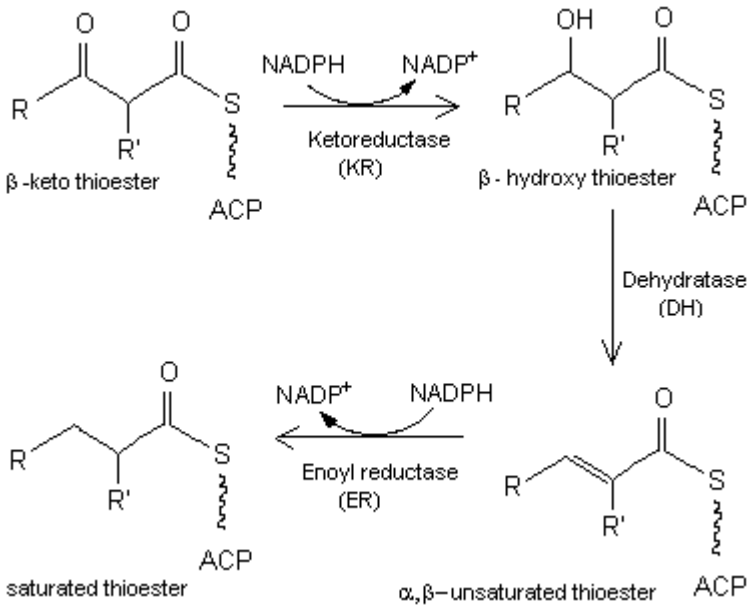

Figure I-7: Series of reductions.

A thioesterase (TE) is required to release the fully grown polyketide chain from the enzyme through cleavage of the thioester. The resultant polyketide may continue as an open chain or cyclizes to a ring structure by lactonization, directed by TE.[I-7]

I.B.2 Combinatorial biosynthesis

The biosynthesis of polyketides is governed by the organized modules present in the domains of PKSs. The structural variety of the polyketides, thus depends on all possible different sequences of the domains in the module, the number of domains present in the module, and post-polyketide synthesis modification through lactonization and tailoring[I-8]. The composition of these domains in a module is the key that determines the structure of each two carbon units within the final polyketide product. A variable set of polyketides can be biosynthesized by modifying the composition of the domains in all

possible permutations and combinations. In recent studies this combinatorial biosynthesis has been used to biosynthesize a set of "unnatural" polyketide natural products[I-8], which would have been very difficult to produce by traditional synthetic chemistry. A combinatorial library of DEB has been prepared by McDaniel *et al*. in 1999 (Figure I-8).

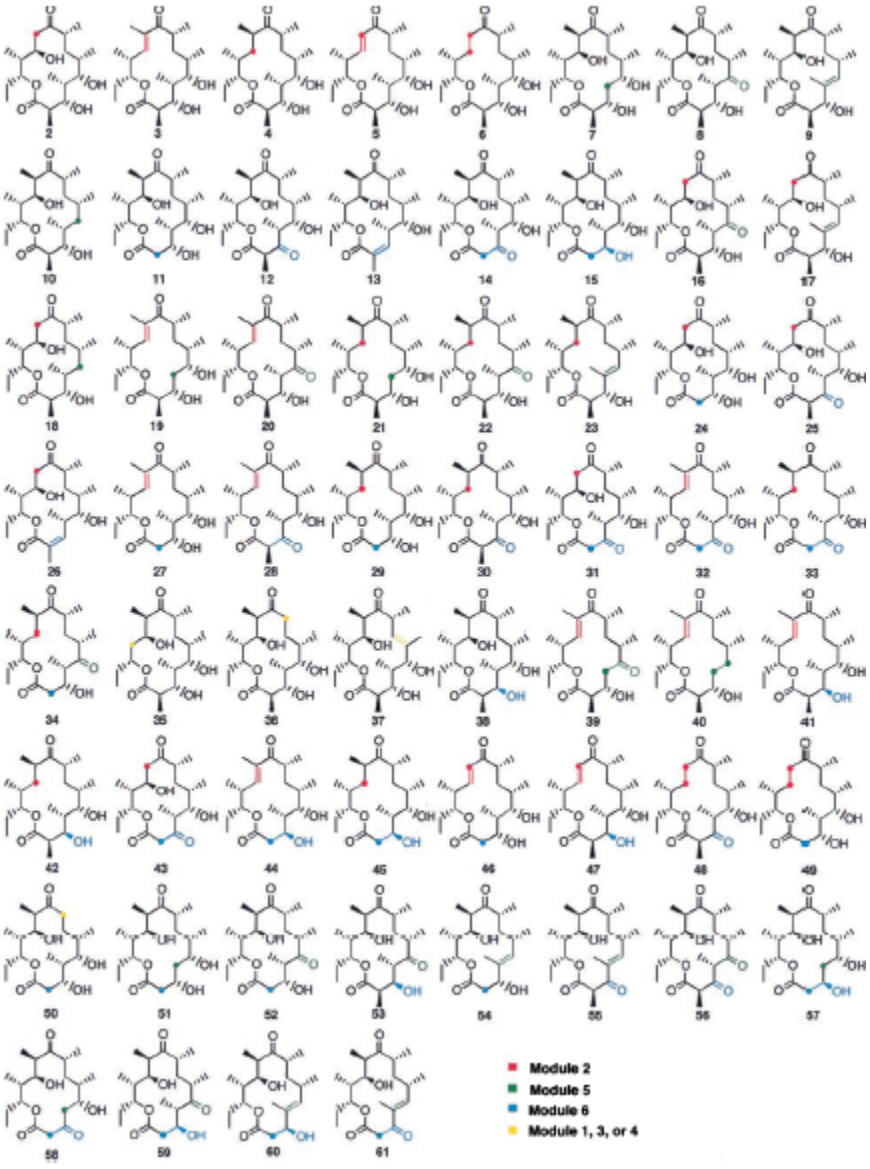

Figure I-8: DEB combinatorial library.[I-8]

The following strategies have been used for genetic manipulation of the arrangement of domains in modules of DEBS, the erythromycin producing PKS: (i) inactivation of individual domains to control the degree of reduction of the β-carbonyl, (ii) insertion of a module to produce longer polyketides, (iii) moving the TE domain upstream in the synthase to produce shorter polyketides, (iv) substitution of AT domains from other known systems to replace methyl malonyl extension monomer unit with malonyl unit, (v) switch two different whole modules to produce a different polyketide, (vi) substitution of KR domains to control the hydroxyl stereochemistry.[I-8, I-10]

Three domains (KR, DH, and ER) govern the degree of reduction of the β-carbonyl formed after the Claisen condensation of the starter and extender units. Inactivation of KR prevents the keto group from undergoing further reduction to a hydroxyl group. Absence of DH leaves the hydroxyl group in place without forming an α, β double bond. Loss of the ER domain leaves the olefin in place instead of reducing it to a saturated hydrocarbon chain.[I-9] If all three domains are present, the removal of KR domain would prevent the action of DH and ER, and loss of DH would leave the function of ER unnecessary. Inactivation of one or more domains in a module affords different polyketides.[I-9]

Alteration of polyketide chain length had successfully been performed so far. Insertion of an extra module between two modules produces an extra ketide in the produced polyketide. A module from rapamycin producing PKS had been inserted into two modules of erythromycin producing DEBS. The resultant product was an octaketide

[I-11], having an extra ketide unit in the chain. Movement of TE domain upstream have been shown to produce shorter polyketide chains.[I-12]

The AT domain in a module is responsible for the specific recognition of an extender unit in chain elongation process. Studies had shown that replacement of AT domain in modules 1 and 2 of DEBS by AT domains from the rapamycin producing PKS successfully loaded malonate instead of methyl malonate as predicted, and the resultant DEB lacked the methyl side chain on the designated C-12. Replacement of the AT domain did not significantly disrupt the chain elongation process.[I-9]

Kao *et al*. had shown that the *R* and *S* stereoisomerism of the hydroxyl groups on the polyketide backbone is governed by the KR domains. This is dictated by the specific orientation of binding between the cofactor NADPH and the substrate carbonyl group.[I-13] These strategies provide the means of alteration, manipulation, and mutation to produce genetically engineered PKSs that allow synthesizing a library of novel "unnatural" polyketides.

I.C Difficidin

The *Bacillus subtilis* strain 168 genome contains a 14-gene operon that shows strong sequence homology to PKSs. Six genes of this large operon showed strong homology to PKS genes, while the rest of them were not related to genes of PKS families.[I-14, I-15]

Difficidin was first isolated from *B. subtilis* strains 39320 and 39374 at Merck & Co. in 1987. Difficidin contains an unusual subunit (C3 in Figure I-9) that is not seen in most other polyketides.

Gene Name	Function based on homology
pksA	Transcriptional Regulator
pksB	Zn-dependent hydrolase
pksC-E	Acyl Transferases (AT)
acpK	Acyl Carrier Protein (ACP)
pksF	Ketosynthase/AT
pksG	Hydroxymethylglutaryl (HMG)-CoA Synthase
pksH-I	Enoyl CoA Dehydratases
pksJ-R	AT-Less Type I PKSs
pksS	Cytochrome P450 (hydroxylase)

Table I-1: Polyketide synthases and their activities based on homology.[I-16]

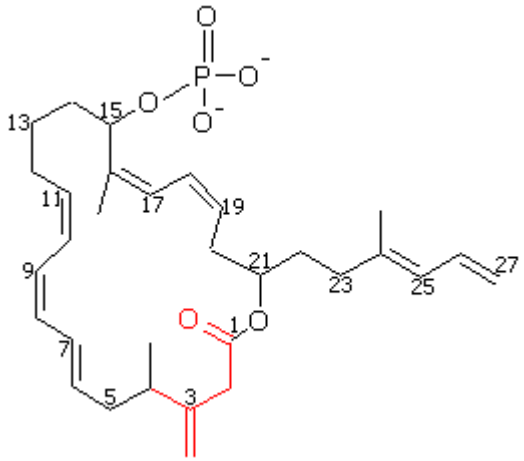

Figure I-9: Difficidin.

Polyketides have a general conceptual formula of CH_3-$(CO\text{-}CH_2)_n$-COOH and all alternate C=O positions are known as ketide positions in this structural formula. Like fatty acid biosynthesis, malonyl groups are the chain extension monomers in polyketide biosynthesis. A simple example of standard polyketide chemistry is shown in Figure I-10. The carbon undergoing reduction is circled in the figure.

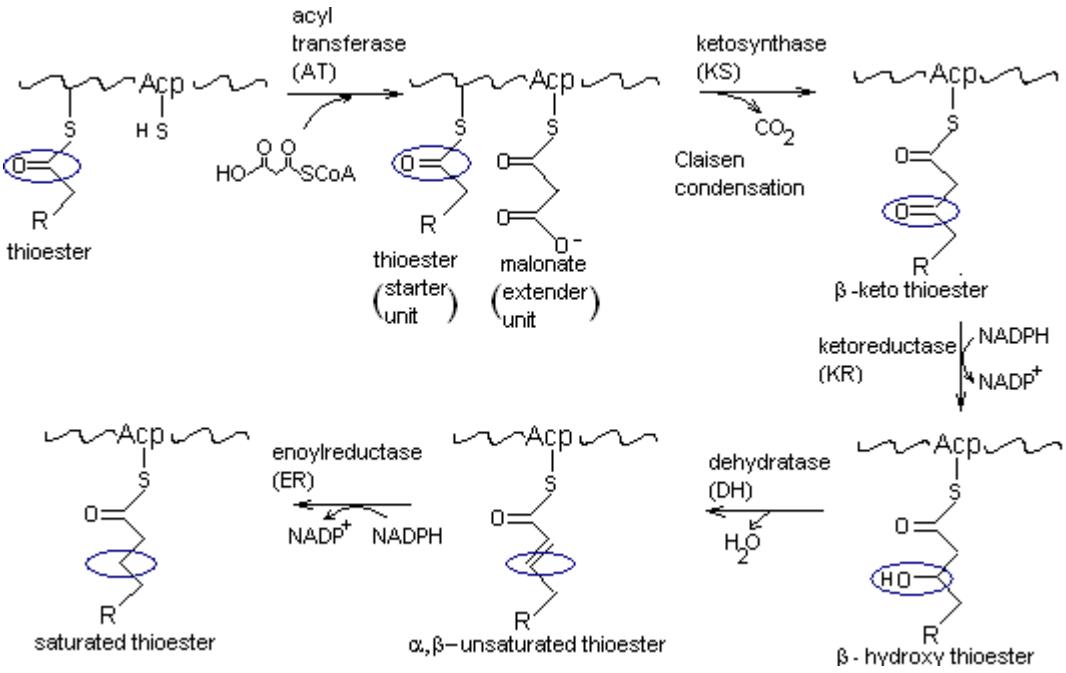

Figure I-10: Pattern of reactions in standard polyketide chemistry.

Instead of having a normal ketide position group, the C3 carbon in difficidin has a branching methylene group, with an exocyclic double bond. Difficidin is an attractive system to study, since this unexplained chemistry could potentially be transferred to combinatorial biosynthesis technologies. The broad goal of research in our laboratory is to understand the biosynthesis of this group at C3. One of two basic possibilities is that a new type of precursor molecule, already containing this group, is attached to the polyketide chain as an alternative to a typical malonyl precursor. The second possibility is that the PKS builds the polyketide backbone of difficidin following the standard pattern and the new carbon atom is attached after the chain is completed. The hypothesis is that the genes *pksFGHI* are responsible for the chemistries predicted by either of these two possibilities.

Other analogous systems such as mupirocin (an antibiotic, produced by *Pseudomomas fluorescens*)[I-17], antibiotic TA (produced by *Myxococcus xanthus*)[I-18], and jamaicamide (a polyketide-peptide neurotoxin, produced by *Lyngbya majuscula*)[I-19], curacin A (an antitubulin natural product produced by *Lyngbya majuscula*)[I-22], onnamides and pederins (both antitumor)[I-23] have been shown to posses unusual subunits similar to that of difficidin. The biosynthetic operons in these polyketides have been sequenced and found to contain homologs of *pksFGHI*.[I-27] Other medicinally important polyketides such as spongistatins and amphidinolides (both anticancer)[I-20, I-21] also contain unusual branching subunits. Therefore, based on these findings it is logical to assume that homologs of *pksFGHI* and *acpK* are generally involved in the biosynthesis of branching ketides in all such polyketides.

I.C.1 Activity of PksD in polyketide biosynthesis

Type I PKSs are arranged into modules, each of which minimally contains a β-ketoacyl synthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP). According to recent findings by Dr. Ben Shen *et al*.[I-24], the leinamycin biosynthetic gene cluster consists of two PKS genes, *lnmI* and *lnmJ*, that encode six PKS modules and none of them contains an AT domain. They showed that *lnmG* provides the missing AT activity to the *lnmI* and *lnmJ* PKS enzymes. LnmG efficiently and specifically loaded malonyl CoA to the ACP domains in all six PKS modules. The sequence analysis of genes *pksJLMNR* denoted that they also lack AT domains[I-24], therefore a homolog of *lnmG* is expected to be needed by this system.

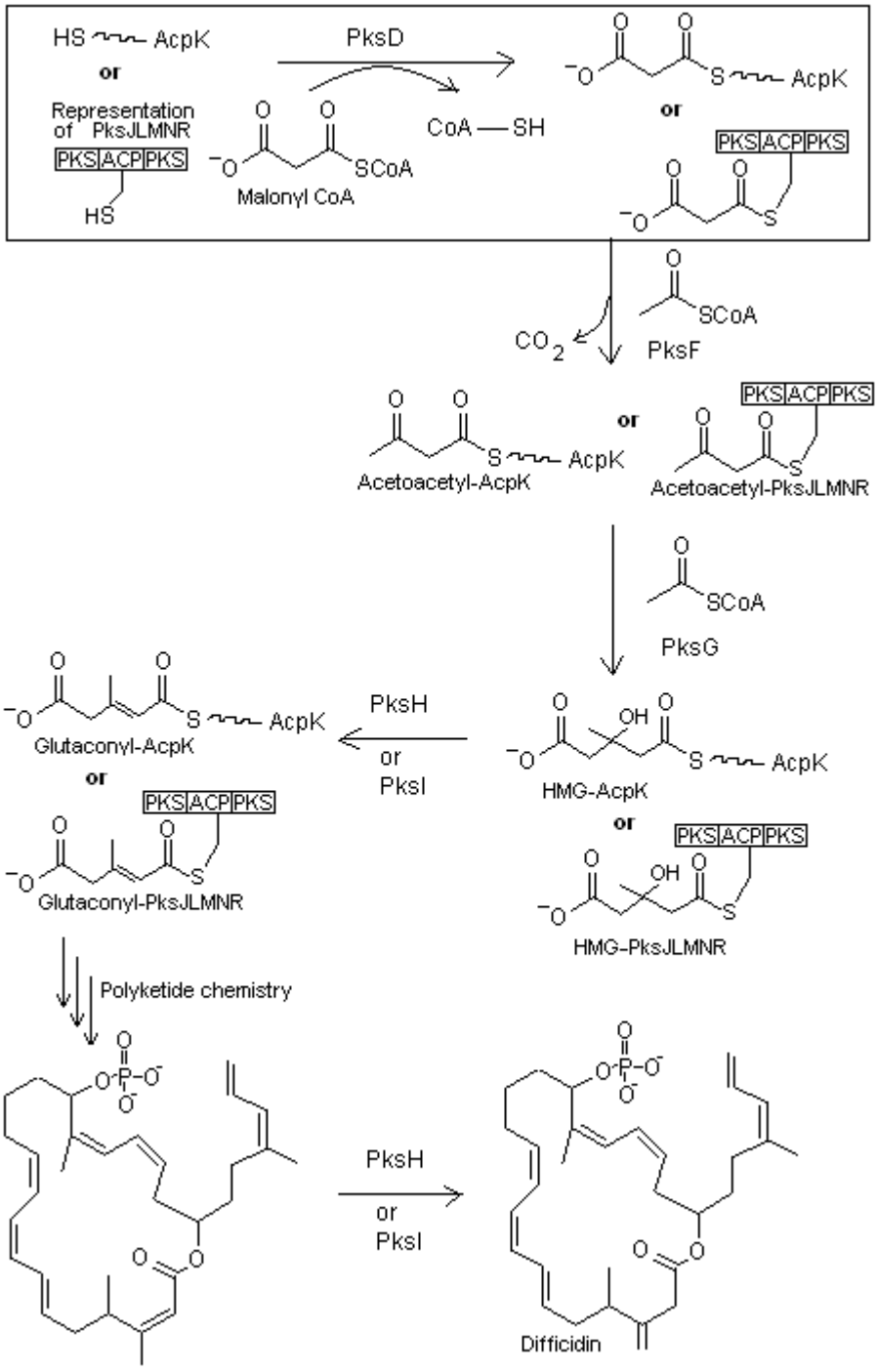

Figure I-11: Hypothesis for PksD activity.

We assume that according to the sequence homology shown in PKS systems (Table I-1), the *B. subtilis* genes *pksCDE* are candidates for *lnmG* homologs. Therefore,

one of the main hypotheses of this project is that PksD possibly installs malonyl CoA or methyl malonyl CoA either to the ACP domains found in PksJLMNR, AcpK or to both (see Figure I-11). One area of study here includes overexpression, purification and determining the enzyme activity of the gene *pksD* from *B. subtilis* strain 168. This work is important because eventual studies on AcpK and PksJLMNR will necessitate having this AT activity on hand in our laboratory.

