WILLIAMS, JAYME, M.S. Overexpression, Purification, and Characterization of AcpK, PksG, and MmgA from *Bacillus subtilis* Strain 168. (2007)
Directed by Dr. Jason J. Reddick. 91 pp.

Polyketides are a diverse class of natural products that have varying therapeutic values. Polyketide synthases (PKSs) use different domains within the larger enzyme to vary the amount of reduction seen at each ketide position. The ability to manipulate PKSs would make the synthesis of unnatural polyketides possible, potentially creating novel biologically active molecules.

*Bacillus subtilis* strains 39320 and 39374 produce the polyketide antibiotic difficidin, which has a subunit that can not be explained by well-understood polyketide synthesis. Studies of the synthesis of difficidin may be useful for diversifying the synthesis of new unnatural polyketides. Sfp, a phosphopantetheinyl transferase (PPTase) activates ACPs. AcpK carries an acyl unit to react with PksG, an enzyme that synthesizes an intermediate that may lead to the unusual subunit. An enzyme expressed during sporulation, MmgA, is homologous to acetyl coenzyme A acetyltransferases.

AcpK, PksG, and MmgA were successfully cloned, overexpressed, and purified from *B. subtilis* strain 168. Sfp was successfully cloned, overexpressed, and purified from a donated overexpression vector. PPT reactions with Sfp and AcpK were successfully performed. Biochemical characterization of PksG was carried out and defined by other unrelated research. MmgA was biochemically determined to be a β-keto acyl thiolase.

# OVEREXPRESSION, PURIFICATION, AND CHARACTERIZATION OF AcpK, PksG, AND MmgA FROM *BACILLUS SUBTILIS* STRAIN 168

by

Jayme Williams

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

2007

Approved by

_____
Committee Chair

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

ACKNOWLEDGEMENTS

I wish to thank my advisor, Dr. Jason Reddick, whose years of teaching, support, and leadership has allowed me to have a wonderful undergraduate and graduate experience. I also would like to thank the other members of my committee, Dr. Gregory Raner and Dr. Johanna Mazlo, for their assistance. Finally, I also acknowledge Dr. Nadja Cech for her guidance and generosity in using her laboratory and instrument to aide in my research.

I wish to express my eternal gratitude to my family, for their emotional and financial support of my academic career. Lastly, I thank my friends and loved ones who have been my rock and my soft place to land.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

## I.A Polyketide Structures and Biosynthesis

Polyketides are a diverse class of natural products that can be produced by certain bacteria, fungi, plants, and animals. These products have a variety of pharmacological applications. Figure I-1 shows examples of polyketides that have therapeutic value including naphthomycin A an antibiotic, dynemycin A an anticancer compound, tetronasin an antiparasitic agent, rapamycin an immunosuppressant, streptazolin an antifungal, and mevinolin a cholesterol lowering compound (Rawlings, 1997).

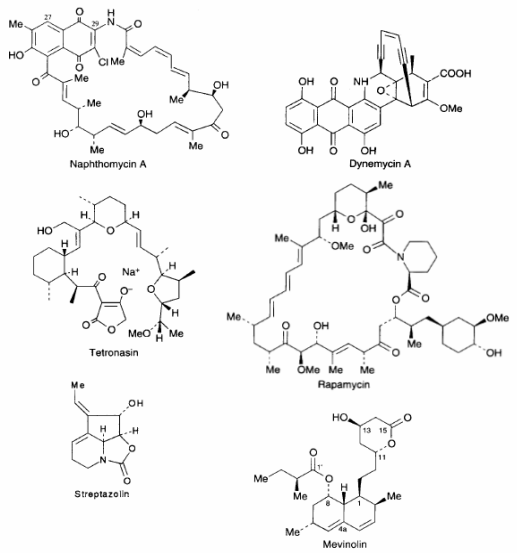


Figure I-1 Examples of polyketides having medicinal value (Rawlings,1997).

Polyketides are assembled by enzymes called polyketide synthases (PKSs). There are three types of PKSs, type I-III, and each assembles the product in different routes, though the chemistry is very similar. Type I PKSs are arranged in modules that contain multifunctional enzymes which produce nonaromatic polyketides. Type II PKSs are made up of a single set of proteins that are iteratively used to synthesize aromatic polyketides (Shen, 2003).

PKSs are comparable to fatty acid synthases (FASs) because the function of enzymes present in each synthase catalyzes similar reactions. Type I FASs are also single, multifunctional, proteins (Hopwood, *et al.*, 1990). FASs initiate fatty acid chain synthesis by a decarboxylative condensation between an acetyl unit (C2 starter unit) and a malonyl unit (C3 extender unit) yielding a C4 product. The acetyl starter unit is transferred from acetyl coenzyme A (CoA) to a domain found in the FAS called the acyl carrier protein (ACP) domain. This transfer is catalyzed by another domain, the acyltransferase (AT) domain.

Decarboxylation occurs after the starter unit moves to the ketosynthase (KS) domain and a extender unit is attached to a newly opened ACP domain. The KS domain catalyzes the decarboxylation step that forms the C4, β-ketone product. This β-ketone undergoes further reductions by other enzymes. These enzymes are a ketoreductase (KR), which yields an alcohol in the β position, a dehydratase (DH), which yields a double bond, and an enoylreductase (ER), which yields a fully saturated methylene group. Elongation occurs when a new extender unit is loaded and another round of condensation and subsequent

reductions occur, extending the chain by two carbons. This elongation and reduction cycle continues until a specific chain length is reached. The chain-length is specified by a final domain, the thioesterase (TE), which hydrolyzes the ACP-thioester, breaking the free fatty acid (Katz, *et al.*, 1993). The initial condensation and sequential reduction reactions catalyzed by the different domains in the FAS can be seen in Figure I-2.

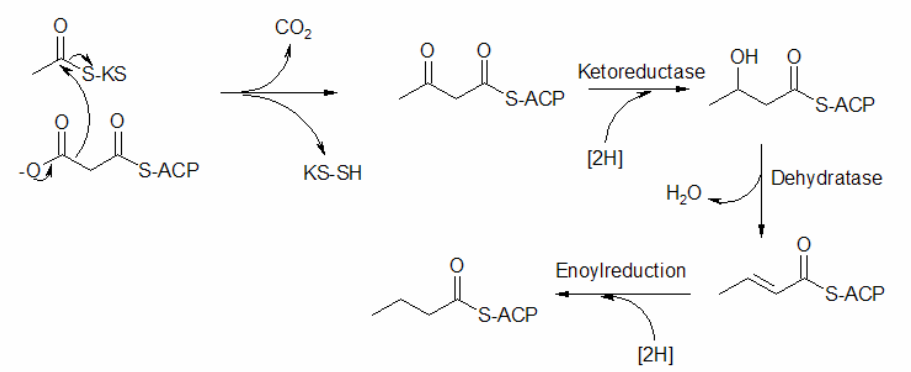


Figure I- 2 Reactions in fatty acid synthesis.

Polyketides are produced by a similar mechanism with four exceptions, the diversity of starter units used, a variable extent of reduction of each β-ketone unit, the introduction of chiral centers by either the β-hydroxyl group or side chains present, and post synthetic modifications such as cyclization, lactonization, and glycosylation, among others (Katz, *et al.*, 1993). These differences are greatly responsible for the diversity seen in polyketide natural products.

One of the most studied PKSs is from the organism *Saccharopolyspora erythraea,* the producer of erythromycin A, an antibiotic useful against Gram-positive bacteria. By the year 1965 the structure and stereochemistry of erythromycin A had been determined (see Figure I-3). Erythromycin A is synthesized in two phases. The first phase involves the PKS portion of the molecule, and the second is post polyketide synthase "tailoring" (Staunton, *et al.*, 1997).

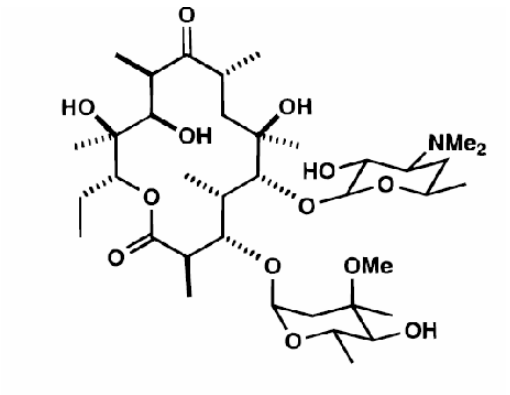


Figure I- 3 Structure of erythromycin A (Staunton *et al.*, 1997).

The polyketide portion of erythromycin A is synthesized using propionate as a starter unit and methylmalonate as extender units. This process forms a precursor to erythromycin A called 6-deoxyerythronolide B (6-DEB). The proteins that synthesize this molecule are called 6-DEB synthase (DEBSI-III) for this reason. DEBSI-III each contains two modules that are responsible for the extension of the polyketide chain by two carbon units until after module 6, when

the TE releases the fully formed chain. Each module contains the necessary domains, the AT, the ACP, and the KS. Other reducing domains may be present which can include, KR, DH, and ER. Each module extends the polyketide chain and reduces the β-ketone until the module-encoded functional group is obtained (Staunton, *et al.*, 1997). The last step catalyzed by DEBS is release by macrolactonization by the TE domain. These modules can be seen in Figure I-4 and the resulting structure of 6-DEB.

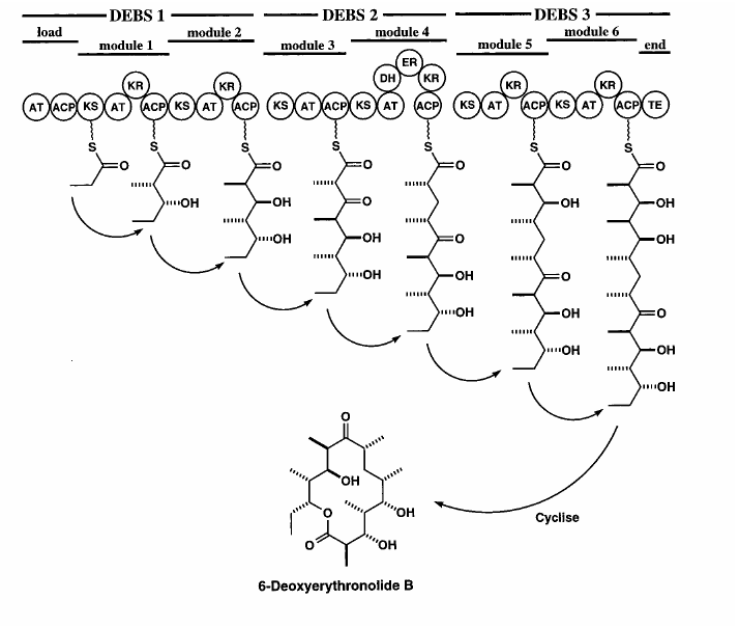


Figure I- 4 DEBS I-III and the structure of 6-deoxyerythronolide B (Staunton, *et al.*, 1997).

## I.B Combinatorial Biosynthesis

Polyketide synthesis occurs through common routes, though the products can be diverse. The structures of polyketides can be correlated directly to the

sequence of modules within their cognate modular PKS. The number of modules, the choice of starter and extender unit, and the domains inside the modules are some of the ways this natural product diversity is explained (McDaniel, *et al.*, 1999).

Research into the manipulation of the biosynthetic pathways of polyketide production has led to the engineered biosynthesis of “unnatural” natural products. With DEBS being one of the most well understood PKS, it has also been the subject of the first studies for this manipulation to yield these products. One of these studies, conducted by McDaniel, *et al.*, AT domains and different β-carbon processing domains from another organism were introduced into DEBS. The combinations of these substitutions yielded polyketides that utilized different starter units and had variably reduced carbons at the ketide positions. With the vast amount of experiments conducted a combinatorial library of over 60 analogs of 6-DEB were produced (McDaniel, *et al.*, 1999).

Other experiments with DEBS involved movement of the TE domain to DEBS I or II, which lead to the expected shortened product, and hybridization of whole modules from other PKSs, leading to chimeric products (Staunton, *et al.*, 1997). Strategies that either add or delete tailoring enzymes can also lead to increased diversity of polyketide products, and replacement of KR domains can change the stereochemistry of the alcohol (Staunton, *et al.*, 1997).

## I.C Difficidin

Though the strain does not produce a polyketide, *Bacillus subtilis* strain 168, a Gram-positive bacterium, contains a 15-gene operon that has six genes with sequence homology to PKSs. The rest of the genes did not show any similarity to PKS genes (see Table I-1) (Kunst, *et al.*, 1997).

Table I-1 *pksX* cluster found in *Bacillus subtil*is strain 168.

| **Gene Name** | **Function Based on Homology** |
|---|---|
| pksA | Transcriptional Regulator |
| pksB | Zn-dependant Hydrolase |
| pksC-E | Acyl Transferase (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) |

*Bacillus subtilis* strains 39320 and 39374 produce the broad spectrum antibiotic difficidin (Zweerink, *et al.*, 1987). Difficidin is of interest because it contains a subunit that cannot be explained by normal polyketide synthesis, being that the C3 position of difficidin has three carbons attached (see Figure I-5).

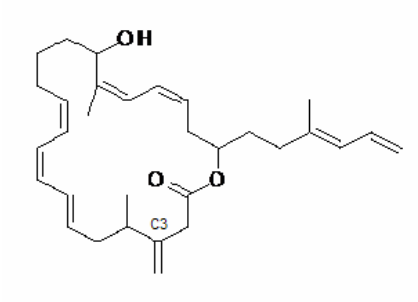


Figure I- 5 Difficidin.

Difficidin contains an external carbon-carbon double bond, instead of a usual polyketide subunit at the C3 position. Usual polyketide subunits are a ketone, an alcohol, a double bond, or a methylene group all with two carbons bound to the central ketide carbon. This branching subunit is an interesting system to study, and the results could potentially be used in combinatorial biosynthesis. Our lab is interested in the formation of this branching ketide, and our hypothesis is that two possibilities could lead to this subunit. The first hypothesis involves the attachment of a new type of precursor molecule that already contains the subunit.

This would occur in the following steps. PksC, an acyltransferase, charges AcpK with a malonyl group. PksF decarboxylates the malonyl group to form acetyl-AcpK. PksG then would condense acetyl-AcpK with acetoacetyl-PksL, which would be contained on one of PksL's thiolation domains, to produce an HMG-PksL intermediate. PksH would then dehydrate the HMG intermediate, creating an unsaturated carbon-carbon double bond. PksI would then decarboxylate this intermediate, leading to a nucleophilic attack on the nascent

polyketide chain contained on the second PksL thiolation domain. This reaction would yield a difficidin precursor with an internal carbon-carbon double bond at the C2-C3 position. To make difficidin, other PKSs within the operon would cyclize the open chain, and an isomerase would externalize the double bond at the C3 position. This hypothesis can be seen in Figure I-6.

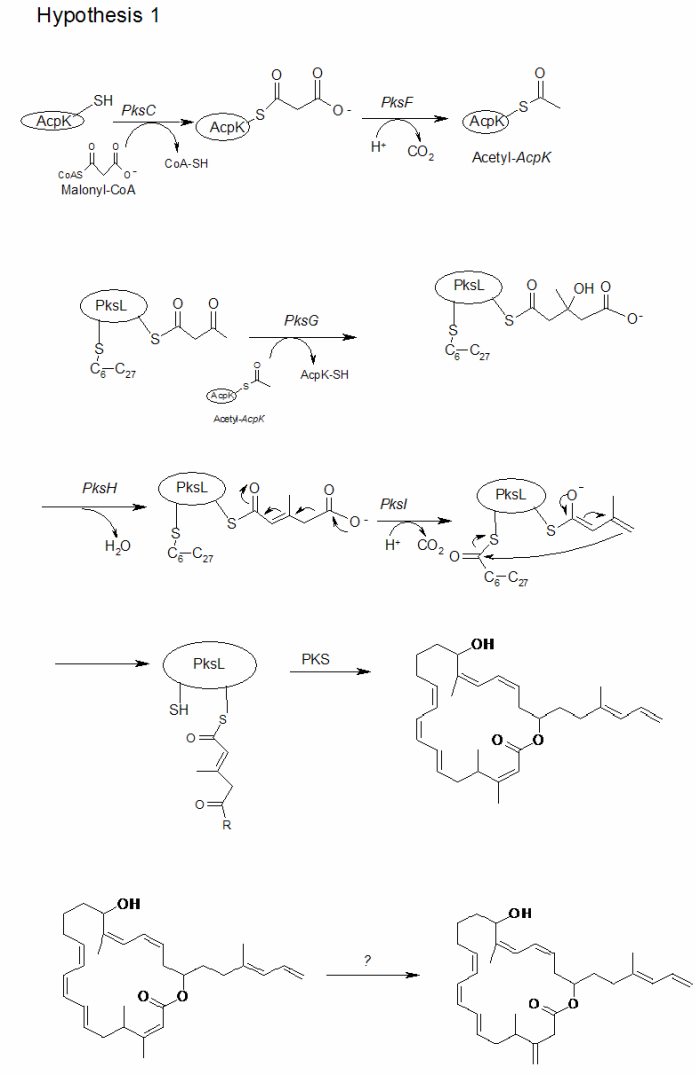


Figure I- 6 First hypothesis for difficidin production.

The second hypothesis is that the polyketide chain is grown through typical polyketide chemistry, and the external carbon atom is added after the backbone is complete. This process would also begin by loading of a malonyl unit onto PksC, and decarboxylation by PksF. In this hypothesis PksG would then condense acetyl-AcpK with a difficidin precursor molecule on PksL that contains a ketone at the C3 position. This would then lead to an HMG-like intermediate. PksH would then dehydrate the intermediate to produce an internal double bond at the C2-C3 position, and PksI would then decarboxylate.

Two possibilities exist for how the branching double bond in difficidin would be formed from this point. One possibility is that the enolate is protonated at the C2 position. In this case the bond at the C3 position would maintain an external double bond, and PKSs can then cyclize the open chain to directly form difficidin. The second possibility is after PksI decarboxylates, and the enolate is formed at the C1-C2 postion, the electron density goes into the branching double bond on the C3 position, causing a proton to be abstracted from a nearby acid. This leads to a fully formed difficidin chain with a carbon-carbon double bond at the C2-C3 position. This difficidin precursor will then be cyclized by other PKSs and the double bond will be externalized by some isomerase. The second hypothesis can be seen in Figure I-7.

