Functional Separation of Multimodular Type I PKS Polypeptides by Utilizing Matched Docking Domains From a Heterologous PKS System

John Kam Yan
Portland State University

Let us know how access to this document benefits you.
Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Recommended Citation

10.15760/etd.135

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Functional Separation of Multimodular Type I PKS Polypeptides by Utilizing Matched Docking Domains From a Heterologous PKS System

by

John Kam Qun Yan JR

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Chemistry

Dissertation Committee:
Kevin A. Reynolds, Chair
Michael Bartlett
Dirk Iwata-Reuyl
Niles Lehman
Kenneth Stedman

Portland State University
2010
Abstract

Bacterial type I modular polyketide synthases (PKS) are large multifunctional enzyme systems responsible for the biosynthesis of complex polyketide natural products such as erythromycin, pikromycin, and borrelidin. Type I systems are comprised of a loading module which generally selects an appropriate acyl group starter unit, and multiple discrete extension modules, responsible for each single round of acyl group incorporation into the final polyketide core structure. These modules can exist naturally as either single discrete polypeptides, such as modules 5 and 6 from the pikromycin PKS (PikA3 and PikA4 respectively), or as multimodular polypeptides fused together by short intrapolypeptide linkers such as the loading module and the first and second extension modules of the erythromycin and pikromycin PKSs (DEBS1 and PikAI respectively). While short peptide linkers between modules on the same polypeptide facilitate the transfer of polyketide intermediates from one module to the next via their close proximity to one another, docking domains found at the C-terminus of one module and the N-terminus of the next subsequent module facilitate the needed protein-protein interactions for the passage of biosynthetic intermediates between modules on separate polypeptides.

The ability to utilize docking domains in place of intrapolypeptide linkers was explored in the pikromycin and erythromycin PKSs by dissecting the tri-modular PikAI and DEBS1 polypeptides with matched docking domains. It has been shown that PikAI can be separated into two proteins at either of these linkers, only when matched pairs of docking domains from a heterologous modular phoslactomycin PKS are used in place of the intrapolypeptide linker. In both cases the yields of pikromycin
produced by the *S. venezuelae* host mutant, which is a PikAI deletion strain were 50% of that of an *S. venezuelae* strain expressing the native trimodular PikAI.

Additionally, expression of module 2 as a monomodular protein fused to a heterologous N-terminal docking domain was also observed to give almost a 10-fold improvement in the *in vivo* generation of pikromycin from a synthetic diketide intermediate.

The utilization of docking domains to separate linked modules was also demonstrated in the erythromycin PKS. Expression of the first protein involved in erythromycin biosynthesis (DEBS1) with the DEBS thioesterase fused to the C-terminal (DEBS1-TE) in *S. venezuelae* results in the production of triketide lactone products. Separation of DEBS1-TE resulted in 50% triketide lactone production, consistent with the observations in the pikromycin system. Published work has shown that the DEBS loading module has relaxed substrate specificity, and is capable of incorporating acetate, butyrate and isobutyrate in addition to the normally observed propionate starter unit, which typically predominates. However, in the current study when the DEBS loading module is separated from module 1 with matched docking domains, a dramatic shift in the starter unit, favoring the isobutyrate derived tri-ketide lactone is observed. This apparent shift in starter unit preference for a dissected PKS system has resulted in insights into the kinetics of acyl group loading, off loading, as well as the hydrolysis and transfer from the AT to ACP domains.

In addition to the separation of multimodular PKS polypeptides with docking domains, it has also been shown that the individual catalytic domains of single discrete module, BorA5 from the borrelidin PKS can be expressed as stand alone proteins
while retaining catalytic functionality *in vitro*. This work has provided a basis for future studies of this module, which has been proposed to function iteratively, catalyzing three rounds of chain elongation.
Acknowledgements

I would like to convey my sincere appreciation and gratitude to Dr. Kevin Reynolds for allowing me to work with him and his research group. It has been his supervision, consistent encouragement, and enthusiastic support that have made this thesis possible. I am also grateful for all the members of the Reynolds group for sharing their many diverse expertise’s with me, this experience has provided an inspiring and thought provoking environment to grow, both as a scientist and a person. I would like to express my dearest thanks to Dr. Galina Florova and Dr. Nadaraj Palanippan for their countless suggestions and advice whenever I needed a fresh perspective. My sincere thanks and appreciation also goes to my advisory committee, Dr. Michael Bartlett, Dr. Dirk Iwata-Reuyl, Dr. Niles Lehman, and Dr. Kenneth Stedman for their generous time, flexibility, constructive criticism and thoughtful suggestions during the various stages of my graduate career.

I would also like to acknowledge the Portland State University chemistry department for providing me with a truly unique and special place to learn and grow these last five years.

Finally I would like to thank friends and family for their love and support. Thank you for your support, encouragement and for always believing in me. Without a doubt, none of this would have been possible without all of you in my life.
Table of Contents

Abstract ........................................................................................................................................... i

Acknowledgements ......................................................................................................................... iv

List of Figures ..................................................................................................................................... vi

List of Schemes ................................................................................................................................. x

List of Tables ..................................................................................................................................... xi

List of abbreviations .......................................................................................................................... xii

Introduction ......................................................................................................................................... 1

Functional dissection of PikAI from the pikromycin type I polyketide synthase .......................................................... 31

Functional modular dissection of the trimodular DEBS1-TE results in an unexpected shift in triketide lactone ratios and provides new insights into aspects of acyl group loading and hydrolysis ................................................................. 59

*In vitro* investigation of the iterative mono module BorA5 from the borrelidin PKS from *Streptomyces parvulus* Tu4055 .......................................................................................................................... 111

Conclusions and recommendations for future work ........................................................................... 134

References .......................................................................................................................................... 139
List of Figures

Figure 1. Examples of anticancer and immunosuppressantmicrobial natural products………………………………………………………………………………2

Figure 2. Examples of polyketide natural products……………………………………4

Figure 3. Examples of non-ribosomal peptide natural products…………………………5

Figure 4. Generic reaction scheme for both fatty acid and polyketide biosynthesis………………………………………………………………………………7

Figure 5. Schematic examples and organization of Type I, II and III PKS………………9

Figure 6. Polyketide synthase domain interactions…………………………………………………………14

Figure 7. Passage of polyketide intermediates between rounds of elongation……….17

Figure 8. Structural basis for docking in modular PKSs………………………………19

Figure 9. Structures of the pikromycin series of antibiotics…………………………….23

Figure 10. Erythromycin PKS…………………………………………………………….27

Figure 11. PCR Recombination strategy…………………………………………………..36

Figure 12. PCR confirmation of plasmids isolated from S. venezuelae……………….48

Figure 13. HPLC and LC-MS analysis of methymycin and pikromycin production...50

Figure 14. Levels of pikromycin products made in the BB138 PikAI deletion strain of S. venezuelae with respective complementation plasmids…………………………51

Figure 15a. Two routes to prime the pikromycin PKS and provide 12 and 14-membered macrocyclic products…………………………………………………52
Figure 15b. LC-MS analysis of BB138/pBK51 grown in the presence of NAC thioester of 2(S)-methyl-3(R)-pentanoic acid…………………………………………53

Figure 15c. LC-MS analysis of BB138/pBK51* grown in the presence of NAC thioester of 2(S)-methyl-3(R)-pentanoic acid…………………………………………54

Figure 15d. LC-MS analysis of BB138/pYJ7* grown in the presence of NAC thioester of 2(S)-methyl-3(R)-pentanoic acid…………………………………………55

Figure 16. Proposed mechanisms for the removal of acyl groups from DEBS
AT₁-ACP₁………………………………………………………………63

Figure 17. PikA3-A4 PKS docking interface structures……………………………..64

Figure 18. Selective fluorescent properties of Thioglo melamide reagents…………76

Figure 19. GC-MS total ion chromatogram of an extract of
S. venezuelae (BB138) carrying pBK1……………………………………………….78

Figure 20a. Ions associated with acetate derived TKL…………………………….79

Figure 20b. Ions associated with propionate derived TKL………………………….80

Figure 20c. Ions associated with isobutyrate derived TKL………………………….81

Figure 21. Triketide lactone production levels for DEBS1-TE expression constructs pBK1, pBK1*, pBK1** and pBK1* scar expressed in BB138……………………………82

Figure 22. GC-MS total ion chromatogram of an extract of S. venezuelae (BB138) carrying pBK1*………………………………………………………………………83
Figure 23. Extracted ion chromatogram of BB138 fermentation carrying expression construct pBK1* scar ................................................................. 85

Figure 24. GC-MS total ion chromatogram of an extract of S. venezuelae (BB138) carrying pBK1** ................................................................. 87

Figure 25. GC-MS total ion chromatogram of an extract of S. venezuelae (AX912) carrying pBK1 ................................................................. 89

Figure 26. GC-MS total ion chromatogram of an extract of S. venezuelae (AX912) carrying pBK1* ................................................................. 90

Figure 27. Triketide lactone production levels for DEBS1-TE expression constructs pBK1, pBK1*, and pBK1** expressed in AX912 ............................................. 91

Figure 28. Production of TKLs derived from cyclic starter units ................. 95

Figure 29. Expression and purification of DEBS AT₅₋ACP₅ (apo and holo) .......... 98

Figure 30. Kinetic analysis of apo DEBS AT₅₋ACP₅ ........................................ 99

Figure 31. Experimentally determined and kinetic model predicted levels of DEBS AT₅₋ACP₅ labeling with 14C propionyl or isobutyryl CoA ................................. 103

Figure 32. Proposed dynamics of propionyl and isobutyryl CoA loading and off loading on DEBS AT₅₋ACP₅ ......................................................... 104

Figure 33. Experimentally observed levels of DEBS AT₅₋ACP₅ labeling with competing acyl CoA substrates (propionyl and isobutyryl CoA) ........................................... 107

Figure 34. Modular organization of the borrelidin PKS ................................... 113

Figure 35. PCR amplification and cloning of BorA5 ....................................... 118
Figure 36. Expression and purification of apo BorA5…………………………….124

Figure 37. Expression and purification of holo BorA5 ACP………………………126

Figure 38. Positive mode mass analysis of BorA5 ACP……………………………127

Figure 39. Proposed reaction between apo BorA5 and stand alone holo BorA5 ACP domain in trans with one another……………………………………………………………129

Figure 40. Positive mode mass analysis of an overnight incubation of apo BorA5 with holo BorA5 ACP and methylmalonyl CoA……………………………………….130
List of Schemes

Scheme 1. Synthesis of NAC diketide thioester.............................................42
List of Tables

Table 1. Kinetic parameters for DEBS AT\textsubscript{L}-ACP\textsubscript{L}.................................100
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACP</td>
<td>Acyl carrier protein</td>
</tr>
<tr>
<td>Ala</td>
<td>Alaine</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>Apra</td>
<td>Apramycin</td>
</tr>
<tr>
<td>ARO</td>
<td>Aromatase</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>CYC</td>
<td>Cyclase</td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DD</td>
<td>Docking domain</td>
</tr>
<tr>
<td>DH</td>
<td>β-hydroxyacyl ACP dehydratase</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DPM</td>
<td>Disintegrations per minute</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>ER</td>
<td>Enoyl ACP reductase</td>
</tr>
<tr>
<td>EtOAC</td>
<td>Ethyl acetate</td>
</tr>
<tr>
<td>FAS</td>
<td>Fatty acid synthase</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
</tr>
<tr>
<td>GC-MS</td>
<td>Gas chromatography - mass spectrometry</td>
</tr>
<tr>
<td>His</td>
<td>Histidine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>hr</td>
<td>Hour</td>
</tr>
<tr>
<td>HRMS</td>
<td>High resolution mass spectrometry</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl thio-b-D-galactoside</td>
</tr>
<tr>
<td>KR</td>
<td>β-ketoacyl ACP reductase</td>
</tr>
<tr>
<td>KS</td>
<td>β-ketoacyl ACP synthase</td>
</tr>
<tr>
<td>LB</td>
<td>Luria broth</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography - mass spectrometry</td>
</tr>
<tr>
<td>LD</td>
<td>Loading domain</td>
</tr>
<tr>
<td>LSC</td>
<td>Liquid scintillation counting</td>
</tr>
<tr>
<td>MeoH</td>
<td>Methanol</td>
</tr>
<tr>
<td>min</td>
<td>Minute</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>NAC</td>
<td>N-acetyl cysteamine</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>O.D</td>
<td>Optical density</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PKS</td>
<td>Polyketide Synthase</td>
</tr>
<tr>
<td>PTLC</td>
<td>Preparative thin layer chromatography</td>
</tr>
<tr>
<td>SDS PAGE</td>
<td>Sodium docecyl sulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris (2-carboxyethyl) phosphine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>TE</td>
<td>Thioesterase</td>
</tr>
<tr>
<td>TKL</td>
<td>Triketide lactone</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin layer chromatography</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 General background on streptomyces and secondary metabolites

The largest genus of the bacterial order Actinomycetales is streptomyces with over 500 species having been reported and described (1, 2). Streptomyces are gram-positive filamentous bacteria with genomes characteristically high in G+C content (3). Streptomyces species are found in virtually all soil environments worldwide and are the producers of chemical metabolites called geosmins that are responsible for the characteristic earthly smell that is associated with soil (4).