I.C.2 Function of PksG in difficidin production

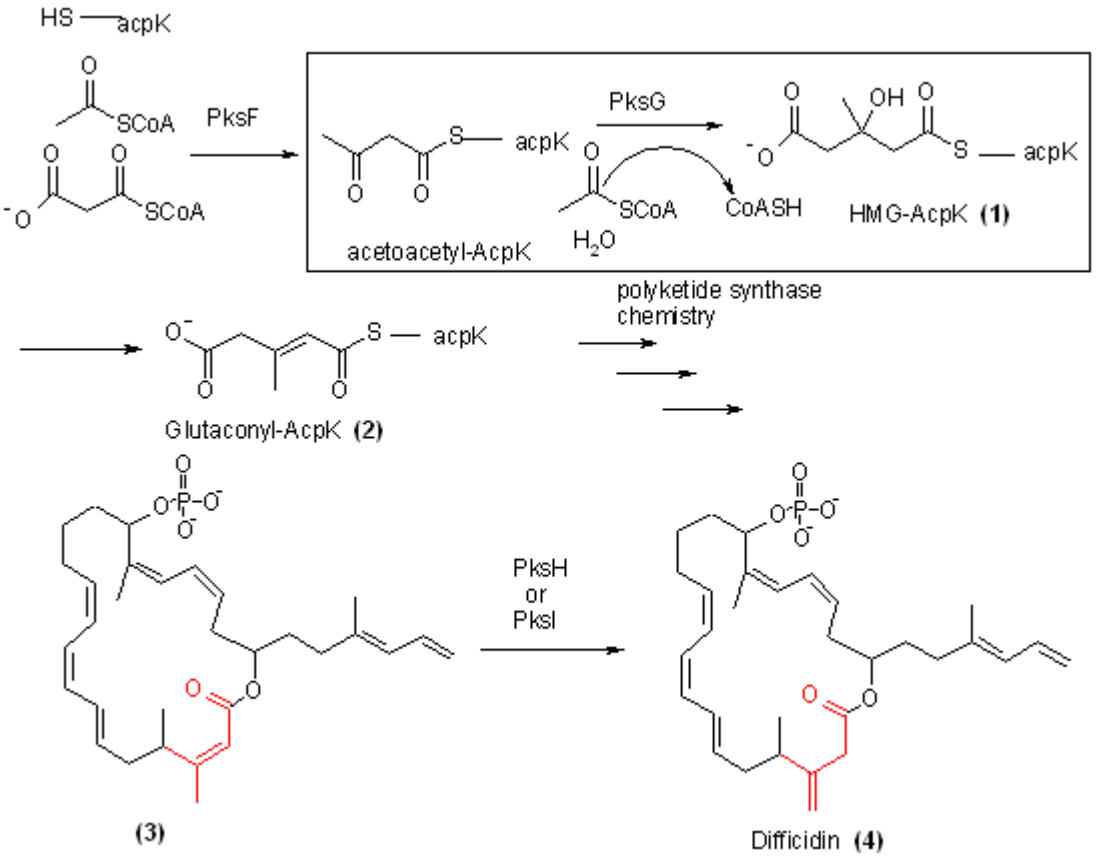

Figure I-12: Hypothesis 1 – role of PksG in unusual precursor formation.

The other goal of the research described in this thesis involves determining the chemical function of PksG in difficidin synthesis. There are two hypotheses involving PksG, the first one involves the use of PksG to synthesize an unusual polyketide precursor, glutaconyl-AcpK (**2** in Figure I-12), which is formed from 3-hydroxy-3-methylglutaconyl-AcpK (HMG-AcpK, **1** in Figure I-12). Then precursor **2** is incorporated into the immature difficidin molecule to form **3** and difficidin (**4** in Figure I-12) synthesis follows by the mechanism of standard polyketide biosynthetic chemistry.

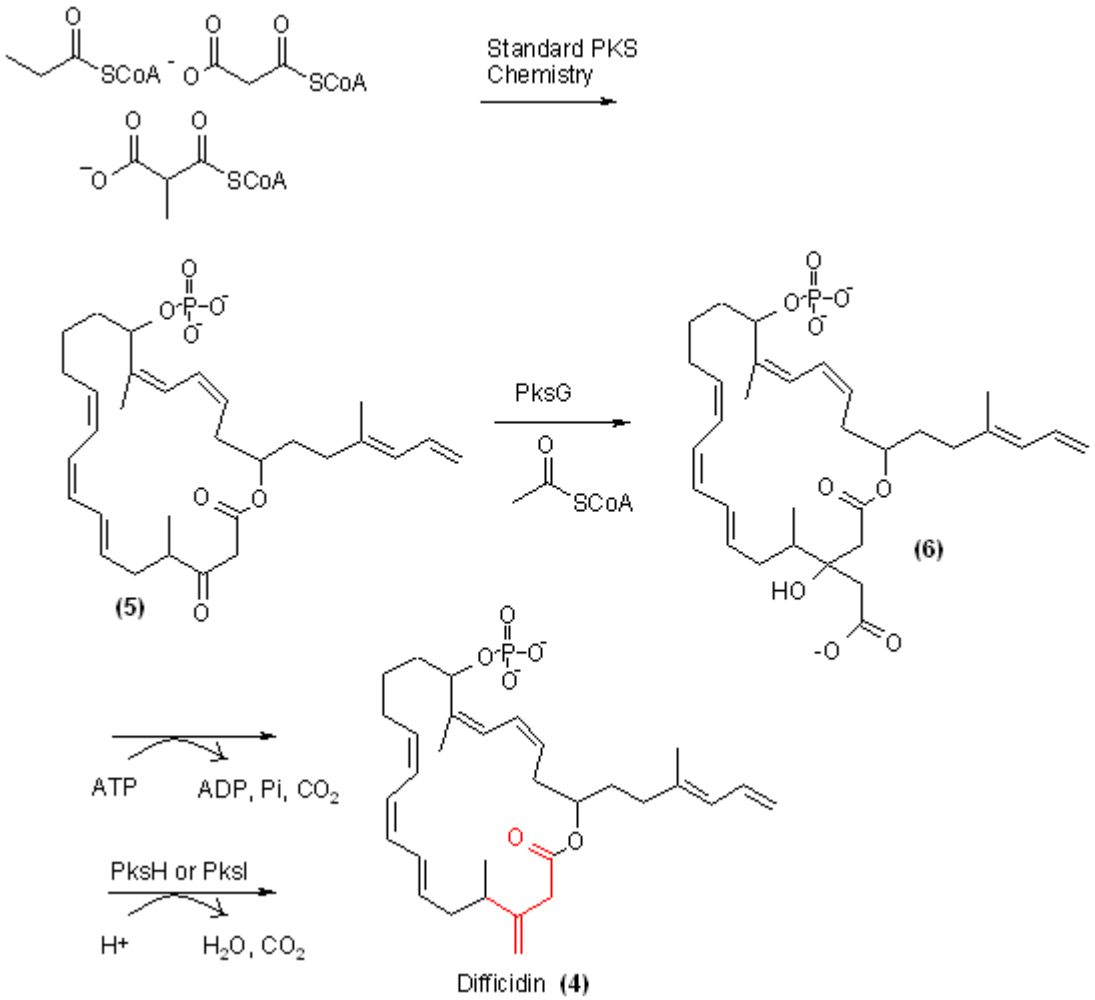

Figure I-13: Hypothesis 2 – role of PksG in unusual precursor formation.

The second hypothesis involves the formation of a ketone intermediate (**5** in Figure I-13) and PksG is then proposed to act on this intermediate in one of the steps towards the formation of difficidin (**4**). Both the hypotheses involve PksG, catalyzing an aldol condensation in the process of difficidin formation.

This project aims to clone, overexpress, purify, and characterize two important genes of polyketide synthases required to synthesize difficidin. The hypotheses sited are based upon the homology found in PKS modules. This research also attempts to sequence a fragment of *pksG* from *B. subtilis* strain 39374, showing experimental proof that homologs of strain 168-*pksG* exist in the unsequenced difficidin producing strain 39374. This research also attempts to construct a mutant lacking the *pksG* gene, in order to investigate the function of PksG in synthesis of difficidin. Upon completion, the knowledge of PksG and associated chemistries are expected to be transferable to combinatorial biosynthesis methods for introduction of branching ketides.

In a recent study in 2006 by Calderone *et al*., characterization of PksCFGHI had been successfully performed, and the experimental data were in concert with our hypotheses. PksC, being homologous to malonyl loading acyltransferase (AT) domain selectively loaded malonyl CoA onto AcpK. PksG, being homologous to hydroxymethylglutaconyl CoA synthase (HCS) performed an aldol condensation between acetyl-AcpK and acetoacetyl-ACP to form HMG-ACP as expected. Product molecules were identified by high resolution mass spectrometry.[I-27] However none of these works was tied with a specific natural product, showing a possible general trend followed by homologous systems. Very recently Sherman *et al*. showed the HCS-like gene activity,

but with homologs from curacin system.[I-26] These all point to a general method revealed by genetics for the biosynthesis of unusual branching ketides.

CHAPTER II

OVEREXPRESSION, PURIFICATION, AND CHARACTERIZATION OF PksD FROM *BACILLUS SUBTILIS* STRAIN 168

II.A Introduction to PksD

II.A.1 Acyltransferase

In polyketide biosynthesis, acyltransferases (AT) load a malonyl unit to the polyketide synthase (PKS) modules. This is the first step in polyketide biosynthesis. PksD, being homologous to acyltransferases, is expected to load malonyl CoA or methyl malonyl CoA to the acyl carrier protein (ACP) domains found in PksJLMNR. PksD may be important in the biosynthesis of the unusual precursor, and as such may also load a malonyl unit onto AcpK. Figure II-1 shows the type of reaction catalyzed by AT.

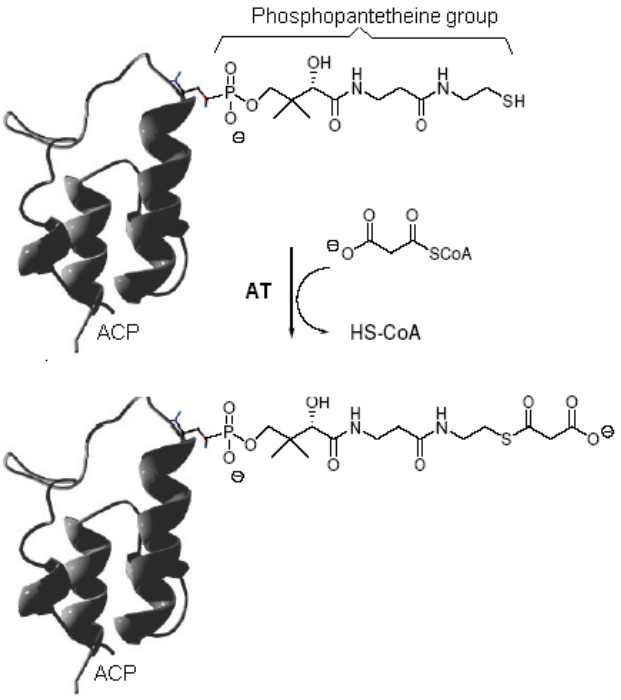

Figure II-1: Transfer of a malonyl unit from malonyl CoA to a holo-ACP, catalyzed by acyltransferase. (Image courtesy: Dr. Reddick's lecture notes)

Recent studies have shown that PksC specifically loaded malonyl CoA onto AcpK. [I-27] Therefore it is reasonable to hypothesize that PksD, being a member of the gene assembly PksCDE would catalyze the same type of malonyl transfer reaction. Polyketide biosynthesis usually occurs in two repeated phases. The first one employs Claisen-like condensation reactions where a starter and an extender unit condense to form β-keto thioesters through decarboxylation. Three other important domains that are responsible for this phase, are found in modular type I PKSs, β-ketoacyl synthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP). In the second phase, the thioester β-ketoacyl-S-ACP is then reduced by other three optional domains, ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER). The degree of reduction of the β-keto group depends on the presence of these domains in a module. The chain is elongated by another extender unit that repeats the condensation and subsequent reduction. A thioesterase (TE) hydrolyzes the thioester to release the polyketide chain from the enzyme.[II-1, II-2]

According to recent findings by Dr. Ben Shen group[II-3], the leinamycin (LNM) biosynthetic gene cluster consists of two PKS genes, *lnmI* and *lnmJ*, that encode six PKS modules and none of them contains the cognate AT domain (Figure II-2). They showed that *lnmG* provides the missing AT activity to the *lnmI* and *lnmJ* PKS enzymes. *LnmG* efficiently and specifically loaded malonyl CoA to the ACP domains from all six PKS modules. The sequence analysis of genes *pksJLMNR* from *B. subtilis* denoted that they lack AT domains[II-3], making this system very similar to the leinamycin system discovered

by Dr. Ben Shen's group. We assume that according to the sequence homology shown in PKS systems, the *B. subtilis* genes *pksCDE* are candidates for *lnmG* homologs.

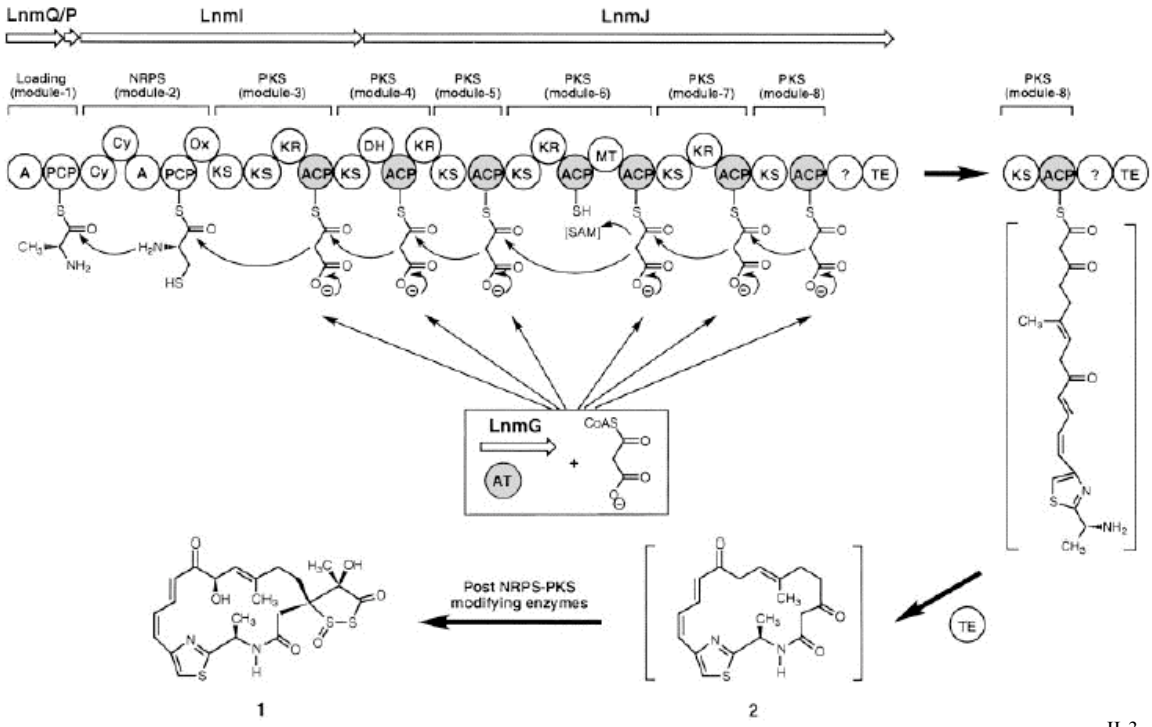

Figure II-2: *LnmG* loads malonyl CoA to the ACP domains from all six PKS modules.[II-3]

II.A.2 Role of AT in synthesis of difficidin

Two different hypotheses have been proposed concerning the synthesis of difficidin. The first involves the synthesis of an unusual precursor. Once this unusual precursor is formed, it is incorporated in the difficidin molecule by a mechanism similar to standard biosynthetic polyketide chemistry. The other hypothesis involves standard polyketide chemistry to be used to generate a difficidin precursor containing a ketone at C3 and then other enzymes (PksFGHI) are proposed to act on this intermediate with acetyl CoA, to eventually form difficidin.

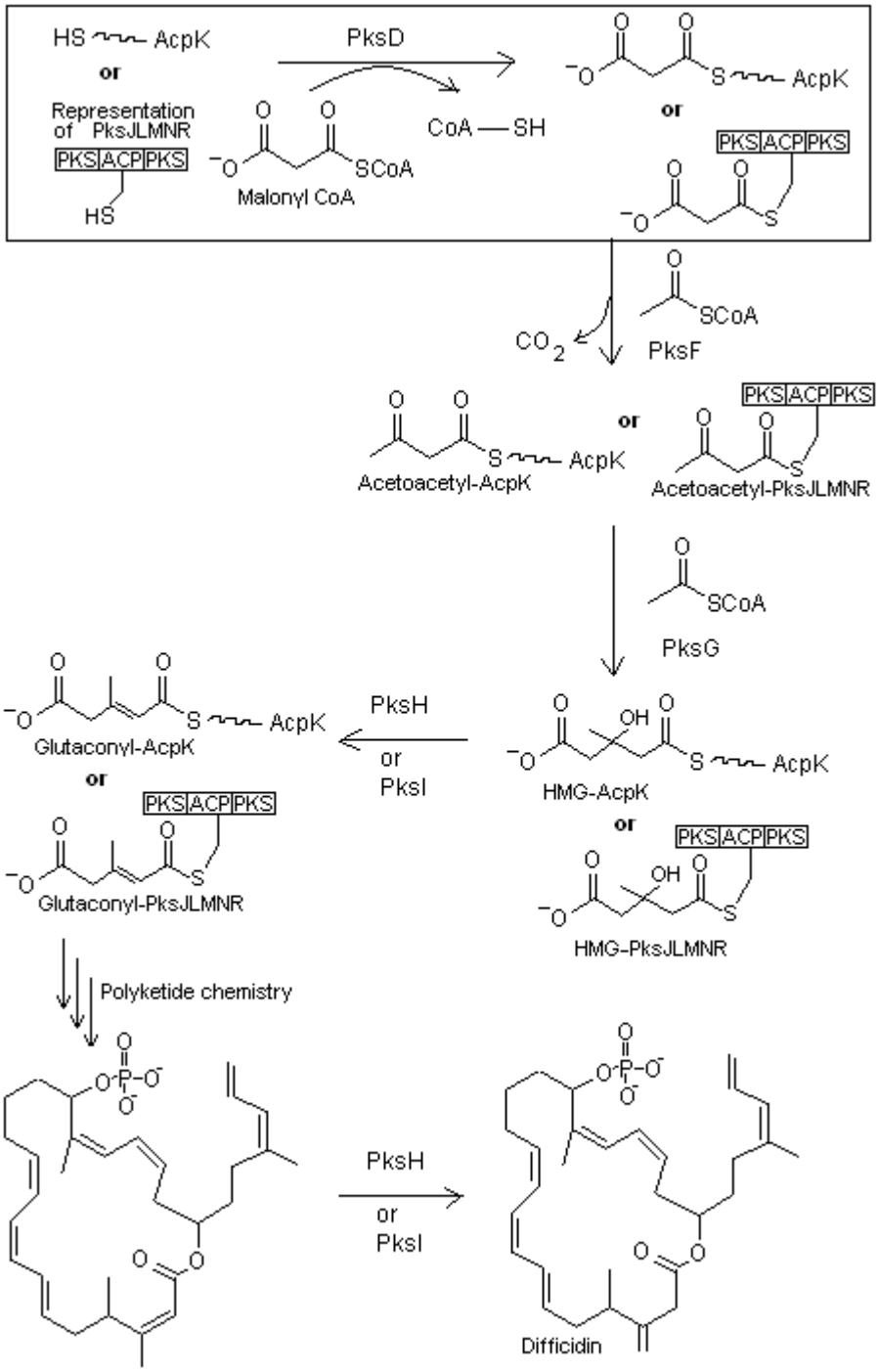

Figure II-3: Hypothesis for PksD activity in difficidin synthesis.

AT, in both of these cases installs the malonyl unit onto the ACP domain of the PKS modules. Figure II-3 shows that PksD, being homologous to AT, loads the malonyl or

methyl malonyl unit either onto the ACP domains of PksJLMNR modules or in AcpK, or to both. Malonyl-ACP undergoes a decarboxylation and condensation reaction to synthesize acetoacetyl-AcpK, which undergoes an aldol condensation reaction with acetyl CoA in the presence of PksG, which is homologous to hydroxymethylglutaryl (HMG) CoA synthase. The product of this reaction is HMG-AcpK. The next proposed step involves dehydration. PksH and PksI are homologous to enoyl CoA dehydratases, therefore one of them is proposed to perform dehydration. Finally the standard polyketide biosynthesis chemistry completes the synthesis of difficidin. In conclusion, PksD is expected to load the malonyl unit to the ACP domains of PksJLMNR or in AcpK, or to both, as necessary to produce difficidin.