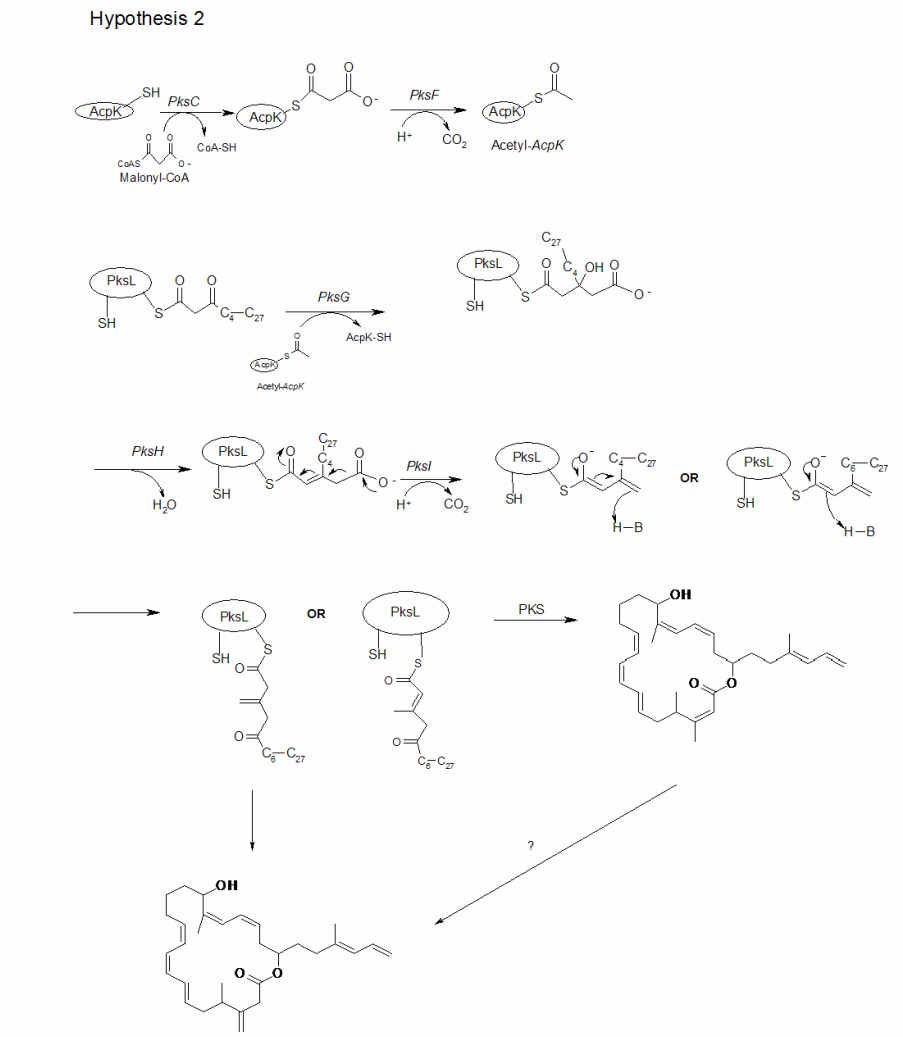


Figure I-7 Second hypothesis for difficidin production.

These two possibilities probably involve the proteins PksFGHI, and have been characterized by other workers (Calderone, *et al.*, 2006). The proteins PksFGHI found in *B. subtilis* 168 are not unique to this organism. Other homologs of these proteins can be found in other species of bacteria that produce polyketides with branching subunits that contain three carbons instead of the usual two carbons. Two specific examples of these polyketides are

curacin A, and myxovirescin (Simunovic, *et al.*, 2006; Gu, *et al.*, 2006). The genome of *Lyngbya majuscula*, the producer of curacin A, and the genome of *Myxococcus xanthus*, the producer of myxovirescin, contains genes that are homologs of to PksFGHI from *B. subtilis* 168. These polyketides can be seen in Figure I-8 and the branching subunits are circled. Curacin A and myxovirescin also have subunits in their structure that include an extra carbon. Research has shown that the homologous genes of PksFGHI, contained in the two organisms, are involved in the synthesis of these subunits (Simunovic, *et al.*, 2006; Gu, *et al.*, 2006).

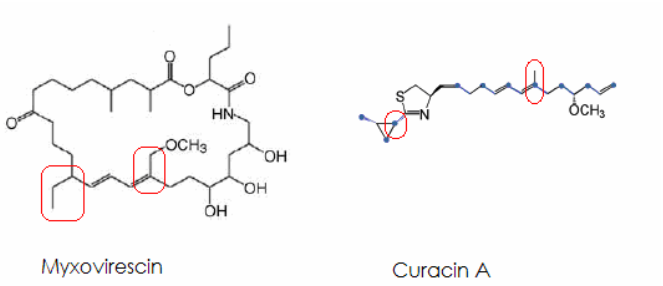


Figure I- 8 Structures of myxovirescin and curacin A with branching ketide positions circled (Simunovic, *et al.*; Gu, *et al.*).

Though reactions catalyzed by PksFGHI have been observed, the question of how the unusual subunit is installed in the full biosynthesis has not been completely elucidated. Research is still needed to determine if an unusual subunit is first synthesized by PksFGHI and then added to the backbone, or if the backbone is synthesized and acted on by PksFGHI after backbone synthesis.

The genome of *B. subtilis* shows that either hypothesis could be correct because there is another protein in the genome of *B. subtilis* 168, PksL, containing two thiolation domains.(Calderone, *et al.*, 2006). The two thiolation domains support the hypothesis that an unusual precursor could be synthesized and then loaded onto PksL which would contain the growing polyketide backbone, and then added to the backbone by other enzymes.

I.B.1 Roles of AcpK and Sfp in polyketide biosynthesis

Type I PKSs are organized in modules that contain a minimum of a β-ketoacyl synthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP). ACPs are synthesized in an apo, inactive, form. In order for the ACP to become active, its holo form, a phosphopantetheinyl prosthetic group from CoA is attached to a conserved serine residue in a post translational modification. This reaction is catalyzed by enzymes called phosphopantetheinyl transferases (PPTases) (Mootz, *et al.*, 2001). This reaction can be seen in Figure I-9.

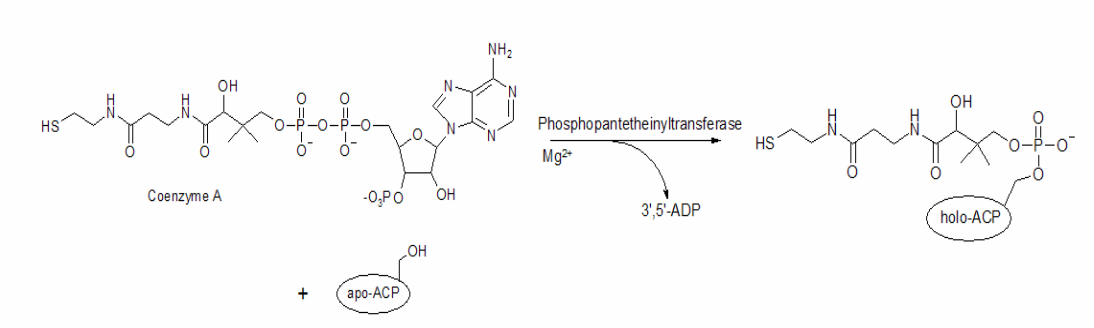


Figure I- 9 Phosphopantetheinylation reaction.

To initiate polyketide biosynthesis, an activated monomer is loaded onto the first holo-ACP. During chain elongation, the chain is docked at an upstream ACP and undergoes nucleophilic attack initiated by the downstream KS domain on the downstream module. This ensures the direction of chain transfer from the upstream ACP to the next downstream KS (Stein, 2005). According to the auxiliary domains present in the module the chain is passed from active site to active site until the full module has been utilized. The finished chain is then extended and processed by the next downstream module. The genome of *Bacillus subtilis* strain 168 contains the *pksX* cluster (refer to Table I-1). The gene *acpK* is part of this operon, and codes for a freestanding ACP (Kunst, *et al.*, 1997). In *B. subtilis* the PPTase enzyme is called Sfp. However, in strain 168 of this bacterium, Sfp is synthesized in an inactive form due to a frame shift mutation (Stein, 2005).

In work by Nakano, *et al.*, a *sfp* gene was cloned from *Bacillus subtilis* that codes for an active protein and in research done by Mootz, *et al.*, Sfp was found to catalyze the phosphopantetheinylation reaction on AcpK (Nakano, *et al.*, 1992; Mootz *et al.*, 2001). This research showed that Sfp will use CoA to phosphopantethinylate AcpK; however, it has been shown that Sfp can also catalyze the same reaction using different acyl CoAs. This ability to make holo-AcpK directly charged with acyl chains will be important to our studies of the synthesis of difficidin and bacillaene. The products that are of interest are acetyl-holo-AcpK, acetoacetyl-holo-AcpK, malonyl-holo-AcpK, and holo-AcpK. Acetyl-

AcpK will be important for functional studies using PksG, which will be discussed in more detail in the following section. For this study overexpression and purification of AcpK, cloned from *B. subtilis* strain 168, and Sfp, from a plasmid donated to the lab from the Zuber group, will be necessary (Nakano, *et al.*, 1992). Finally, the different phosphopantetheinylation reaction conditions will need to be optimized. The information obtained from these reactions will be useful in later studies concerning the enzymology of polyketide biosynthesis in *B. subtilis*.

I.B.2 Function of PksG in difficidin production

Another research goal outlined in this thesis is the determination of the chemical function of PksG in difficidin synthesis. There are two hypotheses concerning PksG, The first involves an acetoacetyl group, which is attached to one ACP domain on PksL, is reacted with acetyl-AcpK in the presence of PksG. This reaction produces an HMG-PksL precursor that would later be introduced into a previously synthesized difficidin chain (see Figure I-10). Then through subsequent reactions difficidin will be synthesized with the exocyclic double bond.

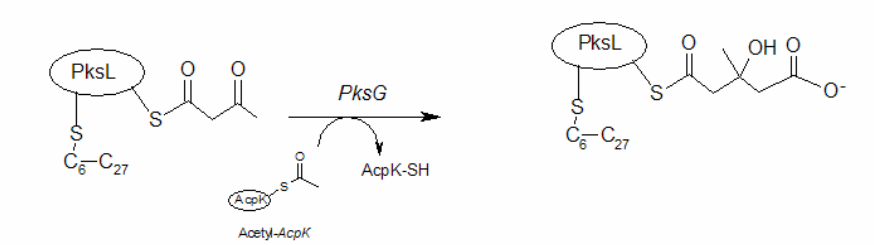


Figure I- 10 PksG function in first hypothesis.

The second hypothesis involves PksG acting on a fully formed difficidin chain on one ACP domain of PksL. PksG would then condense the β-keto group with acetyl-AcpK to form an HMG-like intermediate on the longer polyketide chain (see Figure I-11). Enzymes (PksHI) would then decarboxylate and dehydrate this intermediate to yield a mature difficidin molecule with the exocyclic double bond.

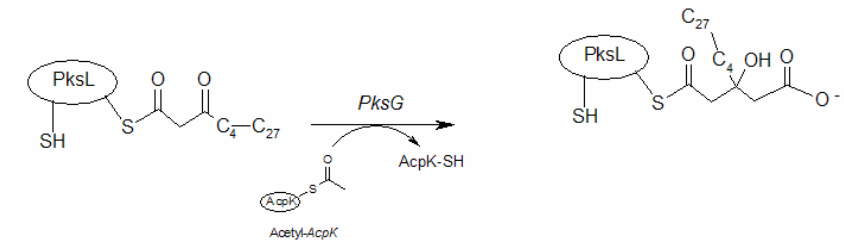


Figure I- 11 PksG function in second hypothesis.

This project aims to determine the activity of PksG from *B. subtilis* 168. In order to successfully carry out this study we will require data obtained from the above section concerning the optimized PPTase reactions with AcpK. In studies performed by Calderone, *et al.* the characterization of PksCFGHI had been

successfully completed prior to this thesis. This research conclusively showed that PksG can act as an HMG-PksL synthase (Calderone, *et al.*, 2006). This result, however, does rule out the second mechanism in Figure I-7.

I.B.3 Functionality of MmgA

*Bacillus subtilis* has long been known to have the ability to sporulate. This is a response to environmental stresses like starvation. The end result is that one cell makes a daughter cell that eventually becomes an endospore, while the mother cell commits most of its resources to the sporulation process and at the end finally lyses releasing the spore. This process can be divided into five phases, with four being under the control of specific sigma factors (Errington, *et al.*, 1993). One of these sigma factors, $\sigma^E$, is used in the intermediate stages of the sporulation process. Bryan, *et al.* have uncovered an operon that is under control of a $\sigma^E$ promoter which they called mother cell metabolic genes, the *mmg* operon (Bryan, *et al.*, 1996). This group reported that there were five genes in this operon, *mmgABCDE*, however the full genome sequence later showed that the operon contains one additional gene under the same control, called *yqiQ*, which could potentially be renamed *mmgF* (Subtilist Web Server, accessed March 12, 2007).

The gene *mmgA* shows acetyl CoA acetyltransferase homology (Bryan, *et al.*, 1996). This type of enzyme catalyzes the condensation of two acetyl CoA units to produce one acetoacetyl CoA unit and a CoA unit. Conversely these

enzymes have been able to catalyze the reverse reaction; the thiolysis of one acetoacetyl CoA, by CoA, to yield two equivalents of acetyl CoA (see Figure I-12).

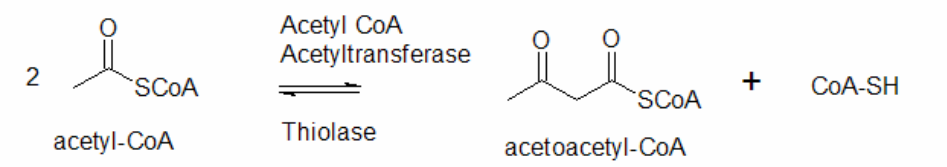


Figure I- 12 Proposed MmgA reaction.

Having an enzyme that has this functionality would be helpful for experiments concerning the specific substrate for PksG in polyketide biosynthesis. There have been two hypotheses proposed, and since there is no direct way to determine which hypothesis of difficidin production is correct, another study will be used. This study requires a PksG$^-$ mutant of *B. subtilis* strain 39374, the producer of difficidin (Zweerink, *et al.*, 1987). PksG, a 3-hydroxy-3-methylglutaryl (HMG) synthase, acts on either the acetoacetyl group attached to one ACP domain on PksL, or on the long chain difficidin precursor molecule. PksG is an important step in the synthesis of difficidin and by removing this enzyme from the *in vivo* difficidin synthesis process it is possible to reveal its role in polyketide biosynthesis by showing new metabolism or demonstrate natural requirements caused by the mutation. Because a PksG$^-$ mutant would not be able to create the HMG intermediate, a labeled replacement

substrate, $^{13}C$ labeled 3-hydroxy-3-methylglutaryl-*N*-acetylcysteamine (HMG-SNAC), will be introduced to a culture of PksG$^-$ *B. subtilis* 39374.

MmgA could be used in a synthesis process to make $^{13}C$ labeled 3-hydroxy-3-methylglutaryl-*N*-acetylcysteamine (HMG-SNAC), as this compound is not available commercially. The synthesis of HMG-SNAC would occur in the following steps. First, labeled pyruvate will be reacted with CoA and $NAD^+$ to produce labeled acetyl CoA, in a reaction catalyzed by commercially available pyruvate dehydrogenase. MmgA would then be used with two labeled acetyl CoA units to produce labeled acetoacetyl CoA. Labeled HMG-CoA will be produced by a HMG-CoA synthase from *Staphylococcus aureus* using labeled acetoacetyl CoA and labeled acetyl CoA (Campobasso, *et al.*, 2004). Finally, HMG-CoA will be reacted with SNAC to make labeled HMG-SNAC (see Figure I-13).

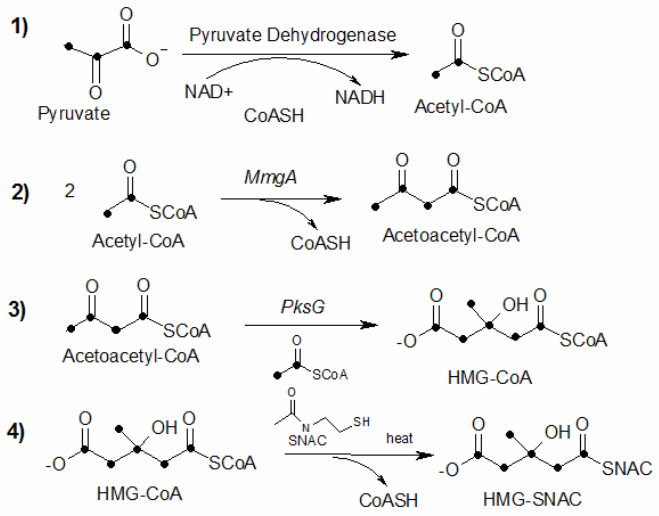


Figure I- 13 Synthesis of labeled 3-hydroxy-3-methylglutaryl-*N*-acetylcysteamine.

If the first hypothesis is correct, the culture will then directly use the labeled HMG-SNAC, and produce a labeled difficidin product, which can be detected using mass spectral analysis. If the second hypothesis is correct, the culture will not be able to use this labeled substrate and we can instead expect a new difficidin derivative having a ketone group at the C3 position. This project has two goals, the first aims to clone, overexpress, purify, and biochemically characterize the enzyme MmgA, and the second is to use MmgA in reactions for the synthesis of labeled substrates.

# CHAPTER II

# OVEREXPRESSION, PURIFICATION, AND REACTIONS INVOLVING ACPK, SFP, AND PKSG

## II.A Introduction to AcpK and Sfp

In polyketide biosynthesis, acyl carrier proteins (ACPs) carry carboxylic acid building blocks and incomplete polyketide chains that, through Claisen condensations, result in the synthesis of a ketide chain. These acyl chains are attached to the ACPs on the terminal thiol of a phosphopantetheine prosthetic group which is attached to a conserved serine residue. ACPs are not synthesized with the phosphopantetheinyl (PPT) moiety attached, so post translational modification is necessary to convert the ACP from its inactive apo form to the active holo form. A second enzyme is needed to carry this conversion out. Phosphopantetheinyltransferases (PPTases) catalyze the transfer of PPT from Coenzyme A to a specific serine residue of the ACP (Lai,*et al.,* 2006). This general reaction can be seen in Figure II-1.

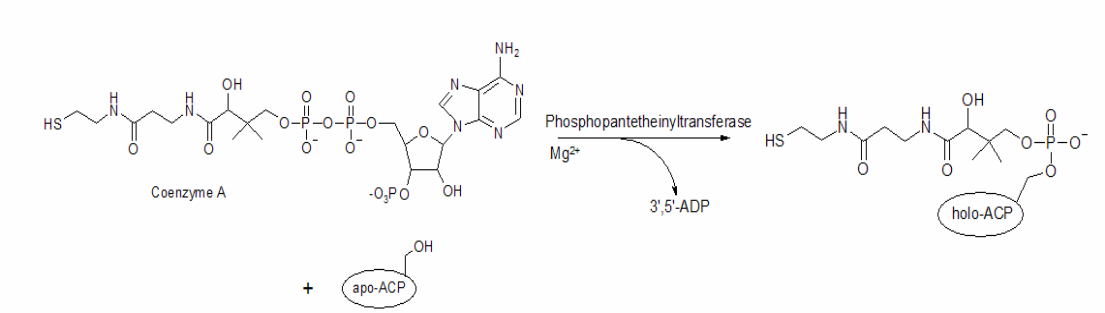


Figure II- 1 Phosphopantetheinyltransferase Reaction.