Microbial organisms have long been used as a source of active pharmaceutical therapeutic agents. Organisms ranging from filamentous fungi, actinomycetes, myxobacteria, and cyanobacteria are some of the most industrious producers of the over 20,000 reported secondary metabolites (5) which are also referred to as microbial natural products. Secondary metabolites, whether they are from microbial or in many cases plant sources, are not required for primary growth of the organism but presumably offer a competitive advantage in potentially harsh and competitive natural environments. Many of these microbial secondary metabolites are defined as antibiotics, which are compounds that inhibit the growth of other microbes. Actinomycetes are responsible for the production of over two thirds of the currently known antibiotics. Although many of the bacteria-derived natural products were initially discovered based on their antibacterial, antifungal, antiviral and antiparasitic activities, many have been developed for other applications. Compounds such as
epothilone A, tubulysin A, and rapamycin, (Figure 1) are currently better known and used for their anticancer and immunosuppressant activities.

Figure 1. Examples of anticancer and immunosuppressant microbial natural products.

1.2. Polyketide and nonribosomal peptide natural products

Generally, bacterial-derived natural products have been categorized according to their biosynthesis as either polyketide or non-ribosomal peptide natural products (6). Polyketide, nonribosomal and mixed polyketide-non-ribosomal natural products are a structurally rich and diverse class of compounds with an enormous range of medically important applications (7). Examples of polyketides include the antibiotics
erythromycin, tylosin, tetracyclines, pikromycin, the antitumor agent doxorubicin and the cholesterol lowering agent lovastatin (Figure 2). These compounds are biosynthesized by large organized sets of enzymes known as polyketide synthases (PKSs) in a manner resembling fatty acid biosynthesis (8), where simple acyl-coenzyme A (CoA) building blocks such as acetyl, propionyl, malonyl, methylmalonyl, and ethylmalonyl CoAs are elongated in a head to tail manner into the final polyketide structure. Natural products with nonribosomal peptide synthase origins such as the siderophores yersiniibactin, enterobactin and coelichelin (Figure 3) are biosynthesized in a similar manner to polyketides by multimodular biocatalysts, except that the building block monomers are amino acids instead of acyl-CoAs. In addition to the 20-proteinogenic amino acids many non-proteinogenic amino acids have also been shown to be utilized; there have been over 400 reported monomers used for nonribosomal peptide building blocks (9, 10).
Figure 2. Examples of polyketide natural products.
Figure 3. Examples of non-ribosomal peptide natural products.
1.3. Polyketide synthases and its relationship to fatty acid synthases

Polyketide natural products represent a large array of structurally diverse and medically important compounds. The enzyme systems which are responsible for their production are comprised of a combination of catalytic domains with specific dedicated functions which together are defined as polyketide synthases (PKSs) (11-15). PKSs are part of secondary metabolic pathways, meaning that expression of the genes likely occurs only under certain environmental conditions (16, 17). Much of the basic information that is known about polyketide biosynthesis has been derived from the mechanistically similar fatty acid synthase (FAS) from primary biosynthetic pathways (17, 18). Both Type I FAS and PKS systems contain specific protein domains (19, 20) that are responsible for the production of their respective fatty acid and polyketide products, these specific domains along with their functional roles are summarized in Figure 4. In the general scheme that is followed by both FAS and PKS biosynthesis, each round of elongation involves a decarboxylative condensation between a ketosynthase (KS) domain bound intermediate and an acyl carrier protein (ACP)/CoA bound extender unit. In FAS systems the keto group on the β carbon is fully reduced before the next round of elongation or chain termination. In contrast, PKS systems control the level of keto group reduction by selectively reducing only at certain positions to either a hydroxyl or alkene or full alkane. This selective control of reduction at specific positions of the polyketide structure helps dictate the rich chemical diversity, which these compounds naturally display.
Figure 4. Generic reaction scheme for both fatty acid and polyketide biosynthesis. Each round of elongation involves the decarboxylative condensation of the KS bound product of the preceding round with either the CoA or ACP bound extender unit. Additional processing can occur involving a combination of ketoreduction (KR), dehydration (DH) and enoyl reduction (ER) to generate the final fatty acid or polyketide product. Figured modified from (21).
FAS systems are classified as either types I or II, depending on the configuration of the functional domains. FASs are classified as either type I if the catalytic domains are covalently linked on the same polypeptide as observed in vertebrate systems or type II if each domain is a discrete protein as found in bacteria. FASs synthesize long fatty acid chains via successive rounds of iterative decarboxylative condensation of short carbon groups from acyl-CoAs in combination with ketoreduction, dehydration and enoyl reduction steps to produce the final fatty acid product (Figure 4). Typically acetyl CoA is incorporated as the starter unit and malonyl CoA the source of the extender units however there have been reported examples of isobutyryl and methyl-butyryl CoA being incorporated as well, generating branched chain fatty acids (22).

1.4. Types of polyketide synthases

PKSs are complex protein systems which have been generally classified into three categories, type I, II and III (Figure 5). This classification is system is primarily based on the mechanistic manner in which the protein machinery incorporates the extender units into the polyketide product and whether the functional domains, which are homologous to those of the FAS enzyme system, operate in a discrete or iterative fashion.
a) Type I modular (non-iterative) PKS

Figure 5a. Schematic example and organization of a type I modular PKS
b) Type II PKS (iterative)

\[
\begin{align*}
\text{E}^\text{KS} \cdot \text{KS} \cdot \text{ACP} & \rightarrow \text{E}^\text{KS} \cdot \text{KS} \cdot \text{ACP} \\
\text{Tetracenomycin C} & \rightarrow \text{Post-PKS} \text{ modification} \rightarrow \text{Release and cyclization}
\end{align*}
\]

Figure 5b-c. Schematic examples and organization of Type II and III PKS.
AT – Acyl transferase, ACP – Acyl carrier protein, DH – Dehydratase, ER – Enoyl reductase, KR – Ketoreductase, KS – Ketoreductase, TE – Thioesterase. Figure adapted from (23).
The linear description of type I modular PKSs resemble that of animal type I FASs in both the domain organization at the primary sequence level as well as the sequence of chemical reactions (24, 23). However, there is one distinction between animal FASs and type I polyketide biosynthesis, which is the iterative nature of fatty acid production, where the same catalytic domains are used repeatedly to produce products of the proper length. In type I modular PKS systems large multifunctional enzyme complexes containing discrete modules are almost always responsible for each single round of polyketide chain elongation (11). A module is defined as a set of enzymatic domains required to carry out a single round of chain elongation.

Contained within each module there are discrete sets of catalytic domains responsible for that particular round of chain elongation and any tailoring activities. At minimum each module contains an acyltransferase (AT) domain responsible for selecting the extender acyl substrate and priming the cognate acyl carrier protein (ACP) domain that holds the growing polyketide chain. The ketosynthase (KS) domain within the module is charged with accepting the polyketide intermediate from the preceding module’s ACP domain and the catalyzation of a Claisen-like condensation between the ACP bound extender acyl substrate and the growing chain. Some PKS modules contain a variety of tailoring domains (KR, ER, DH) responsible for catalyzing keto group reduction, enoyl group reduction, and dehydration activities respectively.

Type II PKS (Figure 5b) are iterative acting complexes of discrete proteins with specific functionalities (25). These systems catalyze iterative decarboxylative condensation reactions usually with malonyl CoA by utilizing two KS-like condensing
enzymes domains. This, along with a discrete ACP where the growing polyketide chain is covalently bound, and a variety of tailoring domains [KR, CYC (cyclase), aromatase (ARO)] direct the production of the final polyketide product (21). The third type of PKS, type III is characterized by the use of a single multifunctional KS-like enzyme active site with the capacity to incorporate acyl CoA extender units and perform all the necessary steps to assemble the polyketide backbone of the defined and proper length (21, 26).

1.5. Protein domain interactions required for polyketide production

An interesting aspect of type I modular PKSs involves the ordered arrangement of the modules. The polyketide chain elongation process is initiated by the priming of a loading module or loading domain (LD) with the appropriate starter unit, which is elongated by each subsequent module with either typically malonyl or methylmalonyl via their respective CoA thioesters into the final polyketide product. In many PKSs, multiple modules (typically two to six) can be housed on a single polypeptide. There are also many PKSs where several or even all of the modules are contained on separate polypeptides.

Perhaps the simplest, yet imperative question about polyketide biosynthesis involves the specificity and fidelity of processing biosynthetic intermediates. Whether the next subsequent module in polyketide biosynthesis is on the same polypeptide or separate polypeptides, an efficient and faithful passage of intermediates is vital. The proper channeling of polyketide biosynthetic intermediates involves specific protein-protein interactions between the individual domains of a single module in order to
prime, extend and tailor the intermediate properly. Additionally the appropriate module-to-module interactions are required to properly and efficiently transfer the growing polyketide backbone from the ACP of one module to the correct KS of the subsequent module. The ACP domain plays a vital role in this process; it is required to interact with all the individual domains of its own module as well as the KS of the subsequent module. During each round of elongation, the growing polyketide backbone is bound to the ACP domain via a prosthetic phosphopantetheine group that is the result of a post translational modification of the protein at a specific serine residue. The phosphopantetheine group has been described to act like a swinging arm, capable of a variety of activities; a) accepting the elongation acyl group from the AT, b) interacting with the KS to facilitate condensation, c) presenting the intermediate group to the various reduction domains (KR, DH, ER) described above that maybe present on the module and d) facilitating passage of the intermediate to the KS of the subsequent module for elongation (Figure 6).
Figure 6. Polyketide synthase domain interactions. ACP bound intermediates must be able to efficiently interact with all the domains of its cognate module as well as the subsequent KS domain. Figure adapted from (24).
The ability of the ACP domain to efficiently interact with the KS and AT domains of its own module is vital for polyketide biosynthesis. While ACP interactions with the reduction domains appear to be dictated by proximity of the domains to the ACP bound intermediate (24), its interactions with the KS and AT appear to be more structured. Specifically matched recognition motifs are likely to be involved in the KS to ACP and AT to ACP interactions, in addition the ACP domain likely possesses inherent conformational dynamics that play an important role in these interactions (27-29).