II.B Production and characterization of pET-28a/*pksD*

II.B.1 Goals

In order to test our above hypothesis for the activity of PksD, our goal was to obtain pure PksD from *B. subtilis* strain 168 and characterize its activity. The first step towards the goal was to clone and overexpress *pksD*, in order to produce a large amount of pure PksD, using the pET-28a plasmid vector. Overexpression using pET plasmids involves the strong bacteriophage T7 promoter and extremely selective and efficient T7 RNA polymerase. In pET-28a, the T7 promoter is placed under the control of the LacI repressor protein, rendering the pET system susceptible to external control. When induced, this causes nearly all the cell's resources to be committed to transcribing the gene under the control of the T7 promoter; thus "overexpressing" the gene. The degree of production can be controlled by altering the concentration of the inducer. [II-7] Here the

target gene being PksD, would be cloned into the plasmid pET-28a using a restriction digest on both pET-28a and PCR-amplified *pksD*. A subsequent ligation using T4 DNA ligase would complete the successful cloning step. Overexpression of a protein involves growing a culture of cells until they reach mid-log phase and then inducing the growth by adding an inducer. Any overexpression system based on LacI repression can utilize isopropyl-beta-D-thiogalactopyranoside (IPTG), which is an artificial inducer of the *lac* operon, resulting in production of large amounts of the gene of interest. Once *pksD* was successfully cloned into pET-28a, the second goal would be to purify the overexpressed PksD. A nickel column would be used for purification because PksD would be cloned in frame with a C-terminal histidine-tag coded by the plasmid. His_6-tagged proteins can be purified by affinity chromatography on a nickel-nitrilotriacetic acid (Ni-NTA) column. NTA, a Ni^{+2}-chelating adsorbent, leaves two vacant sites on the nickel for binding the imidazole nitrogens present in the His_6-tag, allowing rigorous washes with imidazole. In the process other non-specifically bound proteins are washed through the column while the His_6-tag, connected to the C-terminus of the protein, binds to the Ni-ions. An optimal concentration of imidazole is used to elute the desired pure protein.

The final step would be to test PksD for its expected enzyme activity, which is shown in Figure II-4. The first reaction is known as phosphopantetheinylation because the phosphopantetheinyl group from coenzyme A becomes attached to apo-AcpK. This reaction makes use of Sfp, which posses a wide range of ACP-substrate specificity. Sfp is also larger in size in comparison to Acp.[II-4, II-5] Sfp is expected to convert the apo-form of

AcpK into the holo-form which is the active form using Coenzyme A (CoA-SH). The leaving group is 3′, 5′- ADP.

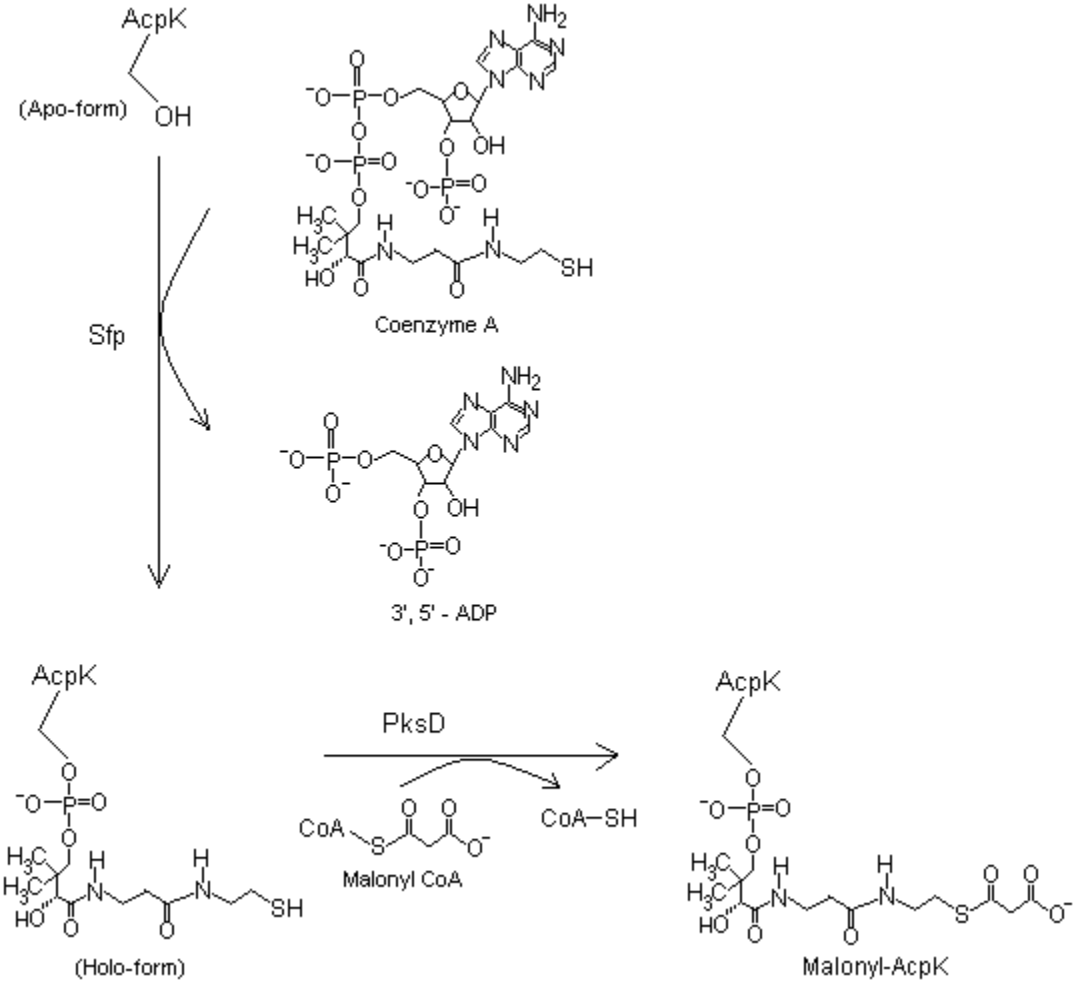

Figure II-4: Hypothesis for enzyme activity of PksD.

We expect that PksD forms malonyl-AcpK from holo-AcpK and malonyl-CoA. Malonyl-AcpK would be detected by Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS). A common analytical tool for detection of

biological molecules is autoradiography, using radioactive substrates. Since we do not have this capability in our lab, we will instead use MALDI-MS.

MALDI-MS is an important analytical technique, which has been successfully used to identify proteins and other large biological molecules such as oligosaccharides and oligonucleotides. MALDI-MS employs laser bombardment on an air-dried sample that was first mixed with a matrix as ionization technique. The matrix helps avoiding the problem of decomposition of sample molecules because of the direct laser.[II-6] The ionized molecules are accelerated in a flight tube under an electric field and different ions travel to the detector at different times depending on their mass to charge (m/z) ratio, resulting in a distinct signal for each molecule. Positive ionization is used in general for protein and peptide analyses.

MALDI-TOF-MS offers another technique in which protein identification is achieved by first digesting the protein into peptide fragments using a proteolytic enzyme such as trypsin, and peptide fragments are then identified by matching them with a library of known digested peptide masses. Here we do not require such a library because we are working with a known protein of known sequence (AcpK).

II.C Results

II.C.1 Cloning of *pksD*

The first step in the cloning process was designing the primers for polymerase chain reaction (PCR). After a couple of unsuccessful attempts, new primers for PCR amplification of *pksD* were designed.

II.C.1.a PCR, restriction digest, and ligation reaction of *pksD*

A PCR reaction was conducted to amplify *B. subtilis* strain 168 genomic DNA as the template and a gene fragment was produced having a sequence that inserts an NcoI site at the start and inserts an XhoI site downstream.

An annealing temperature of 63°C was applied with *Phusion* polymerase. The gene length of *pksD* is 888 base pair (bp). Figure II-5 shows the very bright band that appeared at ~ 900 bp, indicating a successful PCR.

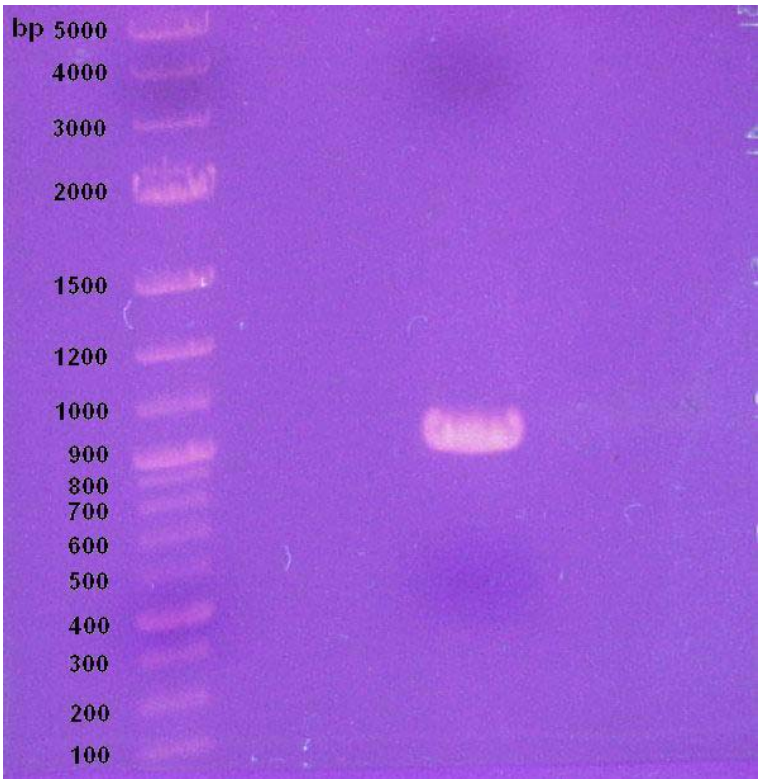

Figure II-5: Agarose gel showing a bright band at 900 bp.

An attempt was made to clone *pksD* into the plasmid vector pET-28a using restriction enzymes NcoI and XhoI, having restriction sites at 296 and 158 base pairs respectively (see Figure II-6, showing a vector map and restriction sites of pET-28a). T4

DNA ligase and T4 DNA ligase buffer were used to ligate the restriction digested plasmid and PCR product. Three different reactions at varying insert/plasmid ratios were performed.

pET-28a(+) sequence landmarks	
T7 promoter	370-386
T7 transcription start	369
His•Tag coding sequence	270-287
T7•Tag coding sequence	207-239
Multiple cloning sites (*Bam*H I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.

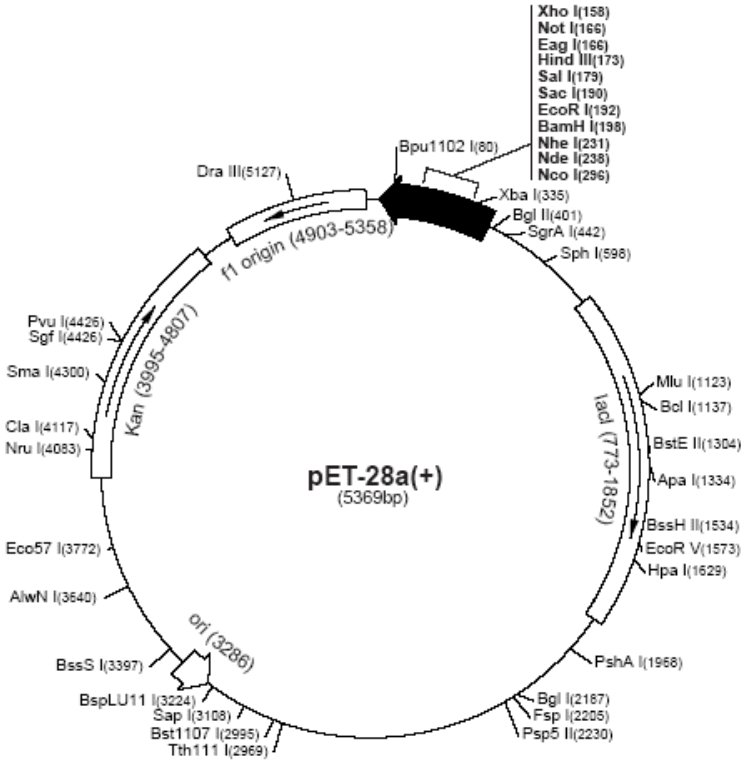

Figure II-6: pET-28a DNA vector map showing restriction locations.[II-7]

II.C.1.b Transformation and purification of the plasmid

Once the ligation/cloning was complete, the three ligation mixes of pET-28a/*pksD* was transformed into *E. coli* DH5α competent cells. Three different kanamycin plates were used to plate cells, which were incubated overnight at 37°C temperature. One of them (number 2) had grown a lot of colonies. Six single transformants were selected from the number 2 kanamycin plate to inoculate six starter cultures that were grown overnight in a shaker at 37°C. All six of the plasmids were purified the following day.

II.C.1.c Gel Electrophoresis

Another restriction digest was performed to screen for plasmids containing the insert. A gel electrophoresis showed successful ligation for all of them but one (see Figure II-7).

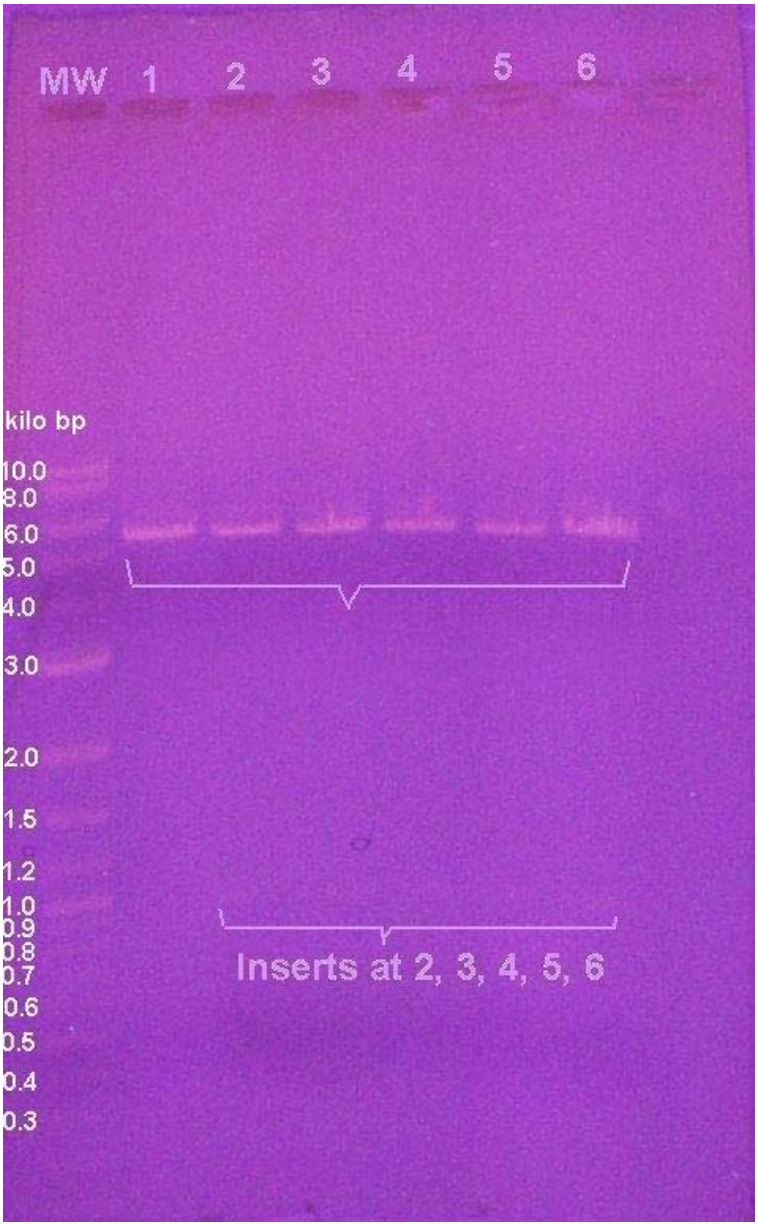

Figure II-7: Agarose gel showing good ligation inserts for (2), (3), (4), (5) and (6). The bands are indicated by brackets because they were not visible in the photo. Bands between 5000-6000 bp are the linearized pET-28a plasmid now lacking the insert (the known size of pET-28a is 5369 bp).

II.C.1.d Retransformation and Re-purification of the plasmid pET-28a/*pksD*

Once the cloning was done successfully, the vector was transformed into *E. coli* BL21(DE3) competent cells. Retransformation of plasmid number 6 was performed because it showed the brightest band in the screening experiment.

II.C.2 Overexpression and purification of pET-28a/*pksD*

Our target gene *pksD* had been successfully cloned in frame with a C-terminal His-tag coded by the plasmid pET-28a. The next step was to overexpress the protein using this clone. The overexpression strain pET-28a/*pksD* was grown in Luria Bertani (LB) medium supplemented with 30 µg of kanamycin per ml at 37°C in a shaker until the optical density at 595 nm (OD_{595}) was in between 0.500 and 0.600. Once the OD_{595} was 0.600, overexpression was induced by adding IPTG to a final concentration of 1.0 mM, and growth was continued for 3 hours. Cells were harvested by centrifugation (6,500 rpm for 30 minutes). Cells were resuspended in 20 mL of Binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris, pH 7.9) and were lysed by sonication. The lysate was cleared by centrifugation (11,000 *g* for 30 minutes).

To be sure that the extract had the desired protein, a Ni-NTA (Nickel-nitrilotriacetic acid) His-Bind resin chromatography was done in a small scale. An SDS-PAGE (Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis) was run on samples from the purification (see Figure II-8). A bright and large band appeared at ~ 31-34 kDa, which is consistent with the expected molecular weight of His_6-tagged PksD, which is 34,877 Da.

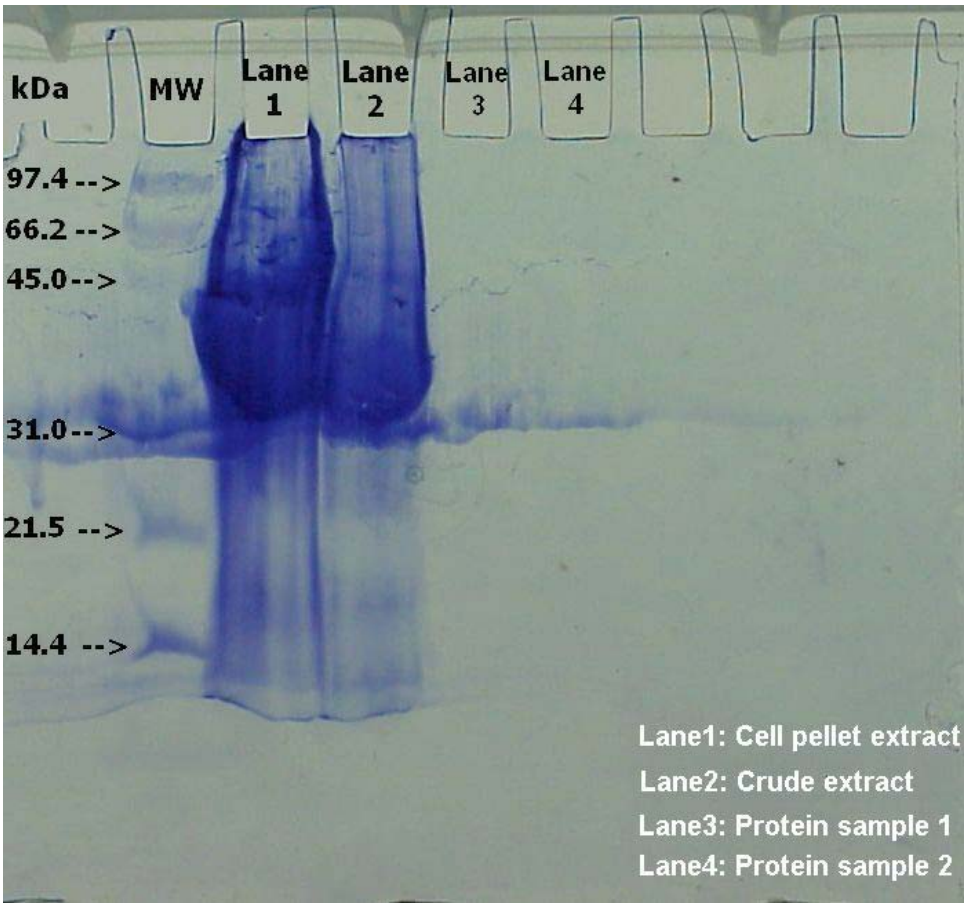

Figure II-8: Gel showing bright band at ~ 31-34 kDa.

Since the first attempt was successful, Ni-NTA chromatography was performed on a larger scale and SDS-PAGE was run on samples taken at various stages of the chromatography (see Figure II-9). This time bright bands appeared again for protein eluents.