The protein AcpK (acyl carrier protein K) is encoded by the genome of *B. subtilis* strain 168 and is part of the *pksX* cluster. Like all ACPs, AcpK also requires a post translational modification to make an active (holo) protein. In studies by Mootz, *et al.*, AcpK had been reacted with Coenzyme A (CoA) in the presence of a known PPTase, Sfp (Mootz, *et al.*, 2001). Sfp is found in the genome of *B. subtilis*, and is necessary for the production of the antibiotic surfactin. Sfp stands for surfactin production, which was a genotypic designation given to the *B. subtilis* strains that were known to produce surfactin and a *sfp*$^{o}$ genotype was given to surfactin non-producing strains. The gene *sfp* was known to be important to the production of surfactin when it was originally studied by Nakano, *et al.*, though the function was not yet determined (Nakano, *et al.* 1992). Later, when this function was elucidated, the gene kept its original designation. The function of Sfp is to phosphopantetheinylate seven PCP (peptidyl carrier protein) domains in three subunits of the surfactin synthase (SrfABC). This protein is also of particular interest because it is able to phosphopantetheinylate PCPs and ACPs domains from other organisms, including *Saccharomyces*

*cerevisiae* and *Escherichia coli*, making it a broadly specific protein (Quadri, *et al.*, 1998).

In the studies by Mootz, *et al.*, Sfp was found to be able to charge the phosphopantetheinyl moiety from CoA to AcpK in the presence of $Mg^{2+}$. (Mootz, *et al.*, 2001) The literature indicates that Sfp is not only capable of putting the usual phosphopantetheinyl chain from CoA on ACPs, but also has broad substrate specificity and can charge ACPs with the phosphopantetheinyl moiety of CoAs that have different acyl groups attached to the terminal thiol. This ability to directly synthesize R-holo-AcpK proteins from apo-AcpK and R-CoAs will be helpful in the studies of understanding the biosynthesis of difficidin.

Difficidin is produced by *B. subtilis* strains 39320 and 39374. (Zweerink, *et al.*, 1987) In normal polyketide chemistry the ketide chain is extended two carbons at a time. Difficidin has an unusual subunit at C3, an exocyclic double bond that can not be explained by normal polyketide chemistry (see Figure II-2). The synthesis of this subunit likely involves homologs of the enzymes PksFGHI, though the sequence in which this happens is still unknown.

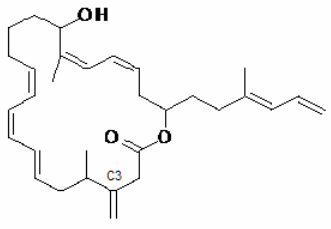


Figure II- 2 Difficidin.

Two hypotheses have been proposed about the synthesis of this subunit in difficidin. The first (Figure II-3) involves the synthesis of an unusual precursor that is incorporated in the difficidin molecule by the polyketide synthases. The other hypothesis (Figure II-4) uses standard polyketide chemistry that generates a difficidin precursor containing a ketone at the C3 position. Other enzymes (PksFGHI) would then act on this precursor, using acetyl-AcpK as the source of the third carbon to form difficidin.

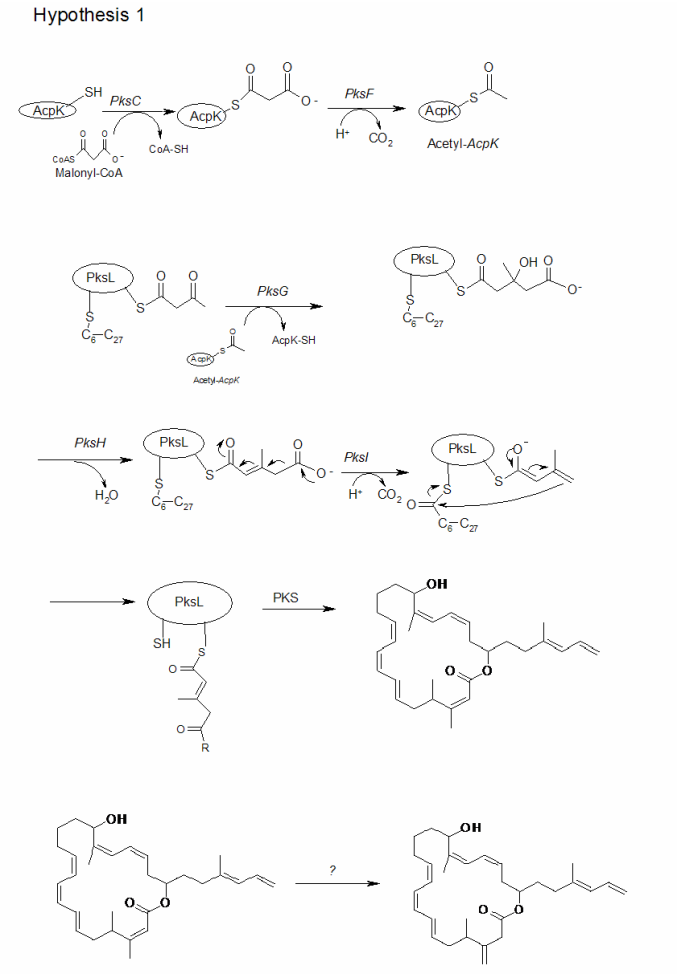


Figure II- 3 First hypothesis of difficidin production.

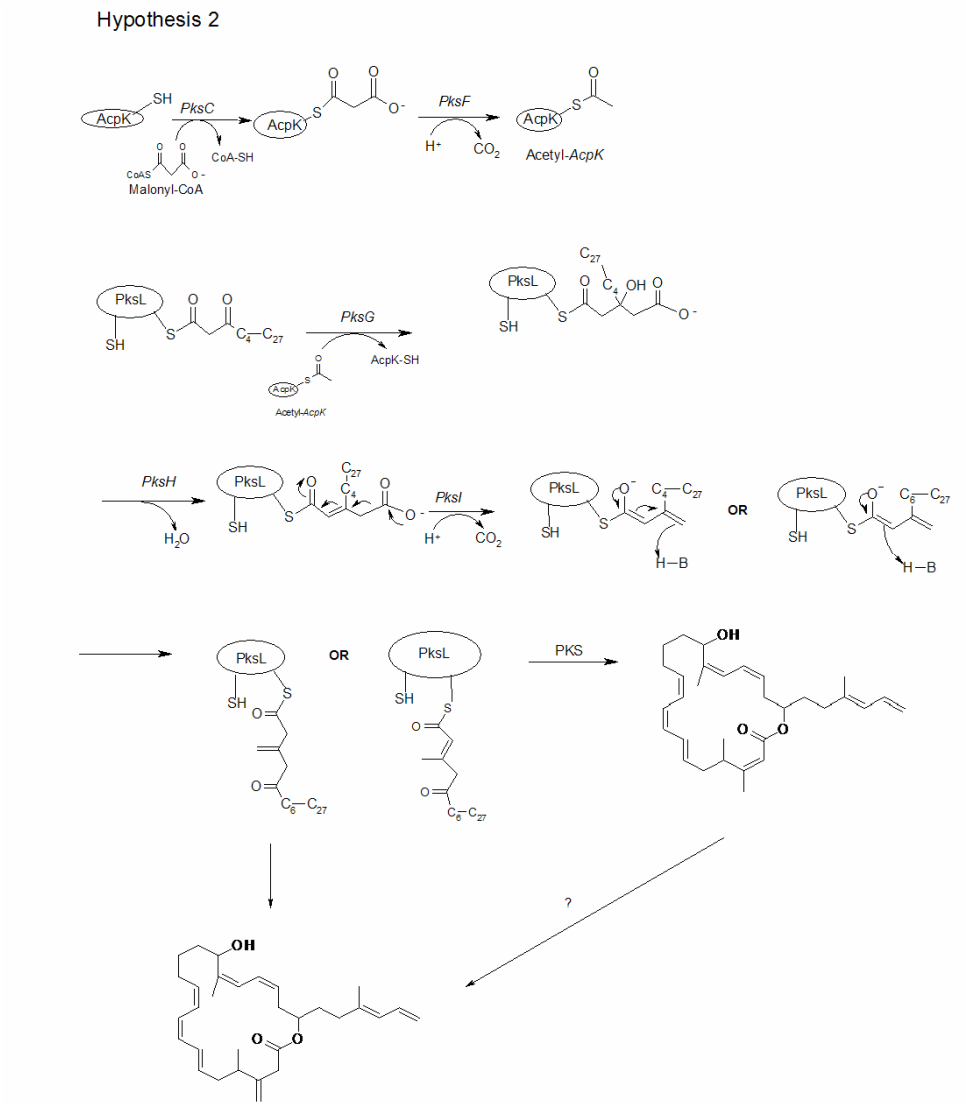


Figure II- 4 Second hypothesis for difficidin production.

PksC, an acyltranferase, uses malonyl-CoA to produce malonyl-AcpK. This is then decarboxylated by PksF to make acetyl-AcpK. PksG uses acetyl-AcpK and PksL with either an acetoacetyl group on one of its thiolation domains, or the ketone with the difficidin precursor to produce a HMG subunit. This then

undergoes a dehydration and decarboxylation by PksH and PksI, respectively. Finally, standard polyketide chemistry is used to produce the difficidin product.

### II.B Introduction to PksG

PksG from *B. subtilis* strain 168 has been studied by Calderone, *et al.* and has found that it catalyzes a reaction similar to HMG-CoA synthases. (Calderone, *et al.*, 2006) PksFGHI are responsible for synthesizing the unusual subunit found in bacillaene (see Figure- 5) (Butcher, *et al.*, 2007). Homologs of PksFGHI in *B. subtilis* 39374 and 39320 may be responsible for the branching subunit found in difficidin. These genes are also contained in the genome for *B. amyloliquefaciens* this species has been known to synthesize difficidin (Chen, *et al.*, 2006). These genes, PksFGHI, are not unique to the genome of *B. subtilis* since homologs are found in the genomes of other organisms such as *Myxococcus xanthus* and *Lyngbya majuscula*. *Myxococcus xanthus* and *Lyngbya majuscula* produce myxovirescin and curacin A, respectively (see Figure II-6) (Simunovic, *et al.*, 2006; Gu, *et al.*, 2006). These polyketides have branching polyketides similar to the one at C3 in difficidin. In the research done on the synthesis of myxovirescin, the PksFGHI homologs TaKC/FXY from *M. xanthus*, are likely responsible for synthesizing the unusual subunits found in the final polyketide in a similar pathway proposed for the *B. subtilis pksX* enzymes (Simunovic, *et al.*, 2006). *M. xanthus* has two locations that have a third carbon on the ketide position, and the genome also has two homologs of PksG, called

TaC and TaF, that lead to an HMG intermediate. *L. majuscula* also has homologs to PksFGHI called CurCDEF that leads to the synthesis of the cyclopropane substituent (Gu, *et al.,* 2006).

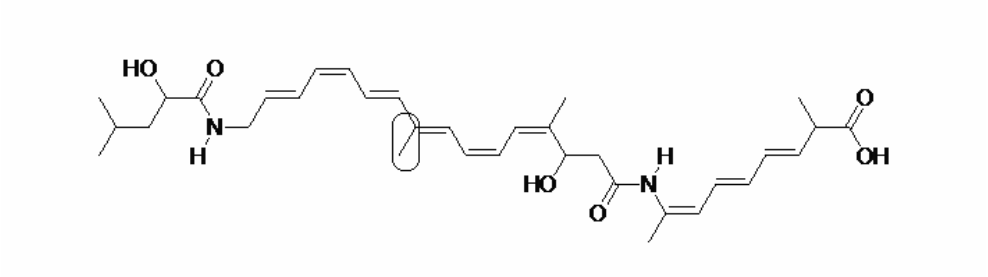


Figure II- 5 Bacillaene, with branching subunit circled.

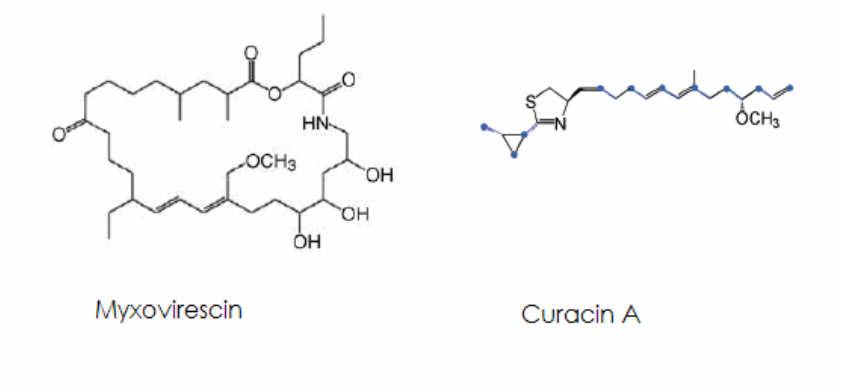


Figure II-6 Structure of myxovirescin and curacin A (Simunovic, *et al.*, 2006; Gu, *et al.*, 2006).

Though the enzymatic activity of PksG from *B. subtilis* 168 has been measured, the timing of its involvement in the overall polyketide biosynthesis is still in question. Studies on the substrate specificity of PksG can help elucidate the order of steps in this biosynthesis. In the two hypotheses, PksG is proposed to react with substrates of vastly different length; a 4 carbon acetoacetyl chain in

hypothesis 1, or a 27 carbon chain in hypothesis 2. In order to study the substrate specificity of PksG, β-keto acyl chains of varying length can be synthesized and loaded onto PksL by Sfp. PksG in the presence of acetyl-AcpK can then react with these different length chains to form the HMG-PksL product (see Figure II-7 for reaction scheme).

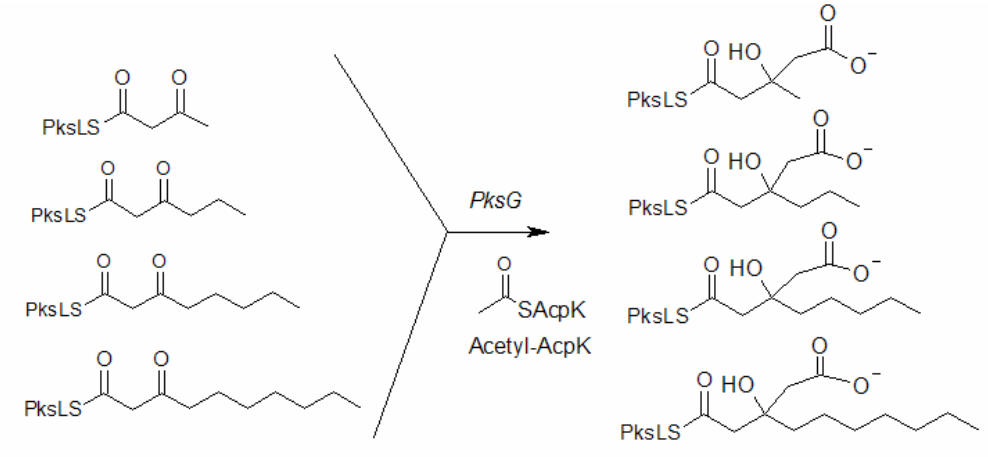


Figure II- 7 Reaction of PksG with β-keto acyl PksL Chains in the Presence of Acetyl-AcpK.

Both the rates of the reactions with different substrates and mass spectral analysis of the products can help elucidate the substrate specificity of PksG. The rates of reaction for each β-keto acyl-PksL can be compared. Higher reaction rates with shorter β-keto acyl chains would show that PksG has a preference for smaller substrates, thereby supporting hypothesis 1. Hypothesis 2 would be supported if PksG had a preference for longer chained substrates. In order to conduct these studies, we require purified PksG, PksL, and AcpK. Longer β-keto acyl chains would have to be synthesized, since they are not commercially

available, and these chains would have to be loaded onto the PksL protein. Acetyl-AcpK would also have to be produced using acetyl-CoA and Sfp as outlined in the studies of AcpK found in II.D.4.a.

Other studies involving PksG would still use the standard reaction seen in Figure II-7, but another enzyme will be used to produce $^{13}$C-labeled long β-keto acyl chains. An acetyl-CoA acetyltranferase can be used to react labeled acetyl-CoA derivatives to make the desired product. These labeled chains would then be loaded onto PksL and, of a mixture containing all the different chain length reactants, reacted with PksG in the presence of excess labeled acetyl-AcpK. The product abundance can be determined my mass spectral analysis. Hypothesis 1 would be supported if the most abundant products formed were from the shorter β-keto acyl chains to produce a shorter chain HMG product. Conversely, hypothesis 2 would be supported if the most abundant products were formed from the longer β-keto acyl chains to form a longer chain HMG product.

## II.C Production and Characterization of AcpK, Sfp, and PksG

### II.C.1 Goals for the Production and Purification of AcpK

In order to test any hypothesis for difficidin production, we require pure holo-AcpK from *B. subtilis* 168, and also pure Sfp. AcpK had successfully been cloned into pQE-60 by Mootz, *et al.* Since this research showed the pQE system yielded a pure protein in high amounts, our goal was to also clone into this same

system using pQE-60 (Mootz, *et al.*, 2001). Overexpression using pQE plasmids involves the strong T5 promoter on the plasmid and T5 RNA polymerase in the host. The plasmid pQE-60 contains a T5 promoter sequence and two lac operator sequences for increased *lac* repressor binding. The LacI repressor protein is encoded by a *trans* plasmid, pREP4, which controls the overexpression of the target protein. Induction of the protein is started by disengaging the LacI protein from the lac repressor by addition of isopropyl-β-D-thiogalactopyranoside (IPTG), an artificial inducer of the lac operon. Once IPTG is added, the cell commits nearly all of its resources to produce the gene under control of the T5 promoter (The Qia*expressionist,* 5$^{th}$ ed.). This is referred to as "overexpressing" the protein. Here the gene under the control of the T5 promoter is *acpK*. This gene would be cloned into pQE-60 using a restriction digest of both pQE-60 and PCR-purified *acpK*, and a ligation would be performed using T4 DNA ligase. This cloned plasmid, pQE-60/*acpK*, would then be transformed into competent *E. coli* cells that contain the plasmid pREP4. Overexpression would then be achieved by growing a culture of cells to mid-log phase and adding the inducer, IPTG.

The second goal would be purification of the overexpressed protein. Because pQE-60 offers the option of including a C-terminal 6x histidine tag, affinity chromatography will be used to purify recombinant His-tagged AcpK. For affinity chromatography we will use nitrilotriacetic acid (NTA) as a $Ni^{2+}$ chelating adsorbent. When a Ni-NTA column is used there are two vacant sites on the $Ni^{2+}$

that can bind the imidazole nitrogens present in the $His_6$-tagged protein (see Figure II-8). Low concentrations of imidazole will remove proteins that have histidine residues, but not a $His_6$-tag, from the column. A final, optimal concentration of imidazole would be used to elute the pure AcpK from the column. The final step would be the enzymatic modification of AcpK by a phosphopantetheinylation reaction with CoA substrates containing different acyl chains. These reactions will be analyzed by Electrospray Ionization Mass Spectroscopy (ESI-MS) and Matrix Assisted Laser Desorption/ Ionization Time-of-Flight Mass Spectroscopy (MALDI-TOF-MS).

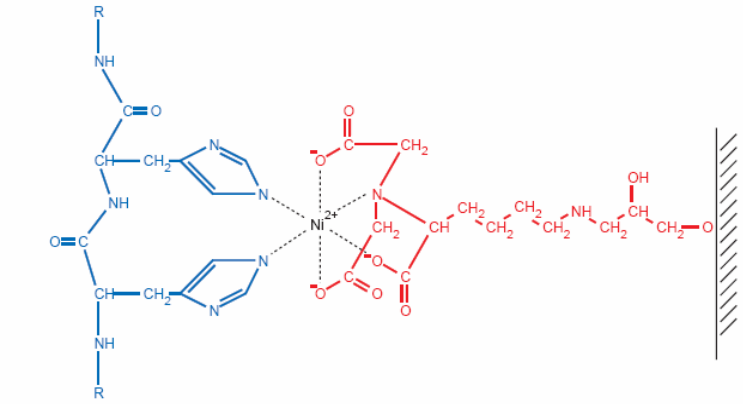

Figure II- 8 Nitrilotriacetic acid chelated to nickel and associated with two histadine sidechains of a protein (The Qia*expressionist,* 5th ed.).