The substrate specificity of the accepting KS domain as well as the conformation of the donating ACP play crucial roles for the efficiency of polyketide chain transfer. In type I modular PKS systems, transfer of the growing native polyketide chain between ACP and KS domains seems to occur efficiently whether modules are linked together on the same polypeptide or if they are disconnected on separate polypeptides. Presumably, when modules are on the same polypeptide connected with a short linker sequence (30) the transfer is facile, as the donating ACP and accepting KS are close together. For sequential modules separated on different polypeptides, structural evidence has shown that key interactions between amino acids at the C-terminal of the ACP and the N-terminal of the KS forming important docking interactions between the two modules (31).
1.6. Linkers and Docking Domains

PKS modules that are housed on the same polypeptide are separated by a short intrapolyptide linker (i.e. between module 1 and 2 of the Pik PKS) (Figure 5a), that is typically about 20-30 amino acids long and is responsible for linking modules within multimodular polypeptides (31, 32). Generally these intrapolyptide linkers are not conserved at either the nucleic or amino acid level but appear necessary for polyketide production process. Conversely, much longer docking domains (DDs, ~100 amino acids) are found at the C-terminus of ACP domains, and the N-terminus of KS domains between both monomodular and multimodular PKS polypeptides such as that between module 2 and 3 of the Pik PKS (Figure 5a). Both linkers and DDs function in mediating the transfer of polyketide intermediate from the ACP of one module to the KS of the subsequent module (32-34) (Figure 7a-b). The term DD has been defined as the stretch of amino acids at the C-terminus of a PKS protein after the end of ACP homology, along with the amino acids found at the N-terminus of the next protein before the beginning of KS homology. Together these peptide regions or DDs are vital to the passage of the growing polyketide chain from one module to the next neighboring module on a separate polypeptide.
Figure 7. Passage of polyketide intermediates between rounds of elongation. The passage of intermediates between ACP and KS domains can be mediated by either (a) short intraprotein linkers or (b) docking domains. Iterative functionality (c) entails the passage of an ACP bound intermediate to the KS of the same module. Figure adapted from (24).
These docking domains (DD) have been shown to be essential in promoting the appropriate protein-protein, and thus module-module, interactions required for the biosynthetic processes (35, 33, 36-38). Based on amino acid sequence analysis, a recent method for categorizing PKS DDs into subclasses has been published (39). Thattai and co-workers have defined specific amino acid interactions between head (ACP bound) and tail (KS bound) DD pairs that have been generally categorized as H1-T1, H2-T2, or H3-T3. These interactions are proposed to give the DD pair a specific affinity for one another that facilitates the proper and efficient passage of biosynthetic intermediates.

Both NMR and x-ray crystallographic structural models of the DD interactions for DEBS2/DEBS3 (32) (Figure 8) and PikAIII/PikAIV (40) junctions respectively have been reported; these two sets of docking domain fall into the category of H1-T1 and H2-T2 respectively. Both of these models depict the DD interactions as being primarily determined by two sets of the four alpha helix interactions. Dimerization elements are generated by two four alpha helix bundles formed by the two C-terminal ACP bound DDs of the PKS homodimer, while a second is formed by the coiled-coil of the N-terminal KS which forms a helix-bundle dimerization element with two helices from the C-terminal ACP bound DD. These interactions play a critical role in both the stabilization of the PKS homodimer and the docking interactions of the C-terminal ACP bound domain with the N-terminal DD of the subsequent KS (32, 40).
Figure 8. Structural basis for docking in modular PKSs. NMR solution structure of the DEBS 2/3 docking domain complex. The dimeric C-terminal ACP DD is shown in blue and yellow (three helices), while the dimeric N-terminal KS DD is shown in green and orange. Two dimerization elements are present, an intertwined four α-helical bundle formed by helices 1, 1’, 2, and 2’, and a coiled-coil motif formed by the N-terminal docking domain. Docking between the two domains, results in formation of a second four α-helical bundle, as indicated. The linker region between helices 2 and 3 is highly mobile, and therefore is represented as a dashed line in the structure. Reproduced with permission from (22).
1.7. Genetic Engineering of PKS pathways

Despite the vast structural diversity that exists with naturally occurring polyketide natural products, there has been a strong focus and desire to generate compounds with even more structural diversity than what nature has already supplied. Approaches that have been successfully used to generate novel natural product inspired compounds include semisynthetic and total synthetic routes using the natural product scaffold. Efforts to generate new compounds using these synthetic routes based on natural product scaffolds have resulted in many natural product inspired compounds that are currently used in pharmaceutical applications (7).

When the linear description of modular type I PKSs was first reported (11, 13) these enzyme systems seemed to be an excellent tool for generating compounds with rich chemical diversity via combinatorial and genetic engineering methods (41-43). Based on the assembly line model for these systems a variety of strategies have been used in attempts to reprogram these pathways at the DNA level, creating hybrid PKS proteins capable of processing a variety of substrates, thus creating engineered natural products. Common approaches such as the inactivation/deletion of individual catalytic domains (AT, ACP, KR, DH, ER) or the insertion of new ones within an intact module have generated over 200 novel structures (44), however often times only trace levels of the new product is produced if any at all from the genetically engineered hybrid systems. The low levels of production come from compatibility issues, which limits the transfer of the growing polyketide chain in a hybrid system, from the donating ACP to the sequential KS domain. Indeed, in model systems it has been
shown that when the non-native KS domain for the donating ACP is replaced with the
naturally accepting KS domain, production of a polyketide product can be restored
(45). This evidence shows that there are intrinsic interactions between naturally
sequential modules (ACP and KS domains), which are important for polyketide
biosynthesis.

Due to the intricacies involved with forming the proper domain-domain
interactions within an engineered single module, there has also been an effort to
generate hybrid PKS pathways by using complete and intact modules as the building
blocks, versus individual catalytic domains, along with matched docking domain pairs
to generate the required module-module interactions (44). The rationale for this
approach has been built on the fact that domains contained within a single module are
already optimized to interact with one another efficiently, which alleviates the need to
engineer the proper domain interactions. However with this approach there are still
considerations that must be met to form efficient hybrid systems, such as a) the
modules as a whole which are being used as the building blocks must exhibit some
substrate flexibility capable of processing a variety of intermediates, and b) there must
be the proper communication/interactions between heterologous modules for efficient
passage of ACP bound intermediates to the subsequent KS of the next module.

1.8. Iterative PKS modules and substrate flexibility

PKS modules such as BorA5 from the borrelidin (46) and AurA from the
aureothin (47) PKSs differ from typical PKS modules in that they are proposed to
function in an iterative fashion. While the mechanism for iteration is unknown, the
idea of a module functioning iteratively suggests that there is inherent substrate
flexibility, something that allows for a variety of substrates to be processed efficiently.
For iterative PKS modules a variety of intramodular events must take place. In
addition to the normal reduction and dehydration processing which the ACP bound
intermediate must undergo (when these domains are present on the module), the ACP
bound intermediates must also be passed back to the KS of the same module (Figure
7c) to undergo further elongations before passage to the subsequent KS of the next
module, which presumably has specificity for a polyketide chain of a certain length to
prevent premature transfer. This is an interesting phenomenon for modular PKS
systems, where it is usually accepted that either linkers or DDs are required for
passage of the growing polyketide chain between ACP and KS domains (Figure 7a-b);
where the presence of intrapolypeptide linkers results in the close proximity of the
ACP donor and KS acceptor and DDs aid in the specific recognition of the proper
ACP – KS interactions (32-34, 48). This means that the KS domains of the iterative
modules are capable of being primed by the ACP of its own module multiple times
with polyketide chains of differing lengths. It is this built in flexibility that offers the
possibility of generating more efficient hybrid systems, capable of possibly generating
a variety of polyketide intermediates.

1.9. Macrolide antibiotics

Polyketide natural products such as pikromycin, methymycin, erythromycin,
azithromycin, and roxithromycin are classified as macrolides because of their large
macrocyclic lactone core structure, to which a deoxysugar is attached. Pikromycin
(Figure 9) was the first isolated macrolide antibiotic reported in 1951 and methymycin was the first to have its structure fully elucidated. Macrolide antibiotics exert their biological activity by binding to the 50S subunit of the bacterial ribosome, blocking the translocation of peptidyl tRNA and therefore inhibiting protein synthesis (49).

Figure 9. Structures of the pikromycin series of antibiotics
1.10. Introduction to pikromycin and erythromycin PKS

Much of the knowledge of polyketide biosynthesis has come from the well-studied model system for type I PKS biosynthesis: 6-deoxyerythronolide B (DEBS) PKS (Figure 10a), which produces the aglycone precursor for the antibiotic erythromycin B from *Saccharopolyspora erythraea* NRRL 2338. Genetic sequence information from DEBS along with the homology that PKSs share with FASs, have allowed for structural dissection (12) and biochemical investigation of the modular subunits that comprise this PKS pathway. This well studied model system has laid much of the groundwork for biochemical, structural, and genetic engineering efforts for other type I modular PKS systems (50).

1.10.1. Pikromycin PKS

The pikromycin and erythromycin PKSs are similar in that they both contain a LD and six extension modules, the first polypeptide houses the LD along with the first two extension modules, and a propionyl acyl group is passed from the LD to the KS of the first extension module (in the Pik system methylmalonyl CoA is initially loaded and decarboxylated to yield the propionyl group, while in the DEBS system propionyl CoA is loaded directly). Despite the similarities in modular organization, these two PKS systems produce different compounds, as a result of tailoring domain placement within the modules, which mediates the degree of reduction during each round of elongation and the activities of post PKS enzymes.

The pikromycin/methymycin (Pik) PKS from *Streptomyces venezuelae* ATCC 15439 produces four members of the pikromycin series of macrolides, methymycin,
neomethymycin, pikromycin and narbomycin (Figure 9). The Pik PKS is unique in its ability to produce both 12- and 14- member macrolactones. This is accomplished via thioesterase activity after either five or six rounds of chain elongation, followed by cyclization. The biosynthetic gene cluster encoding the Pik PKS responsible for biosynthesis of these macrolides was cloned and reported in 1998 (51). This gene cluster consists of 18 genes that are divided into five loci based on their biosynthetic function. *PikA* contains the PKS genes, *pikB* is responsible for desoamine biosynthesis, *pikC* encodes a cytochrome P450 hydroxylase, *pikD* encodes a transcriptional activator regulating the biosynthetic pathway, and *pikR* is a resistance gene to yield self-protection from the produced antibiotics (52).

The six Pik extension modules along with the LD and thioesterase domain are housed on four separate polypeptides (Figure 5a) that are responsible for the biosynthesis of both 12 and 14-membered ring macrolides 10-deoxymethnolide and narbonolide. Additional modification by a glycosyl transferase results in glycosylated macrolides, followed by hydroxylation results in the final corresponding antibiotics methymycin and pikromycin.

### 1.10.2. DEBS PKS

The DEBS PKS (Figure 9a) from *Saccharopolyspora erythraea* is remarkably similar to the Pik system. However in the DEBS system the six extension modules are housed on three discrete polypeptides (versus four in the case of Pik). DEBS1 is similar to PikA1, in that it is a trimodular protein, containing the loading domain (LD) and the first two extension modules on the same polypeptide.
Although the natural starter unit for erythromycin biosynthesis is propionyl CoA, DEBS LD displays a substantial amount of substrate flexibility. There have been several examples both \textit{in vitro} and \textit{in vivo} where acetyl, butyryl, and isobutyryl CoA in addition to its natural starter unit, propionyl CoA have been successfully processed by the erythromycin PKS enzymes (53-56). Previously reported \textit{in vitro} and \textit{in vivo} engineering work on the DEBS PKS (54, 55, 57) has shown that fusion of the DEBS TE domain to the C-terminal of DEBS1 yields a catalytically fusion protein DEBS1-TE (Figure 10b). This fusion protein is capable of loading a starter unit and processing two rounds of chain elongation before the terminal TE domain catalyzes release and cyclization of the predicted triketide lactone polyketide product.
**Figure 10. Erythromycin PKS.** a) Modular organization of the erythromycin PKS. b) DEBS1 – TE fusion protein functions to generate triketide lactone products.
1.11. Research objectives

In many PKSs, multiple modules (typically two to six) can be housed on a single polypeptide. There are also many PKSs where several or even all of the modules are contained on separate polypeptides. The six extension modules of the pikromycin PKS are housed on a combination of multimodular (PikAI and PikAII) and monomodular (PikAIII and PikAIV) polypeptides (Figure 5a). In contrast the six extension modules of DEBS, are all housed on three bimodular polypeptides (Figure 10a) (11-13).

The overall goal of this work involves development of strategies to generate efficient hybrid PKS systems. One such strategy, involves the utilization of matched DDs to pair modules from heterologous systems together. However, before DDs can be utilized, the first step in this approach requires that the specific impact on polyketide production when matched docking domains pairs are substituted for the natural intrapolyptide linkers be examined. The impact on the efficiency of polyketide production will be examined in both the pikromycin (PikAI) and DEBS (DEBS1-TE) PKSs when heterologous DDs have been substituted for the native linker peptides between the modules of these multimodular polypeptides. This work will provide insight into the feasibility of using matched DD pairs to mix and match modules from heterologous PKS systems by determining the relative impact on the overall polyketide production when multimodular PKS proteins are dissected using DD pairs.