The Ni-NTA His-bind resin chromatography was performed at room temperature to avoid protein precipitation, which we found to occur under normal refrigerated conditions. PksC, a close relative of PksD by homology showed the same problem of precipitation while elution. PksC was purified in our lab by Philip MacArthur, one of my colleagues. A dialysis of the eluted protein PksD was performed overnight in 50 mM tris-

HCl buffer at pH 8.8. The dialysed protein was concentrated by centrifugation using a Vivaspin 5000 Molecular Weight Cut Off (MWCO) concentrator for one hour at maximum speed. Sterile glycerol was added to the pure concentrated protein to the final concentration of 10% to keep the protein stable at -80° storage. The volume of glycerol plus concentrated PksD was 320 µL.

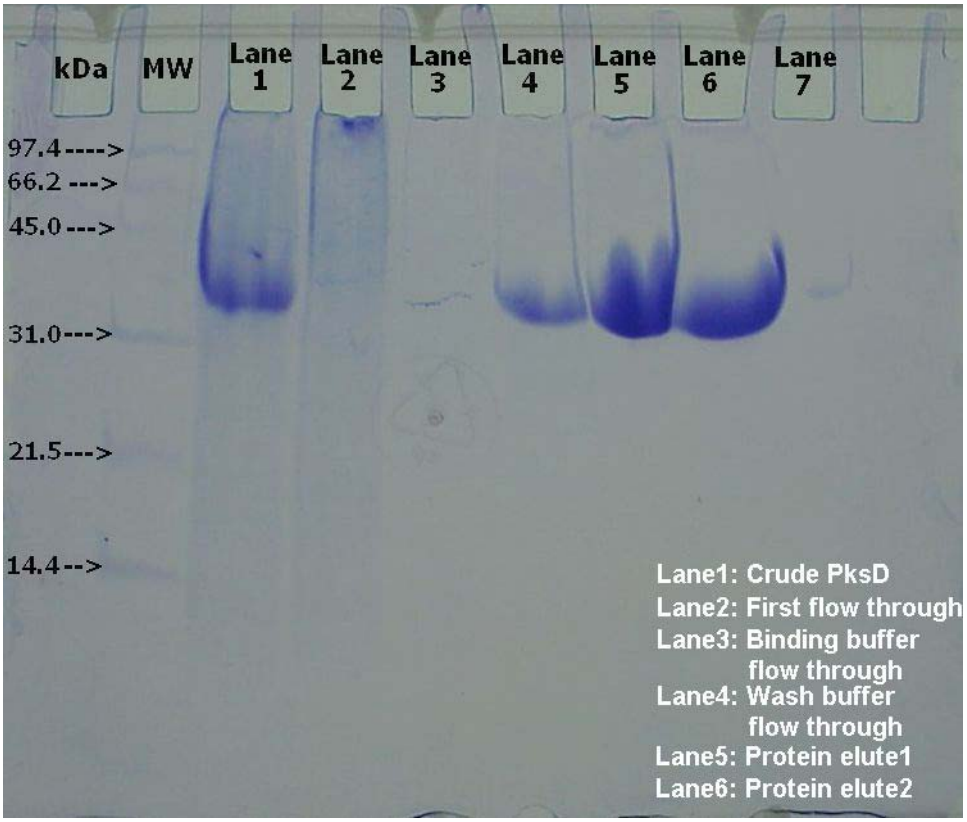

Figure II-9: Gel showing bands for large amount of proteins at ~ 31-34 kDa.

II.C.3 Protein concentration determination

The concentration of PksD in all samples were measured spectrophotometrically using the Bradford method, and with Bovine Serum Albumin (BSA) as the standard.

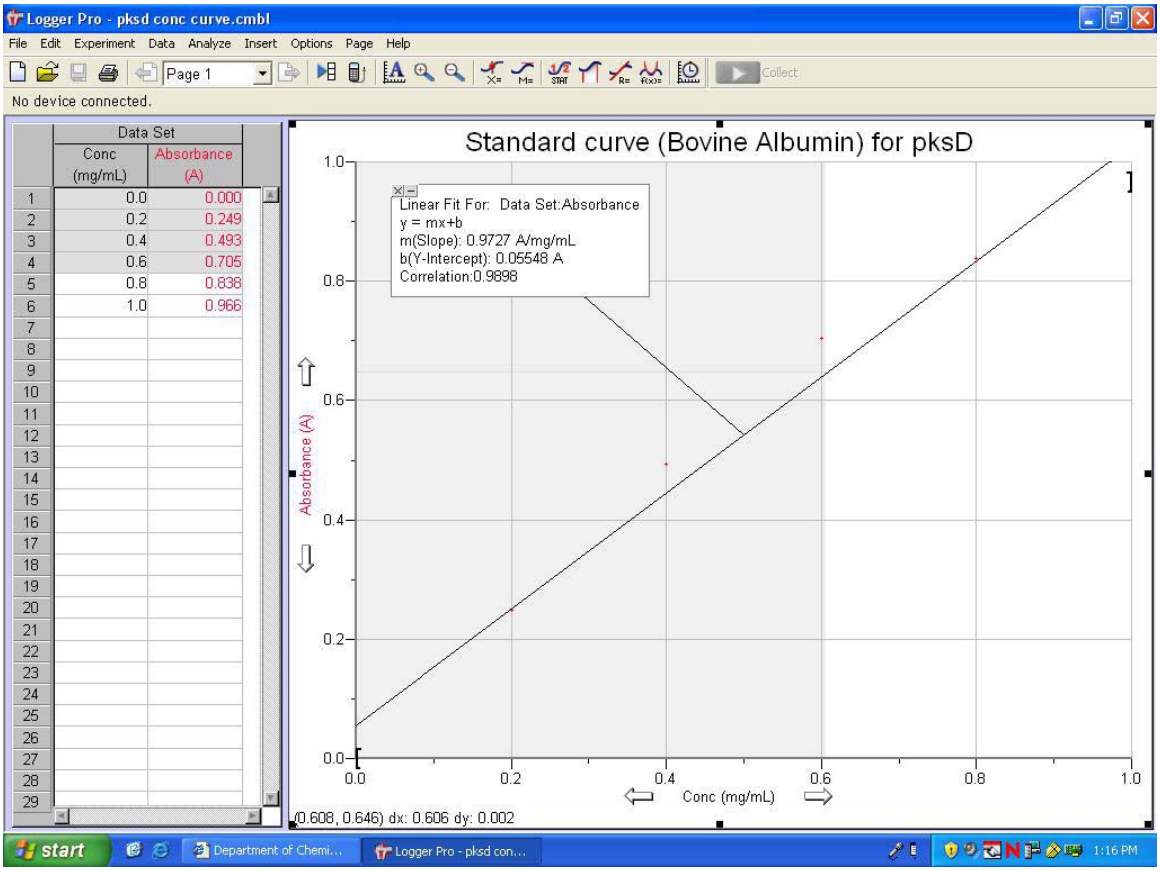

Figure II-10: BSA standard curve.

BSA standard, available commercially at a concentration of 2.0 mg/mL was used for the standardization curve (Figure II-10). OD_{595} was measured at varying concentrations of albumin. Ten, 100, and 1000 fold diluted PksD solution were prepared because we were measuring the concentration of PksD for the first time. Using the standard curve, a 100 fold diluted PksD sample was shown to have a concentration of 0.608 mg/mL, showing that the actual concentration of PksD was 60.8 mg/mL. This showed that ~39 mg pure PksD protein can be prepared from a 1L culture.

II.C.4 Testing PksD activity

The hypothesis for enzyme activity of PksD (see Figure II-4) requires the holo-form of AcpK as a substrate to react with PksD in the malonyl transfer reaction. The reaction for changing the inactive apo-AcpK to the active holo-AcpK is known as phosphopantetheinylation because the phosphopantetheinyl moiety from coenzyme A becomes attached to apo-AcpK.[II-8] Sfp, having a wide range of ACP-substrate specificity, acts as a catalyst in this reaction. The leaving group is 3′, 5′- ADP. Then, PksD is expected to catalyze the formation of malonyl-AcpK from holo-AcpK and malonyl CoA.

Product	Expected mass
Apo-AcpK	10561 Da
Holo-AcpK	10901 Da
Acetyl-holo-AcpK	10944 Da
Malonyl-holo-AcpK	10985 Da

Table II-1: Theoretical mass of whole AcpK proteins.

Before starting experiments with PksD, it was necessary to validate MALDI-MS as a suitable analytical tool for measuring any changes on AcpK. To perform the validation test, the broad substrate specificity of Sfp for CoA-thioesters was exploited. It was also important to generate a standard of expected products before performing experiments with PksD.

Low resolution MS data were obtained (Figure II-11 – II-13) repeatedly from the samples of phosphopantetheinylation reaction, showing insufficient evidence to conclude for the formation of holo-AcpK. While receiving low resolution MS data for the whole AcpK, we decided to experiment with trypsin digested AcpK.

Products	Expected mass of tryptic fragment
Apo-AcpK	1919 Da
Holo-AcpK	2259 Da
Malonyl-AcpK	2344 Da

Table II-2: Theoretical mass of trypsin digested AcpK proteins.

Sequence alignment homology shows that the tryptic peptide fragment $F^{26}EPEDQLVELGADSVDR^{42}$ has a single serine residue Ser^{39}, which is the probable site of reaction. However we did not observe any evidence of phosphopantetheinylation in any tryptic fragment from MALDI-MS analysis (Figure II-14 – II-16).

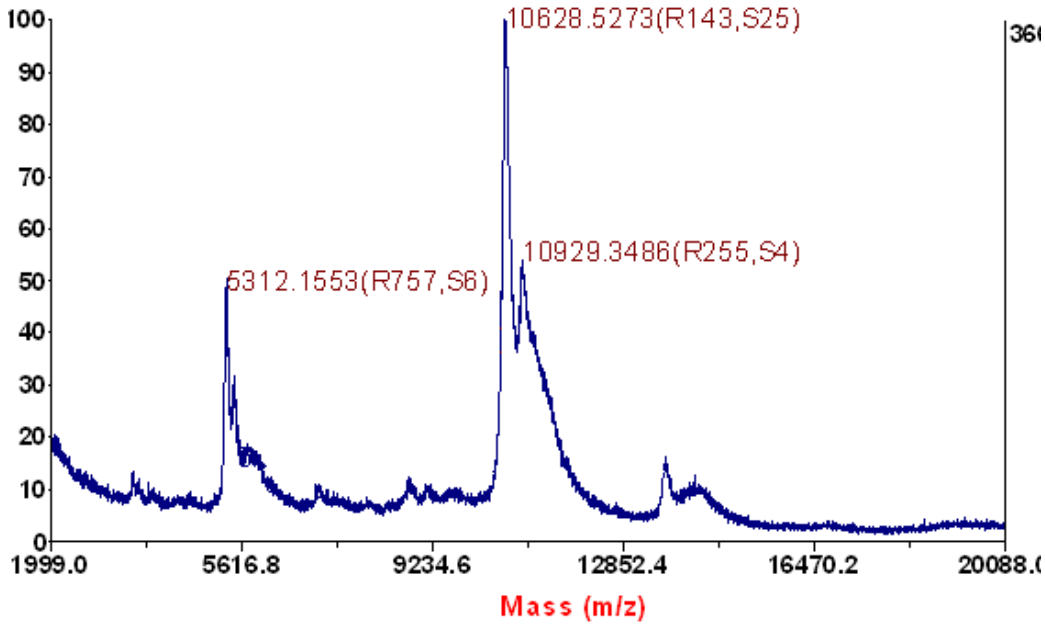

Figure II-11: Mass spectrum of control (apo-AcpK).

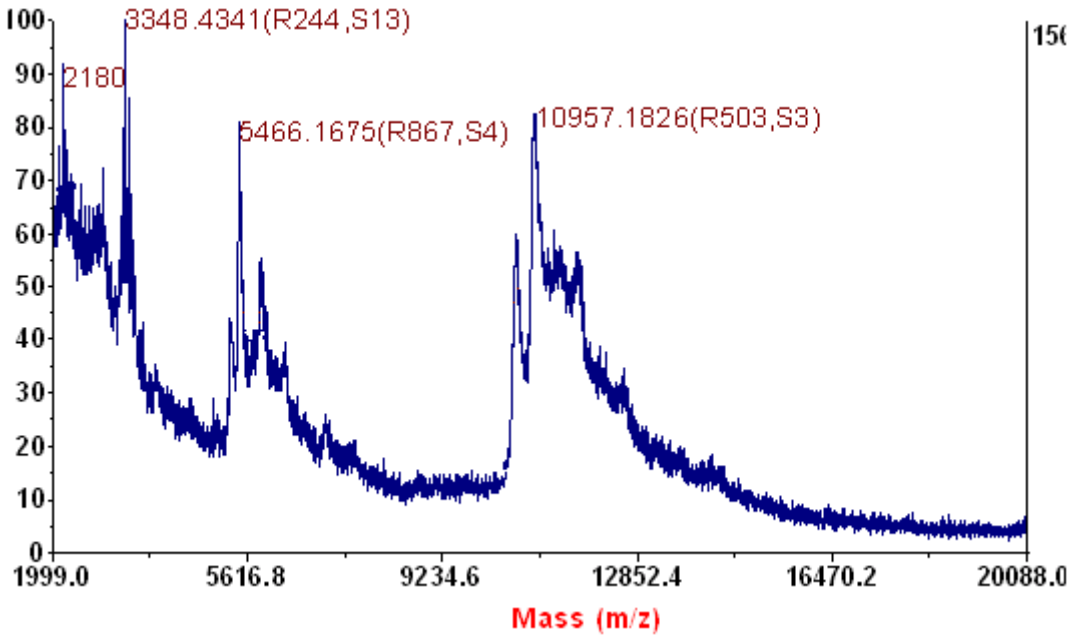

Figure II-12: Spectrum of holo-AcpK.

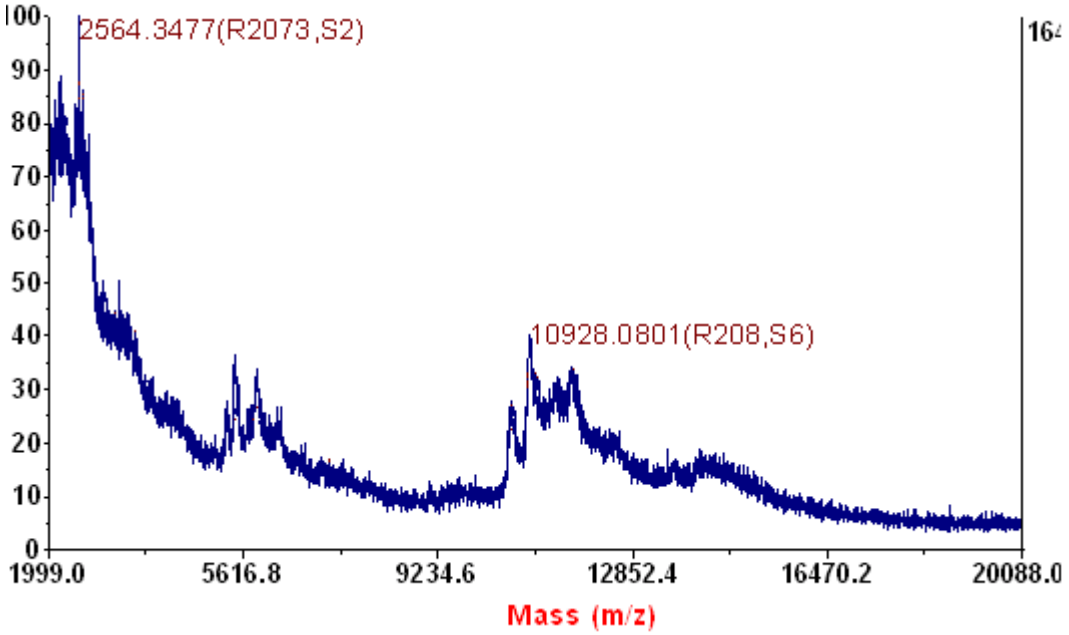

Figure II-13: Spectrum of holo-AcpK treated with PksD + Malonyl CoA.

Spectra of trypsin-digest experiments on phosphopantetheinylation reaction are shown below.

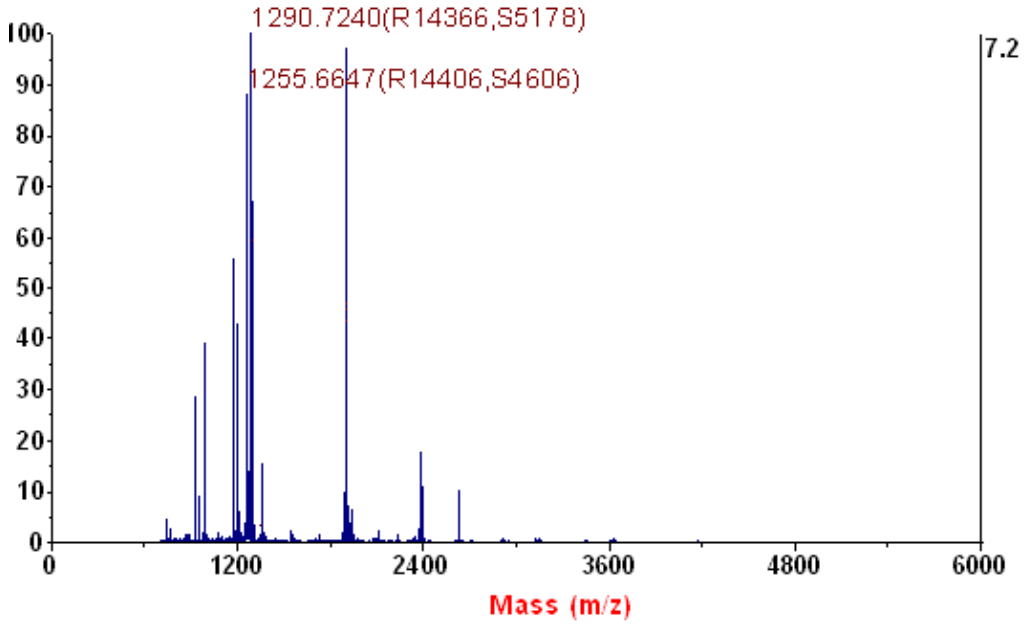

Figure II-14: Spectrum of trypsin digested apo-AcpK + Sfp in matrix A.

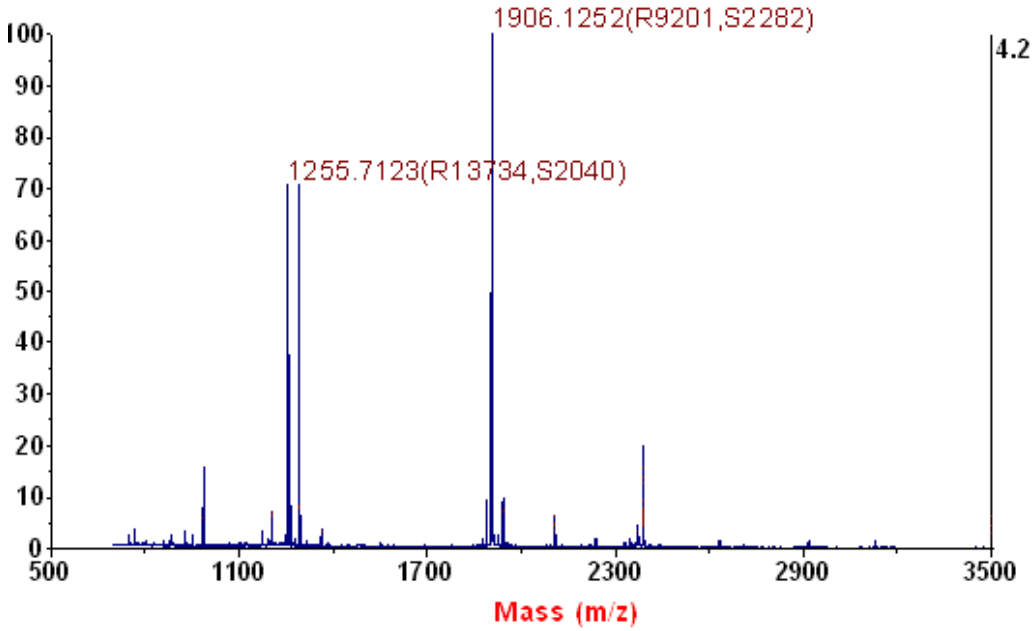

Figure II-15: Spectrum of trypsin digested apo-AcpK + Sfp + CoA in matrix A.