II.C.2 Goals for the Production and Purification of Sfp

In order to perform the phosphopantetheine transfer reaction, large amounts of pure Sfp would be needed. During the development of the work for this thesis, two methods were used to obtain purified Sfp. Sfp had been

previously cloned into the overexpression vector pUC8 by Nakano, *et al.*, and this vector was obtained as a gift from the Zuber group. (Nakano, *et al.*, 1992) In our first procedure we used this pUC8 system for this enzyme by modifications of the method of Nakano, *et al.* (Nakano, *et al.*, 1992). This method involved introducing the pUC8-sfp vector into competent *E. coli* MV1190 cells, a system which automatically overexpressed Sfp without the need for chemical induction. After the cells had grown overnight, ammonium sulfate precipitation was used to purify Sfp.

Because the first method of purification of Sfp did not yield a highly pure product, a second method was used. To improve the efficiency and yield of Sfp we decided to clone Sfp into a His-tag system. The vector chosen was pET-28a, a vector that places the gene of interest under the control of the T7 promoter. This promoter is under control of the *lacI* gene, which produces the LacI repressor protein (pET System Manual, 11$^{th}$ ed.). After addition of the inducer IPTG, the cell then overexpresses the target Sfp protein.

To clone the *sfp* gene into pET-28a, a PCR of the pUC8/*sfp* plasmid would have to be performed, and a restriction digest is required of the pET-28a vector, and the similarly digested *sfp* PCR product. This product is then ligated into the plasmid by T4 DNA ligase, and transformed into competent *E. coli* cells. Overexpression is performed by growing a culture of cells to mid-log phase, and induction with IPTG. Purification would then be performed by Ni-NTA affinity

chromatography. Sfp would then be used with purified AcpK in phosphopantetheinylation reactions.

### II.C.3 Goals for the Production and Purification of pBAD/*pksG*

To test our hypotheses for PksG activity it was necessary to have a large amount of pure PksG. To accomplish this, PksG will need to be cloned into an overexpression vector, overexpressed in culture, and finally purified. In prior work done by another graduate student, Kevin Williams, attempts at cloning PksG into the pET-28a overexpression system were not successful. For this reason it was thought that PksG is toxic to *E. coli*. Because of this hypothesis, it was determined *pksG* needed to be placed under a more tightly regulated expression system, so the pBAD system was chosen for this reason.

The pBAD system uses the araBAD promoter, which more tightly regulates the expression of the target protein. The protein AraC creates a DNA loop that prevents transcription. Arabinose binds to the AraC protein and releases the DNA loop that stopped the transcription from occurring, thus making arabinose the “inducer” for overexpression of the target PksG protein (Guzman, *et al.*, 1995). The pBAD system was used, though in later studies done by Calderon, *et al.* PksG was cloned and overexpressed using the pET system, showing that PksG is not a toxic protein (Calderone, *et al.*, 2006).

Once *pksG* is in an overexpression vector, it is possible to obtain large amounts of the protein. To do this a culture of cells will be grown to mid-log

phase and induced using arabinose. To purify PksG, a Ni-NTA column will be used because the protein was cloned in frame with a C-terminal $His_6$-Tag coded by the plasmid. The purified protein will then be used for activity assays with the enzyme's suspected substrates.

## II.D Results

### II.D.1.a Cloning of pQE-60/*acpK*

To clone *acpK* into pQE-60, primers were designed for polymerase chain reaction (PCR). Using *B. subtilis* strain 168 as the template DNA and Vent polymerase, a PCR reaction was performed with 60 $^{o}$C as the annealing temperature. The gene length of *acpK* is 246 base pairs (bp). The PCR reaction was determined to be successful because a band appeared on an agarose gel between 200-300 bp. Once *acpK* had been successfully amplified, a restriction digest was performed on both the PCR product and the pQE-60 plasmid. The restriction enzymes used for the digest of the PCR product was *NcoI* and *BglII*, while the plasmid was digested with *BamHI* and *NcoI*. *BglII* and *BamHI* leave the same overhang when it cleaves DNA, which made it possible to still ligate the two fragments together using T4 DNA ligase. Multiple ligation mixtures were used that contained different ratios of insert to vector, and the ligation reaction mixtures were transformed into competent DH5α cells. The cells were then plated on ampicillin plates and incubated overnight at 37$^{o}$C. The following day ten transformants were chosen to inoculate ten starter cultures that were grown

overnight in 5mL of Luria Bertani (LB) media containing 25µg/mL of ampicillin. The cultures were shaken overnight at 37$^{o}$C, and all plasmids were purified the following morning except for one culture which did not grow.

II.D.1.b Gel Electrophoresis

To test for insert acceptance into all nine purified plasmids, nine PCR reactions. The primers used were from the PCR amplification of *acpK* from *B. subtilis* strain 168. The appearance of a band with the same number of base pairs as *acpK*, ~250 base pairs (bp), on an agarose gel would indicate a successful ligation. A gel showed five out of nine plasmids contained the desired gene in the pQE-60 vector.

II.D.1.c Retransformation and Re-purification of pQE-60/*acpK*

Once the cloning step was completed successfully, one cloned vector was chosen to be transformed into competent *E. coli* strain M15[pREP4] cells. The cells were then grown at 37$^{o}$C containing kanamycin and ampicillin. Kanamycin and ampicillin selection was used because the pQE-60 plasmid confers ampicillin resistance, while the pREP4 vector confers kanamycin resistance, so it is possible to have a culture of cells replicating the two plasmids.

II.D.1.d Overexpression and Purification of pQE-60/*acpK*

The target gene, *acpK*, had been successfully cloned in frame with a C-terminal His-tag coded by the plasmid pQE-60. The next step was to overexpress the protein using the cloned plasmid. The strain pQE-60/*acpK* was grown in LB media containing 25µg/mL of ampicillin and 30µg/mL of kanamycin in a 37 $^{o}$C shaker until mid-log phase was reached. The culture was then induced and allowed to overexpress for three hours and the cells were harvested. Once harvested the cells were resuspended and lysed to extract the protein from inside the cell and put it in solution.

Once the protein was in solution, the crude extract was applied to a Ni-NTA affinity column and various concentrations of imidazole were used to wash proteins lacking His-tags from the column, and an optimal concentration was used to elute AcpK. The eluted protein was then dialyzed overnight to remove excess imidazole and salt. The following day the dialyzed protein was concentrated by centrifugation using a Vivaspin 5000 Molecular Weight Cut Off (MWCO) concentrator. The protein was stored at -80$^{o}$C with 10% glycerol. An SDS-PAGE (Sodium Dodecyl Sulfate PolyAcrilamide Gel Electrophoresis) was performed on the crude AcpK extract, eluents from the various imidazole washes, the eluted protein, and the concentrated protein. This SDS-PAGE can be seen in Figure II-9 which shows a bright band below the 14.4 kilodalton (kDa) molecular weight marker. The expected mass of AcpK is 10561 Daltons and would be expected to migrate below the lowest molecular weight marker.

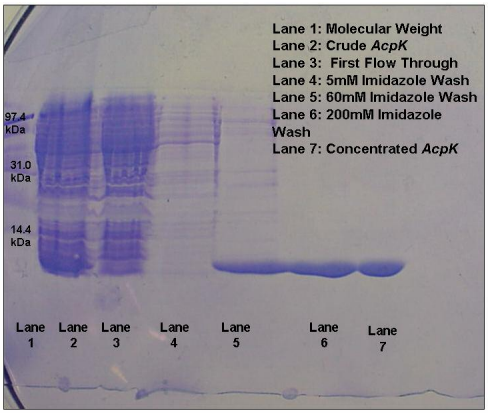

Figure II-9 SDS-PAGE of AcpK column chromatography purification.

II.D.1.d Protein Concentration Determination

The concentration of AcpK was determined by the Bradford method using Bovine Serum Albumin (BSA) as the standard. BSA is available commercially at a concentration of 2.0mg/mL, and a standard curve was generated by varying the concentrations of albumin. The average amount of AcpK obtained from a 1L culture was ~900µg.

II.D.2.a Overexpression and Purification of Sfp by a Previously Published Method

In previously published work, the overexpression and purification procedures for Sfp had been studied and optimized by Nakano, *et al.* (Nakano, *et al.*, 1992). Their procedure was followed to overexpress and purify the Sfp enzyme with the one exception of not using a DEAE column chromatography step after ammonium sulfate precipitation. An overexpression vector that already contained the *sfp* gene, pUC8/*sfp*, was donated to the lab. This vector was

transformed into competent MV1190 *E. coli* cells, and were grown overnight on plates that contained 25µg/mL of ampicillin. To overexpress Sfp, a culture of cells was grown overnight and harvested the next day. After a sonication step the protein was ready to be purified.

To purify the Sfp enzyme, ammonium sulfate precipitation was used. Four different percent concentrations of ammonium sulfate were used: 35%, 50%, 65%, and 80%. The final ammonium sulfate concentration precipitated the Sfp protein. After centrifuging the 80% solution, the supernatant was discarded, and the pellet was retained. This pellet was resuspended in buffer (20 mM Tris·HCl, pH 8.0) and dialyzed overnight in the same buffer the pellet was resuspended in. The enzyme was concentrated the next day, and glycerol was added to the concentrated protein at 10% concentration for $^{-}80\ ^{o}C$ storage.

II.D.2.b Cloning of pET-28a/ *sfp*

To clone the *sfp* gene into pET-28a, a PCR was performed using the pUC8/*sfp* plasmid as the template. Primers incorporated restriction sites for *NcoI* and *XhoI*, upsteam and downstream of the insert, respectively. The PCR analyzed by agarose gel electrophoresis showed two bands on the gel, one at the correct size for *sfp* and one having a smaller size. The correctly sized PCR product was purified by low melting agarose gel electrophoresis with the extraction of the band having the correct size. To clone the amplified gene into pET-28a, a restriction digest was performed using *NcoI* and *XhoI* on both the

correctly sized PCR product and the pET-28a plasmid. DNA ligase was used to ligate the restriction digested plasmid and PCR product together. Multiple ligation reactions were performed that contained different insert to plasmid ratios.

The ligation reactions were transformed into competent *E. coli* DH5α cells and plated. After incubation overnight, only two transformants had grown. These transformants were inoculated into a starter culture of LB containing 30μg/mL of kanamycin. The cultures were allowed to shake overnight at 37 $^{o}$C. The following day the plasmids were purified. To test for insert acceptance into the pET-28a plasmid, a second PCR was performed using Vent polymerase and the purified plasmids as the template DNA. Agarose gel electrophoresis confirmed both plasmids contained the *sfp* gene. Once the cloning of *sfp* into pET-28a was successful, one vector was chosen to be transformed into competent *E. coli* BL21(DE3) cells.

II.D.2.c Overexpression and Purification of pET-28a/*sfp*

The gene, *sfp*, had been successfully cloned in frame with a C-terminal His-tag coded for by the plasmid pET-28a. The next step was overexpression of the protein. Using the pET-28a/sfp overexpression strain, a culture of cells was grown in LB media containing 30μg of kanamycin per mL of culture. The culture was shaken at 37 $^{o}$C until mid-log phase was reached, at which time the culture was induced and grown for three hours. The cells were then harvested,

resuspended, and sonicated. After a final centrifugation step, the protein was purified by Ni-NTA chromatography.

To purify the enzyme the crude extract was applied to an Ni-NTA column. Various concentrations of imidazole were used to was nonHis-tagged proteins from the column, and an optimal concentration was used to elute Sfp. The eluted protein was then dialyzed overnight. The following day the dialyzed protein was concentrated by centrifugation. The concentration was then determined using the Bradford method. The average amount of Sfp obtained was ~1125µg per 1L of culture. To be sure the correct protein was overexpressed, the various washes from the Ni-NTA column, and the concentrated, dialyzed eluent were analyzed by SDS-PAGE. The gel showed a band at the expected mass of Sfp (~28 kDa), so it was determined the overexpression/purification was successful. Sfp was then ready to undergo reactions with AcpK and different acyl CoAs.

### II.D.3 Cloning, Overexpression and Purification of PksG

The gene *pksG* from *B. subtilis* strain 168 had been cloned and inserted into the pBAD overexpression plasmid by a former lab colleague, Kevin Williams. This plasmid was retransformed into the *E. coli* TOP10 strain and were plated and incubated for 37$^{o}$C overnight. To overexpress PksG, a culture was started which contained 25µg/mL ampicillin, and was allowed to grow until mid-log phase. The culture was then induced with 0.01 M arabinose. Growth was continued for three hours after which the cells were harvested by centrifugation.

The pellet was resuspended and 200μL of protease inhibitor was added, and the suspension was then sonicated to lyse the cells. The protocols for the cloning, overexpression, and resuspention of PksG were provided by Kevin Williams. After sonication and centrifugation the supernatant was then applied to an Ni-NTA column to purify the PksG protein. Different concentrations of imidazole were used to elute nonHis-tagged proteins from the column, and a final concentration of imidazole was used to elute PksG from the column. The eluate was dialyzed in buffer overnight and the following morning the diasylate was concentrated until the volume was between 500μL-750μL. Bradford reagent was used to determine the concentration of the concentrated PksG. To the concentrated PksG enough glycerol was added for a final 10% concentration, and the protein was stored at -80 $^{o}$C.

II.D.4.a Testing AcpK and Sfp Activity

AcpK is synthesized in the apo, inactive, form. The protein Sfp is a 4'-phosphopantetheine transferase (PPTase). In order to be functional, the holo form requires a phosphopantetheine appendage on a conserved serine residue. The function of Sfp is to transfer the 4'-phosphopantetheine moiety of coenzyme A (CoA) to the serine residue of ACPs and PCPs making the holo, active, forms of these proteins. The literature also indicates that Sfp can also transfer phosphopantetheine thioesters from a wide range of acyl CoAs as well.

In order to study the *pks* proteins, we were interested in exploiting this broad CoA and acyl CoA substance tolerance. The specific CoAs tested were; CoA, acetyl CoA, acetoacetyl CoA, and malonyl CoA. These expected AcpK modifications were analyzed by Electrospray Ionization Mass Spectral (ESI-MS) analysis and Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI/TOF-MS). These methods were also used to confirm the molecular weight of apo-AcpK from the overexpression previously performed. The expected masses of all the CoA's can be seen in Table II-1.

Table II- 1Theoretical mass of whole AcpK proteins.

| Product | Expected Mass |
|---|---|
| Apo-AcpK | 10561 Da |
| Holo-AcpK | 10901 Da |
| Acetyl-holo-AcpK | 10944 Da |
| Malonyl-holo-AcpK | 10986 Da |
| Acetoacetyl-holo-AcpK | 10985 Da |

Before the phosphopantetheinyl reactions were carried out the mass of the apo-AcpK protein was analyzed by ESI-MS. The spectrum obtained from this experiment can be seen in Figure II-10. The molecular weight obtained, 10577 Da, is consistent with the mass of apo-AcpK+16 Da. Multiple experiments have also shown the correct mass of apo-AcpK, and also other variations observed during experimental assays.

The phosphopantetheinyl reactions were prepared that contained Sfp, AcpK, acyl CoAs, and enough buffer (50 mM Tris·HCl, 12.5 mM $MgCl_2$, pH 8.0) to make a 500 µL reaction volume. The reactions were incubated at 37°C and were analyzed by ESI-MS and MALDI/TOF-MS. The mass data obtained showed optimal reaction times for all of the different phosphopantetheinylation reactions to occur (Table II-2). The optimal reaction time found for acetoacetyl-CoA was used in the following PksG activity assays.

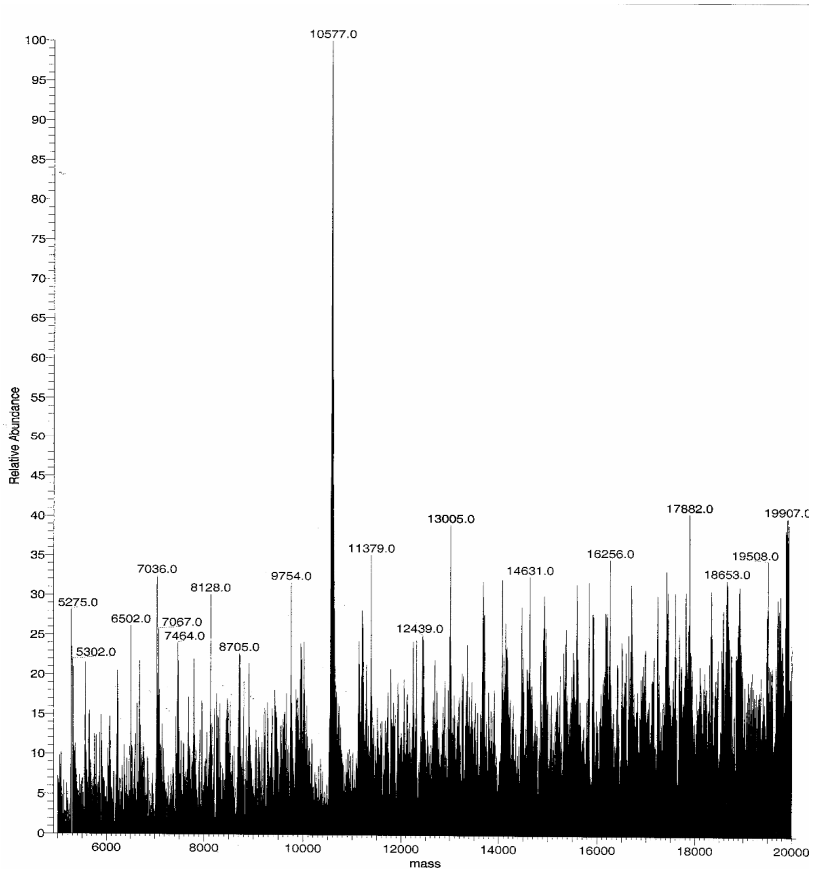

Figure II-10 Mass spectrum showing mass of apo-AcpK+16 Daltons.

Table II- 2 Optimal incubation times for AcpK phosphopantetheinylation reactions.

| Product | Optimal incubation time (hours) |
|---|---|
| Holo-AcpK | 2 |
| Acetyl-holo-AcpK | 1 |
| Malonyl-holo-AcpK | 1 |
| Acetoacetyl-holo-AcpK | 0.5 |

II.D.4.b Testing PksG Activity

Because the function of PksG was unknown, the correct combination of thioester pairs, CoA, AcpK, or PksL, which would yield the HMG product was unclear. For this reason, the two simplest substrates were reacted first, acetoacetyl CoA and acetyl CoA. According to results obtained we planned on progressively studying more difficult systems. The products and substrates of acetoacetyl CoA and acetyl CoA PksG reactions were analyzed by High Pressure Liquid Chromatography (HPLC) analysis. Results from these experiments showed PksG did not synthesize the expected HMG-CoA from these substrates. It was then decided to check acetoacetyl-AcpK and acetyl CoA as substrates for PksG to determine if the enzyme could synthesize HMG-AcpK. These reactions were carried out by spectrophotometric analysis.