A second strategy for generating hybrid PKSs capable of producing novel products involves utilizing iterative PKS modules. By examining the substrate
flexibility of BorA5, an iterative module from the borrelidin PKS to utilize a variety of substrates it will also be possible to use this module’s intrinsic processing flexibility as a tool to generate efficient hybrid PKS systems. In the long term, utilizing the implicit broad substrate specificity that iterative modules such as BorA5 exhibits may make it amenable to pair with heterologous modules by utilizing the specificity and exclusiveness that matched DD pairs offer, in forming the necessary interactions for the passage of polyketide biosynthetic intermediates.

**Specific goals:**

**Pikromycin PKS**

1. To determine whether the trimodular PikA1 polypeptide can be functionally dissected into two separate polypeptides at both the loading domain-module 1 and module 1-module-2 junctions with the docking domain pair from the phoslactomycin (Plm) PKS.
2. To determine the necessity of either intrapolypeptide linkers or docking domains for efficient polyketide production by deletion of the natural linker between module 1 and module 2.
3. To examine the efficiency of loading a synthetic mimic of the natural diketide intermediate to a series of constructs expressing Pik module 2 in a) its native context, covalently linked to module 1 which has been catalytically inactivated, b) when module 2 is expressed as a separate polypeptide from the
loading domain and module 1 with docking domains in between and c) when only module 2 is expressed with the N-terminal Plm DD.

**Erythromycin (DEBS) PKS**

1. To validate the use of Plm docking domains to separate the multimodular DEBS1-TE polypeptide at the loading domain-module 1 and module 1-module 2 junctions.
2. To examine the effect on triketide lactone production when the docking domain pair between the PikA3-PikA4 polypeptides (modules 5-6) from the Pik PKS to dissect the DEBS1-TE multimodular polypeptide.

**Borrelidin (BorA5) iterative PKS module**

1. To examine the substrate specificity and flexibility of BorA5 to catalyze multiple elongations in an *in vitro* setting when the entire module is expressed as a single protein or when the module is dissected into discrete domains and expressed as separate polypeptides.
2. Functional dissection of PikAI from the pikromycin type I polyketide synthase

2.1 Summary

The pikromycin polyketide synthase in *Streptomyces venezuelae* is comprised of a loading module and six extension modules, which generate both 12- and 14-membered macrolactone products. The first polypeptide of this system, PikAI is a multimodular component of this PKS and houses both the loading domain and the first two extension modules. The modules of this protein are joined together by two short intraprotein amino acid linker sequences. This work has shown that PikAI can be separated into two proteins at either of these linkers, while retaining the ability to produce significant quantities of polyketide product. However the ability to produce methymycin and pikromycin is only retained when matched pairs of docking domains (DDs) from a heterologous modular phoslactomycin PKS are used in place of the intraprotein linker. Complete deletion of the linker region without replacement with DDs resulted in no detectable polyketide production. In both cases, separations at the LD-module 1 and module 1-module 2 junctions the yields of pikromycin products produced by the *S. venezuelae* mutants were 50% of that of an *S. venezuelae* strain expressing the native trimodular PikAI. This observation provides the first demonstration that such separations do not dramatically impact the efficiency of the entire in vivo biosynthetic process.

Expression of module 2 as a monomodular protein fused to a heterologous N-terminal docking domain was also observed to give almost a 10-fold improvement in the in vivo generation of pikromycin from a synthetic diketide intermediate. These
results demonstrate the utility of DDs to manipulate biosynthetic processes catalyzed by modular PKSs and the quest to generate novel polyketide products. This chapter has been published as a paper in *ChemBioChem* 2009, 10(9), 1537-43 (58).

### 2.2 Introduction

The length and structure of PKS intrapolyptide linkers and docking domains (DD) are different (32, 38) but their functional role is equivalent; they mediate transfer of a growing polyketide chain from the C-terminal ACP of one module to the N-terminal KS of the next module. This functional equivalency raises the question of whether they can be exchanged for one another without significantly reducing overall production levels, as well as what are the consequence of deleting the linker sequence entirely between modules without the substitution of DD.

Replacement of docking domains with an intrapolyptide linker would permit an increase in the number of modules on a single polypeptide. An increase from dimodular to tetramodular polypeptides in the DEBS system has been accomplished via a different approach of fusing the N- and C-terminal docking domains (59). Other examples of functional PKS polypeptide fusions have been reported in the aureothin (60), borrelidin (46), and rapamycin (61) systems. The fusion of two of the three polypeptides (DEBS1+2 or DEBS 2+3) together yielded comparable production levels but fusion of all three polypeptides resulted in barely detectable levels when compared to the wild type system (59). In the aureothin and borrelidin examples functional fusion of AurA and BorA5 monomodules with their respective adjacent PKS polypeptides AurB and BorA4 and or BorA6 provided evidence of their iterative
functionality. In the case of rapamycin, the second module of this PKS was fused between the first and second module of DEBS, producing a small amount (5%) of the expected tetraketide product (61). With the exception of the DEBS polypeptide fusions these examples yielded a significant drop in production levels compared to control experiments with native modules.

Conversely, replacement of intraprotein linkers with a matched pair of docking domains from a heterologous system should permit separation of a multimodular polypeptide into monomodules. Such an experiment has been reported for both the multimodular components of the rifamycin (62) and epothilone (63) PKS proteins, when reconstructing these complex pathways in *E. coli*. For the epothilone case, EpoD which contains 4 modules was split into two modules using matched DD pair from the stigmatellin PKS (63). For the rifamycin case the tetramodular RifA was split into two proteins using a matched DD from the erythromycin PKS (62). In both cases the yields of the fully extended polyketide products were very low (10 µg/L for epothilone C). The multimodular separation was also needed to address problems with expression of soluble multimodular PKS component. Therefore the effect of multimodular separations on polyketide production in hosts where the multimodular PKS is expressed as a soluble protein is unknown.

This question has been investigated in the pikromycin producing *Streptomyces venezuelae* host, using a heterologous DD pair from the phoslactomycin (Plm) PKS, (64) to replace the two PikAI intraprotein regions that link the loading domain and first two extension modules. The results show that such separations lead to only modest decreases in the overall pikromycin yields.
2.3 Materials and methods

2.3.1 Commercially purchased materials

All chemicals and antibiotics were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Restriction endonucleases and T4 DNA ligase was purchased from New England Biolabs (Ipswich, MA). PCR primers were custom synthesized from Integrated DNA Technologies (Coralville, IA). PCR reactions were carried out with the GC Rich PCR System from Roche (Basel, Switzerland).

2.3.2 Experimental Design

As described above in the introduction, the Pik PKS from *S. venezuelae* consists of four separate polypeptides responsible for macrolide biosynthesis. The first polypeptide (PikAI) is a trimodular protein expressing the loading domain and the first two extension modules covalently linked together (Figure 5a). The experimental design employed was built on an established plasmid-based complementation system in which pDHS722 drives expression of PikAI in BB138 (a PikAI deletion strain of *S. venezuelae*). The expression construct pDHS722 is a complementation plasmid that fully restores pikromycin production to BB138 (65) and was used as a starting point for the generation of variants pJY12, pJY7 and pJY9 (see results and discussion). In an analogous manner to which the pDHS722 variants were generated a series of pBK51 variants (pBK51* and pBK51* scar) was generated for diketide feeding experiments. Plasmid pBK51 is a complementation plasmid that does not restore
pikromycin production in BB138 due to the mutation of the active site cysteine residue in the KS1 domain to alanine.

2.3.3 Plasmids, strains and culture conditions.

All constructed complementation plasmids were derivatives of pDHS722, a PikA1 complementation plasmid that functions as a conjugative Streptomyces-E. coli shuttle plasmid. All plasmid DNA manipulations were transformed and purified from E. coli and the proper plasmid sequences confirmed by restriction digest, PCR screening and plasmid sequencing before transformation in S. venezuelae. Transformants were cultivated on LB (Laurie Broth) agar media and plasmids isolated from LB liquid media, both supplemented with the appropriate antibiotics.

2.3.4 DNA manipulation.

Variants of pDHS722 were prepared by a series of PCR reactions and subsequent PCR targeting recombination crossovers in E. coli, before final transformation into S. venezuelae (Figure 11). All DNA sequencing was performed at Oregon Health and Sciences University, (MMI research core facility) and the data analyzed with Accelyrs DS Gene Software. DNA encoding the docking domain pair between modules 1 and 2 of phoslactomycin (Plm) PKS was PCR amplified from cosmid 10B4 (64) with primers PLMDDF (5’-CTG GGG CAC CTC GAT TTC GG-3’) + PLMDDR (5’-GTG GTG CTT CGT CTC GGC GT-3’). The PCR program consisted of an initial hold for 2 minutes at 96°C, followed by 30 cycles of 96 °C, for 30 seconds, 58 °C, for 30 seconds, 72 °C, for 1 minute. This was followed by a final
extension period at 72 °C for 5 minutes. The PCR product was gel purified with use of Qiagen Gel Extraction Kit (Valencia, CA) and cloned into TOPO TA 2.1 cloning vector and subsequently transformed in One Shot Mach1 E. coli cells from Invitrogen (Carlsbad, CA). The inserts were verified by plasmid sequencing with manufacture supplied M13F and M13R primers, yielding construct pJY1.

**Figure 11. Recombination strategy.** General schematic diagram of PCR templates and PCR products used for recombination in pDHS722 and pBK51. Successful recombinants were verified by PCR screening and restriction enzyme analysis.
The apramycin $\text{aac (3) IV}$ resistance marker was amplified from $\text{pIJ773 (66)}$, using primers $\text{PlmDD-ApraF}$ (5' - CTG ATG CAG TTC ATC ATC ACC GAG TTG GGG GAC ATC TGA GTA *TAC GGT TCA TGT GCA GCT CCA TC - 3') + $\text{PlmDD-Apra-R}$ (5' - GAA CAG GGC GCG ACG GGA ACA GGA GTG GTG GCA GAG GAT GTA *TAC TGT AGG CTG GAG CTG CTT C - 3'), (bold, italicized sequences represent annealing sites for the antibiotic marker and * indicates location of unique $\text{Bstz17I}$ restriction sites introduced for later removal of this marker). The PCR program consisted of an initial hold at 94 °C, for 2 minutes, followed by 10 cycles of 94 °C for 45 seconds, 55 °C for 30 seconds, 72 °C, for 1 minute. This was followed by another 15 cycles of 94 °C for 45 seconds, 58 °C for 30 seconds, 72 °C for 90 seconds. This was followed by a final extension period at 72°C for 10 minutes. The resulting PCR product was recombined with $\text{pJY1}$ via established procedures (66). The resulting recombinant plasmid $\text{pJY2}$ was isolated from $\text{E. coli}$ with a GeneJET Plasmid Miniprep Kit, Fermentas (Glen Burine, MD) and confirmed by sequencing. Construct $\text{pJY2}$ was used as a template to generate PCR products with the needed 39nt sequence complementary to the desired point of recombination between the linked $\text{ACP}$ and $\text{KS}$ domains of $\text{PikA1}$. PCR products for the exchange of DNA sequences encoding the intramodular linkers between modules 1-2 and LD-module 1, for that coding for the Plm docking domain pair, were amplified with primer pairs $\text{PikA1F}$ (5' - GTC TTC GAC CAC CCG ACG CCG CTG GCC CTC GTG TC G CTG CTG GGG CAC CTC GAT TTC GG - 3') + $\text{PikA1R}$ (5' - GGG GTA GCG GCA GCT CAT CGC GAC GAT CGC GAT CGG ATC GTG GTG CTG CTC GGC GT - 3') and $\text{LDSepF}$ (5' - TTC GAC TTC CCC ACC CCC GAG GCT CTC GCG GAG CAG
CTG CTG GGG CAC CTC GAT TTC GG - 3') + LDsepR (5' - GGG CAG GCG GCA GGC CAT GCC GAC GAT CGC CAC CGG CTC GTG GTG CTT CGT CTC GGC GT - 3') respectively (primer sequences in bold italicized font represent annealing sites on the template pJY2 and remaining 5' region represent the complementary 39nt of the PikAI region for recombination). The PCR program consisted of an initial hold at 94 °C, for 2 minutes, followed by 15 cycles of 94 °C for 45 seconds, 54 °C for 30 seconds, 72 °C, for 90 seconds. This was followed by another 15 cycles of 94 °C for 45 seconds, 58 °C for 30 seconds, 72 °C for 90 seconds. This was followed by a final extension period at 72 °C for 10 minutes. PCR products were introduced to pDHS722 by homologous recombination via established protocols (67). Recombinant pDHS722 plasmids, were isolated as described above, treated with restriction enzyme Bstz71I to remove the apra marker, ligated overnight at 12°C with T4 DNA ligase, transformed into E. coli and single colonies screened for both (+) Amp and (-) Apra, yielding JY7 and JY12 respectively. The plasmid pJY9, which has had the region encoding the intrapolyptide linker between modules 1 and 2 deleted, was also generated by homologous recombination via established protocols. The apramycin aac (3) IV resistance marker amplified from pIJ773 was used for deletion of the linker encoding region with primer Pik linkAF (5’-GTC TTC GAC CAC CCG ACG CCG CTG GCC CTC GTG TCG CTG TGA* TAC TGA TCA TGT GCA GCT CCA TC -‘3) + Pik linkAR (5’- GGG GTA GCG GCA GCT CAT CGC GAC GAT CGC GAT CGG ATC CAC CTA TCG GCT TCG CGG CTG TCC GCC CGG TA*T ACT GTA GGC TGG AGC TGC TTC - ‘3). Bold and italicized regions represent annealing regions for the resistance marker and * indicates location of the unique Bstz171 restriction sites
introduced for later removal of this marker. A TGA stop codon (underlined and bold text) was introduced at the 3’ end of the module 1 ACP encoding region in the forward primer. The reverse primer contained the complementary sequence to the associated RBS (underlined) and start codon (CAC) of the N-terminal Plm DD (underlined and bold text) fused to the beginning of the KS module 2 encoding region. The PCR program consisted of an initial hold at 94 °C, for 2 minutes, followed by 15 cycles of 94 °C for 45 seconds, 54 °C for 30 seconds, 72 °C, for 90 seconds. This was followed by another 15 cycles of 94 °C for 45 seconds, 58 °C for 30 seconds, 72 °C for 90 seconds. This was followed by a final extension period at 72 °C for 10 minutes. This PCR product was again recombined into pDHS722 via established protocols and the apramycin aac (3)Iv resistance marker was removed by restriction enzyme digestion with Bstz17I as before and re-ligated to yield the final pJY9 plasmid.