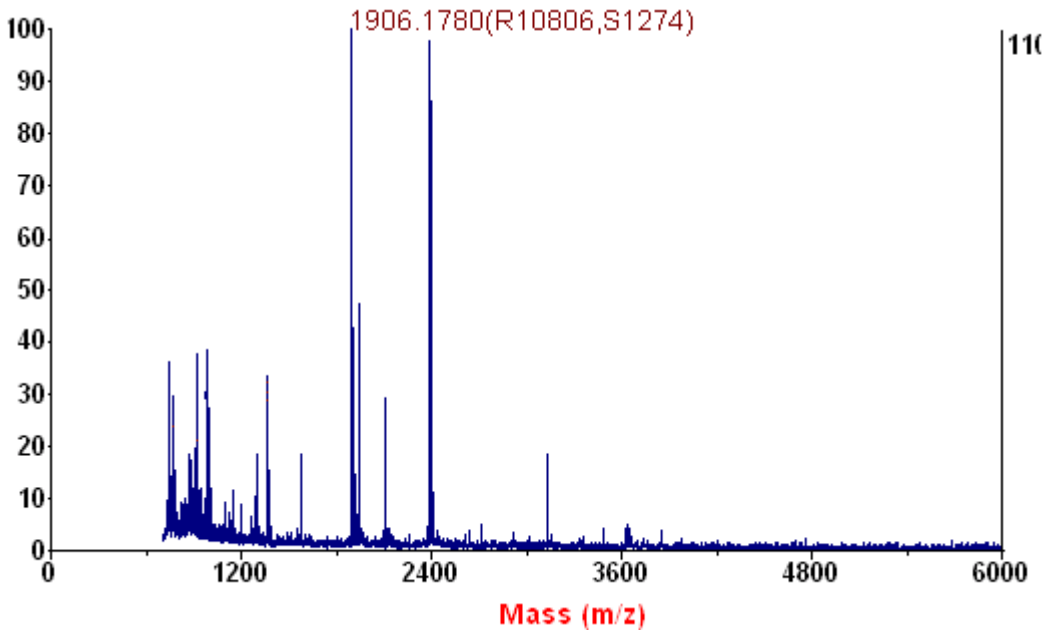

Figure II-16: Spectrum of trypsin digested apo-AcpK + Sfp + CoA in matrix B.

II.D Conclusion

Purified His-Tagged PksD from *B. subtilis* strain 168 could successfully be produced with the use of pET-28a overexpression plasmid vector. Ligation was quite efficient with T4 DNA ligase. Regulation of overexpression of PksD-pET-28a was performed by IPTG induction. Overexpressed PksD was purified by Ni-NTA chromatography because PksD was cloned to include a C-terminal His-Tag coded by the pET-28a plasmid. 1M imidazole was used to successfully elute PksD, yielding large amounts of pure PksD.

MALDI-MS studies were conducted in two steps. The first step was to demonstrate the presence of holo-AcpK, which was the main substrate of the next PksD-catalyzed reaction. If phosphopantetheinylation happens, a mass increase of 340 Da is expected in the product molecule. It was clear that the mass was increasing, but the mass

amount was inconsistent in each run by the MALDI-MS, even from the same sample. Since we could not consistently produce good MS-data for holo-AcpK, further reactions with PksD could not be evaluated. Attempts to solve the resolution problem by using trypsin were unsuccessful because the laser was likely cleaving the phosphopantetheinyl group from the active serine[39].

II.E Experimental section

II.E.1 Cloning and overexpression of PksD

The PCR to amplify PksD used *B. subtilis* strain 168 genomic DNA as the template and 5′ AAG GTT ATA CCC ATG GAT GAA CCG CTT G 3′ (inserts an NcoI site at the start) and 5′ GAA AGA CAT ATG CTC GAG TCT TGT AAA CTT CCT TTC TG 3′ (inserts an XhoI site downstream) as the primer pair. *Phusion* polymerase was used to conduct the PCR with the following conditions: 2 minutes at 98°C; 30 cycles of 30 seconds at 98°C, 30 seconds at 63°C, and 1 minute at 72°C; then a hold of 5 minutes at 72°C. The resulting PCR product and plasmid pET-28a were digested separately with restriction enzymes NcoI and XhoI for 3.5 hours. A low melting 1.0% agarose gel electrophoresis was performed to separate the digest fragments by size and visualized with UV light. Plasmid and PCR product/insert fragments were excised and melted in a hot water bath to perform ligation reaction using T4 DNA ligase and T4 DNA ligase buffer. The following three different reactions at varying insert/plasmid ratios were performed overnight:

The insert: plasmid ratio for the 1st reaction was 5µL:1µL, for the 2nd reaction 5µL:5µL and for the 3rd reaction 1µL:5µL. Three overnight ligation reactions of pET-

28a/PksD were transformed into *E. coli* DH5α competent cells and plated onto three LB-agar plates containing kanamycin of final concentration 30µg/mL. Ligation #2 had the most number of colonies. Six single transformants were selected and starter cultures were grown overnight in 5mL LB with kanamycin of final concentration 30µg/mL in a 37°C shaker, having 225 rpm. All six of the plasmids were purified from these cultures by a Qiagen Miniprep kit the following day, according to the manufacturer's instruction manual. These plasmids were screened by another restriction digest using the same restriction enzymes as above. The digests were performed for 3.5 hours and a 1.0% agarose gel electrophoresis was performed to screen for the inserts in the plasmid vector. Once the cloning was done successfully, the vector was transformed into *E. coli* BL21(DE3) competent cells. Retransformation of plasmid #6 was performed because it showed the brightest band in the screening experiment, confirming the correct size and orientation of *pksD* in plasmid pET-28a.

One colony of this clone was selected and grown overnight in a starter culture of LB and 30µg/mL concentration of kanamycin. One liter of LB containing 30µg/mL concentration of kanamycin was inoculated with 2mL of starter culture and grown in a 37°C shaker (225 rpm) until OD_{595} reached between 0.500 and 0.600. IPTG was added to a final concentration of 1.0mM and growth was continued for another 3 hours. After 3 hours, the cells were harvested by centrifugation (6,500 rpm for 30 minutes). The pellet was stored at -80°C.

II.E.2 Purification of PksD

Cells obtained from 500mL LB culture were resuspended in 20mL of binding buffer (5mM imidazole, 500mM NaCl, 20mM Tris-HCl, pH 7.9) and were lysed by sonication on ice for 3 minutes. The lysate was cleared by centrifugation (11,000 rpm for 30 minutes) and added to the Ni-NTA chromatography column containing 2mL column bed. Since we found that precipitation of eluted PksD occurred at 4°C, the chromatography was conducted at room temperature. After loading the crude extract, the column was washed by ten volumes of binding buffer. Six volumes of wash buffer (60mM imidazole, 500mM NaCl, 20mM Tris-HCl, pH 7.9) was used to remove non-specific proteins from the column. Six volumes of 1M imidazole was used to elute PksD from the column. Snakeskin® pleated dialysis tubing, having 7000 MWCO was used for overnight dialysis of the eluted protein in 4L volume of 50mM Tris and 2mM DTT (Dithiothreitol) at pH 8.8. The dialysed protein was concentrated by centrifuging using a Vivaspin 5000 MWCO concentrator for one hour at maximum speed. A 12% SDS-PAGE was run to estimate the protein purity.

II.E.3 Measurement of concentration of PksD

The concentration of PksD was measured spectrophotometrically using the Bradford method. Samples of Bovine Serum Albumin (BSA) were used in the standard curve, employing a range of concentrations between 0 mg/mL and 1 mg/mL. OD_{595} was measured at varying concentration of albumin. Ten, 100, and 1000 fold diluted PksD solution were prepared because we were measuring the concentration of PksD for the

first time, and was expected to be at concentration, outside of the useful range for Bradford method.

II.E.4 Preparation of AcpK/pQE60 M15[pREP4]

AcpK/pQE60 M15[pREP4] was overexpressed in 1L LB, ampicillin concentration of 25μg/mL and kanamycin concentration of 30μg/mL. The culture was grown until OD_{595nm} reached 0.6-0.7 and IPTG was added to a final concentration of 0.25 mM. The culture was grown for 3 hours after induction and cells were harvested by centrifugation. Purification of AcpK was performed as above for PksD, except that AcpK was eluted with 200mM imidazole, and performed at 4°C. AcpK was dialyzed overnight at 4°C in 50mM Tris buffer at pH 8.0. Plasmid and protocols for overexpression and purification of AcpK/pQE60 M15[pREP4] were provided by Jayme Williams, one of my colleagues.

II.E.5 Preparation of Sfp/pET-28a BL21(DE3)

Overexpression of Sfp/pET-28a BL21(DE3) was performed in 1L LB and kanamycin concentration of 30μg/mL. The culture was grown until OD_{595} was 0.5 to 0.6 and IPTG was added to final concentration of 1mM. The culture was grown overnight after induction and cells were harvested by centrifugation. Purification of Sfp was performed as above for AcpK. Sfp was dialyzed overnight at 4°C in 25 mM Tris buffer at pH 8.0. Plasmid and protocols for overexpression and purification of Sfp/pET-28a BL21(DE3) were provided by Jayme Williams, one of my colleagues.

II.E.6 MALDI-TOF MS analysis

Concentration of AcpK and Sfp were measured spectrophotometrically as above using the Bradford method and BSA standard curve. Reactions were prepared in which the final AcpK and Sfp concentrations were 20μM and 0.8μM, respectively. The reaction buffer composition was 50mM Tris and 12.5mM $MgCl_2$ at pH 8.0. Reactions were prepared by mixing the buffer, AcpK, Sfp and Coenzyme A (CoA). A control reaction lacking CoA was prepared by adding the buffer, AcpK, and Sfp in proportion. Both of the reaction and the control were incubated at 37°C for 2 hours. Matrix A for MALDI samples was prepared by adding 8mg α-cyano-4-hydroxycinnamic acid 99% (CHCA) and 1mL of matrix A diluent (0.3% aq. trifluoroacetic acid in 50% acetonitrile), then vortexed for 20 seconds, followed by centrifugation, and using the supernatant. Matrix B was prepared by adding 8mg sinapic acid 98% and 1mL of matrix B diluent (0.3% aq. trifluoroacetic acid in 30% acetonitrile), then vortexed for 20 seconds, followed by centrifugation, and using the supernatant.

MALDI samples were prepared by mixing 24μL matrix A and 1μL reaction sample. A calibration sample was also prepared by mixing 24μL matrix and 1μL calibration mixture (commercially purchased). Each sample (0.5μL) was spotted on the MALDI sample loading plate and were dried at room temperature. The mass of the various forms of AcpK in the samples was determined using MALDI-MS, on an Applied Biosystems 4700 Proteomics Analyzer at the University of North Carolina Greensboro. Positive-ion mass spectra were recorded in both the linear and reflective modes. For the

next step involving PksD, the final concentration of 1.2μM of PksD was added to the reactions obtained from the phosphopantetheinylation step and incubated for another hour.

In a different approach towards characterizing PksD, samples were digested by 0.2μL trypsin for 1 hour. The reaction buffer was composed of 50mM Tris and 12.5mM $MgCl_2$ at pH 8.0. In order to generate estimation standard for the expected products, three different control reactions were prepared by adding trypsin just with the buffer, AcpK + buffer, and AcpK + Sfp + buffer. An internal calibration was done and masses of the fragments were detected by MALDI-MS.

CHAPTER III

CHARACTERIZATION OF PksG IN *B. SUBTILIS* BY KNOCKOUT MUTATION

III.A Introduction

Even after a decade since the full genome sequence of *B. subtilis* strain 168 had been reported [III-1], there is no genetic data available for *B. subtilis* strain 39374, which is a known difficidin producer. *B. subtilis* strain 168 is known to contain a *pks* operon and has never been shown to synthesize difficidin. However, difficidin was isolated from *B. subtilis* strain 39374. While we have assumed that this strain has *PksFGHI* homologs, but this strain has not yet been sequenced. One goal of this project was to sequence a portion of the *pks* operon from *B. subtilis* strain 39374.

Cloning, overexpression and purification of PksG from strain 168 had already been conducted in our lab.[III-2] In order to characterize the function of PksG in the synthesis of difficidin, two hypotheses have been proposed. Our main hypothesis was that PksG has a role for the biosynthesis of the unusual subunit at C3 in difficidin. The first hypothesis (Figure III-1) involves PksG in the synthesis of an unusual polyketide precursor, glutaconyl-AcpK. Then this precursor is incorporated into the immature difficidin molecule, and difficidin synthesis is completed by the mechanism of standard polyketide biosynthetic chemistry. The second hypothesis (Figure III-2) involves the formation of a ketone intermediate and PksG is then proposed to act on this intermediate in one of the steps towards the formation of difficidin.

Both of these hypotheses involve PksG playing an important role in biosynthesizing difficidin. If PksG is essential to synthesize an important intermediate in difficidin production, then construction of a mutant of PksG (39374ΔPksG) is expected to disrupt the formation of difficidin, no matter which hypothesis is correct.

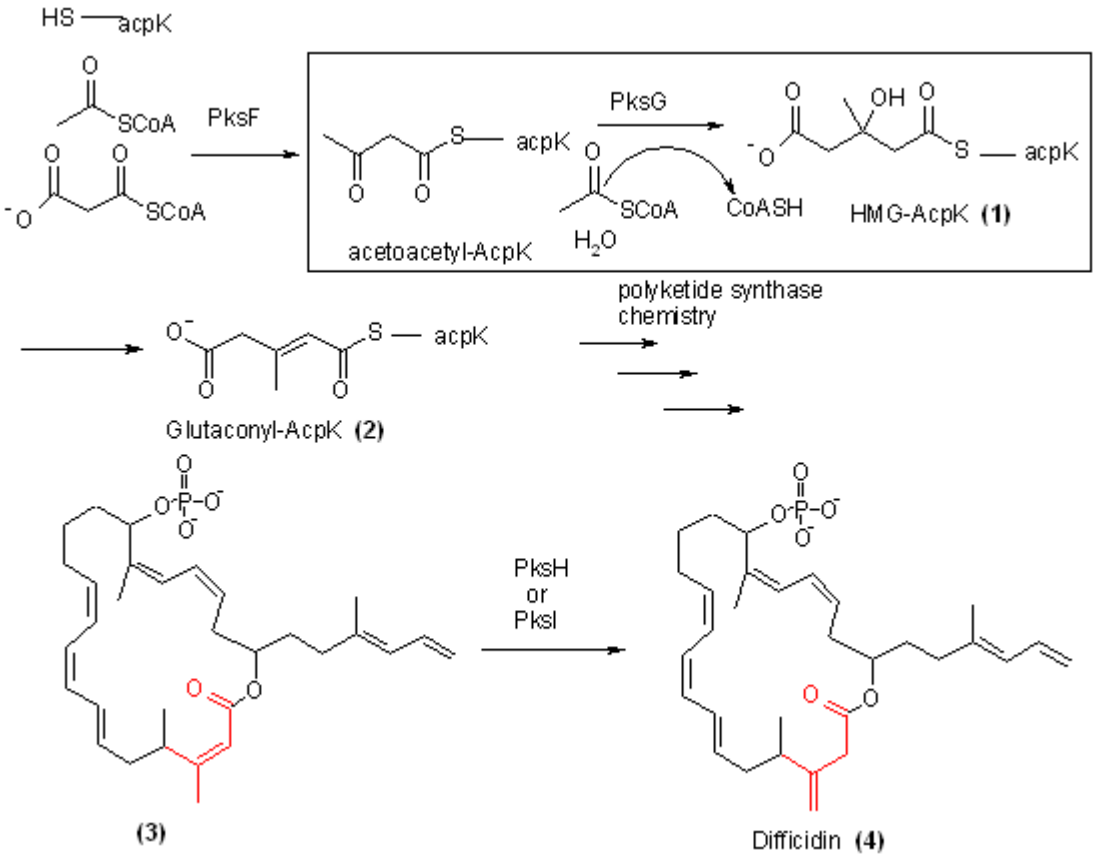

Figure III-1: Unusual precursor formation in difficidin - hypothesis 1.

If hypothesis 1 is correct and PksG catalyzes the formation of glutaconyl-AcpK (**2** in Figure III-1), then the mutation of *pksG* would stop production of **2**. If **2** is an essential intermediate for difficidin production, then the disruption of *pksG* would halt difficidin biosynthesis. This would leave unfinished polyketides attached to each polyketide system

module, and after *in situ* hydrolysis, linear polyketides that are fragments of the final product would be released.[III-3] Therefore, if hypothesis 1 is correct, small fragments of linear polyketides corresponding to difficidin should be detected in extracts of 39374ΔPksG by Liquid Chromatography-Mass Spectrometry (LC-MS).

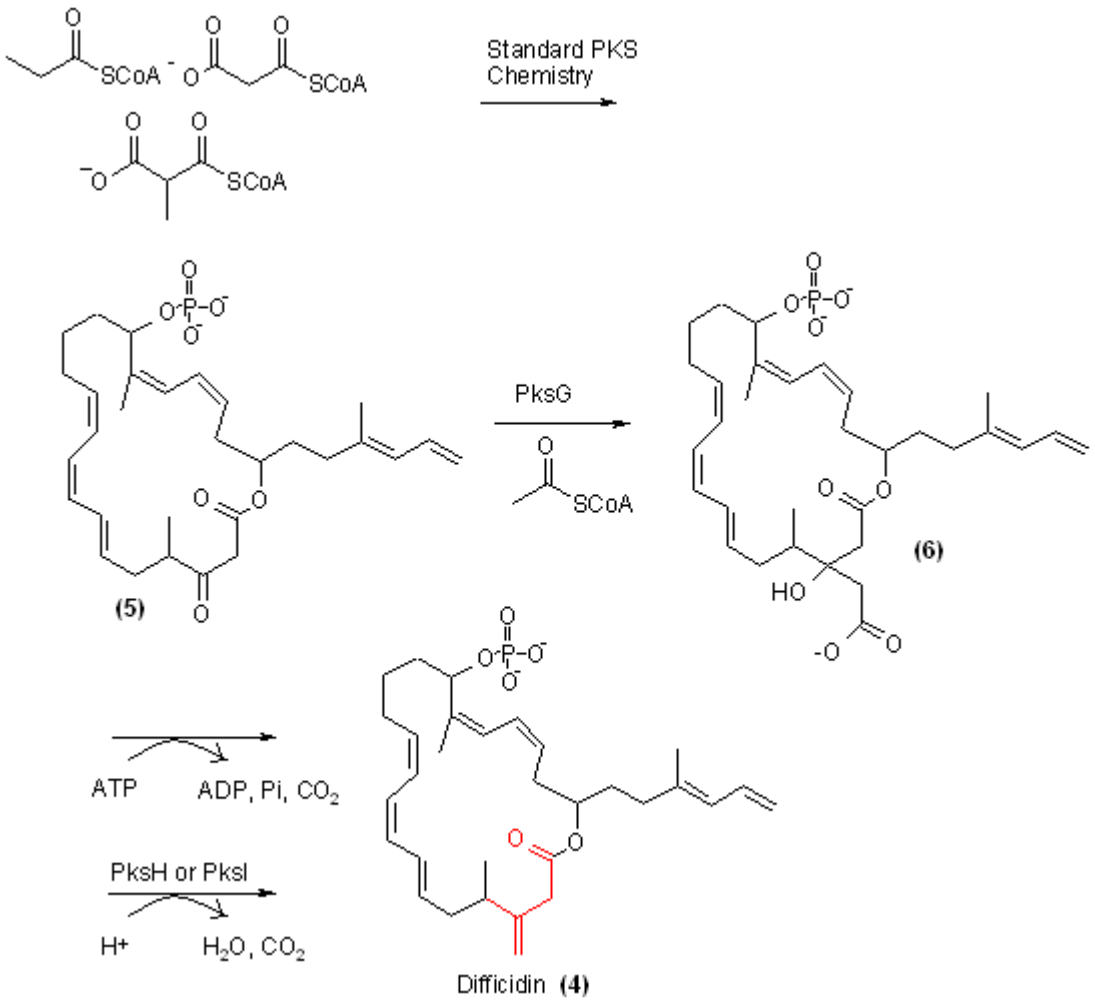

Figure III-2: Unusual precursor formation in difficidin - hypothesis 2.

Alternatively, a ketone intermediate (**5** in Figure III-2) is synthesized in hypothesis 2 and PksG catalyzes the aldol condensation of **5** and acetyl CoA. The mutant

lacking *pksG* would not catalyze the aldol condensation. So if hypothesis 2 is correct, **5** instead of difficidin would be detected by LC-MS.