For the spectrophotometric analysis the absorbance of the reaction was measured at 300 nm, because the acetoacetyl group on acetoacetyl-AcpK has a characteristic absorbance at that wavelength in the presence of $Mg^{+2}$. A loss of absorbance was expected due to acetoacetyl-AcpK being depleted during a

synthesis of the HMG-AcpK product. Reactions were carried out discontinuously every ten minutes for 1.5 hours. The reaction showed no enzymatic activity by the spectrophotometric method when acetoacetyl-AcpK and acetyl-CoA were used as substrates. In our next approach we were going to also monitor reactions using MADLI/TOF-MS; however, during these PksG studies another group determined the functionality of PksG. Calderone, *et al.* found PksG synthesizes HMG-PksL from the substrates acetoacetyl-PksL and acetyl-AcpK (Calderone, *et al.*, 2006). PksL is a protein found in the genome of *B. subtilis* and contains two tandem ACPs. Reactions involving variations of the published reaction were to be attempted; however, Calderone, *et al.* were able to determine PksG function before these reactions could be carried out. A figure showing the reactions attempted and the one found to work by the other researchers can be seen in Figure II-11.

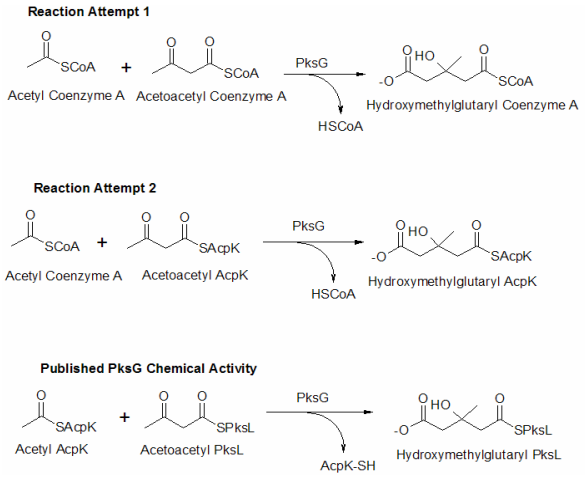


Figure II- 11 Tested reactions involving PksG and published function of PksG (Calderone, *et al.,* 2006).

## II.E Conclusions

### II.E.1 AcpK and Sfp Conclusions

Purified His-Tagged AcpK from *B. subtilis* strain 168 was successfully produced with the use of pQE-60 overexpression vector and M15[pREP4] *E. coli* cells. Purified His-Tagged Sfp, which was amplified from the pUC8/*sfp* overexpression vector, could also be produced using the pET-28a plasmid. For both of these genes, ligation was efficient with T4 DNA ligase, and regulation of overexpression was performed using IPTG induction. Ni-NTA chromatography was used to purify the AcpK and Sfp proteins due to the C-terminal His-Tag coded for by the pQE-60 plasmid for AcpK and the pET-28a plasmid for Sfp. We were able to obtain large amounts of pure AcpK and Sfp from these systems.

ESI-MS studies confirmed that Sfp has a wide range of ACP-substrate specificity, and conclusively determined the synthesis of apo-AcpK made by the overexpression strain pQE-60/*acpK*. We were able to use ESI-MS to determine the optimal reaction times for CoA and the various acyl CoAs.

### II.E.2 PksG Conclusions

Purified His-Tagged PksG from *B. subtilis* strain 168 can be produced with the use of pBAD overexpression vector. Regulation of overexpression of pBAD/*pksG* was performed using arabinose. Overexpressed PksG was purified using a Ni-NTA column since PksG was synthesized with a C-terminal His-Tag encoded by the pBAD plasmid. To elute pure PksG from the column, 200 mM

imidazole was used. Reactions performed that involved the PksG enzyme showed no activity. The activity that was reported by Calderone, *et al.* showed PksG reacting with acetyl-AcpK with acetoacetyl-PksL to form HMG-PksL. PksG apparently requires very specific protein-based substrates. The work performed by Calderone, *et al.* proved the activity of PksG being that of a HMG-PKS synthase (Calderone, *et al.*, 2006). Though the chemical activity and protein scaffold specificity of PksG has been determined, the acyl substrate specificity has not yet been determined. While chemical activity has already been shown, there is still question about its timing in polyketide synthesis, so we still require PksG for experiments that explore our hypotheses.

**II.F Experimental**

II.F.1.a Cloning and Overexpression of AcpK

The PCR to amplify AcpK used *B. subtilis* strain 168 genomic DNA as the template and 5' TAT CCA TGG ATA AAC AGA GAA TCT TTG 3' (inserts a *NcoI* site at the start) and 5' TAT AGA TCT GGC AGA TTG CAC TTT GTC 3' (inserts a *BglII* site downstream) as the primer pair. (Mootz, *et al.*, 2001) Vent polymerase was used to conduct the PCR with the following conditions: 94 $^{o}$C for 2 minutes; 30 cycles of 94 $^{o}$C for 30 seconds, 60 $^{o}$C for 30 seconds, and 65 $^{o}$C for 1.5 minutes. The resulting PCR product and pQE-60 plasmid were digested separately using the restriction enzymes *NcoI* and *BglII* for the PCR product, and *Bam HI* and *NcoI* for the pQE-60 plasmid. The digestion was carried out for 3.5

hours after which a 1.0% low melting agarose gel electrophoresis was performed to separate the digest fragments by size and visualized with UV light. Plasmid and PCR product fragments were removed from the gel and melted in a 75 $^{o}$C. These gel/DNA melts were used directly in ligation reactions using T4 DNA ligase and T4 DNA ligase buffer. Multiple ligation reactions were performed using different insert to plasmid ratios. These reactions were then transformed into competent *E. coli* DH5α cells and plated on LB plates containing 25μg/mL of ampicillin.

The following day 10 transformants were selected and were used to inoculate 10x5 mL cultures with ampicillin (25μg/mL) grown overnight in a 37$^{o}$C shaker. Nine cultures were purified (one did not grow) by the Qiagen Miniprep kit. Each plasmid was screened for insert acceptance by a second PCR reaction using the respective purified plasmids as the template, and Vent polymerase with the same program seen above. A 1.0% agarose gel electrophoresis was used to screen for the inserts in the vector. One successful clone was chosen to be transformed into competent *E. coli* M15(pREP4) cells.

One colony was then used to make a starter culture, 5 mL of LB containing 30μg/mL of kanamycin and 25μg/mL of ampicillin. One liter of LB containing 30μg/mL of kanamycin and 25μg/mL of ampicillin was inoculated with this starter culture and grown in a 37 $^{o}$C shaker. Growth was continued until an $OD_{595}$ of 0.600-0.700 was reached. IPTG was added to a final concentration of 0.25 mM to induce the culture and growth was continued for three hours. After

the three hours the cells were harvested by centrifugation (7500*g* for 30 minutes) and the resulting pellet was stored at -80$^{o}$C.

II.F.1.b Purification of AcpK

Cells from 1 L of LB culture were resuspended in 20 mL of binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris·HCl, pH 7.9) and were lysed by sonication on ice for three minutes. The lysate was cleared by centrifugation (11000*g* for 30 minutes) and then applied to a Ni-NTA column containing 2mL of column bed, and the chromatography was conducted at 4$^{o}$C. After the loading of the crude extract the column was washed with 20mL of binding buffer. Wash buffer (12mL) (60 mM imidazole, 500 mM NaCl, 20 mM Tris·HCl, pH 7.9) was used to remove nonHis-tagged proteins from the column. The eluent buffer (12mL) which contained 200 mM imidazole, 500 mM NaCl, and 20 mM Tris·HCl, pH 7.9, was used to elute AcpK from the column. Snakeskin® pleated dialysis tubing, having a 7000 Molecular Weight Cut Off (MWCO), was used for overnight dialysis of the eluted protein in 4L of buffer (50 mM Tris·HCl, pH 7.5) kept at 4$^{o}$C.

The dialyzed protein was concentrated by centrifuging using a Vivaspin 5000 MWCO concentrator until the volume was between 500µL-750µL. The concentration of AcpK was measured spectrophotometrically using the Bradford method. Samples of Bovine Serum Albumin (BSA) were used for a standard curve having a range of concentrations between 0mg/mL to 1mg/mL and the $OD_{595}$ was recorded at the different concentrations. Concentrated AcpK was

then added to the Bradford reagent and the absorbance was measured to determine the protein concentration.

II.F.2.a Overexpression, and Purification of Wild-Type Sfp

Sfp had previously been cloned into the pUC8 plasmid by Nakano, *et al.* (Nakano, *et al.*, 1992). This plasmid was donated by Prof. Zuber (Oregon Health and Science University) and the overexpression and purification techniques outlined by his group were used. First pUC8/*sfp* was transformed into competent *E. coli* MV1190 cells and plated on LB plates that contained 25μg/mL of ampicillin. The next day a colony was used to inoculate a starter culture (5mL LB with 25μg/mL of ampicillin). The following day, 1 L of 2xYT media was inoculated with the starter culture and the culture was shaken overnight at 37$^{o}$C. The cells were then harvested by centrifugation (7500*g* for 30 minutes). The pellet was then resuspended in buffer (20 mM HEPES, 100 mM KCl, 1 mM EDTA, and 10% glycerol, pH 8.0) and sonicated on ice for three minutes. To clear the lysate a second centrifuge step was used (11000*g* for 30 minutes).

The supernatant was then fractioned by successive ammonium sulfate precipitations. The first three ammonium sulfate concentrations, 35%, 50%, and 65%, were used to precipitate nonspecific proteins from the crude extract. For each step, the appropriate amount of ammonium sulfate was added to raise the percent concentration to the desired concentration all while incubation on ice for two hours. After each precipitation the solution was then centrifuged for 10

minutes at 10000*g*. In the first three steps the supernatant was retained, while the pellet formed after centrifugation was discarded. After addition of ammonium sulfate to 80% and incubation for two hours as before and being centrifuged, the pellet was retained and resuspended in buffer (20 mM Tris·HCl, pH 8.0). Once the pellet was resuspended, it was dialyzed and concentrated by the same method used in the AcpK procedure seen previously, with the exception of the dialysis buffer being 20 mM Tris·HCl, pH 8.0.

II.F.2.b Cloning, Overexpression, and Purification of Sfp

The PCR to amplify Sfp from the pUC8/*sfp* plasmid as the template DNA was carried out using 5' TCT ACC ATG GAG ATT TAC GCA ATT TAT ATG GAC CG 3' (inserts a *NcoI* restriction site at the start) and 5' GAG CCT CGA GTA AAA GCT CTT CGT ACG AGA CC 3' (inserts a *XhoI* restriction site down stream) as the primer pair. *Phusion* polymerase was used along with the following conditions: 30 seconds at 98$^{o}$C; 30 cycles of 10 seconds at 98$^{o}$C, 30 seconds at 63$^{o}$C, and 15 seconds at 72$^{o}$C; and finally a hold of 5 minutes at 72$^{o}$C. The product from this PCR reaction, and the pET-28a plasmid were digested with *NcoI* and *XhoI* for 3.5 hours in separate reactions. A 0.7% low melting agarose gel was used to separate the digested fragments and to visualize with UV light. The desired fragments were excised and placed in a hot water bath to melt the gel. The solution containing the digested PCR fragment and digested plasmid

were used in ligation reactions with T4 DNA ligase and T4 DNA ligase buffer. The ligation reactions were performed using varied insert:plasmid ratios.

The ligation reactions were transformed into competent *E. coli* DH5α cells and plated on LB-agar plates containing 30μg/mL kanamycin. Two cultures were made from the two colonies that grew on one plate, 5mL LB containing 30μg/mL of kanamycin and shaken at 37$^{o}$C. Once the starter cultures had grown overnight, the plasmids were purified from the cultures using the Qiagen Miniprep Kit following the procedure outlined in the manufacturer's instruction manual. To screen for insert acceptance, a second PCR reaction was performed. The parameters for the reaction were the same used in the first PCR reaction, with the only exception being the template DNA used was the purified plasmids. A 1.0% agarose gel electrophoresis was performed to visualize the PCR reaction mixtures. A successful clone was transformed into competent *E. coli* BL21(DE3) cells.

One colony from this retransformation was grown overnight in a starter culture of LB and a concentration of 30μg/mL of kanamycin, and the following day 2 mL was added to 1L of LB containing 30μg/mL of kanamycin. This was shaken at 37$^{o}$C until the $OD_{595}$ reached 0.500-0.600. Once this absorbance was reached 1.0 mM of IPTG was added to the culture and was shaken overnight. The following morning the cells were harvested by centrifugation (7500*g* for 30 minutes). The pellet was then stored at -80$^{o}$C until it was used for purification. To purify Sfp the same procedure outlined in II.F.1.b for the purification of AcpK

was used, with the only exception being the dialysis buffer was 50mM Tris·HCl, pH 7.5.

II.F.2.c Assays involving AcpK and Sfp

Reactions involving AcpK, Sfp, and acyl CoAs were conducted and the determination of phosphopantetheinyl reaction products was characterized by ESI-MS and MALDI/TOF-MS. The reactions were conducted using various amounts of AcpK, Sfp, and acyl CoAs in buffer (50 mM Tris·HCl, 12.5 mM $MgCl_2$, pH 8.0) in a total reaction volume of 500µL. After different incubation times, see Table II-2, the samples were prepared for analysis by ESI-MS. First the reaction was filtered through a Fisherbrand nylon 0.45 µm syringe filter to remove any particulates from the reaction solution. The filtered reaction was then mixed with methanol to a 1:1 ratio of methanol:reaction. Finally, enough glacial acetic acid was added to give a 0.1% solution. ESI-MS was carried out by a direct infusion method and analyzed in positive ion mode. The peaks obtained were then deconvoluted to determine the molecular weight of the proteins in the different reactions.

For analysis by MALDI/TOF-MS, the reactions were prepared in the same manner as in the ESI-MS experiments. To prepare the reactions for analysis, 1µL of each reaction was added to 24µL of Matrix A (8mg of α-cyano-4-hydroxycinnamic acid 99% dissolved in 1 mL of 0.3% aq. trifluoroacetic acid in 50% acetonitrile). Matrix A was prepared by vortexing for 20 seconds, followed

by centrifuging, and using the supernatant. Each sample was spotted (0.5μL) on a MALDI sample loading plate and analyzed on an Applied Biosystem 4700 Proteomics Analyzer. Positive-ion mass spectra were recorded in the linear mode.

II.F.3.a Overexpression and Purification of PksG

PksG had been previously cloned into the pBAD vector by Kevin Williams. The pBAD/*pksG* plasmid was transformed into competent *E. coli* TOP10 cells by the method outlined in the manufacturer's manual and plated on LB-agar plates (25μg/mL ampicillin). One colony from this plate was chosen and grown overnight in a starter culture, 5mL LB containing 25μg/mL ampicillin. All 5mL of the starter culture was added to 1L of LB containing 25μg/mL ampicillin, and was shaken at 37$^{o}$C until the $OD_{595}$ reached 0.500-0.600 (mid-log phase). At this point arabinose (0.01 M final concentration) was added to the culture and shaken for three hours at 37$^{o}$C. The culture was harvested by centrifugation (7500g for 30 minutes), and was stored at -80$^{o}$C until purification was performed. To perform the purification, the same method outlined in II.F.1.b for the purification of AcpK was used, with the only exceptions being the addition of 200μL of protease inhibitor (HALT™ Protease Inhibitor Cocktail) added prior to sonication and the dialysis buffer was 25mM Tris·HCl, pH 7.5.

II.F.3.b Activity Assays with PksG

HPLC analysis was performed on a Varian Prostar HPLC and a Varian Microsorb $C_8$ column (250mm in length and a 4.6mm inner diameter). The two different solvents used were methanol and 100mM potassium phosphate, pH 7.0, containing 2 mM tetrabutylammonium phosphate (TBAP). The gradient used for the HPLC analysis was 0% methanol increasing to 90% methanol over 20 minutes. This was held at 90% methanol 10 minutes and the methanol was decreased to 0% over 5 minutes. The column was equilibrated in the phosphate buffer for 5 minutes. A flow rate of 1 mL/minute was used, and the wavelength monitored was 219nm. Acetoacetyl CoA, acetyl CoA, and HMG CoA were analyzed by this method to determine a standard elution time for each molecule. To test for PksG activity using the HPLC, one reaction containing all reactants and enzyme was prepared, and four control reactions were also mixed. The full reaction contained 4mM acetoacetyl CoA, 4mM acetyl CoA, 16.5µg of PksG, and enough buffer (50mM Tris·HCl, 12.5mM $MgCl_2$, pH 8.0) to yeild a 50µL reaction volume. The controls were prepared the same way, but different components were omitted.

To spectrophotometrically monitor the PksG reactions using acetoacetyl-AcpK or acetoacetyl CoA and acetyl CoA, a Genesys10 spectrophotometer was used. To synthesize acetoacetyl-AcpK, apo-AcpK was mixed with acetoacetyl CoA and Sfp in reaction buffer, and was incubated for 30 minutes at 37$^{o}$C, according to the reaction protocol determined in section

II.D.4.a. Three reactions were then prepared. The first reaction contained 1mM acetoacetyl-AcpK, 1mM acetyl-CoA, and 3.3µg of PksG. As a control for the first reaction 1mM acetoacetyl-AcpK was mixed with 1mM acetyl CoA, but no PksG was added. The last reaction involved mixing 1mM acetoacetyl CoA, 1mM acetyl CoA, 3.3µg PksG in buffer (100mM HEPES, 0.1mM EDTA, 20mM $MgCl_2$, pH 8.0). These reactions were incubated at 37$^{o}$C and every ten minutes 25µL of the reaction was removed and dissolved in 475µL of buffer (100mM HEPES, 0.1mM EDTA, 20mM $MgCl_2$, pH 8.0) to make the final acetoacetyl CoA and acetyl CoA concentration equal to 0.05mM. Each of these samples was placed in a quartz cuvette with a 10.00mm pathlength, and the absorbance was recorded at 300nm.

# CHAPTER III

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

## III. A Introduction to MmgA

### III. A.1 MmgA and Proposed Activity

Certain bacteria, including *Bacillus subtilis*, when experiencing nutrient starvation, have the ability to sporulate. Sporulation is considered a "last ditch" response when other survival strategies are not effective. Such strategies can include activating flagellar motility to seek out nutrient sources, secretion of antibiotics to decrease competition by other organisms, and the secretion of hydrolytic enzymes to seek extracellular proteins and polysaccharides, among others. The signal to begin sporulation is the phosphorylation of a protein called Spo0A. Once Spo0A is phosphorylated, the expression of genes used in the sporulation process is initiated (Stephens, *et al.,* 1988).

An operon used in sporulation from *Bacillus subtilis* strain 168 was discovered in 1996. This operon was called the mother cell metabolic genes (*mmg*) and is under the control of the $\sigma^E$ factor, which is used to express genes at the intermediate stages of sporulation. The operon includes *mmgA-E* and the gene *yqiQ*, with *mmgA-C* having sequence homology to genes used in fatty acid

metabolism (Bryan, *et al.*, 1996; Subtilist Web Server, accessed March 12, 2007). The genes *mmgDE* and *yqiQ* show homology the genes used in propionate metabolism in E. coli with the products of these reactions yielding citric acid cycle substrates. The gene *mmgA,* specifically, is an acetyl coenzyme A (CoA) acetyltranferase homolog (Bryan, *et al.*, 1996). Acetyl-CoA acetyltransferases catalyze the synthesis of acetoacetyl CoA by the condensation of two molecules of acetyl CoA. These types of enzymes have also been shown to catalyze the reverse reaction, called a thiolysis reaction, which uses one CoA to cleave an acetoacetyl CoA unit into two acetyl CoA molecules. These reactions are shown in Figure III-1.