PBK51* was generated in the same manner as pJY7 but using pBK51 (encoding PikAI with a Cys to Ala mutation on KS1) as the final recombination plasmid. The plasmid pJY7*, which expresses only the C terminal PLM DD fused to module 2 of PikAI was generated from pJY7 by the PCR-targeted Streptomyces gene replacement method (66). For this the DNA encoding the polypeptide containing just the loading domain, module 1 and the N-terminal PLM DD in pJY7 was replaced with the kanamycin resistance marker. This marker was amplified from pKD13 (67) with primers PikA1scarF (5’ - GTT GGG TGA AAG CGC GGC TTC CGG AGA CGG AGC CGG ATG ATT CCG GGG ATC CGT CGA CC - 3’) + PikAlscarR (5’ - GCG ACG GGA ACA GGG GTG GTG GCA GAG GAT GTA TAC TCA TGT AGG CTG GAG CTG CTT C - 3’) (bold sequences represent annealing sites for the antibiotic
marker with remaining 5’ regions representing recombination crossover sites). The PCR program consisted of an initial hold at 94 °C, for 2 minutes, followed by 10 cycles of 94 °C for 45 seconds, 50 °C for 45 seconds, 72 °C, for 90 seconds. This was followed by another 15 cycles of 94 °C for 45 seconds, 55 °C for 45 seconds, 72 °C for 90 seconds. This was followed by a final extension period at 72 °C for 10 minutes. Successful replacement in pJY7 was followed by FLP mediated excision of the marker region generating pJY7*, which encodes for the expression of PikAI module 2 with the C-terminal docking domain of Plm2-3 at the N terminus.

2.3.5 Transformation of S. venezuelae

Generated plasmid complementation variants of pDHS722 were individually transformed into the PikA1 deletion strain (BB138) of S. venezuelae by established protocols (68) with slight modifications. BB138 protoplasts were prepared by standard methods. Frozen protoplasts were thawed quickly by hand warming, followed by centrifugation at 3000 rpm for 5 minutes in an Eppendorf Minispin Micro Centrifuge. Following centrifugation the supernatant was gently removed by pipet, leaving approximately 100 µL total volume. To the cells 10 µL (500-600 µg) of the respective plasmid DNA was added, followed immediately by 600 µL of a sterilized 25% 8000 PEG solution in P buffer. P (basic) buffer consisted of 103g sucrose, 0.25g K₂SO₄, 2.02g MgCl₂•6H₂O, and 2mL of trace element solution in 800mL H₂O. Trace element solution was prepared by dissolving 40mg ZnCl₂, 200mg FeCl₃•6H₂O, 10mg CuCl₂•2H₂O, 10mg MnCl₂•4H₂O, 10mg NaB₄O₇•10H₂O, 10mg (NH₄)₆Mo₇O₂₄•4H₂O
in 1L of H₂O. Both solutions were sterilized separately. After sterilization to 80mL of P basic buffer was added 1mL 0.5%KH₂PO₄, 10mL 3.68% CaCl₂·2H₂O, and 10mL 5.73% TES (pH 7.2) to give the final P buffer solution.

The cell mixture solution was gently pipeted up and down three times and plated onto R2YE agar (10.3% sucrose, 0.025 K₂SO₄, 1.1% MgCl₂, 1% glucose 0.01% Difco Casaminoacids, 0.005% yeast extract, 1% trace element solution, 2.2% agar) plates. Plates were allowed to incubate at room temperature overnight and subsequently overlaid with 500µg of thiostrepton in 1mL of LB media. Plates were incubated at 30°C for 4-5 days until single colonies were observed. These colonies were subsequently grown on SPA agar (0.1% yeast extract, 0.1% beef extract, 0.2 % tryptone, 1% glucose, 1.5% agar) plates (50mg mg/L thiostrepton) for 3 days. Spores from these plates were scrapped for SCM liquid (1.5% soluble starch, 2% soytone, 0.01% CaCl₂, 0.15% yeast extract, 1.05% MOPS, pH 7.2) media fermentations and propagation of S. venezuelae strains carrying the desired complementation plasmids (68).

### 2.3.6 Synthesis of diketide NAC thioester of 2(S)-methyl-3(R)-hydroxypentanoic acid

Synthesis of the NAC thioester of 2(S)-methyl-3(R)-hydroxypentanoic acid was completed as detailed in scheme 1 following previously published methods (69, 70). All reactions were carried out under nitrogen using dry solvents under anhydrous conditions, unless otherwise noted. Normal phase flash column chromatography was carried out using Davisil® silica gel (100-200 mesh). Preparative thin-layer
chromatography (PTLC) separations were carried out on 1 mm, or 2 mm E. Merck silica gel plates (60F-254). ¹H NMR spectra were recorded on Tecmag Libra-modified NM-500 MHz or Bruker AC-F 300 MHz spectrometers and calibrated using residual undeuterated solvent as an internal reference. High-resolution mass spectra (HRMS) were recorded on a Micromass LCT Electrospray mass spectrometer performed at the Mass Spectrometry & Proteomics Facility (The Ohio State University).

Scheme 1. Synthesis of NAC diketide thioester. Reagents and conditions: a) DBBT, DIEA, CH₂Cl₂, CH₃CHO; b) TBSOTf, 2,6-Lutadiene CH₂Cl₂; c) LiOH, THF/H₂O (3:1); d) ACHN(CH₂)₂SH, EDCI, DMAP, CH₂Cl₂; e) 48% HF, CH₃CN, H₂O.

Generation of compound 2 (Aldol)

To a stirred solution of Evans’ acyl-oxazolidinone 1 (0.32g, 1.29 mmol) in CH₂Cl₂ (3 ml, 0.3 M) at 0°C were added diisopropylethylamine (1.30 ml, 1.29 mmol) and dibutylboron triflate (1.0 M in CH₂Cl₂, 1.30 ml, 1.29 mmol). The resulting reaction mixture was stirred at 0°C for 30 min and then cooled to -78°C.
Propionaldehyde (0.05 g, 0.86 mmol) in CH₂Cl₂ was added and the mixture was stirred at -78 °C for 1 h and then allowed to warm to 0°C. After stirring for 2 hrs at 0°C the reaction was quenched by the addition of phosphate buffer pH 7 (1 ml). The reaction mixture was poured into a flask containing MeOH (4.3 ml) at 0°C and treated with precooled 30% H₂O₂ (5.4 ml) and stirred at 0°C for 1 hr. MeOH was removed by rotary evaporation, saturated aqueous NaHCO₃ was added and the resultant aqueous layer was extracted with CH₂Cl₂ (3 x 5 ml) and purified by flash silica gel chromatography (silica gel, 10% EtOAc/hexanes) to afford the aldol, 2: ¹H NMR (500 MHz, CDCl₃) δ 7.37-7.20 (m, 5H), 4.74-4.70 (m, 1H), 4.26-4.19 (m, 2H), 3.90-3.86 (m, 1H), 3.82-3.78 (m, 1H), 3.27 (dd, J = 3.5, 13.5 Hz, 1H), 2.88 (m, 1H), 2.80 (dd, J = 10, 13.5 Hz, 1H), 1.65-1.45 (m, 2H), 1.26 (d, J = 7.5 Hz, 3H), 0.99 (t, J = 7.5 Hz, 3H).

**Generation of compound 3 (Protected Aldol)**

To a solution of 2 (1.45 g, 4.97 mmol) in CH₂Cl₂ at 0°C was added 2,6-lutidine (1.73 ml, 14.9 mmol). After stirring for 5 min at that temperature, tert-butylidimethylsilyl trifluoromethane sulfonate (1.71 ml, 7.5 mmol) was added dropwise and the reaction mixture was stirred at 0°C for 20 min after which time no starting material was detected by TLC. Saturated aqueous NH₄Cl was added. The organic phase was separated, and the aqueous layer was extracted with CH₂Cl₂ (3 x 30 ml). The combined organic extracts were dried over anhydrous Na₂SO₄, concentrated and purified by flash silica gel chromatography (silica gel, 1% EtOAc/hexanes) to afford 3: ¹H NMR (500 MHz, CDCl₃) δ 7.36-7.22 (m, 5H), 4.63-4.59 (m, 1H), 4.19-
4.15 (m, 2H), 3.97 (q, \(J = 5.0\) Hz, 1H), 3.91-3.87 (m, 1H), 3.31 (dd, \(J = 3.0, 13.5\) Hz, 1H), 2.77 (dd, \(J = 10.0, 13.5\) Hz, 1H), 1.58-1.55 (m, 2H), 1.21 (d, \(J = 7.0\) Hz, 3H), 0.92-0.88 (m, 12H), 0.03 (d, \(J = 9.5\) Hz, 6H).

**Generation of compound 4 (hydrolysis of chiral auxiliary)**

To a suspension of 3 (0.05 g, 0.123 mmol) in a mixture of THF (0.96 ml) and water (0.33 ml) at 0°C was added LiOH•H₂O (0.10 g, 0.246 mmol) and 30 % H₂O₂ (0.06 ml, 0.49 mmol). The reaction mixture was allowed to stir at 0°C for 1 hr after which it was quenched by addition of sodium sulfite solution (1.5 M). The mixture was concentrated and aqueous phase was washed with dichloromethane and then carefully acidified to pH ~2 with 1N aqueous HCl. The mixture was extracted with diethyl ether and the combined organic extracts were dried over anhydrous sodium sulfate and concentrated *in vacuo*. The crude product was subjected to flash column chromatography (silica gel, 10% EtOAc/hexanes) to give 4: \(^1\)H NMR (500 MHz, CDCl₃) \(\delta\) 3.93 (q, \(J = 5.5\) Hz, 1H), 2.65-2.60 (m, 1H), 1.60-1.47 (m, 2H), 1.14 (d, \(J = 7.5\) Hz, 3H), 0.93-0.90 (m, 12H), 0.09 (d, \(J = 9.5\) Hz, 6H).