III.B Sequencing of the *pksG* homology from strain 39374

Cloning of *pksG* started with degenerate primers (DEGEN) (see Table III-1). The degenerate primers were first reported by Edwards *et al.*[III-15] Degenerate primers are used because of the lack of sequence data of 39374. LPYEDPV and MVKGAHR, two conserved amino acid sequences that have been found for hydroxymethylglutaryl CoA synthases (HMGCS) in polyketide synthesis pathways, are the basis for the design of these primers.[III-8] Since any particular amino acid can be coded by more than one codon, a degenerate mixture of primers can be used to amplify a segment of *pksG* from strain 39374 even if sequence knowledge is lacking for 39374. PCR using DNA sequence derived from the two conserved amino acid sequences should produce a product of length 680 bp. PCR with these primers successfully amplified a gene length approximately of 700 bp. The PCR produced a fragment which contained *pksG* and had single deoxyadenosine (A) at the 3′ end of the product. The *Taq* polymerase naturally adds these 3′ deoxyadenosine to the ends of PCR products, which are needed for the *TOPO* ligation. The *pBAD–TOPO* vector is linearized and the topoisomerase is attached to the vector. The vector contains single, overhanging 3′ deoxythimidine (T) residues. The 3′ overhangs allow the PCR inserts to ligate efficiently with the vector. However, the PCR product can be cloned in either direction because the overhangs are the same on each end of the product. The actual cloning is done by simply adding the PCR product to the *pBAD-*

TOPO vector, and incubating at room temperature. The topoisomerase then ligates the PCR insert into the vector. This procedure is very efficient, taking only a few hours.

Primer	Sequence
pksG MUTIN UP	5' GCCAAGCTTACTTAGGCCTCAACCGGAAC 3'
pksG MUTIN DN	5' CGGGATCCCAATGACAAATCGGCATCAC 3'
pksG DEGEN UP	5' CTNCCNTAYGARGAYCCCGT 3'
pksG DEGEN DN	5' NCKRTGNGCNCCYTTNACCAT 3'
lacI UP	5' ATGGCGGAGCTGAATTACAT 3'
lacI DN	5' GCATTAATGAATCGGCCAAC 3'
pBAD UP	5′ ATGCCATAGCATTTTTATCC 3′
pBAD DN	5′ GATTTAAATCTGTATCAGGC 3′

Table III-1: List of sequence of primers,

where N = A, G, C, T; Y = C or T; R = A or G; K = G or T.

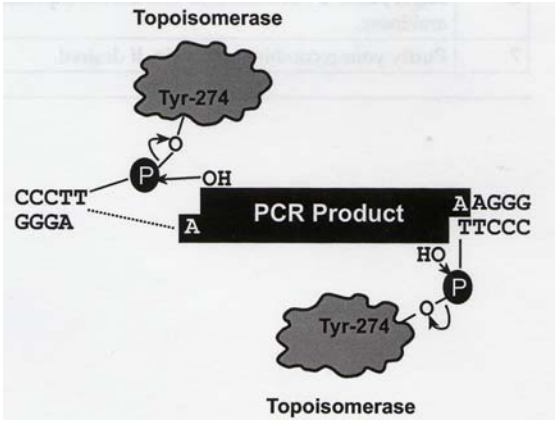

Figure III-3: Cloning by topoisomearse.[III-4]

Topoisomerase I is naturally used in replication and transcription to unwind supercoiled DNA. The cleavage occurs at a specific site because the enzyme recognizes the sequence 5′-CCCTT[III-4]. Once the topoisomerase is in the correct location a tyrosyl residue (Tyr-

274) acts as a nucleophile, and attacks the phosphate group in the DNA backbone and creates a phosphodiester linkage between the topoisomerase and PCR insert. Once the cleavage is complete, the ligation can occur between the bonds that were originally cleaved. The same chemistry occurs in DNA replication. Ligation with a new acceptor DNA can occur as long as the acceptor DNA has a 5′ hydroxyl tail complementary to the cleavage product. The phosphotyrosyl bond between the DNA and enzyme can subsequently be attacked by the 5′ hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase. After cloning, the plasmid pBAD/*pksG* was purified and sent for sequencing.

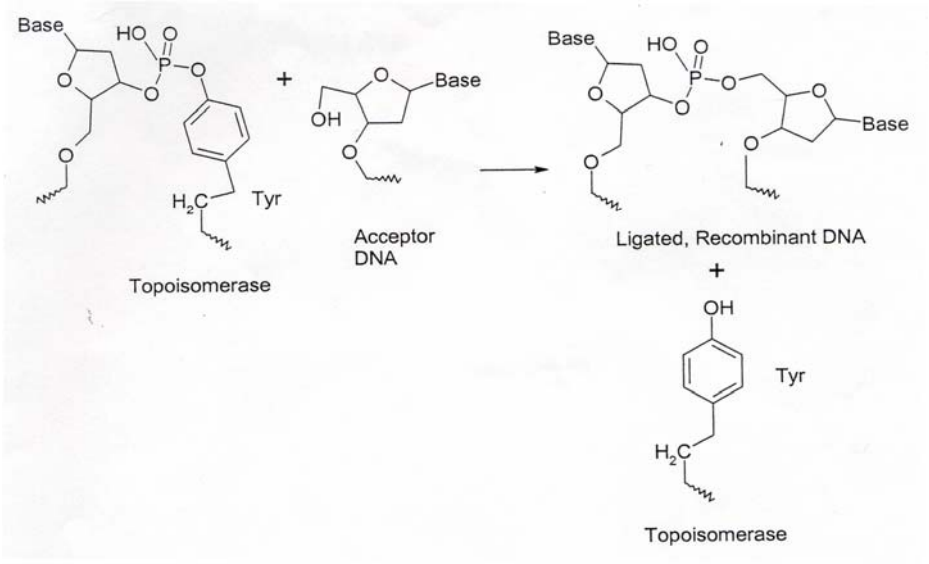

Figure III-4: The reaction catalyzed by topoisomerase.[III-5]

III.C Construction of 39374ΔPksG

III.C.1 Goals

In previous work PksG was successfully cloned from strain 168 genomic DNA, which was used because complete genomic data is available for this strain. The strain

39374 has not yet been sequenced, however it was assumed that it had a *pksG* homolog. Difficidin was first isolated from this *B. subtilis* strain 39374. Since *pksG* is assumed to be present in 39374, a mutant lacking the *pksG* gene (39374ΔPksG) could be synthesized from strain 39374. Once we could synthesize the mutant PksGΔ39374, the role of PksG in the production of difficidin can be investigated.

III.C.2 pMUTIN4

pMUTIN4 is a vector (see Figure III-5), which was designed for systematic gene inactivation in *B. subtilis*.[III-6] pMUTIN4 allows the knockout mutation of any known coding sequence. A section of a gene of interest is inserted into the multiple cloning site (MCS) and then transformed into the host strain. After a successful transformation, the recombinant plasmid integrates into the chromosome by homology to the insert and a knockout mutation is produced.

With the knowledge of sequencing of the *pksG* homolog, new primers complementary to this were designed to clone a ~380 bp fragment into pMUTIN4. All pMUTIN plasmids contain a *lacZ* reporter gene (see Figure III-5) that allows the measurement of gene expressions in *B. subtilis*[III-9]. A tightly regulated inducible p_{spac} promoter allows 100-fold induction of genes downstream from the mutation. pMUTIN has an erythromycin-resistance gene (Em^R) for selection in *B. subtilis* and a β-lactamase gene (Ap^R) for selection in *E. coli*. pMUTIN also contains a modified *lacI* gene that becomes expressed in *B. subtilis*. *lacI* represses P_{spac} and therefore P_{spac} can be induced by IPTG.[III-10]

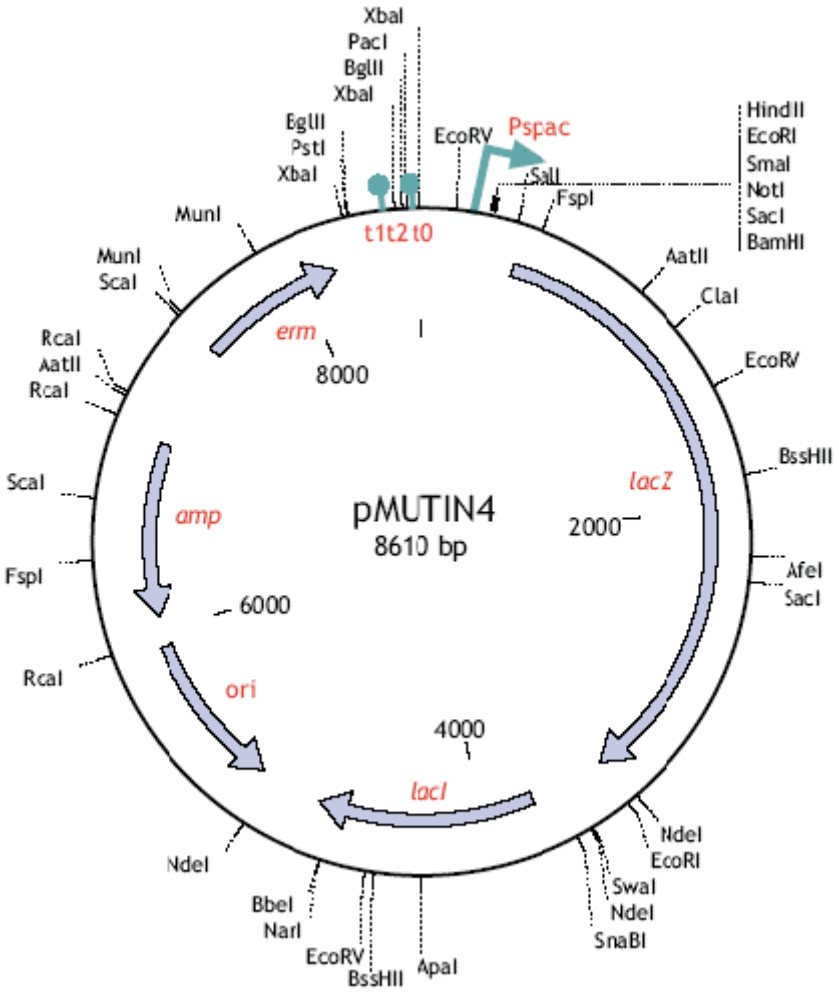

Figure III-5: Map of plasmid pMUTIN4.[III-6]

pMUTIN can be inserted into the target gene of interest by a crossover reaction (see Figure III-6). Two restriction enzymes are employed to form an internal fragment of the target gene, follwed by PCR amplification and cloning in pMUTIN at the multicloning site. The resulting plasmid is used to transform the *B. subtilis*. Once the homologous recombination occurs between the promoter and the *lacZ* gene, the target gene is disrupted. The disruption occurs because pMUTIN cannot replicate in *B. subtilis*,

and the antibiotic pressure can be released only through recombination with the genomic DNA which leads to insertional mutagenesis.

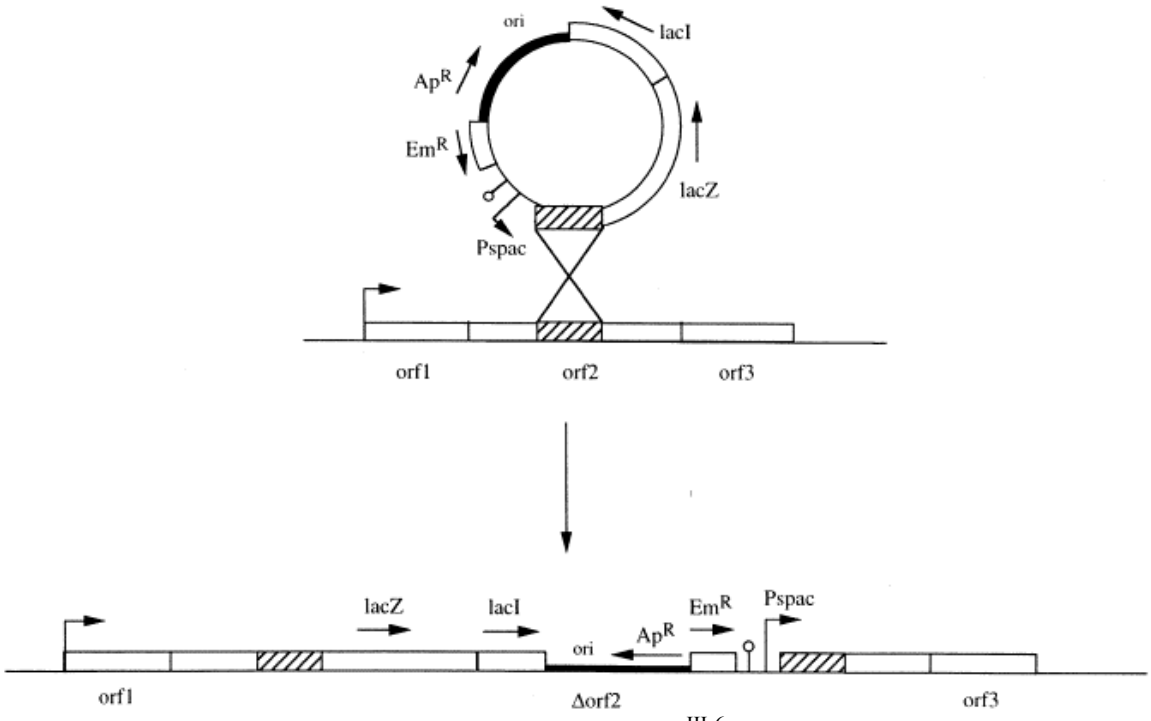

Figure III-6: Integration of pMUTIN into target gene.[III-6]

B. subtilis does not naturally contain *lacI*, but a successful mutagenesis incorporates *lacZ*, *lacI*, and erythromycin resistance genes. A screening PCR is conducted by amplifying *lacI* from genomic DNA to validate if the PksGΔ39374 mutation occurred.

III.D Results

III.D.1 Sequencing of *pksG* from strain 39374

A successful PCR was performed on 39374 genomic DNA with Taq polymerase, using the degenerate primers, and the gel is shown in Figure III-7.

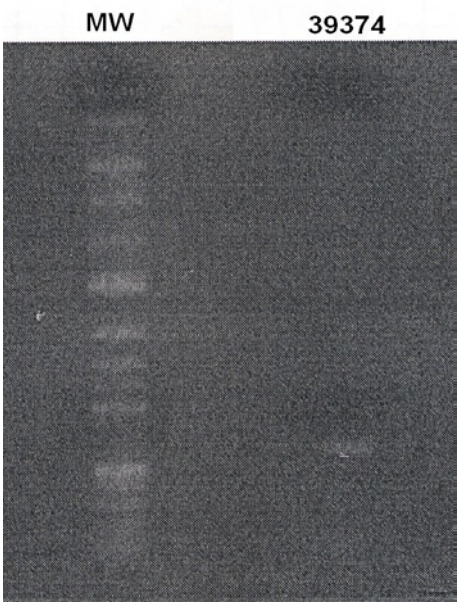

Figure III-7: Agarose gel using Taq on *B. subtilis* strain 39374, showing a band at ~650 bp.

Once the TOPO cloning was done, the pBAD-insert vector was transformed into the *E. coli* TOP10 strain. Six single transformants were selected from the LB-ampicillin plate, and then six starter cultures were grown. Purification of the *pBAD* plasmid was done for all six starter culture using the QIAPrep plasmid purification kit. A PCR was performed using TOPO cloning components and subsequently an agarose gel electrophoresis was performed. A bright band appeared at around 650 bp for just one of them. This indicated that a *pksG* homolog must be present in strain 39374 because the experimental band size matched with that of the calculated size. A figure is not provided since the photograph of the gel was not clear.

Retransformation of plasmid number 2 was performed because it showed a successful ligation of *PksG* in plasmid *pBAD*. Absorbance of the purified *pBAD* plasmid at 260 nm was 0.052. Therefore the final concentration of the DNA was 260.0 µg/mL.

The whole 50 µL sample was sent for sequencing. Below is the analyzed sequence data from strain 39374.

The analyzed DNA sequence data for PksG 39374 was:

TACGAGGATCCCGTTACATTCGGCGTTAACGCGGCAAAGCCGATTATCGACC
GATTGACAGAGGCTGAAAAAGACAGAATCGAGCTGCTGATTACCTGCTCGGA
GTCCGGCATTGACTTCGGAAAATCACTCAGTACATATATTCATGATCACTTAG
GCCTCAACCGGAACTGCCGTTTATTTGAACTGAAACAAGCCTGTTATTCAGGA
ACGGCGGGCTTGCAGATGGCGCTTAACTTTATTTTGTCGCAGACATCCCCCGG
AGCGAAGGCGCTTGTCGTCGCGACGGATATTTCCCGGTTTTTAATTGCCGAAG
GCGGCGATGCGTTAAGCGAGGATTGGTCTTACGCAGAACCGAGCGCCGGAGC
GGGTGCCGTAGCCTTATTGATCGGTGAAAATCCCATCGTATTTCAGGCGGATG
CGGGCGCCAACGGCTATTACGGCTATGAGGTGATGGATACATGCCGTCCGAT
CCCGGACAGTGAAGCCGGTGATGCCGATTTG

The translated sequence was:

YEDPVTFGVNAAKPIIDRLTEAEKDRIELLITCSESGIDFGKSLSTYIHDHLGLNRN
CRLFELKQACYSGTAGLQMALNFILSQTSPGAKALVVATDISRFLIAEGGDALSE
DWSYAEPSAGAGAVALLIGENPIVFQADAGANGYYGYEVMDTCRPIPDSEAGD
ADLSLMSYLDCCEQTFREYKNRVPGADYKETFHYLAFHTPFGGMVKG

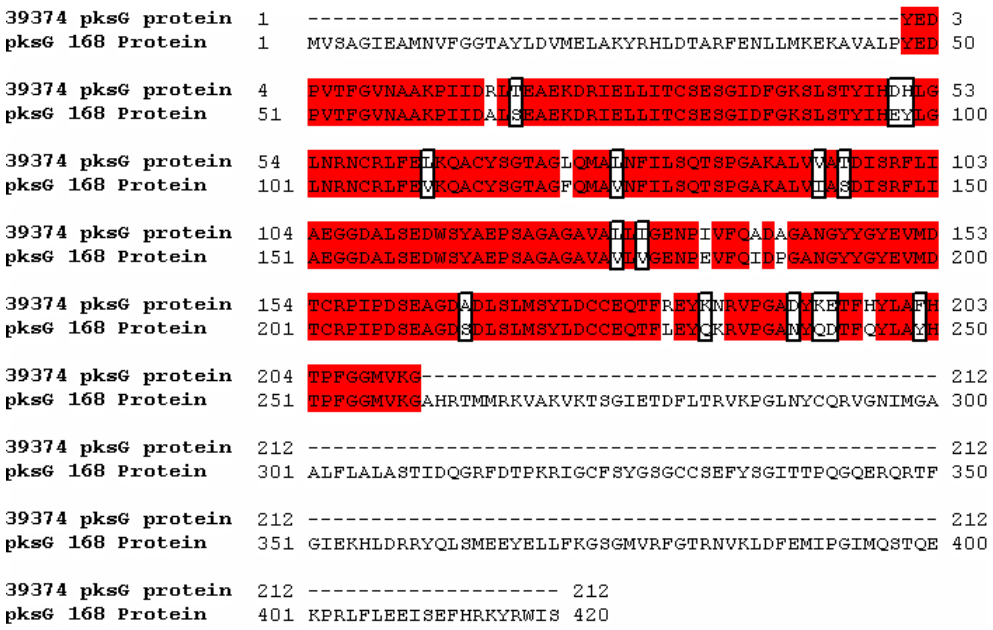

Figure III-8: Sequence alignment of PksG-39374 with PksG-168.

The sequence alignment of PksG from 39374 with PksG from 168 matched quite well except very few mismatches (Figure III-8), confirming the presence of *pksG* homolog in *B. subtilis* strain 39374.

III.D.2 Construction of pMUTIN4-*pksG*

The cloning process started with PCR amplification of 39374 genomic DNA with Taq polymerase and *pksG* pMUTIN primers (Table III-1). This was to prepare a approximately 350 bp fragment of the gene having correct sequence to be digested by the restriction enzymes, leading to ligation with the plasmid vector. Agarose gel electrophoresis photograph showed a band at approximately 350 bp pair region which was the expected size of the PCR product.