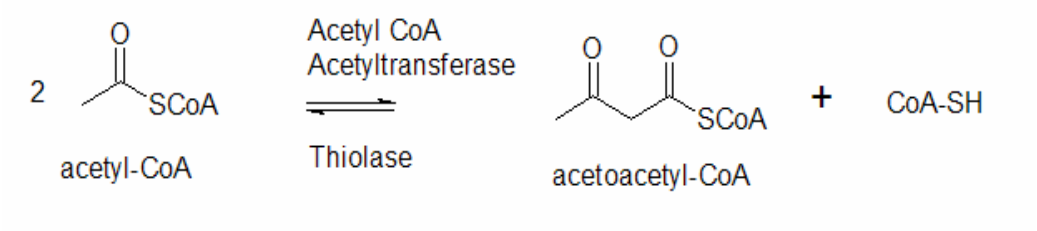


Figure III- 1 Reaction of acetyl coenzyme A acetyltransferases.

Because MmgA has not been biochemically proven to be an acetyl-CoA acetyltransferase, it is an interesting enzyme to study. Along with the determination of the proposed functionality of the enzyme an investigation of its substrate specificity will also be investigated. Other enzymes of the same class as MmgA have been able to synthesize β-keto acyl CoAs longer then acetoacetyl CoA (Antonenkov, *et al.*, 1997). Also, we intend to investigate whether MmgA

can thiolyze long chain β-keto bonds. (see Figure III-2). MmgA may cleave longer β-keto acyl bonds due to the presence of branched or odd chain fatty acid bonds produced by the organism. However, in order to study these reactions, the synthesis of long chain β-ketos will have to be performed, since these compounds are not available commercially.

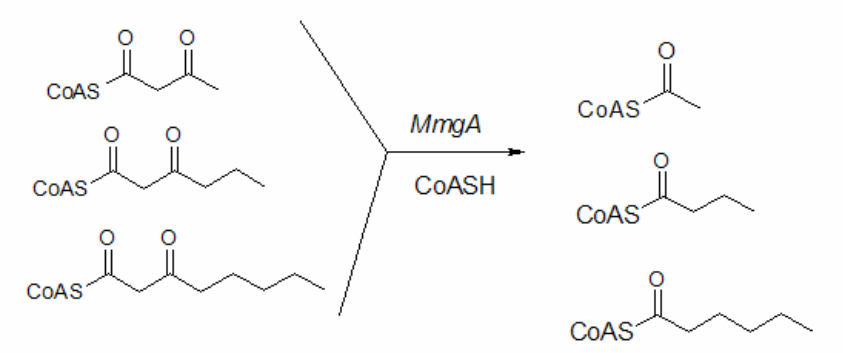


Figure III- 2 Thiolysis of different length β-keto acyl coenzyme A bonds using MmgA.

III.A.2 MmgA and the Synthesis of Difficidin

Once the functionality of MmgA has been determined, the protein can be used as a synthetic tool to aid our investigation of the biosynthesis of difficidin. Two different hypothesis have been proposed for the synthesis of difficidin. The first hypothesis involves PksG synthesizing an HMG precursor molecule that can then be reacted with other enzymes and finally added to the nascent difficidin chain by polyketide synthases (see Figure III-3). The second hypothesis involves a full-length difficidin intermediate that is synthesized by standard polyketide

chemistry. This intermediate would contain a ketone at the C3 position which is modified by other enzymes (PksFGHI) and a unit of acetyl-AcpK to form difficidin. These two hypotheses can be seen in the following figures (Figure III-3, Figure III-4).

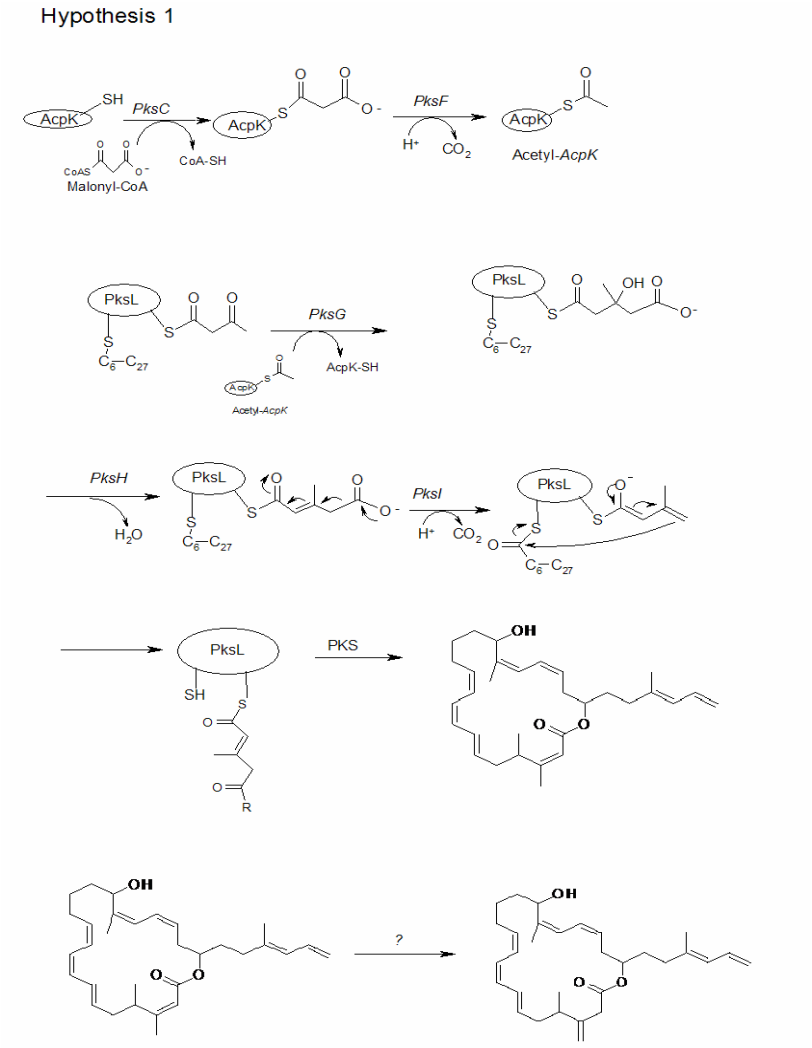


Figure III- 3 First hypothesis for difficidin production.

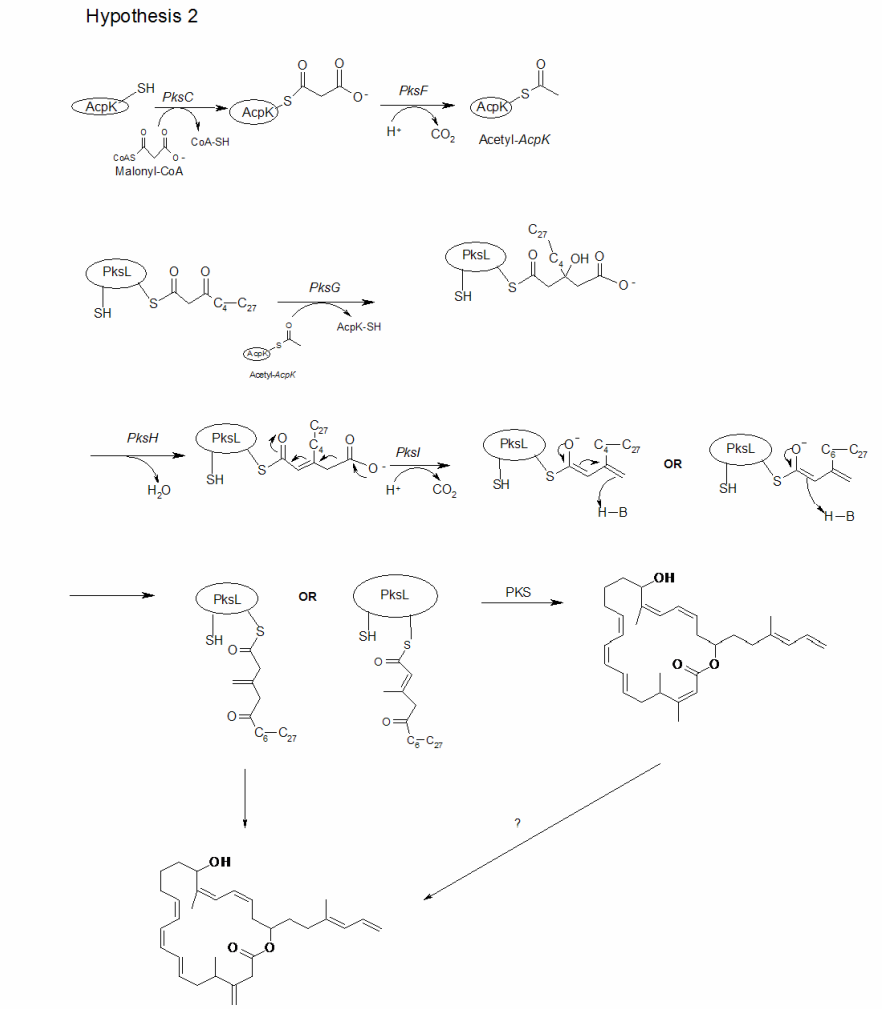


Figure III- 4 Second hypothesis for difficidin production.

Since PKSs have extremely large masses, from 100 to 10,000 kiloDaltons, it is very difficult to determine which hypothesis is valid (Khosla, *et al.*, 1999). One approach can utilize a PksG$^-$ mutant of *B. subtilis* strain 39374, the producer of difficidin (Zweerink, *et al.*, 1987). PksG, a 3-hydroxy-3-methylglutaryl (HMG) synthase, acts on either the acetoacetyl group attached to one thiolation domain on PksL, or on a long-chain difficidin precursor molecule. PksG is an important

step in the synthesis of difficidin and by removing this enzyme from the difficidin synthesis process it is possible to reveal the biochemical role of this enzyme. A PksG$^-$ mutant would not be able to synthesize the HMG intermediate, so a labeled replacement substrate, $^{13}$C labeled 3-hydroxy-3-methylglutaryl- N-acetylcysteamine (HMG-SNAC), will be introduced to a culture of the PksG$^-$ *B. subtilis* 39374 mutant. If the first hypothesis is correct, the culture will then directly use the labeled HMG-SNAC, and produce a labeled 3-keto difficidin precursor product, which can be detected using mass spectral analysis. If the second hypothesis is correct, the culture will not be able to use this labeled substrate and we can instead expect a new difficidin derivative having a ketone group at the C3 position.

Labeled HMG-SNAC is not commercially available, so a synthesis in our lab is necessary. The synthesis of HMG-SNAC would occur in the following steps (Figure III-5). First, labeled pyruvate will be reacted with CoA and $NAD^+$ to produce labeled acetyl CoA, in a reaction catalyzed by a commercially available pyruvate dehydrogenase. MmgA would then be used with two labeled acetyl CoA units to produce labeled acetoacetyl CoA. Labeled HMG-CoA will be produced by a HMG-CoA synthase from *Staphylococcus aureus* using labeled acetoacetyl CoA and labeled acetyl CoA (Campobasso, *et al.*, 2004). Finally, HMG-CoA will be reacted with SNAC to make labeled HMG-SNAC.

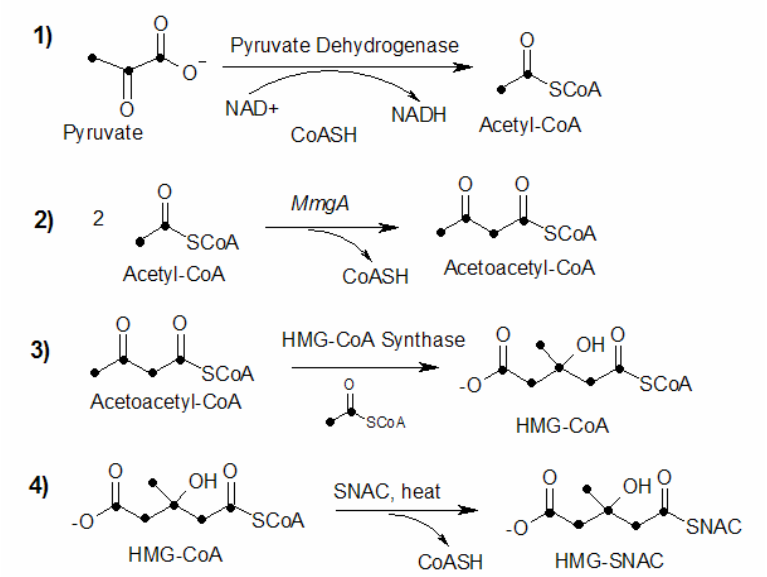


Figure III- 5 Synthesis of labeled 3-hydroxy-3-methylglutaryl-*N*-acetylcysteamine.

## III.B Production and Characterization of pET-28a/*mmgA*

### III.B.1 Goals

In order to test our hypothesis for the activity of MmgA, pure MmgA would be needed from *B. subtilis* strain 168. The first step was to clone and overexpress *mmgA* using the pET-28a plasmid. Overexpressions with pET plasmids involve the bacteriophage T7 promoter and T7 RNA polymerase. For pET-28a the LacI repressor protein controls the T7 promoter, making the pET system able to be externally controlled (pET System Manual, 5th ed.). Once induction occurs, the cell begins to use most of its resources to transcribe the gene under control of the T7 promoter, known as “overexpressing” the gene. The amount of production can in principle be controlled by changing the amount of inducer added. MmgA in this case would be the gene of interest, and would

be cloned into the pET-28a plasmid by restriction digesting the pET-28a vector and PCR-amplified *mmgA*. This will be followed by performing a subsequent ligation using T4 DNA ligase. Overexpression of the protein involves growing a culture of cells to mid-log phase and then inducing the growth by adding an inducer. A LacI repression system, like the pET vector, can use isopropyl-beta-D-thiogalactopyranoside (IPTG) as an inducer. IPTG is an artificial inducer of the *lac* operon and addition of this inducer leads to the production of large amounts of the protein of interest. Once *mmgA* is successfully cloned into pET-28a, the next goal would be to purify the overexpressed MmgA.

Nickel-affinity purification would be used because the protein would be cloned with a C-terminal histidine-tag encoded by the plasmid. $His_6$-tagged proteins can be purified by affinity chromatography using a nickel-nitrilotriacetic acid (Ni-NTA) column. NTA is a $Ni^{+2}$-chelating adsorbant which leaves two empty sites on the nickel for binding nitrogens of the $His_6$-tag. Washes with low imidazole concentrations remove nonHis-tagged proteins from the column, while the $His_6$-tag of the protein associates strongly with the $Ni^{2+}$ on the column. A buffer with high imidazole concentration is used to elute the purified protein from the column.

The last goal would be to characterize the enzymatic activity of MmgA in both the condensation and thiolysis directions. We expect that MmgA can catalyze the condensation of two acetyl-CoA molecules to form acetoacetyl-CoA, or thiolyze the β-keto bond of acetoacetyl-CoA in the presence of CoA to form

acetyl-CoA. UV-Vis spectrophotometric analysis will be used to test the kinetics of the reactions, and High Performance Liquid Chromatography (HPLC) analysis will be used to conclusively identify the substrates and products. UV-Vis spectrophotmetric analysis has been used to test for acetyl-CoA acetyltranferase activity in the past. The β-keto group of acetoacetyl coenzyme A, in the presence of $Mg^{2+}$, absorbs around 300 nm.

**III.C Results**

III. C.1 Cloning of *mmgA*

Primers for the polymerase chain reaction (PCR) were designed and used to amplify *mmgA* from the *B. subtilis* 168 genome. The gene fragment produced had sequence coding on each end that inserts an *NcoI* site at one end and an *XhoI* site on the opposite end. *MmgA* has a gene length of 1179 base pairs (bp). A band at ~1200 bp appeared on an agarose gel, indicating the PCR was successful.

Insertion into the pET-28a vector was attempted using *NcoI* and *XhoI* on both the plasmid and *mmgA* PCR product. T4 DNA ligase was used to ligate the restriction digested products together. Different ratios of insert:plasmid were used in six separate reactions. Once the ligation/cloning step was complete the ligation mixture was transformed into competent *E. coli* DH5α cells. The transformed cells were then plated on six agar plates containing kanamycin (30μg/mL) and were incubated overnight at 37°C. The following morning each

plate had several transformants. Colonies from one plate were selected and six starter cultures were grown overnight. The plasmids were then purified from the cultures the next day and another restriction digest was performed to test for insert acceptance into the plasmid. An agarose gel showed that all of the plasmids contained an insert and the first plasmid was chosen to be transformed into competent *E. coli* BL21(DE3) cells.

### III.C.2 Overexpression and Purification of pET-28a/*mmgA*

The gene of interest, *mmgA*, had been successfully cloned into pET-28a which codes for a C-terminal His-tag. Our next step was to overexpress the protein from the clone. The overexpression strain pET-28a/*mmgA* was grown in Luria Bertani (LB) medium containing 30µg of kanamycin per mL at 37 $^{o}$C in a shaker until the optical density reached between 0.400 and 0.450 at 595nm. The culture was then moved to an 18$^{o}$C shaker until it reached an absorbance between 0.500 and 0.600 at 595nm. Once this absorbance was reached the culture was induced with 1.0mM IPTG. The culture was then shaken overnight, and the cells were harvested the next day by centrifugation (9950*g* for 30 minutes). The cells were then resuspended in 20mL of Binding buffer (5mM imidazole, 500mM NaCl, 20mM Tris, pH 7.9) and lysed by sonication. The lysate was cleared by centrifugation (11000*g* for 30 minutes).

Because inclusion bodies were found to be a problem, a low incubation temperature was determined to be necessary since whole cell SDS-PAGE

(Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophorisis) samples of cultures grown at 37$^{o}$C showed a band with high expression with a molecular weight consistent with MmgA, but none of this material was found in the postlysis supernatant.

The centrifuged lysate was then applied to a Ni-NTA (Nickel- nitrilotriacetic acid) column and was kept at 4$^{o}$C. Different concentrations of imidazole, 5mM and 60mM was used to wash nonHis-tagged proteins from the column. A final buffer containing 200mM imidazole was used to elute the protein. An SDS-PAGE was taken of samples taken from various stages of the chromatography. The expected mass of MmgA is 42.1 kDa. The SDS-PAGE gel shown in Lane 6 of Figure III-6 shows a large band below the 45.0 kDa molecular weight marker. The eluted protein from the column was dialyzed overnight in buffer containing 25 mM Tris HCl, pH 7.5. The protein was then concentrated the next morning by centrifugation using a Vivaspin 5000 Molecular Weight Cut Off (MWCO) concentrator. Sterile glycerol was then added to the protein to a final concentration of 10% for -80$^{o}$C storage. The concentration of the protein was then determined spectrophotomectrically using the Bradford method. The average concentration of MmgA was ~ 675µg per 1L of culture.