**Generation of compound 5 (NAC thioester of 4)**

A solution of carboxylic acid 4 (0.37 g, 1.49 mmol) in anhydrous methylene chloride (7.5 ml) was cooled to 0 °C for 15 min. N-acetyl cysteamine (47 mg, 0.394 mmol) was added to this solution, followed by 4-(N,N-dimethylamino) pyridine (0.21 g, 1.8 mmol), and N-(3-dimethylaminopropyl)-N’-ethylcarbodiimide hydrochloride (0.04 g, 0.30 mmol). The mixture was allowed to warm to room temperature and
stirred overnight. Saturated aqueous NH₄Cl solution (7 ml) was added and the organic phase was separated. The aqueous phase was extracted with 3 x 15 ml of ether and the combined organic phases were dried over anhydrous sodium sulfate and concentrated \textit{in vacuo}. The residue was purified by column chromatography (silica gel, 30\% EtOAc/hexanes) to afford 5: \(^1\)H NMR (500 MHz, CDCl\(_3\)) \(\delta 5.85\) (br, 1H), \(3.91\) (q, \(J = 6.0\) Hz, 1H), \(3.48-3.40\) (m, 2H), \(3.01\) (t, \(J = 6.0\) Hz, 2H), \(2.82-2.77\) (m, 1H), \(1.97\) (s, 3H), \(1.66\) (s, 1H), \(1.57-1.46\) (m, 2H), \(1.17\) (d, \(J = 6.5\) Hz, 3H), \(0.91-0.87\) (m, 12H), \(0.04\) (d, \(J = 9.0\) Hz, 6H).

**Generation of compound 6 (deprotection of 5)**

To a solution of the protected NAC thioester (0.24 g, 0.68 mmol) in acetonitrile (6 ml, 0.1 M) and water (1.3 ml, 0.6M) was added hydrofluoric acid (1.6 ml, 48\% wt in H\(_2\)O). After stirring for 2 hrs at room temperature the reaction was cool to 0°C and pH was adjusted to 7.5 using saturated aqueous NaHCO\(_3\). Acetonitrile was removed by rotary evaporation and the resultant aqueous layer was extracted with EtOAc (3 x 10 ml). The combined organic extracts dried over anhydrous Na\(_2\)SO\(_4\), concentrated and purified by flash silica gel chromatography (silica gel, 10\% MeOH/CH\(_2\)Cl\(_2\)) to afford the final diketide NAC thioester 6: \(^1\)H NMR (500 MHz, CDCl\(_3\)) \(\delta 6.02\) (br, 1H), \(3.85-3.32\) (m, 1H), \(3.48-3.40\) (m, 2H), \(3.07-2.98\) (m, 2H), \(2.76-2.71\) (m, 1H), \(2.61\) (s, 1H), \(1.96\) (s, 3H), \(1.56-1.41\) (m, 2H), \(1.20\) (d, \(J = 7.0\) Hz, 3H), \(0.96\) (d, \(J = 8.0\) Hz, 3H). HRMS calculated for C\(_{10}\)H\(_{19}\)NO\(_3\)S + Na 256.098; found 256.097.
2.3.7 Polyketide production analysis

Polyketide production was analyzed from 10 mL SCM media supplemented with thiostrepton (50 mg/L) fermentations carried out at 30°C with agitation (220 rpm) for 72 hours. After fermentation the medium was separated by centrifugation at 10,000 rpm for 10 min. The fermentation media was extracted 3X with 10 mL of EtoAC, dried by rotary evaporation, and re-suspended in 1mL of MeOH. Extracts of BB138/(pJY7, pJY12, pJY9, pBK51, pBK51* and pJY7*) were analyzed for pikromycin production by reverse phase HPLC on an Agilent 1100 series system with a 5µm Discovery HS C18 reverse phase column (25 x 4.6 cm, Supelco), with a linear gradient system from solvent A (20:80 acetonitrile: water -10mM ammonium acetate) to solvent B (80:20 acetonitrile: water -10mM ammonium acetate) over 50 min at a flow rate of 1 mL/min with UV detection at 220nm. Relative amounts were determined by integrating and averaging the peak areas for pikromycin produced in 3 separate extracts and compared to extracts of BB138/pDHS722. Extracts were also analyzed by LC-MS to correlate the observed retention times with the appropriate masses for both methymycin and pikromycin. LC-MS analysis was carried out using 20 µL of each extract under the same solvent system conditions at a rate of 0.3 mL/min with a Surveyor HPLC system (ThermoFinnigan) connected to a diode array detector equipped with a 2.1 µm Discovery HS C18 reverse phase column (25 x 4.6 cm Supelco). Mass spectra were collected on an LCQ quadrupole ion trap (ThermoFinnigan) mass spectrometer equipped with an electrospray ion source in
positive mode (Oregon Health and Sciences University, Mass Spectrometry core facility).

Diketide feeding experiments to strain BB138 carrying plasmids pBK51, pBK51* or pJY7* were completed by initially starting with an overnight seed culture. 100 µL of this culture (1%) was added to fresh 10 mL SCM fermentations with addition of 50mg mg/L thiostrepton. After agitation for 10 hours at the above conditions, the diketide was added from a stock solution in MeOH to a final concentration of 1mM, (MeOH alone was used a negative control), and the fermentation process permitted to continue for approximately additional 60 hours. Analyses of polyketide production were carried out as described above.

2.3.8 Plasmid confirmation

In order to determine that no cross contamination had occurred and that each transformation of BB138 carried the intended plasmid only, plasmid DNA was isolated from S. venezuelae and confirmed by PCR screening with primers specifically designed for plasmid sequencing (Figure 12). PCR amplification of the region encoding the natural intrapolypeptide linker yields a product that is approximately 400 base pairs while the recombinant pDHS722 variants are approximately 400 base pairs larger, at approximately 800 base pairs.
Figure 12. PCR confirmation of plasmids isolated from *S. venezuelae*. Plasmids pDHS722 and pJY7 were isolated from strain BB138 and used as templates for PCR confirmation along side authentic pDHS722 and pJY7 standards. PCR amplification was completed with specifically generated sequencing primers previously used for plasmid sequencing. Lanes 1 and 2 are PCR products from authentic pDHS722 and pJY7 respectively. Lanes 3 and 4 are PCR products from BB138 purified plasmids pDHS722 and pJY7 respectively. The PCR product of pJY7 is approximately 800 base pairs compared to pDHS722, which is approximately 400 base pairs.
2.4 Results and Discussion

2.4.1 Consequences of substituting linkers for DD in PikAl

Triplicate liquid fermentations of each *S. venezuelae* BB138 carrying pDHS722, pJY12, pJY7 or pJY9 respectively were carried out in order to determine if these three complementation variants produced the expected antibiotic products. Detection and quantification of methymycin and pikromycin products was performed using HPLC analysis combined with UV and MS (m/z = 470 and 526 in positive mode), and confirmed with authentic standards (Figure 13a-c). Production of methymycin and pikromycin were observed in both BB138/pJY12 and BB138/pJY7, and was approximately 50% of that observed for BB138/pDHS722 (Figure 13). However, in the case of BB138/pJY9 where no linker or DD pair is present between modules 1 and 2 only trace amounts of the polyketide products were detected, <0.1% compared to intact PikA1 (pDHS722) (Figure 14). This dramatic decrease in pikromycin yields likely indicates that in the separated modules without DDs, protein recognition between the ACP of module 1 and the KS of module 2 is not sufficient to permit transfer from module 1 to 2. Alternatively, the expression of a soluble separate module 2 may be lower without either the natural linker or in the case of pJY7 an N-terminal DD being present. Indeed, reported work by Khosla et al. on *in vitro* expression of individual PKS monomodule (33, 71-73) and specifically Pik module 2 (74) has shown that an N-terminal stretch of amino acids (linker or DD) is required for the expression and catalytic activity of these proteins. The trace levels of pikromycin
detected in BB138/pJY9 fermentations demonstrate some level of expression of this protein.

Figure 13. HPLC and LC-MS analysis of methymycin and pikromycin production. a) Overlaid HPLC trace for fermentations of BB138 carrying the indicated complementation plasmids. b) Selected ion monitoring for the expected macrolide products methymycin and pikromycin eluting at approximately 15 and 21 min respectively. c) Mass spectra of the products methymycin and pikromycin (m/z = 470 and 526) in positive mode respectively.
Figure 14. Levels of pikromycin products made in the BB138 PikAI deletion strain of *S. venezuelae* with respective complementation plasmids. Red block shapes indicate the Plm1-Plm2-3 docking domain pair from the phoslactomycin system. Diketide feeding experiments were carried out only with BB138 carrying constructs expressing PikAI in which module 1 was either inactive (through mutational inactivation of the KS domain, indicated by an X), absent (pJY7*) or missing the C-terminal DD (pJY9): N/A (not applicable) and N/D (not detectable).
2.4.2 Directed biosynthesis, feeding NAC thioester to module 2

Fermentations of BB138 carrying pBK51, pBK51* or pJY7* did not produce any detectable levels of methymycin or pikromycin, consistent with either loss of the loading domain and module 1 of PikAI, or mutational inactivation of module 1. These fermentations were also performed in the presence of the NAC thioester of 2(S)-methyl-3(R)-hydroxypentanoic acid, the natural diketide intermediate with production monitored by LC-MS (Figure 15a-d).

Figure 15a. Two routes to prime the pikromycin PKS and provide 12 and 14-membered macrocyclic products. The natural process occurs with generation of enzyme-bound propionate starter unit of the loading domain from methylmalonyl CoA. An alternate, less efficient route which can occur in the absence of the loading domain and module 1 or mutational inactivation of module 1 is transfer of the diketide intermediate, 2(S)-methyl-3(R)-hydroxypentanoic acid (activated as a NAC thioester) onto the KS domain of module 2.
Figure 15b. LC-MS analysis of BB138/pBK51 grown in the presences of NAC thioester of 2(S)-methyl-3(R)-pentanoic acid.
Figure 15c. LC-MS analysis of BB138/pBK51* grown in the presence of NAC thioester of 2(S)-methyl-3(R)-pentanoic acid.
Figure 15d. LC-MS analysis of BB138/pYJ7* grown in the presences of NAC thioester of 2(S)-methyl-3(R)-pentanoic acid.
Fermentations were conducted three times; product yields and standard deviation were calculated relative to those obtained with BB138/pDHS722 (Figure 14). Levels of pikromycin production in the cases of BB138/pBK51 (Figure 15b) were extremely low, much less than 1% of that observed in the control culture. This result contrasts with previous experiments, which have demonstrated that for the DEBS PKS, the natural diketide intermediate is able to load onto the second module of a dimodular polypeptide in an efficient manner (75). This observation with DEBS has been reported in a number of hosts and in this laboratory BB138/pBK3 has been used, (76) which produces the truncated DEBS PKS triketide lactone product to show that it also functions similarly in *S. venezuelae*. It has also been shown that feeding the natural triketide intermediate to either BB138 or BB138/pBK51 gives pikromycin yields of 10% of that observed for control (77). The low levels of pikromycin production were increased to about 2% of the control with BB138/pBK51* (Figure 15c), indicating a modest improvement in diketide utilization may result from expressing module 2 as a monomodular polypeptide. Notably, a significant increase was seen with BB138/pJY7* in which the pikromycin yields were about 8% (Figure 15d) of the control and comparable to that observed with feeding the triketide intermediates. Thus removal of the preceding module rather than separation of the two modules is more effective at improving utilization of the diketide intermediate (improvements in utilization of diketides for erythromycin biosynthesis have also been observed upon removal of module 1 of DEBS) (53). This result may indicate that even as separate polypeptide the loading domain and module 1 may negatively impact loading of diketide onto module 2. Alternatively, the process may simply be inefficient.
regardless of the presence of the preceding module and the improvements may reflect better expression of the new monomodular module 2. The efficient utilization of the natural diketide by a separated module 2 was only observed when there was an N-terminal docking domain. There was no improvement in the trace levels of pikromycin made by BB138/pJY9 when grown in the presence the diketide intermediate (suggesting that that the separate module 2 without an N-terminal DD is either catalytically impaired or poorly expressed).