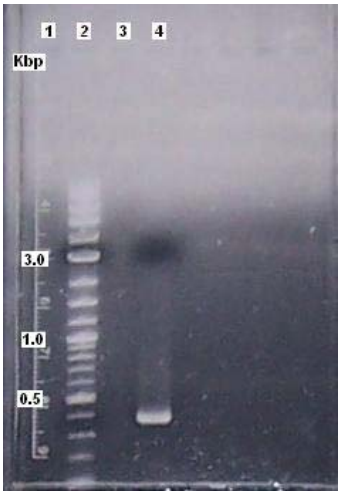

Figure III-9: Agarose gel showing a PCR product at ~350 bp. Lane 2 is molecular weight (MW) marker; lane 4 is the PCR with 39374 genomic DNA (band ~350 bp).

Since restriction enzymes BamH1 and HindIII do not have a common compatible buffer, a sequential restriction digest had to be conducted. Both plasmid and PCR insert were digested with BamH1 and purified separately. Another digestion by HindIII was performed on both the purified products. A low melting gel electrophoresis was

performed to confirm the size and orientation of the products (Figure III-10). This generated the plasmid vector and PCR amplified insert for ligation.

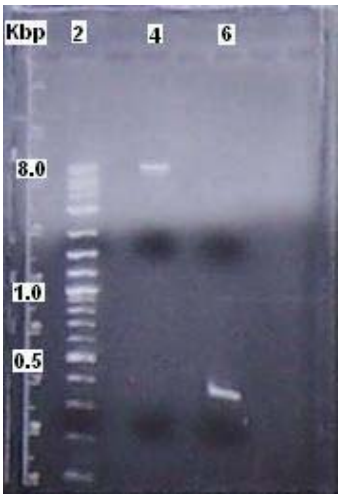

Figure III-10: Low melting agarose gel showing bands on lane 2 is MW marker; lane 4 is digested pMUTIN plasmid (band at ~8600 bp); lane 6 is digested PCR product of 39374 genomic DNA (band at ~350 bp).

Three sets of ligation reactions with different plasmid:insert ratio were set overnight. In order to screen for the insert into the plasmid, restriction digest and regular gel electrophoresis were performed to screen for the inserts in the plasmid. Three out of six samples were successful and had PCR insert in the plasmid (see Figure III-11).

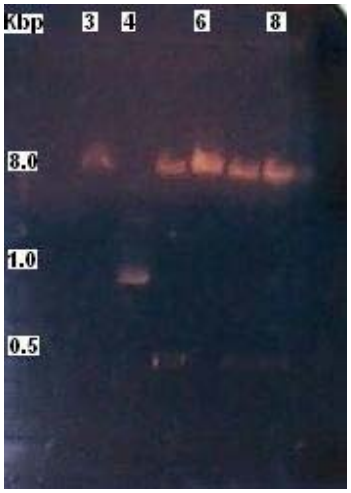

Figure III-11: Agarose gel showing bands for plasmids and inserts. Lane 1 is MW marker; lanes 3-8 are different ligation reactions (bands for plasmid at ~8600 bp); lanes 5, 7 and 8 had insert in the vector (band at ~350 bp).

Once the gene of interest was successfully cloned into pMUTIN4, it was ready for transformation into *B. subtilis* strain 39374. The transformation steps employed were to take advantage of the natural competence of *B. subtilis*. Transformation of *B. subtilis* required preparation of growing media for strain 39374. Problems were encountered during the procedure for the transformation of *B. subtilis*. At first, host strains had poor growth in SpC media (composition described in experimental section) prepared for the transformation procedure. Transformation was carried out in the media and protocols specified by Spizizen and coworkers.[III-11] A control procedure was also performed without the introduction of DNA, along a second experiment using all components necessary for transformation. Figure III-12 showed a promising result, having no colonies on the control experiment, but individual colonies on the plate which was spread a mixture of competent strain 39374 and pMUTIN4-*pksG* DNA.

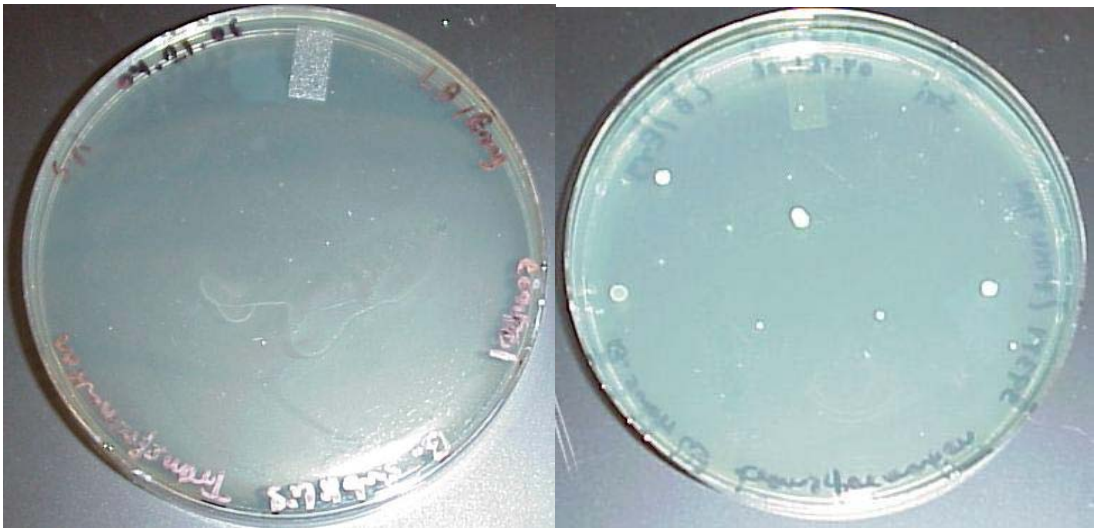

Figure III-12: The control reaction (left) shows no colonies, while the 39374 + pMUTIN4/*pksG* containing plate (right) shows colonies.

Individual colonies were collected and starter cultures were set overnight. DNA was isolated the next day. PCR reaction of purified DNA was performed with *pksG* DEGEN UP and *lacI* DN primers using Phusion polymerase. The size of the linearized gene fragment bound by the two primers *pksG* DEGEN UP and *lacI* DN was expected to be approximately 5000 bp. Figure III-13 showed the agarose gel showing bands for the possible recombination/mutation at ~5000 bp.

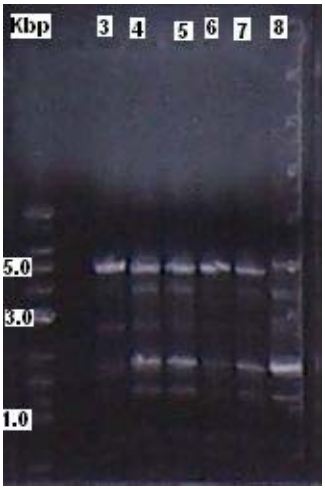

Figure III-13: Agarose gel showing bands at ~5000 bp. Lane 1 is MW, lane 3-8 are six different samples of possible mutant lacking *pksG*.

The stack of bands below 5000 bp for every sample of DNA was unexpected and could not be explained. However it is worth mentioning that the wildtype strain 39374 genomic DNA yielded the same results from PCR with the same primers. Therefore, we could not unambiguously demonstrate that the integration occurred.

III.D.3 *Bacillus amyloliquefaciens*

In June 2006, Chen *et al*. confirmed that *Bacillus amyloliquefaciens* strain FZB42 contains *pks* gene clusters, having compared that with the model *B. subtilis* strain 168 (Figure III-14).[III-14]

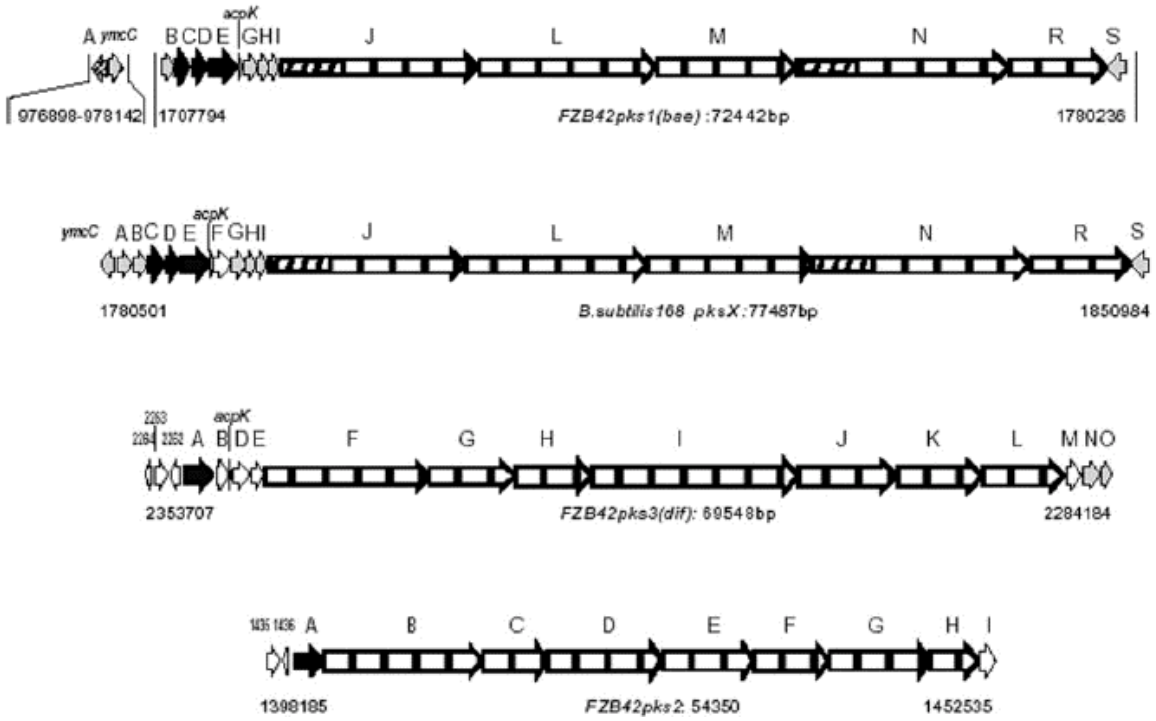

Figure III-14: Organization of pks gene clusters in *B. amyloliquefaciens* strain FZB42 and model *B. subtilis* strain 168.[III-14]

As mentioned the DNA sequence alignment of *pksG* homolog in *B. subtilis* 39374 with that of *B. subtilis* 168 matched quite well, with several mismatches (~ 23%), shown in Figure III-15. The protein sequence derived from this data was shown in Figure III-8, showing the presence of *pksG* homolog in 39374. When a DNA sequence alignment was performed with *pksG*-39374 and *baeG*-FZB42, just eleven nucleotide mismatch (~ 2%) was observed (Figure III-16). Protein sequence alignment of *pksG*-39374 with *baeG*-FZB42, the *pksG* homolog in *B. amyloliquefaciens* showed 100% match (Figure III-17).

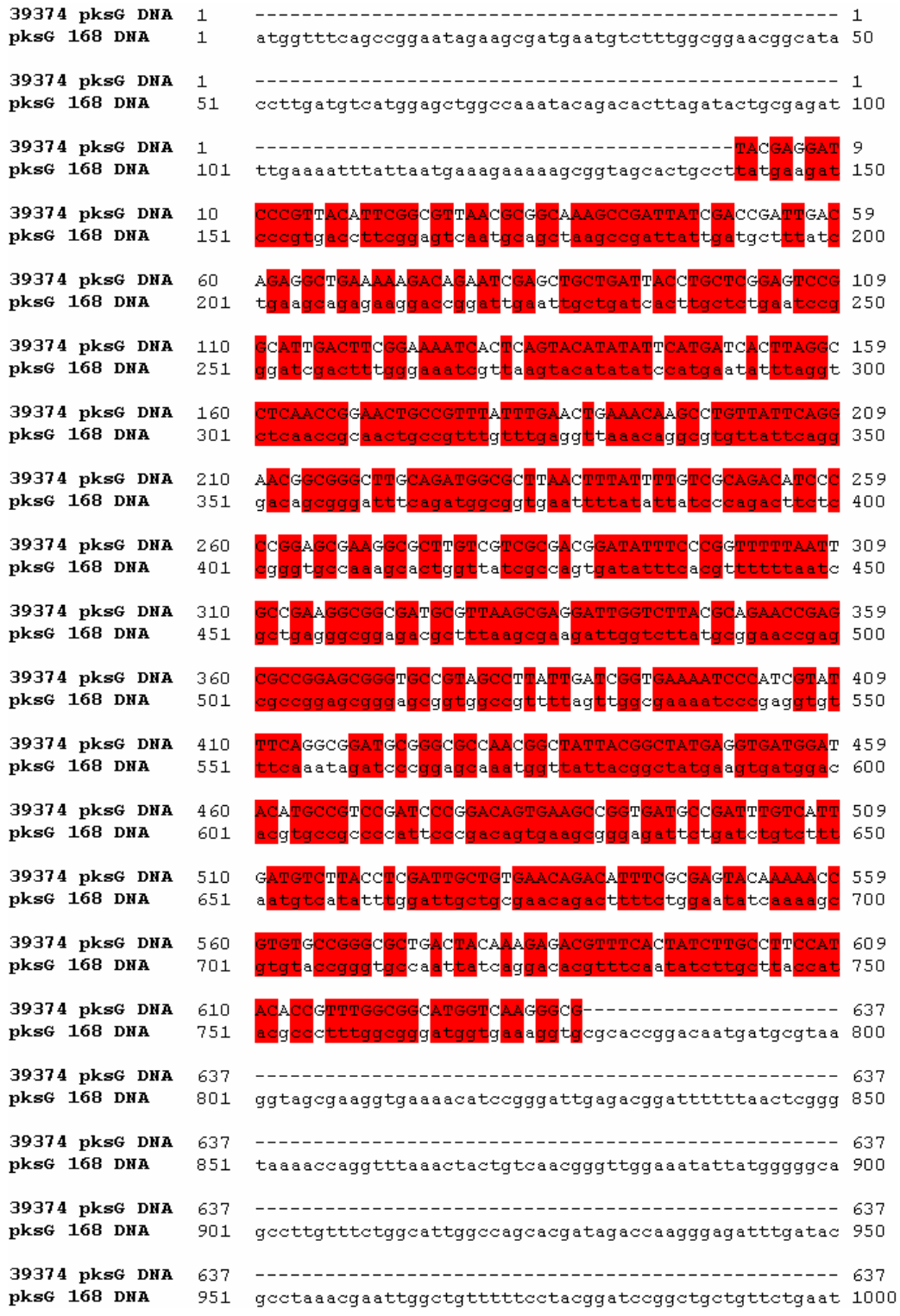

Figure III-15: DNA sequence alignment – *pksG*-39374 vs. *pksG*-168.

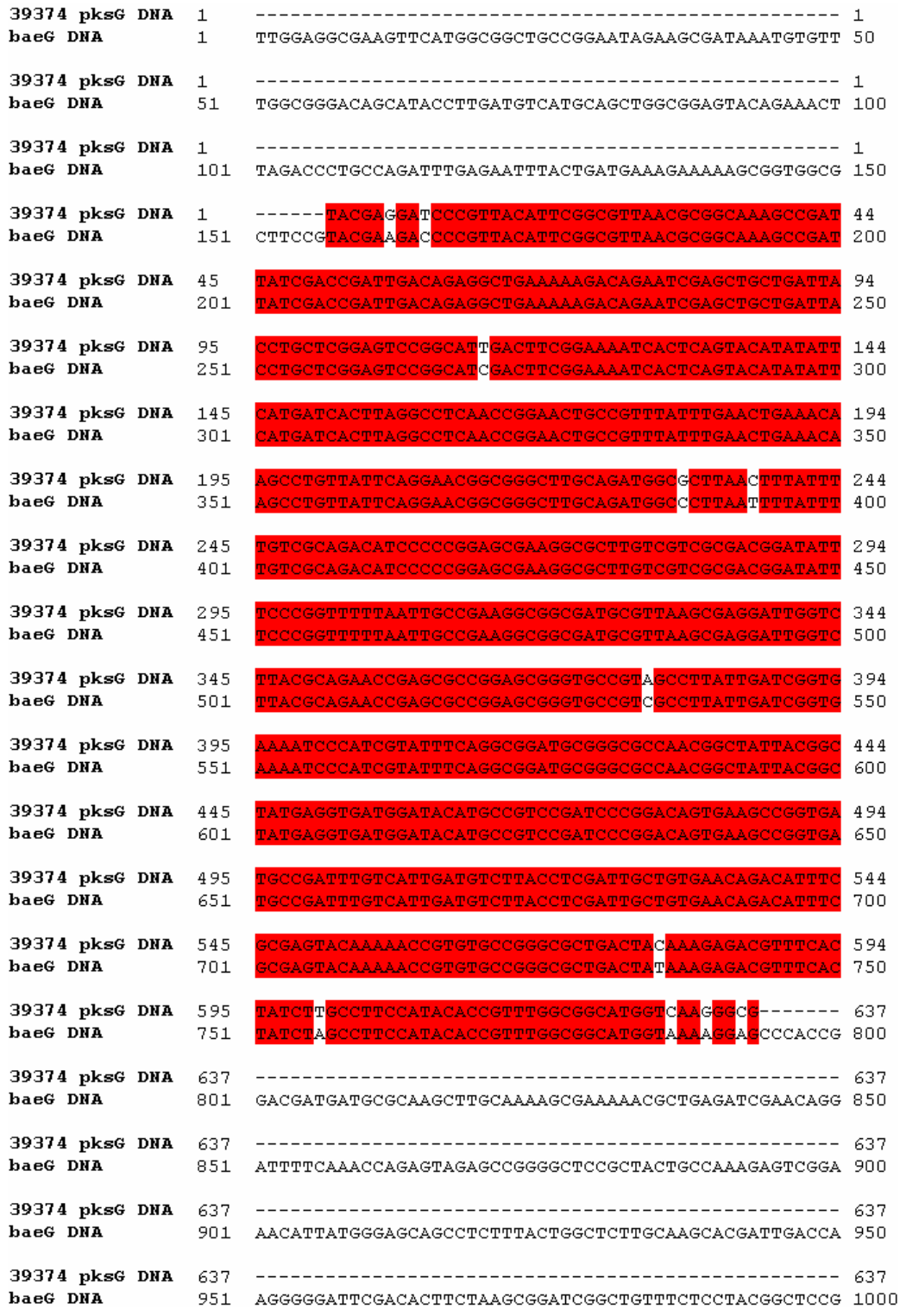

Figure III-16: DNA sequence alignment – *pksG*-39374 vs. *baeG*-FZB42.

```
39374 pksG protein  1   ------------------------------------------------- 1
baeG Protein        1   LEAKFMAAAGIEAINVFGGTAYLDVMQLAEYRNLDPARFENLLMKEKAVA 50

39374 pksG protein  1   --YEDPVTFGVNAAKPIIDRLTEAEKDRIELLITCSESGIDFGKSLSTYI 48
baeG Protein        51  LPYEDPVTFGVNAAKPIIDRLTEAEKDRIELLITCSESGIDFGKSLSTYI 100

39374 pksG protein  49  HDHLGLNRNCRLFELKQACYSGTAGLQMALNFILSQTSPGAKALVVATDI 98
baeG Protein        101 HDHLGLNRNCRLFELKQACYSGTAGLQMALNFILSQTSPGAKALVVATDI 150

39374 pksG protein  99  SRFLIAEGGDALSEDWSYAEPSAGAGAVALLIGENPIVFQADAGANGYYG 148
baeG Protein        151 SRFLIAEGGDALSEDWSYAEPSAGAGAVALLIGENPIVFQADAGANGYYG 200

39374 pksG protein  149 YEVMDTCRPIPDSEAGDADLSLMSYLDCCEQTFREYKNRVPGADYKETFH 198
baeG Protein        201 YEVMDTCRPIPDSEAGDADLSLMSYLDCCEQTFREYKNRVPGADYKETFH 250

39374 pksG protein  199 YLAFHTPFGGMVKG----------------------------------- 212
baeG Protein        251 YLAFHTPFGGMVKGAHRTMMRKLAKAKNAEIEQDFQTRVEPGLRYCQRVG 300

39374 pksG protein  212 ------------------------------------------------- 212
baeG Protein        301 NIMGAASLLALASTIDQGGFDTSKRIGCFSYGSGCCSEFYSGVVTPEGQA 350

39374 pksG protein  212 ------------------------------------------------- 212
baeG Protein        351 RQHSFKIEEHLNNRYRLSMEEYEELFKGSGLVKFGTRNVKLDVNMIPNIL 400

39374 pksG protein  212 -------------------------- 212
baeG Protein        401 DHHTGSPRLYLEEITEFHRKYRWIS  425
```

Figure III-17: Protein sequence alignment – PksG-39374 vs. baeG-FZB42.