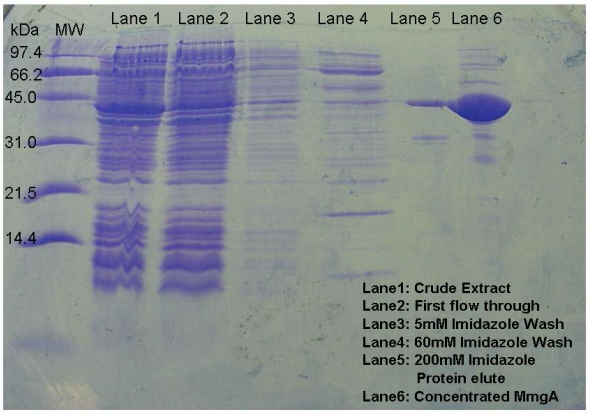


Figure III-6 Sodium dodecyl sulfate polyacrylamide gel electrophorisis of column chromatography of MmgA purification.

III.C.3 Testing MmgA Activity

III.C.3.a Spectrophotometric Analysis of MmgA Reaction Kinetics

Our hypothesis is that MmgA catalyzes the synthesis of an acetoacetyl coenzyme A, or the CoA-dependant thiolysis of acetoacetyl CoA with the production of two acetyl CoAs. To test for this activity, and to also measure the kinetics, a spectrophotometric method was used. The reaction is able to be monitored spectrophotometrically due to $Mg^{+2}$ being present in the buffer used for the reactions. The enolate form of the β-keto group on acetoacetyl CoA absorbs at 300 nm, and this form is stabilized by $Mg^{+2}$ (Pantazaki, *et al.*, 2005).

The simplest case of rate kinetics for enzymes is single substrate rate kinetics where the substrate in the presence of the enzyme is converted to product, $E + S \leftrightarrow E \bullet S \leftrightarrow E + P$. The Michaelis-Menten equation for a single substrate reaction is $v = \frac{V_{\max}[S]}{K_M + [S]}$ where $V_{max}$ is the maximum rate an enzyme

can catalyze a reaction at a specific enzyme concentration, and $K_M$ is the substrate concentration the results in a reaction rate at $1/2V_{max}$. Acetyl CoA acetyltranferases catalyze a multi-substrate reaction, and are known for having a "ping-pong" mechanism, which leads to a more complicated rate equation, $v = \frac{V_{\max}[A][B]}{K_M{}^A[B] + K_M{}^B[A] + [A][B]}$, where A is substrate1 and B is substrate2. Rate data from this type of reaction would be laborious to obtain; however, it is possible to create a pseudo-single substrate reaction if one substrate is held at a high, saturating concentration such as if acetoacetyl CoA were held at saturating conditions then the rate of the reaction would depend on the concentration of CoA. The rate equation in this case would simplify to, $v = \frac{V_{\max,app}[A]}{K_m{}^A{}_{app} + [A]}$. Pseudo-single substrate kinetics can be used to determine an apparent $V_{max}$ for the reaction and an apparent $K_M$ for both substrates, acetoacetyl CoA and CoA.

To accomplish this, the concentration of acetoacetyl CoA was held constant at 60μM, while the concentration of CoA was varied, and the rate of reaction was recorded for each varied concentration. The reactions were carried out for 0.3 minutes at a constant temperature of 30°C. The $K_{M,app}$ for CoA was 52μM, and the $V_{max,app}$ was found to be 2.2μM/sec from a double reciprocal plot of 1/substrate concentration versus 1/ average rate of reaction (see Figure III-7).

The kinetic parameters for acetoacetyl CoA were obtained in a similar manner. The concentration of CoA was held constant at 70μM, and the concentration of acetoacetyl CoA was varied. The amount of enzyme used in

each 1mL total reaction volume was 1.0µg of MmgA. Concentrations above 70µM CoA showed substrate inhibition since the rate of the reaction decreased with concentrations of CoA above 70µM.

Figure III-7 shows the graph of the acetoacetyl CoA rate data obtained. Though there are six points on the graph, only four were used because they are in the linear range of the graph. The kinetic parameters found for acetoacetyl CoA were a $K_{M,app}$ of 770µM and a $V_{max,app}$ of 10µM/sec. The trend that is seen in Figure III-7 shows that as the points approach the reciprocal velocity axis the line began to level out. This shape shows a classic plot shape for substrate inhibition. In another study, done by Middleton, an acetoacetyl CoA thiolase from rat liver showed substrate inhibition by CoA (Middleton, 1974). In this report a double reciprocal plot of 1/velocity versus 1/concentration of CoA was given. The plot obtained started to curve up as the line approached the 1/velocity axis (see Figure III-8). This led Middleton to conclude that CoA does inhibit the thiolase enzyme (Middleton, 1974). A similar shape can be seen in the graph in the acetoacetyl CoA plot in Figure III-7, which suggests acetoacetyl CoA is inhibiting the reaction as well as CoA for the protein MmgA. Once the kinetic constants had been estimated for the reaction the identification of the products of the thiolysis reaction would be carried out using High Performance Liquid Chromatography (HPLC) analysis.

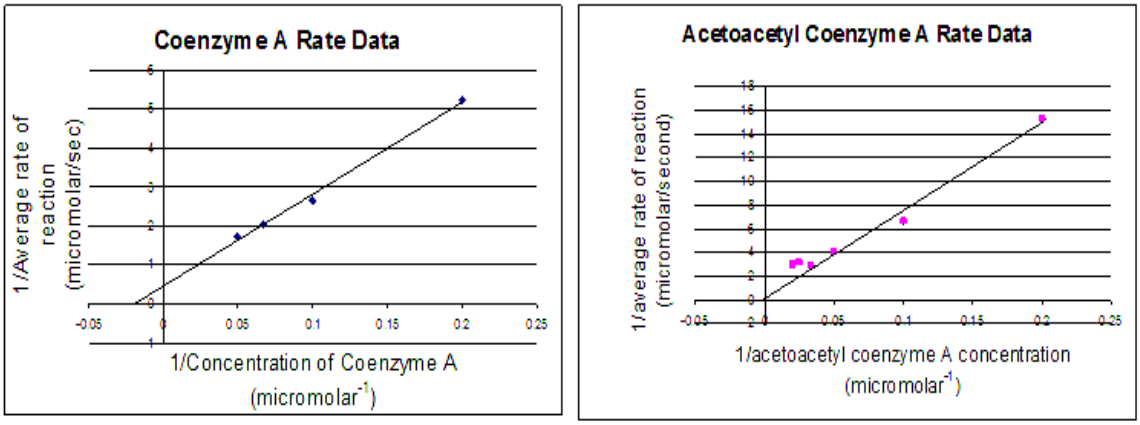


Figure III- 7 Double reciprocal plot of coenzyme A and acetoacetyl coenzyme A thiolysis rate data.

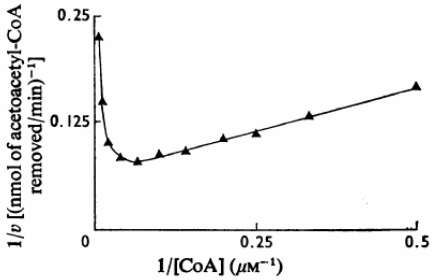


Figure III-8 Inhibition of rat liver acetoacetyl coenzyme A thiolase by coenzyme A (Middleton, 1974).

The condensation reaction was carried out in the spectrophotometer using the same buffer and temperature as the thiolysis reaction. The condensation reaction uses acetyl CoA as the substrate. Reactions containing 60 µM acetyl CoA were carried out over a 30 minute time period, which showed no acetoacetyl CoA being formed. To see if MmgA had a broader substrate specificity, reactions containing butyryl CoA and acetyl CoA were performed. Varying concentrations of acetyl and butyryl CoA were used. Enzyme (1.0µg.mL) was

added, and the reaction was monitored over 30 minutes to see if MmgA could synthesize a longer chain β-keto acyl CoA. The spectrophotometric assays did not show any activity for this reaction, and it was decided HPLC analysis would be used to conclusively determine whether MmgA was able to successfully perform a condensation reaction.

### III.C.3.b High Performance Liquid Chromatography (HPLC) of MmgA Reaction Products

Once the spectrophotometric measurements had been taken, the next step was to use HPLC to confirm the products of the MmgA reactions. CoA, acetoacetyl CoA, acetyl CoA, and butyryl CoA were used to determine the elution times for each substrate. Reactions in both the thiolysis direction and the condensation direction were conducted, with samples being run on the HPLC at different times during the reaction. The standard elution times can be seen in Table III-1.

Table III- 1 Standards of MmgA substrates and elution times.

| Standard | Elution Time (minutes) |
|---|---|
| Coenzyme A | 3.57 |
| Acetyl Coenzyme A | 8.95 |
| Acetoacetyl Coenzyme A | 10.07 |
| Butyryl Coenzyme A | 20.5 |

In the thiolysis reaction, the peak associated with acetoacetyl CoA, after three hours of reaction time, completely disappeared, and a peak with the same

retention time as acetyl CoA appeared gradually over the same three hour period. This process can be seen in the following figure.

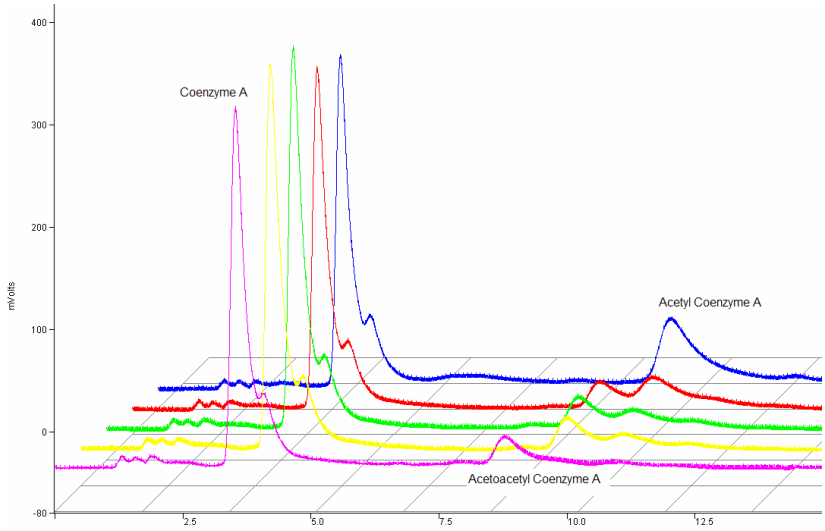


Figure III- 9 Thiolysis reaction showing loss of acetoacetyl coenzyme A peak and appearance of acetyl coenzyme A peak over three hours.

For the condensation reaction, the acetyl CoA peak did not decrease; however, a peak associated with CoA did begin to appear after an hour of reaction time. When left to react overnight the CoA peak increased and the acetyl CoA peak did decrease, but a peak associated with acetoacetyl CoA was never detected. This would not be an unexpected result with an enzyme that shows a ping-pong type mechanism, a mechanism other acetyl CoA acetyltransferases have been reported to undergo (Pantazaki *et al.*, 2005). Acetyl CoA would enter the active site of the enzyme and an active site cysteine would then initiate a nucleophilic attack on the carbonyl carbon, releasing CoA.

This would also form an enzyme-acetyl intermediate. If the enzyme was able to form acetoacetyl CoA then another unit of acetyl CoA would enter the active site and through a Claisen like condensation releasing acetoacetyl CoA. Figure II-10A shows acetyl-CoA in reaction buffer before addition of MmgA and Figure III-10B shows the condensation reaction after being incubated at 37$^{o}$C overnight. Reactions involving butyryl CoA also showed the development of a peak associated with CoA, but a condensation product was never detected.

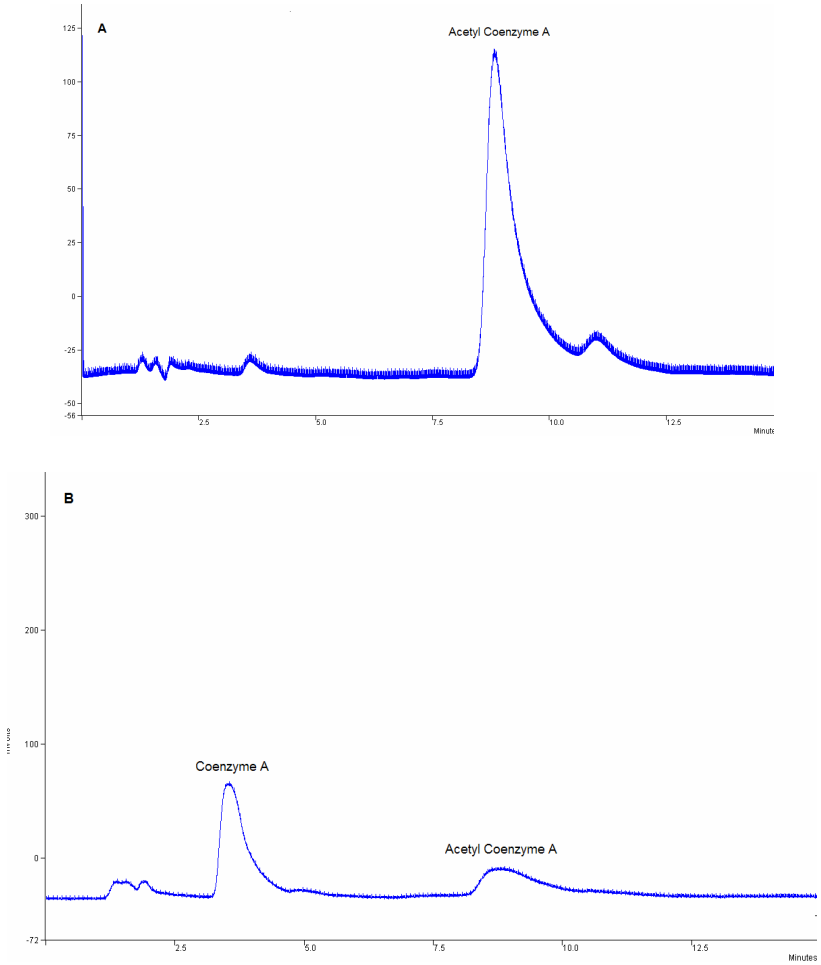


Figure III-10 Condensation reaction (A) before MmgA is added and (B) after MmgA is added and incubated overnight.

### III.D Conclusion

Purified His-Tagged MmgA from *B. subtilis* strain 168 was successfully produced with the use of pET-28a overexpression plasmid vector. IPTG induction was used to regulate the overexpression of MmgA-pET-28a, and a lower incubation temperature was at least partially effective in reducing inclusion bodies. A Ni-NTA column was used to purify MmgA since it was cloned with a C-terminal His-Tag encoded by the pET-28a plasmid.

MmgA activity was studied in two ways; spectrophotometrically, and by HPLC. The spectrophotometric measurements also gave the kinetic constants for the enzyme reactions. The loss of absorbance at 300nm was proportional to the loss of acetoacetyl CoA. This loss of substrate correlated with the proposed function of the enzyme. These rate studies also showed that both acetoacetyl CoA and CoA exhibit substrate inhibition at high concentrations.

For HPLC analysis, an isocratic method was used and standards of CoA, acetyl CoA, and acetoacetyl CoA were analyzed. The retention times of these standards were used to compare to determine the identity of the components of MmgA reactions catalyzed. The HPLC analysis showed a loss of acetoacetyl CoA, and the production of acetyl CoA, with the reaction going to completion after three hours. Both the spectrophotometric analysis and HPLC confirmation demonstrated that MmgA is a thiolase. MmgA was shown not to produce acetoacetyl CoA in the condensation reaction, or a longer chain β-keto from reactions with butyryl CoA, though the production of CoA was noted over time.

Though MmgA does not show condensation activity it does show activity in the thiolation direction. This makes MmgA ineffective for the proposed synthesis of labeled HMG-SNAC (see Figure III-5). Though MmgA will not be used for this synthesis, other enzymes that show this activity from other organisms can easily be obtained and utilized for the synthesis studies. Because the activity of MmgA had not been determined, the results from this thesis are of scientific value.

The researchers who discovered this operon proposed the physiological function of the *mmg* operon was to gain energy from fatty acid metabolism. The importance of MmgA in this energy harvesting processes can be hypothesized from the results presented in this work. The operon containing *mmgA* can be seen in Figure III-11. One gene, *yqiQ*, has not been officially designated an *mmg* gene; however, the location of transcription termination sequence indicates that it is likely part of the *mmg* operon. Based on this signal and the expected functions described below we suggest that *yqiQ* be renamed *mmgF* and *yqiQ* will be referred to as *mmgF* for all further discussions.

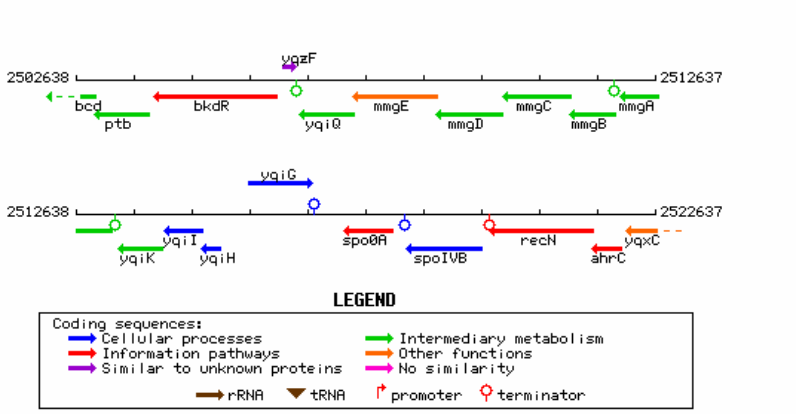


Figure III-11 Portion of *Bacillus subtilis* genome containing *mmg* operon (Subtilist Web Server, accessed March 12, 2007).

The genes *mmgB* and *mmgC* are homologs of 3-hydroxybutyryl-CoA dehydrogenases and acyl-CoA dehydrogenases, respectively. These enzymes are expected to oxidize a fatty acyl-CoA to a β-keto fatty acyl CoA. MmgA can then cleave the β-keto bond yielding acetyl CoA. MmgD shares homology with citrate synthase III which catalyzes the condensation of acetyl-CoA and oxaloacetate in the presence of water to yield citrate and CoA. This citrate is a substrate for the citric acid cycle, therefore this is an apparent path fatty acids can be harvested for energy metabolism. In the case of even chain fatty acids, this pathway may be used; however, *B. subtilis* is known to synthesize branched chain and odd chained fatty acids and would be available to the organism for energy (Oku, *et al.*, 1988; Kaneda, *et al.*, 1967). In this case the first two steps involving MmgABC would remain the same with the exception being the product of the MmgA reaction would be a propionyl-CoA unit instead of acetyl-CoA (see Figure III-12).

The next enzyme, MmgD, would then synthesize a methylcitrate from propionyl-CoA and oxaloacetate. MmgE shares strong sequence homology to *prpD* from *E. coli* as shown by a Basic Local Alignment Search Tool (BLAST) search of the protein sequence of *mmgE* on the Colibri web server (Colibri Web Server, accessed March 12, 2007). The *prpD* gene is located in a propionate catabolism operon and catalyzes the reaction of methylcitrate to methylaconitate which in a subsequent step is rehydrated synthesizing methylisocitrate (Brock, *et al.*, 2002). Finally, *mmgF* shows strong homology to the gene *prpB* of *E. coli* also located in the propionate metabolism operon. The enzyme PrpB is a methylisocitrate lyase which breaks methylisocitrate into pyruvate and succinate (Brock, *et al.*, 2002). Succinate is a well known citric acid cycle intermediate. Whether or not MmgA is able to cleave a longer chain β-keto acyl CoA is still to be determined. The rest of the operon would also need to be characterized in order to determine how *B. subtilis* uses this operon in sporulation. The hypothetical pathways that lead to the citric acid cycle from fatty acid metabolism using this operon can be seen in Figure III-12.