2.5 Conclusion

These experiments have provided an additional in vivo example that a multimodular PKS can be separated into individual modules by replacing the intrapolypeptide linker with a heterologous DD pair. Moreover, these observations demonstrate for the first time that the multimodular separation of these complex multifunctional proteins can lead to only a modest decrease in the overall production of the final polyketide product compared to the intact protein. Additionally, this work has demonstrated that simply separating multimodular PKS proteins without DDs results in a tremendous drop in polyketide production. Such dissected systems that utilize DD interactions, make multimodular PKSs more amenable for the creation of hybrid PKSs and potentially the creation of new polyketide products. However, there are still many challenges associated with a) the appropriate molecular recognition between the ACP and KS domains, and b) the substrate specificity of the catalytic domains. An alternative approach to generating novel polyketide products is feeding NAC thioesters of analogs of polyketide intermediates. This approach also faces the
challenge associated with substrate specificity of the catalytic domains, and inefficiencies of loading onto the KS of the appropriate module. In the case of PikAI it has been shown that the \textit{in vivo} utilization of the natural NAC- diketide intermediate by a multimodular protein is very inefficient, but can be improved significantly by removal of the preceding modules and expression of the appropriate loading module with an N-terminal docking domain.
3. Functional modular dissection of the trimodular DEBS1-TE results in an unexpected shift in triketide lactone ratios and provides new insights into aspects of acyl group loading and hydrolysis

3.1 Summary

The ability of DEBS 1, the first of the three polypeptides from the 6-deoxyerythronolide B (DEBS) polyketide synthase of *Saccharolyspora erythraea* to produce triketide lactone products when the DEBS thioesterase domain is covalently fused to the C-terminal of the protein has been reported in both *in vivo* and *in vitro* experiments. Plasmid-based expression of the intact DEBS1-TE fusion protein in a variety of *streptomyces* hosts, including *S. venezuelae* results in efficient triketide lactone production, where the observed preference is for a product derived from a propionate starter unit. In addition to the propionate-derived compound, acetate, butyrate, and isobutyrate derived compounds have also been observed. Previous results showed that PikAI from the pikromycin/methymycin PKS can be separated at the loading domain-module 1 and module 1-module 2 junctions with a matched pair of docking domains from the phoslactomycin PKS. This separation method was applied to the DEBS1-TE PKS polypeptide to examine if another PKS multimodular protein could be efficiently separated with docking domains. The phoslactomycin docking domain pair was again utilized for this separation of DEBS1-TE as well as another docking domain pair, PikA3-PikA4 from the pikromycin PKS.
This work has shown that dissection of DEBS1-TE at the two module-module junctions also results in only a modest decrease in the overall yields of polyketide products (triketide lactones) when compared to intact DEBS1-TE. However an unexpected shift in triketide lactone ratios favoring the isobutyrate derived product was observed. These observations along with radiolabeling and acyl CoA hydrolysis experiments with the stand alone DEBS loading domain (AT\textsubscript{L}-ACP\textsubscript{L}) have led to new insights into a) the rate of AT acylation and the subsequent hydrolysis of the acyl group off of the AT domain and b) relative rates for the subsequent transfer step of an AT bound acyl group to the holo ACP domain has been investigated.

3.2 Introduction

Much of the information to date regarding structure, function, and domain organization of type I modular polyketide synthases has been derived from genetic and biochemical studies involving the 6-deoxyerythronolide B (DEBS) PKS from *Saccharopolyspora erythraea*, as described in chapter 1. This PKS produces the macrocyclic lactone core of the antibiotic erythromycin. The genes encoding the PKS enzymes responsible for the biosynthesis of DEBS were the first type I PKS cluster to be sequenced in its entirety and as a result has been one of the most well studied PKS systems (78, 11-13). The information, which the DEBS PKS has revealed in regards to functional domains and module organization, has laid the foundation for virtually all analyses of modular PKSs (79). One of the key factors which has made the DEBS PKS such a well studied system is the ability for either all or part of the DEBS
biosynthetic gene cluster to be successfully expressed in several heterologous hosts other than *Saccharopolyspora erythraea*, including *S. venezuelae* (75, 80-83).

One of the unique features of the DEBS PKS is the ability of the loading domain AT to utilize a variety of different starter units. Despite the strict substrate specificity that extension module AT domains typically display (79, 55) PKS loading domain ATs from the DEBS and avermectin systems have relaxed substrate specificity capable of incorporating a variety of starter units (80, 84-86). Although the natural starter unit for erythromycin biosynthesis is propionate, the DEBS AT_L-ACP_L and specifically the AT domain been shown to have a certain amount of flexibility. Other starter units such as acetyl, butyryl, isobutyryl, cylopropionate, and cyclobutyryl have been shown to be processed by DEBS into their final corresponding products, albeit with much lower production levels (55, 56, 80, 84, 85, 87). The observation of novel erythromycin analogs not only means that the AT_L has relaxed substrate specificity but also that the catalytic domains in downstream modules to some extent do as well. This relaxed processing capacity has made DEBS one of the most genetically engineered and manipulated PKSs. Despite the seemingly broad substrate flexibility, which DEBS exhibits, not all down-stream PKS modules have the ability to process a variety of different substrates. The inefficiency of alternate substrate processing and the presumed blockage of PKS protein active sites is one of the major hurdles for genetic engineering, combinatorial, and precursor-directed biosynthetic efforts.

One possible mechanism to remove acyl intermediates, which can not be properly processed has been attributed to the activity of specific thioesterases, which have been proposed to act as housekeeping or editing enzymes (88-90). Genes
encoding an additional thioesterase enzyme to the one generally required for the release of the final polyketide product have been reported in the PKS biosynthetic clusters of many of many polyketide pathways; such as erythromycin from *Saccharopolyspora erythraea* (91), rifamycin from *Amycolatopsis mediterranei* (92), tylosin from *Streptomyces fradiae* (93), and pikromycin/methymycin from *Streptomyces venezuelae* (51). These enzymes hydrolyze acyl groups of mis-acylated ACP domains by cleavage of the acyl group from the 4’-phosphopantetheine prosthetic arm of the ACP domain (94, 95) (Figure 16a). Improperly acylated ACP domains can arise from the premature decarboxylation of the selected extender unit, resulting in an acyl-ACP which blocks polyketide biosynthesis and must be removed for the biosynthetic process to proceed. *In vitro* evidence from the tylosin (96) and Pik (94) PKS systems has shown that the TEII enzyme for each respective system, can process the removal of acyl groups bound to ACP domains. However these enzymes possess higher affinities for acyl groups, which lack the carboxyl group needed for polyketide decarboxylative chain elongation.
Figure 16. Proposed mechanisms for the removal of acyl groups from DEBS AT$_1$-ACP$_1$. Proposed mechanisms for removal of loaded acyl groups a) TE cleavage of ACP bound acyl intermediate b) water induced hydrolysis of AT bound acyl intermediate. c) Proposed dynamic equilibrium of an acyl group between the AT and ACP domains of a PKS.

Published work has shown that when the DEBS TE domain is covalently fused to the C-terminal of DEBS1, this fusion protein is capable of loading a starter unit and catalyzing two rounds of elongation before the terminal TE domain catalyzes the formation of the predicted triketide lactone (TKL) product (54, 80, 84, 97) (Figure 10b). Based on the similarities in the modular organization of DEBS1 and PikAI from the Pik PKS as well as an established method to screen for triketide lactone production, it was sought to expand on the successful results from the experiments described in chapter 2, separating multimodular PKS proteins with matched docking domains. The goals of this work were to determine whether or not replacing the natural intrapopolypeptide linkers in DEBS1 between the loading domain and module 1 and module 1 and module 2 with the Plm1-Plm2 DD pair would again result in an efficient dissected PKS system. This work was accomplished by generating variants
of expression construct pBK1 (76), which expresses DEBS1-TE and produces acetate, propionate, and isobutyrate derived triketide lactone, when expressed in *S. venezuelae*. In addition to the Plm DD pair, the impact on production when another set of docking domains, PikA3-A4 DD from the pikromycin PKS was examined by insertion of this pair between the first and second extension module of DEBS1. This particular set of DDs, for which a crystal structure (Figure 17) (PDB ID 3F5H) was recently published, it was selected based on intriguing binding data showing that this pair exhibited the highest binding affinity for one another when compared to the other DEBS and Pik docking domain pairs (40).

![Figure 17. PikA3-A4 PKS docking interface structures. a) Highlighted PikA3 ACPdd/PikA4 KSdd pair (PDB id 3F5H). b) Three PikA3/PikA4 docking dimers are shown. The docking interaction formed by neighboring dimers is highlighted. (Reproduced with permission from (48) License Number: 2427290325871)](image-url)
By using the Plm docking domain pair previously used in the Pik system (chapter 2) to dissect the trimodular DEBS1-TE polypeptide, a dramatic change in the ratio of triketide lactone products was observed. Investigations into this phenomenon have revealed a second mechanism for the removal of acyl groups from the loading di-domain, which occurs via hydrolysis of acyl groups bound to the AT domain (Figure 16b). In this proposed mechanism an acyl group bound to the active site serine of an AT domain can be hydrolyzed yielding the carboxylic acid of the bound acyl group and the unmodified serine residue of the AT domain. The data from this work demonstrates that different acyl groups such as propionate and isobutyrate transfer from the AT to the ACP at different rates and that the AT to ACP transfer is reversible.

For a loading domain AT\textsubscript{L}-ACP\textsubscript{L}, the position of the equilibrium between either the AT or ACP bound form (Figure 16c) depends on two factors that are important for the normal polyketide biosynthetic process; one is the rate of transfer for acyl groups bound to the AT domain to the \textit{holo} ACP domain and the second is the rate of transfer to the subsequent KS domain of the first extension module. This equilibrium between the AT and ACP domains determines the levels of acyl groups which are either cleaved from the ACP domain by thioesterases or hydrolyzed from the AT domain.
3.3 Materials and Methods

3.3.1 Commercially purchased materials

All chemicals and antibiotics were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Ipswich, MA). PCR primers were custom synthesized from Integrated DNA Technologies (Coralville, IA). PCR reactions were carried out with the GC Rich PCR System from Roche (Basel, Switzerland).

3.3.2 Plasmids, strains and culture conditions

All constructed plasmids were derivatives of pBK1, a DEBS1+TE expression plasmid (76). Construct pBK1 was derived from pSE34, a multicopy shuttle plasmid containing the \textit{ermE}* promoter, provided by Pfizer (Groton, CT). All plasmid manipulations were completed in \textit{E. coli} hosts. Transformants were cultivated on LB/agar media and plasmids isolated from LB liquid media, both supplemented with appropriate antibiotics. Generated variants of the pBK1 were transformed into BB138 or AX912 by established protocols (68) onto R2YE plates. Transformants were subsequently grown on SPA agar plates (50mg mg/L thiostrepton). SCM liquid media was used for fermentation and propagation of \textit{S. venezuelae} strains.

3.3.3 DNA manipulation

Variants of pBK1 were prepared by a series of PCR reactions and subsequent PCR targeted recombination crossovers in \textit{E. coli}, before final transformation into \textit{Streptomyces venezuelae}. Previously reported construct pJY2 (58) was used as a
template for the amplification of PCR products containing the encoding region for the apramycin resistance marker and the Plm DD pair along with the needed 39 nt complementary to the desired point of recombination between the linked ACP and KS domains of DEBS1. PCR products for the exchange of DNA sequences encoding the intramodular linkers between the LD-module 1 and modules 1-2 of DEBS1 for that coding for the Plm docking domain pair, were amplified with primer pairs DEBS LD sep F (5’ – CAC CCG ACT CCG CGT GCG CTC GCG GAA GCA CTC GCG GCG CTG GGG CAC CTC GAT TTC GG - 3’) + DEBS LD sep R (5’ – GGG CAG CCG GCA GGC CAT CGC GAC GAC CGC GAC CGG TTC GTG GTG CTT CGT CTC GGC GT - 3’) and DEBS M1-M2 sep F (5’ – GAC GTG CGG ACG CTG GCG GCG CAC CTG GCC GCC GAA CTC CTG GGG CAC CTC GAT TTC GG - 3’) + DEBS M1-M2 sep R (5’ – GGGC AGC CGG CAC GCC ATG CCG ACG ATC GCG ATC GGC TCG GTG GTG CTT CGT CTC GGC GT - 3’) respectively. The primer sequences in bold/italic represent annealing sites on the template pJY2 and remaining 5’ region the needed complementary 39nt of the DEBS1 region for recombination.