III.E Conclusion

Successful PCR of 39374 genomic DNA with degenerate primers supported the presence of *pksG* in *B. subtilis* strain 39374. Sequence data of PksG-39374 matched quite well with PksG-168, which proves that strain 39374 has a *pksG* homolog. With the acquired knowledge of the sequence of *pksG* homolog in 39374, plasmid pMUTIN4 containing a fragment of the *pksG* homolog from 39374 was successfully generated and evidence was sited by gel electrophoresis. However it could not be inferred that the mutagenesis 39374ΔPksG occurred, as the experimental data obtained was inconclusive. More studies will be required for successful generation of recombinant 39374ΔPksG.

Since the mutant could not be constructed, we were not able to determine the role of PksG in difficidin production.

Protein sequence alignment of the *pksG* homolog in 39374 with the *pksG* homolog in *B. amyloliquefaciens* strain FZB42 showed that they are 100% match. Therefore it was concluded that *B. subtilis* strain 39374 may instead be *Bacillus amyloliquefaciens*, as sequence data we obtained for *pksG* homolog was identical to this other organism. This calls into question the taxonomic identity of the difficidin producing ATCC strains 39320 and 39374, and efforts have recently begun in our laboratory to investigate this question.

III.F Experimental section

III.F.1 Sequencing: Cloning of strain 39374 in *pBAD* plasmid

The primer sequences of *pksG* DEGEN and pBAD are listed in Table III-1. PCR of *B. subtilis* strain 39374 genomic DNA was performed with *Taq* polymerase and *pksG* DEGEN primers under the following conditions: 2 minutes at 94°C; 30 cycles of 30 seconds at 94°C, 30 seconds at 58°C, and 1 minute at 72°C; then a hold of 5 minutes at 72°C. TOPO cloning was performed by adding the PCR product to the pBAD-TOPO vector, and incubating at room temperature. Once the cloning was done, the vector was transformed into the *E. coli* TOP10 strain. Six single transformants were selected from the ampicillin plate, and then six starter cultures were grown.

III.F.2 Sequencing: Purification of the plasmid pBAD/*pksG*

Purification of the pBAD plasmid was done for all six starter cultures using the QIAPrep plasmid purification kit. A PCR was performed using the TOPO cloning kit and subsequently 1.0% agarose gel electrophoresis was performed to screen for successful PCR.

III.F.3 Sequencing: Retransformation and Re-purification of the plasmid *pBAD/pksG*

A 5 mL of LB starter culture was shaken overnight in a 37°C incubator. Purification was performed using the same *QIAPrep* plasmid purification kit. Absorbance of the *pBAD* plasmid was measured at 260 nm. Final concentration of DNA was calculated before sending the DNA to sequencing. The whole 50 µL sample was sent for sequencing to SeqWright, Inc. (Houston, TX). Sequencing data was analyzed by *Bioedit* and *Chromas Lite* software packages.

III.F.4 Mutant lacking *pksG* construction: Cloning of *pksG* homolog from strain 39374

PCR of *B. subtilis* strain 39374 genomic DNA was conducted with Taq polymerase and pMUTIN primers under the following conditions: 2 minutes at 94°C; 30 cycles of 30 seconds at 94°C, 30 seconds at 60°C, and 1 minute at 72°C; then a hold of 5 minutes at 72°C. The sequence of pksG pMUTIN primers are listed in Table III-1. pMUTIN primers insert a BamH1 and HindIII restriction sites at the ends of the PCR product. 1.0 % agarose gel electrophoresis was performed to confirm the presence of correct gene fragment. The PCR product was purified using Qiagen PCR purification kit.

Plasmid vector pMUTIN4 was transformed into *E. coli* strain DH5α competent cells and plated on LB agar containing 25µg/mL ampicillin. Overnight 5mL starter cultures containing 25µg/mL ampicillin were purified using Qiagen Miniprep plasmid purification kit.

A sequential restriction digest was conducted on both the plasmid and PCR product. First, each was digested with the BamH1 restriction enzyme. The digested PCR product was purified using regular Qiagen PCR purification kit and the plasmid was purified by an ethanol precipitation method.

The ethanol precipitation method involved addition of 80µL of freshly prepared precipitation solution (3.0µL of 3M sodium acetate at pH 4.6 + 62.5µL of non denatured 95% ethanol + 14.5µL of deionized water) in each tube. Mixtures were vortexed and allowed to stand at room temperature for 15 minutes. DNA was peletted by centrifuging at maximum speed for 20 minutes at room temperature. The ethanol solution was removed carefully from each tube and then 250µL of 70% ethanol was added to each of them, followed by centrifugation for 5 minutes at maximum speed. Ethanol was again removed carefully from each of the tubes and then allowed them to air dry for 15 minutes at room temperature.

Both of the purified products were digested with HindIII. A 1.0% low melting gel electrophoresis was conducted. Both plasmid and PCR product/insert fragment were excised and melted in a hot water bath to perform ligation reaction using T4 DNA ligase and three sets of ligation reactions having different plasmid:insert ratios were set overnight at 16°C. All three overnight ligation reactions of pMUTIN4-39374 were

transformed into *E. coli* DH5α competent cells and transformation mixtures were plated onto three LB-agar plates containing ampicillin of final concentration 25μg/mL, and grown overnight at 37°C. Six single transformants were selected and starter cultures were grown overnight in 5mL LB with ampicillin at a final concentration of 25μg/mL in a 37°C shaker with 225 rpm. All six of the plasmids were purified by a *Qiagen Miniprep* kit the following day. Another restriction digest with BamH1 and BglII restriction enzymes for 3:30 hours and 1.0% agarose gel electrophoresis were performed.

III.F.5 Mutant lacking *pksG* construction: Transformation of *B. subtilis* strain 39374

III.F.5.a The following were the composition of the growth media used.

10X T-Base: 150mM ammonium sulfate, 800mM potassium phosphate dibasic, 440mM potassium phosphate monobasic, 35mM sodium citrate.

SpC medium: 100mL 1X T-Base, 1.0mL 50% glucose, 1.5mL 1.2% magnesium sulfate, 2.0mL 10% yeast extract, 2.5mL 1% casein hydrolysate.

SpII medium: SpC + 0.5mL 100mM calcium chloride.

SpII + EGTA: SpII + 2mM EGTA.

The protocol of transformation *B. subtilis* involved first growth of a colony of strain 39374 in SpC medium in a 37°C shaker incubator for 5 hours. Two mL of the grown 39374 culture was added to 18mL of SpII and shaking/incubation continued for 90 minutes. The culture was centrifuged to pellet cells and 1067mL of supernatant was saved as this contained the competence factors. Cells were resuspended in the amount of this saved supernatant and 267mL of 50% glycerol was added to it. For transformation, 500μL of competent 39374 + 500μL of SpII-EGTA + 5μL of pMUTIN4-*pksG* were

incubated for 30 minutes in 37°C. Another reaction was performed as control which had 39374 and SpII-EGTA, but was lacking pMUTIN4-*pksG*. Cells were plated on LB-agar containing 0.3µg/mL erythromycin.

III.F.5.b Preparation of media described by Spizizen[III-11]

Composition of the minimal medium #1: 0.2% ammonium sulfate; 1.4% dipotassium phosphate; 0.6% monopotassium phosphate; 0.1% sodium citrate-$2H_2O$; 0.02% magnesium sulfate.$7H_2O$; glucose to a final concentration of 0.5%; 50µg/mL L-tryptophan; 0.02% casein hydrolysate.

Composition of the minimal medium #2: 0.2% ammonium sulfate; 1.4% dipotassium phosphate; 0.6% monopotassium phosphate; 0.1% sodium citrate-$2H_2O$; 0.02% magnesium sulfate.$7H_2O$; glucose to a final concentration of 0.5%; 5µg/mL L-tryptophan; 0.01% casein hydrolyzate; 5µM/mole $MgSO_4.7H_2O$.

III.F.5.c Protocol of transformation by Spizizen [III-11]

Starter culture of 1 colony of strain 39374 was grown in 3mL medium#1 for ~16 hours. 2.5mL of medium#1 was inoculated with 20µL starter culture and grown for 4 hours. Cells were pelleted and resuspended in 2.5mL medium#2. In a new tube 100µL of this culture was mixed with 900µL of medium#2 in order to have 10-fold dilution and 5µL pMUTIN4-*pksG* was added, followed by incubation for 90 minutes at 37°C shaker. Cells were pelleted, resuspended in 100µL supernatant, and plated on LB-agar-erythromycin plates.

III.F.6 Isolation of recombinant genomic DNA

Individual colonies were selected and a starter culture was set overnight in LB and erythromycin concentration of 0.3µg/mL. Overnight cultures were purified by using a Wizard® Genomic DNA purification kit. PCR was performed with purified DNAs using Phusion polymerase and *pksG*-DEGEN Up + *lacI* DN primers with the following conditions: 2 minutes at 98°C; 30 cycles of 30 seconds at 98°C, 30 seconds at 60°C, and 2:30 minutes at 72°C; then a hold of 5 minutes at 72°C. The PCR reactions were subsequently analyzed by 0.7% agarose gel electrophoresis.

III.F.7 Analysis of *B. subtilis* culture for difficidin

Landy medium was prepared by the following composition. Per liter: 20g glucose, 5g L-glutamic acid, 0.5g $MgSO_4$, 0.5g KCl, 1.0g KH_2PO_4, 0.15mg $Fe_2(SO_4)_3.6H_2O$, 5.0mg $MnSO_4.H_2O$, 0.16mg $CuSO_4.5H_2O$. The pH was maintained at 6.0. Several flasks containing 50mL landy media were prepared and autoclaved. A starter culture of native strain 39374 was set in tryptic soy broth (TSB) medium. A starter culture of one colony of 39374ΔpksG candidate was set in TSB and 0.3µg/mL erythromycin. 100µL of each overnight starter culture was added to two different flasks and allowed to grow at 32°C with shaking for 2 days.

III.F.8 Preparation of samples and protocol for LC-MS

1mL of each grown culture was collected in different 15mL culture tubes. HPLC grade methanol (2mL) was added, vortexed for 30 seconds, and centrifuged for 5 minutes. 1mL supernatants were placed into autosampler vials and introduced to the LC-MS

instrument. Data were collected on a LCQ-Advantage ion trap mass spectrometer coupled to a HP1100 HPLC instrument using a 5mM Prevail C-18 column from Alltech. The following was the gradient elution condition: begin at 95:5 1% acetic acid:acetonitrile, with the acetonitrile increasing to 95% over a 35 minute linear gradient at a flow rate of 0.2 mL/min.

REFERENCES

I-1. Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. "An unusually large multifunctional polypeptide in erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*," *Nature* **1990**, *348*, 176.

I-2. Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B. "Cloning and heterologous expression of the epothilone gene cluster," *Science* **2000**, *287*, 640.

I-3. Hopwood, D. A. "Genetic contributions to understanding polyketide synthases," *Chem. Rev.* **1997**, *97*, 2465.

I-4. Wawrik, B.; Kerkhof, L.; Zylstra, G. J.; Kukor, J. J. "Identification of unique type II polyketide synthase genes in soil," *Appl. Environ. Microbiol.* **2005**, *71*(5), 2232.

I-5. Pereda, A.; Summers, R. G.; Stassi, D. L.; Ruan, X.; Katz, L. "The loading domain of the erythromycin polyketide synthase is not essential for erythromycin biosynthesis in *Saccharopolyspora erythraea*," *Microbiol.* **1998**, *144*, 543.

I-6. Staunton, J.; Wilkinson, B. "Biosynthesis of erythromycin and rapamycin," *Chem. Rev.* **1997**, *97*, 2611.

I-7. Cane, D. E.; Walsh, C. T.; Khosla, C. "Harnessing the biosynthetic code: combinations, permutations, and mutations," *Nature* **1998**, *282*, 5386, 63.

I-8. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; FU, H.; Betlach, Melanie; Betlach, Mary; Ashley, G. "Multiple genetic modifications of the erythromycin

polyketide synthase to produce a library of novel “unnatural” natural products," *Proc. Natl. Acad. Sci. USA.* **1999**, *96*, 1846.

I-9. Katz, L. “Manipulation of Modular Polyketide Synthases,” *Chem. Rev.* **1997**, *7*, 2557.

I-10. Khosla, C. "Harnessing the biosynthetic potential of modular polyketide synthases,” *Chem. Rev.* **1997**, *97*, 2577.

I-11. Rowe, C. et al; “Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase,” *Chem. Biol.* **2001**, *8*, 475.

I-12. Walsh, C. “Combinatorial biosynthesis of antibiotics: challenges and opportunities,” *ChemBioChem.* **2002**, *3*, 124.

I-13. Kao, C. M.; McPherson, M.; McDaniel, R.; Fu, H.; Cane, D.; Khosla, C. “Alcohol stereochemistry in polyketide backbones is controlled by the β-ketoreductase domains of modular polyketide synthases,” *J. Am. Chem. Soc.* **1998**, *120*, 2478.

I-14. Scotti, C.; Piatti, M.; Cuzzoni, A.; Perani, P.; Tognoni, A.; Grandi, G.; Galizzi, A.; Albertini, A. M. "A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases," *Gene* **1993**, *130*, 65.

I-15. Albertini, A. M.; Caramori, T.; Scoffone, F.; Scotti, C.; Galizzi, A. "Sequence around the 159-degrees region of the *Bacillus subtilis* genome - the *pksX* locus spans 33.6 kb," *Microbiol.-UK* **1995**, *141*, 299.

I-16. Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G. "The complete genome sequence of the Gram positive bacterium *Bacillus subtilis*," *Nature* **1997**, *390*, 249.

I-17. El-Sayed, A. K.; Hothersall, J.; Cooper, S. M.; Stephens, E.; Simpson, T. J.; Thomas, C. M. "Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586," *Chem. Biol.* **2003**, *10*, 419.

I-18. Paitan, Y.; Alon, G.; Orr, E.; Ron, E.Z.; Rosenberg, E. "The first gene in the biosynthesis of the polykedtide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase," *J. Mol. Biol.* **1999**, *286*, 465.

I-19. Edwards, D.J.; Marquez, B.L.; Nogle, L.M.;McPhail, K.; Goeger, D.E.; Roberts, M.A.; Gerwick, W.H. "Structure and biosynthesis of the jamaicamides: new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscule*," *Chem. Biol.* **2004**, *11*, 817.

I-20. Pettit, G. R.; Cichacz, Z. A.; Gao, F.; Herald, C. L.; Boyd, M. R.; Schmidt, J. M.; Hooper, J. N. A. "Antineoplastic agents: isolation and structure of spongistatin-1," *J. Org. Chem.* **1993**, *58*, 1302.

I-21. Kobayashi, J.; Ishibashi, M.; Nakamura, H.; Ohizumi, Y.; Yamasu, T.; Sasaki, T; Hirata, Y. "Amphidinolide A, a novel antineoplastic macrolide from the marine dinofalgellate *Amphidinium sp*," *Tetrahedron Letters* **1986**, *27*, 5755.

I-22. Chang, Z.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J.; Sherman, D. H.; Gerwick, W .H. "Biosynthetic pathway and gene cluster analysis of curacin

A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*," *J. Nat. Prod.* **2004**, *67*, 1356.

I-23. Piel, J.; Butzke, D.; Fusetani, N.; Hui, D.; Platzer, M.; Wen, G.; Matsunaga, S. "Exploring the chemistry of uncultivated bacterial symbionts: antitumor polyketides of the pederin family," *J. Am. Chem. Soc.* **2005**, *68*, 472.

I-24. Cheng, Y. Q.; Tang, G. L.; Shen, B. "Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis," *Proc. Natl. Acad. Sc. USA*, **2003**, *100*(6), 3149.

I-25. Mantle, P.G.; Macgeorge, K.M. "Pseudomonic acid biosynthesis: the putative role of 3-Hydroxy-3-methylglutarate," *J. Chem. Soc. Perkin Trans. I* **1991**, *2*, 255.

I-26. Gu, L.; Jia, J.; Liu, H.; Hakansson, K.; Gerwick, W. H.; Sherman, D. H. "Metablic coupling of dehydration and decarboxylation in the Curacin A pathway: functional identification of a mechanistically diverse enzyme pair," *J. Am. Chem. Soc.*, ASAP(**2006**).

I-27. Calderone, C. T.; Kowtoniuk, W. E.; Kelleher, N. L.; Walsh, C. T.; Dorrestein, P. C. "Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the *pksX* pathway in *Bacillus subtilis*," *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8977.

II-1. Hopwood, D. A. "Genetic contributions to understanding polyketide synthases," *Chem. Rev.,* **1997**, *97*, 2465.

II-2. Staunton, J.; Weissman, K. J. "Polyketide biosynthesis: a millennium review," *Nat. Prod. Rep.* **2001**, *18*, 380.

II-3. Cheng, Y. Q.; Tang, G. L.; Shen, B. "Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis," *Proc. Natl. Acad. Sc. USA*, **2003**, *100*(6), 3149.

II-4. Mootz, H. D.; Finking, R.; Marahiel, M. A. "4′-Phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*," *J. Biol. Chem.*, **2001**, *276*(40), 37289.

II-5. Reuter, K.; Mofid, M. R.; Marahiel, M. A.; Ficner, R. "Crystal structure of the surfactin synthetase-activating enzyme Sfp: a prototype of the 4′-phosphopantetheinyl transferase superfamily," *EMBO J.*, **1999**, *18*, 6823.

II-6. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. "Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers," *Anal. Chem.* **1991**, *63*, 1193.

II-7. *Novagen, pET System Manual.*, 11th ed.

II-8. Lambalot, R. H.; Walsh, C. T. "Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase," *J. Biol. Chem.*, **1995**, *270*(42), 24658.

III-1. Harwood, C. R.; Wipat, A. "Sequencing and functional analysis of the genome of *Bacillus subtilis* strain 168," *FEBS Letters*, **1996**, *389*, 84.

III-2. Williams, Kevin; *Thesis research* **2005**.

III-3. Yu, T. W.; Shen, Y.; Doi-Katayama, Y.; Tang, L.; Park C.; Moore, B. S.; Hutchinson, C. R.; Floss, H. G. "Direct evidence that the rifamycin polyketide

synthase assembles polyketide chains processively," *Proc. Natl. Acad. Sci.* **1999**, *96*, 9051.

III-4. Shuman, S. "Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase," *J. Biol. Chem.* **1994**, *51*, 32678.

III-5. Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry* 5 ed. 2002.

III-6. Vagner,V.; Dervyn, E.; Ehrlich, S. D. "A vector for systematic gene inactivation in *Bacillus subtilis*," *Microbiol.* **1998,** *144*, 3097.

III-7. "A vector useful for gram-positive genomics," *Bacillus Genetic Stock Center* **2001**.

III-8. Mantle, P. G.; Macgeorge, K. M. "Pseudomonic acid biosynthesis: the putative role of 3-hydroxy-3-methylglutarate," *J. Chem. Soc. Perkin Trans. I* **1991**, *2*, 255.

III-9. Perkins, J. B.; Youngman, J. C. "Construction and properties of Tn917-*lac*, a transposon derivative that mediates transcriptional gene fusions in *Bacillus subtilis*," *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 140.

III-10. Yansura, D. G.; Henner, D. J. "Use of the *Escherichia coli lac* repressor and operator to control gene expression in *Bacillus subtilis*," *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 439.

III-11. Anagnostopoulos C.; Spizizen J. "Requirements for transformation in *Bacillus subtilis*," *J. Bacteriol.* **1960**, *81*, 741.

III-12. Young, F. E.; J. Spizizen. "Physiological and genetic factors affecting transformation of *Bacillus subtilis*," *J. Bacteriol.* **1961**, *81*, 823.

III-13. Bott, K. F.; Wilson, G. A. "Development of competence in the *Bacillus subtilis* transformation system," *J. Bacteriol.* **1967**, *94*(3), 562.

III-14. Chen, X. H.; Vater, J.; Piel, J.; Franke, P.; Scholz, R.; Schneider, K.; Koumoutsi, A.; Hitzeroth, G.; Grammel, N.; Strittmatter, A. W.; Gottschalk, G.; Sussmuth, R. D.; Borriss, R. "Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42," *J. Bacteriol.* **2006**, *188*, 4024.

III-15. Edwards *et al. Chem. Biol.* **2004**, *11*(6), 817.