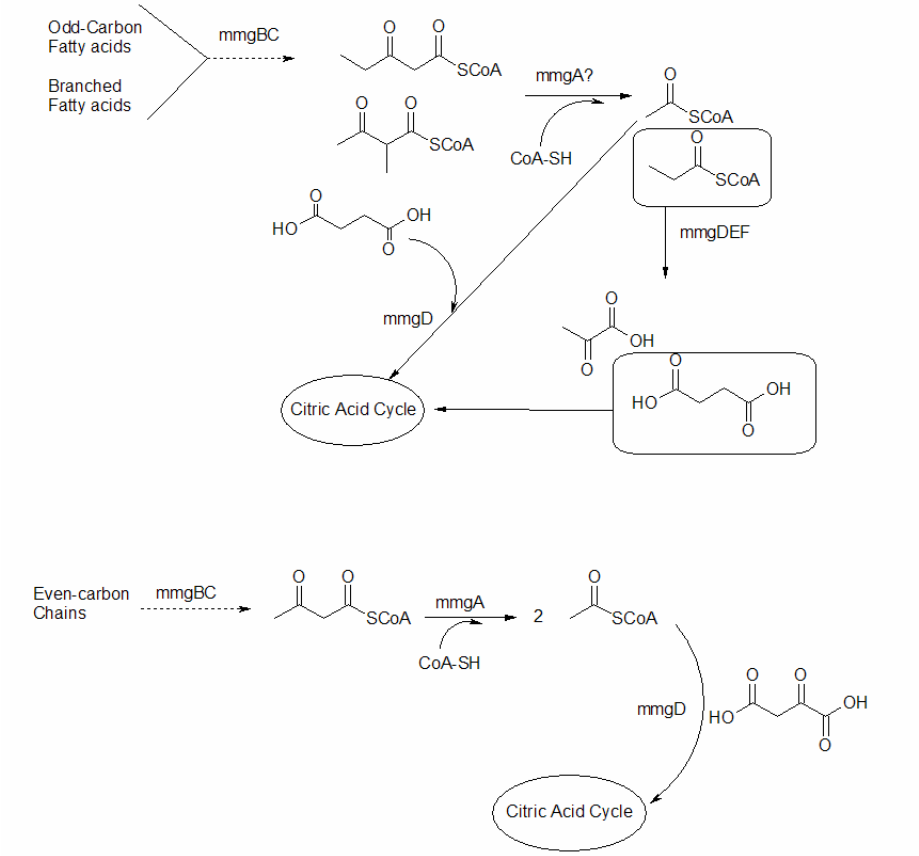


Figure III-12 Hypotheses for fatty acid catalysis pathway in the *mmg* operon.

## III.E Experimental

### III.E.1 Cloning and Overexpression of MmgA

MmgA was PCR amplified using the primers 5' AAA CCA TGG GGA AAA CAG TCA TTG TAA GTG CTG 3' (contains a *NcoI* restriction site) and 5' CAT CTC GAG ATG AAC CTG CAC TAA GAC G 3' (contains an *XhoI* restriction site). *Phusion* polymerase was used along with the following conditions: 30 seconds at 98 $^{o}$C; 30 cycles of 10 seconds at 98 $^{o}$C, 30 seconds at 61 $^{o}$C, and 30 seconds at 72 $^{o}$C; and finally a hold of 5 minutes at 72 $^{o}$C. The product from the PCR reaction and the pET-28a plasmid were digested with *NcoI* and *XhoI* for 3.5

hours in separate reactions. A 0.7% low melting agarose gel was used to separate the digested fragments and to visualize with UV light. The desired fragments were then removed from the gel and placed in a hot water bath to melt the gel. This solution containing the digested PCR fragment and the digested plasmid were used in ligation reactions with T4 DNA ligase and T4 DNA ligase buffer. Six ligation reactions were performed, three with the insert plasmid ratio of 5 µL:1µL, and the second three with the ratio of 5µL:2µL.

The ligation reactions were transformed into competent *E. coli* DH5α cells and plated on LB-agar plates containing 30µg/mL of kanamycin. Six starter cultures, 5mL LB containing 30µg/mL of kanamycin shaken at 37$^{o}$C, were made from colonies grown on various plates from the ligation. After the starter culture had grown overnight, the plasmids were purified from the cultures using the Qiagen Miniprep kit following the procedure outlined in the manufacturer's instruction manual. To screen for insert acceptance, a second restriction digest was performed with the same restriction enzymes used in the first digest reaction. After the plasmids had digested for 3.5 hours, 1.0% agarose gel electrophoresis was performed to visualize the plasmids with the accepted insert. A successful clone was transformed into competent *E. coli* BL21(DE3) cells.

One colony from this retransformation was grown overnight in a starter culture of LB and a concentration 30µg/mL of kanamycin. Two mL of the starter culture was then added to 1 liter of LB containing 30µg/mL of kanamycin, and was shaken at 37$^{o}$C until the $OD_{595}$ reached 0.400-0.450. Once this OD was

reached the culture was shaken at 18°C until an $OD_{595}$ of 0.500-0.600 is observed. 1.0 mM of IPTG was then added and the culture was then allowed to grow overnight. The following morning the cells were harvested by centrifugation (9950*g* for 30 minutes). The pellet was then stored at −80°C until it was used for purification.

III.E.2 Purification of MmgA

The pelleted cells obtained from the overexpression were resuspended in 20 mL of binding buffer (5mM Imidazole, 500mM NaCl, 20mM Tris·HCl, pH 7.9) and lysed by sonication on ice for 3 minutes. The lysate was then removed by centrifugation (11000*g* for 30 minutes) and applied to a Ni-NTA chromatography column containing 2mL of column bed. The chromatography was conducted at 4°C. The crude extract was loaded first, followed by a wash of 20mL of binding buffer. Wash buffer, 12mL, (60mM Imidazole, 500mM NaCl, 20mM Tris·HCl, pH 7.9) was added to remove nonHis-tagged proteins from the column. The Eluent buffer, 12mL, (200mM Imidazole, 500mM NaCl, 20mM Tris·HCl, pH 7.9) eluted MmgA from the column. Snakeskin® pleated dialysis tubing with a 7000 MWCO was used to dialyze the eluent overnight in 4L of buffer containing 25mM Tris·HCl, pH 7.5. The dialyzed protein was then concentrated by using a Vivaspin 5000 MWCO concentrator until the volume read 500µL to 750µL. An SDS-PAGE was used to determine the purity of MmgA. The concentration of the protein was then measured spectrophotometrically using the Bradford method.

The absorbance of the reagent plus the enzyme at 595nm was recorded. Using a standard curve of absorbance versus protein concentration, the concentration of MmgA obtained from the overexpression/purification was determined.

III.E.3 Spectrophotometric Analysis of MmgA Activity

Using the Cary BIO 100 UV·Vis spectrophotometer reaction mixtures containing CoA, acetoacetyl CoA, and MmgA were analyzed. To determine the reaction constants for coenzyme A a constant concentration of acetoacetyl CoA was used in each reaction(60μM), and varying concentrations of CoA (5μM, 10μM, 15μM and 20μM), MmgA was added to give a final concentration of 1.0μg/mL. In a quartz cuvette with a 10.00mm pathlength the acetoacetyl CoA and MmgA were added into enough buffer (100mM Tris, 25mM $MgCl_2$, pH 8.1) so that the total reaction volume equaled 1.0mL. The cuvette was then placed in the spectrophotometer with a constant temperature of 30 $^{o}$C. To begin the reaction the appropriate amount of CoA was added to the reaction. The reaction was then monitored at 300 nm for 0.3 minutes using the program Win UV Kinetics Application.

Once the program had finished recording the absorbance versus time, the data was then exported into a Microsoft Excel spread sheet. A linear curve fit was used to obtain an equation from each reaction. All reactions were done in triplicate to obtain an average rate. The absorbance of acetoacetyl CoA, when placed in the spectrophotometer in buffer without CoA of MmgA, decreased over

time. To account for this in the rates of each reaction, acetoacetyl CoA in buffer was placed in the spectrophotometer, at the same reaction concentration, and the rate of decay obtained from an average of three readings was subtracted from each rate obtained for all the reactions before the reaction rates were averaged. A reported $\varepsilon_{300}$ of 3600 $M^{-1}cm^{-1}$ for acetoacetyl CoA was used to calibrate the absorbance data. (Kornblatt, *et al.*, 1971) A double reciprocal plot of the rate of reaction versus the concentration of CoA was used to determine the kinetic constants $K_M$ and $V_{max}$.

Once the constants for CoA had been determined, the kinetic parameters for acetoacetyl CoA were measured. For these reactions, the concentration of CoA was held constant (70μM) and the acetoacetyl CoA concentration was varied (5μM, 10μM, 20μM, 30μM, 40μM, and 50μM). The same spectrophotometric parameters were used as in the previous reactions: 10.00mm quartz cuvette, 30$^oC$, absorbance monitored at 300nm. Acetoacetyl CoA was added to buffer (100mM Tris, 25mM $MgCl_2$, pH 8.1) and 1.0μg of MmgA. The reaction was started with the addition of CoA and the reaction was monitored for 0.3 minutes. After the data was imported into a Microsoft Excel spread sheet and a linear curve fit was applied, the rate was recorded. All concentrations were repeated at least three times, and the rate obtained for the specific concentrations was an average of the values. To account for any decay in absorbance signal, acetoacetyl CoA was added to the cuvette at the different concentrations in the reactions without CoA or MmgA and the change in

absorbance was recorded a minimum of three times. These values were then averaged. The rate of decay associated with the different concentrations of acetoacetyl CoA was subtracted from the reaction rates before the individual reaction rates were averaged. After averaging, a reported $\varepsilon_{300}$ of $3600 M^{-1}cm^{-1}$ for acetoacetyl CoA was used to obtain concentration versus time for the reaction (Kornblatt, *et al*., 1971).

The condensation reaction using acetyl CoA (60μM) and MmgA (1.0μg) in enough buffer (100mM Tris, 25mM $MgCl_2$, pH 8.1) to make 1mL total reaction volume was measured. A 10.00mm quartz cuvette was used in the reactions and each reaction was held at 30°C. The reaction was started with the addition of MmgA, and the absorbance was monitored at 300nm for 30 minutes. After 30 minutes there was no increase in absorbance. We therefore analyzed the same reaction by HPLC as described below.

The synthesis of a long chain β-keto acyl CoA was carried out by reacting acetyl CoA with butyryl CoA in the presence of MmgA. The reaction was started with the addition of MmgA, and the absorbance was monitored at 300nm for 30 minutes. After 30 minutes there was no increase in absorbance. We therefore analyzed the same reaction by HPLC as described below

III.E.4 HPLC Analysis of MmgA Reaction Products

MmgA reaction products and standards were analyzed using a Varian Prostar HPLC. For all reactions and standards a buffer (95% 200mM

Ammonium acetate, pH 6.0; %5 acetonitrile) was conducted at 1mL/min for 15 minutes with absorbance being monitored at 261nm. The column used was a Waters NovaPak C18 reverse phase column of 4µm particle size, 150mm in length, and a 3.9 mm inner diameter (Burns, *et al.*). Standards of CoA, acetyl CoA, acetoacetyl CoA, and butyryl CoA were used to determine the elution time of the compounds. The thiolysis reaction was conducted in an HPLC sample vial and contained 70µM CoA, 50µM acetoacetyl CoA, and 0.25 µg of MmgA in enough buffer (100mM Tris, 25mM $MgCl_2$, pH 8.1) to make 1 mL total reaction volume. Samples were run through the HPLC were before addition of MmgA, 30 minutes after addition of MmgA, 2 hours, 3 hours, and 4 hours after addition of MmgA. The reaction was incubated at 37$^{o}$C between HPLC runs. Formation of reaction products was determined by comparison of chromatographic peaks from the reaction and the standards.

The condensation reaction was carried out in an HPLC sample vial containing 60µM acetyl CoA, 1.0µg of MmgA, and enough buffer (100mM Tris, 25mM $MgCl_2$, pH 8.1) to give a total of 1mL reaction volume. Before MmgA was added to the reaction, acetyl CoA in buffer was applied to the column. Once the reaction had been started with the addition of the enzyme, a sample was analyzed every hour for 2 hours, and an overnight sample was analyzed the next day. The reaction was incubated at 37$^{o}$C between all timed runs on the HPLC. Reaction products were determined by comparing the obtained chromatographic peaks to those of the standard peaks.

For HPLC analysis of the condensation reaction involving butyryl CoA, 60μM acetyl CoA, and 50μM butyryl CoA were used. MmgA (1.0μg) was used in the reaction with enough buffer (100mM Tris, 25mM $MgCl_2$, pH 8.1) to make 1 mL total reaction volume. The mixture of butyryl CoA and acetyl CoA was added to the column before addition of MmgA using the program of 15 minutes at 95% 200mM ammonium acetate, 5% acetonitrile followed by an increase of acetonitrile to 60% over five minutes and was held constant at this percentage for another five minutes. The acetonitrile was then increased to 80% over five minutes and was again held constant for another five minutes. The flow rate used was 1 mL per minute. The butyryl CoA reaction was run before addition of MmgA, and every hour up to three hours after addition of MmgA, and an overnight sample was applied to the column. Between all HPLC trials the reaction was incubated at 37$^{o}$C. Signals from the chromatogram were assigned by comparison to authentic standards.

# REFERENCES


Antonenkov, V.; Van Veldhoven, P.; Waelkens, E.; Mannaerts, G. "Substrate specificities of 3-oxoacyl-CoA thiolase A and sterol carrier protein 2/3-oxoacyl-CoA thiolase purified from normal rat liver peroxisomes," *J. Biol. Chem.*, **1997**, *272*, 26023.

Brock, M.; Maerker, C.; Schutz, A.; Volker, U.; Buckel, W. "Oxidation of propionate to pyruvate in *Escherichia coli* involvement of methylcitrate dehydratase and aconitase," *Eur. J. Biochem.,* **2002**, *269*, 6184.

Bryan, E.; Beall, B.; Moran, C. "A $\sigma^E$-dependant operon subject to catabolite repression during sporulation in Bacillus subtilus," *J. Bacteriol.*, **1996**, *178*, 4778.

Butcher, R.; Schroeder, F.; Fischbach, M.; Straight, P.; Kolter, R.; Walsh, C.; Clardy, J. "The identification of bacillaene, the product of the *pksX* megacomplex in *Bacillus subtilus*," *PNAS¸* **2007***, 104,* 1506.

Burns, K.; Gelbaum, L.; Sullards, C.; Bostwick, D.; May, S. "Iso-coenzyme A," *J. Biol. Chem.*, **2005**, *280*, 16550.

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

Campobasso, N.; Patel, M.; Wilding, I.; Kallender, H.; Rosenburg, M; Gwynn, M. “Staphylococcus aureus 3-hydroxy-3methylglutaryl-CoA synthase: crystal structure and mechanism,” *J. Biol. Chem.,* **2004**, *M407882200.*

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

Colibri World-Wide Web Server. http://genolist.pasteur.fr/Colibri/ (accessed March 12, 2007). Data Release R5.1.

Errington, J. “Bacillus subtilis sporulation: regulation of gene expression and control of morphogenesis,” *Microbiol. Rev.*, **1993**, *57*, 1.

Gu, L.; Junyong, J.; Liu, H.; Hakansson, K.; Gerwick, W.; Sherman, D. “Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme Pair,” *J. Am. Chem. Soc.*, **2006**, *128*, 9014.

Guzman, L.; Belin, D.; Carson, M.; Beckwith, J. “Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter,” *J. Bacteriol.*, **1995**, *177*, 4121.

Hopwood, D.; Sherman, D. “Molecular genetics of polyketides and its comparison to fatty acid biosynthesis,” *Annu. Rev. Genet.*, **1990**, *24*, 37.

Katz, L.; Donadio, S. “Polyketide synthesis: prospects for hybrid antibiotics,” *Annu. Rev. Microbiol.*, **1993**, *47*, 875.

Kornblatt, J.; Rudney, H.; “Two forms of acetoacetyl coenzyme A thiolase in yeast,” *J. Biol. Chem.*, **1971**, *246*, 4417.

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

McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; 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.

Middleton, B. “The kinetic mechanism and properties of the cytoplasmic acetyl-coenzyme A thiolase from rat liver,” *Biochem. J.,* **1974**, *139*, 109.

Mootz, H.; Finking, R.;Marahiel, M. “4’-Phosphopantetheine transfer transfer in primary and secondary metabolism of *Bacillus subtilis*,” *J. Biol. Chem.*, **2001**, *376*, 37289.

Nakano, M.; Corbell, C.; Besson, J.; Zuber, P. “Isolation and characterization of Sfp: a gene that functions in the production of the lipopepetide biosurfactant, surfactin, in *Bacillus subtilis*,” *Mol. Gen. Genet.*, **1992**, *232*, 313.

*Novagen, pET System Manual.*, 11th ed.

Oku, H.; Keneda, T. "Biosynthesis of branched-chain fatty acids in *Bacillus subtilis*," *J. Biol. Chem.*, **1988**, *263*, 18386.

Pantazaki, A.; Ioannou, A.; Kyriakidis, D. "A thermostable β-ketothiolase of polyhydroxyalkanoates (PHAs) in *Thermus thermophilus*: purification and biochemical properties," *Mol. Cell. Biol.,* **2005**, *269*, 27.

*Qiagen,* The Qia*expressionist*™, 5th ed.

Quadri L.; Weinreb P.; Lei M, Nakano M.; Zuber P.; Walsh C. "Characterization of Sfp, a *Bacillus subtilis* phospsopantetheinyl transferase for peptidyl carrier protein domains in peptide synthesis," *Biochemistry*, **1998**, *37*, 1585.

Rawlings, B. "Biosynthesis of polyketides," *Nat. Prod. Rep.*, **1997**, 14*,* 523.

Shen, Ben. "Polyketide biosynthesis beyond the type I, II, and III polyketide synthase paradigms," *Curr. Opin. Chem. Biol.*, **2003**, *7*, 285.

Simunovic, V.; Zapp, J.; Rachid, S.; Krug, D.; Meiser, P. Muller, R. "Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthase, 3-hydroxy-3-methylglutaryl-CoA synthases, and *trans*-acting acyltransferases," *ChemBioChem.*, **2006**, *7*, 1206.

Staunton, J.; Wilikinson, B. "Biosynthesis of erythromycin and rampamycin," *Chem. Rev.*, **1997**, *97*, 2611.

Stein, T. “*Bacillus subtilis* antibiotics: structures, syntheses, and specific functions,” *Mole. Micro.*, **2005**, *56*, 845.

Stephens, C. “Bacterial sporulation: a question of commitment?” *Cur. Biol.*, **1998**, *8*, R45.

Subtilist World-Wide Web Server. http://genolist.pasteur.fr/SubtiList/ (accessed March 12, 2007).Data Release R16.1.

Zweerink, Z.; Edison, A. “Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*,” *J. Antibiot. (Tokyo)*, **1987**, *40*, 1692.