The PCR program consisted of an initial hold at 94 °C, for 2 minutes, followed by 15 cycles of 94 °C for 45 seconds, 54 °C for 30 seconds, 72 °C, for 90 seconds. This was followed by another 15 cycles of 94 °C for 45 seconds, 58 °C for 30 seconds, 72 °C for 90 seconds. This was followed by a final extension period at 72 °C for 10 minutes. PCR products were introduced to pBK1 by homologous recombination via established protocols (66). Recombinant pBK1 plasmids were isolated as before, treated with Bstz711 to remove the apramycin marker, ligated overnight at 12°C, transformed into
E. coli. Single colonies were screened for both (+) Amp and (-) Apra, yielding pBK1* and pBK1** respectively.

Plasmid pBK1* scar, which expresses only modules 1 and 2 of DEBS1 was generated from pBK1* by the PCR-targeted Streptomyces gene replacement method (66). For this the DNA encoding the polypeptide containing the loading domain and the N-terminal PLM DD in pBK1* was replaced with the kanamycin resistance marker. This marker was amplified from pKD13 (66) with primers DEBS LD scar F (5' – GGG CAG CCG GCA CGC CAT GCC GAC GAT CGC GAT CGG CTC GATT CCG GGG ATC CGT CGA CC - 3') + DEBS LD scar R (5' - GCG ACG GGA ACA GGG GTG GTG GCA GAG GAT GTA TAC TCA TGT AGG CTG GA GCTG CTT C - 3') (bold/italic sequences represent annealing sites for the antibiotic marker with remaining 5' regions representing recombination crossover sites). The PCR program consisted of an initial hold at 94 °C, for 2 minutes, followed by 10 cycles of 94 °C for 45 seconds, 50 °C for 45 seconds, 72 °C for 90 seconds. This was followed by another 15 cycles of 94 °C for 45 seconds, 55 °C for 45 seconds, 72 °C for 90 seconds. This was followed by a final extension period at 72 °C for 10 minutes. Successful replacement in pBK1* was followed by FLP mediated excision of the marker region generating pBK1* scar, which encodes for the expression of DEBS1 module 1 with the C-terminal docking domain of Plm2-3 at the N-terminus.

Expression construct pJY105 which expresses DEBS1-TE with the PikA3-A4 DD pair in place of the intrapolypeptide linker between module 1 and module 2 was generated in a manner analogous to the above mentioned complementation constructs. The PikAIII-PikAIV DD was amplified with primers PikA3A4 DD For (5’-
CACGAGGCAC- 3’) + PikA3A4DD Rev (5’- 
CCGGATTCCGCCGCGAA – 3’) from template pDHS0031 (D.H. Sherman, 
unpublished). The PCR program consisted of an initial hold for 2 minutes at 94°C, 
followed by 30 cycles of 94 °C for 30 seconds, 58 °C for 30 seconds, 72 °C for 1 
minute. This was followed by a final extension period at 72 °C for 5 minutes. The 
PCR products were purified and cloned into TOPO GW cloning kit (Invitrogen) 
yielding pJY101. The apramycin aac(3)IV resistance marker was amplified from 
pIJ773 (66), using primers PikA3A4DD apra rec For (5’ - GCC CTG ATC CGG ATG 
GCT CTC GGC CCC CGT AAC ACC TGA GTA *TAC GGT TCA TGT GCA 
GCT CCA TC – 3’) + PikA3A4 apra rec Rev (5’ - TGC GCG GGA TGC GCG GGG 
TGC GGC GCG TGG GGC AGG ACC GTA *TAC TGT AGG CTG GAG CTG 
CTT C – 3’), (bold sequences represent annealing sites for the antibiotic marker and * 
indicates the locations of Bstz17I restriction sites introduced for later removal of this 
marker). The PCR program consisted of an initial hold at 94 °C, for 2 minutes, 
followed by 10 cycles of 94 °C for 45 seconds, 55 °C for 30 seconds, 72 °C, for 1 
minute. This was followed by another 15 cycles of 94 °C for 45 seconds, 58 °C for 30 
seconds, 72 °C for 90 seconds. This was followed by a final extension period at 72°C 
for 10 minutes. The resulting PCR product was again recombined as described above 
with pJY101 via established procedures. The resulting recombinant plasmid pJY102 
was isolated with the GeneJET Plasmid Miniprep Kit (Fermentas) and confirmed by 
sequencing. Construct pJY102 was used as a template to generate a PCR product with 
the needed 39nt complementary to the desired point of recombination between the
linked ACP module 1 and the KS of module 2 of DEBS1. PCR products for the exchange of DNA sequences encoding the intramodular linkers between modules 1 and 2 for that coding for the PikAIII/PikAIV DD pair, were amplified with primer pairs DEBS A3-A4 rec For (5' - CAC CCG ACT CCG CGT GCG CTC GCG GAA GCA CTC GCG GCG CAC GAG GCG TAC CTC GCA CC - 3') + DEBS A3-A4 rec Rev (5' - GGG CAG CCG GCA GGC CAT CGC GAC GAC CGC GAC CGG TTC TCC TGC CGA CGG TCG GCC CGG - 3') (primer sequences in bold italicized font represent annealing sites on the template pJY102 and remaining 5' region represent the complementary 39nt of the DEBS1 region for recombination).

The PCR program consisted of an initial hold at 94 °C for 2 minutes, followed by 10 cycles of 94 °C for 45 seconds, 52 °C for 45 seconds, and 72 °C for 90 seconds. This was followed by another 15 cycles of 94 °C for 45 seconds, 58 °C for 45 seconds, and 72 °C for 90 seconds. This was followed by a final extension period at 72 °C for 10 minutes. PCR products were gel purified and introduced to pBK1 by homologous recombination via established protocols. Recombinant pBK1 plasmids, were isolated as before, treated with Bstz71I to remove the apramycin marker, ligated overnight at 12°C, transformed into E. coli. Single colonies were screened for both (+) Amp and (-) Apra, yielding pJY105.

3.3.4 Triketide lactone production analysis

Triketide lactone production was analyzed by a series of 10 mL SCM media fermentations with the addition of thiostrepton (50 mg/L) of either BB138 or AX912 transformed with the appropriate DEBS1-TE expression plasmids (pBK1, pBK1*,
pBK1**, pBK1* scar and pJY105). Fermentations were carried out at 30°C with agitation (220 rpm) for 72 hours. After fermentation the entire culture media was extracted 3X with 10 mL of ethyl acetate and dried. Extracts were treated with 200 µL of Trisil-Z (Pierce – Thermo Scientific) derivatization reagent for 20 minutes at 60°C to form their trimethyl silyl esters (TMS) derivative. Reactions were quenched with 1 mL of water and allowed to cool to room temperature. The reaction was extracted 3X with 1mL CHCl₃. The collected CHCl₃ extracts were concentrated to a final volume of 0.5 mL. Derivatized extracts of BB138 or AX912 carrying the appropriate plasmids were analyzed by GC-MS, on an Agilent Technologies 6890N GC system along with an Agilent Technologies 5973 mass selective detector. Triketides were identified by characteristic fragment ions at m/z = 73, 115, 130 and M⁺ (dependent on starter unit).

Feeding experiments with cyclopropane, or cyclobutane carboxylic acids to S. venezuelae carrying plasmid pBK1 or pBK1* were completed by first preparing an overnight seed culture under the above described fermentation conditions. 100 µL of this culture (1%) was added to fresh SCM media fermentations supplemented with thiostrepton (50mg/L) and allowed to grow under normal conditions for approximately 10 hours. Carboxylic acids were added to a final concentration of 4mM and allowed to continue growing under the previously described conditions.

### 3.3.5 DEBS AT₅-ACP₅ expression and purification

The DEBS AT₅-ACP₅ didomain was over expressed as a C-terminal Hisₖ-tagged protein and purified from *E. coli* using the previously reported expression
plasmid pBK12 (94). Expression was carried out in either BL21 DE3 chemically competent cells (Invitrogen) or BAP1 cells. BAP1 E. coli cells are a derivative of BL21 DE3 cells, which contains a chromosomally integrated copy of the sfp gene encoding the requisite 4'-phosphopantetheinyl transferase, required for posttranslational modification of the ACP domain to the holo form. BAP1 E. coli cells were provided as a gift from Professor Chaitan Khosla’s laboratory, Stanford University.

Overnight E. coli seed cultures harboring pBK12 were grown in LB medium supplemented with 100 µg/mL ampicillin at 37°C with agitation. 1000 mL LB medium cultures supplemented with ampicillin (100 µg/mL) were inoculated with 5% (v/v) of the overnight seed culture and grown at 37°C with agitation to an O.D of approximately 0.6 (approximately 2.5 hours). Cultures were allowed to cool to room temperature and expression was induced by the addition of 0.5 mM isopropyl thio-β-D-galactoside (IPTG). Cultures were allowed to grow for an additional 16 hours at room temperature with agitation. Cells were harvested by centrifugation and either stored at -80°C or immediately lysed for protein purification. His-tagged protein was purified following the protocol outlined in Qiagen’s QiAexpressionist handbook under native conditions with the following noted modifications. E. coli cells were re-suspended in 25 mL of BugBuster Protein Extraction Reagent (Novagen). Cells were disrupted by sonication and allowed to incubate on ice with slow agitation for at least 1 hour. The mixture was then subjected to centrifugation and the soluble fraction (supernatant) was incubated with 1 mL of Ni-NTA agarose (Qiagen) on ice with gentle agitation for at least 1 hour. The mixture was then transferred to a Polyprep
Chromatography Column (BioRad) and allowed to equilibrate at 4°C. The cell free extract (CFE) was drained and the Ni-NTA resin washed with 10 column volumes of wash buffer 1 (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM imidazole, pH 8.0) followed by 10 column volumes of wash buffer 2 (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 20 mM imidazole, pH 8.0). His-tagged protein was eluted from the Ni-NTA resin with 3mL of elution buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 250 mM imidazole, pH 8.0). 5 elution fractions were collected containing between 500-600 µL total volume. Each of the 5 fractions along with the initial load, column flow through and the two column washes were analyzed by SDS-PAGE gel electrophoresis using NuPAGE Novex 4-12% Bis-Tris gels (Invitrogen). Expression and purification was confirmed by the presence of a large band between 50-75 kDa, DEBS AT$_L$-ACP$_L$ has a molecular weight of approximately 59 kDa. The target protein was further purified to homogeneity by size exclusion chromatography using an AKTA FPLC System (GE Health care, formally Amersham Biosciences) equipped with a Superdex GL 200 sizing column and a UV detector set at 280nm. Fractions containing the protein were confirmed by SDS PAGE gel electrophoresis and concentrated with Amicon Ultra Centrifuge columns (Milipore) and stored at -80°C for future use.

3.3.6 Radioactive labeling of DEBS AT$_L$-ACP$_L$

DEBS LD $^{14}$C radiolabeling experiments were carried with both apo and holo AT$_L$-ACP$_L$ using $^{14}$C propionyl and $^{14}$C isobutyryl CoA substrates purchased from American Radiolabeled Chemicals. Both materials were supplied with specific
activities of 50 μCi/μmol. DEBS LD (4 μM) was incubated under four different sets of conditions: a) $^{14}$C propionyl CoA (100 μM) + unlabeled propionyl CoA (100 μM), b) $^{14}$C propionyl CoA (100 μM) + unlabeled isobutyryl CoA (100 μM), c) $^{14}$C isobutyryl CoA (100 μM) + unlabeled isobutyryl CoA (100 μM), d) $^{14}$C isobutyryl CoA (100 μM) + unlabeled isobutyryl CoA (100 μM), all in 100 mM NaH$_2$PO$_4$, 1 mM TCEP, 1 mM EDTA, 20% glycerol, pH 7.4. The extent of protein labeling was determined by removing 25 μL aliquots at time points t= 0, 5, 7.5, 15, and 30 minutes. Protein samples were precipitated with ice cold 10% TCA (400 μL) supplemented with 40 μg of BSA, acting as a carrier protein. Samples were centrifuged at 14,500 rpm (14,000g) for 10 minutes, washed with another portion of 10% TCA (200 μL) and centrifuged for an additional 5 minutes. Precipitated protein samples were re-suspended in a solution of 2% SDS, 20 mM NaOH (50 μL) for liquid scintillation counting.

3.3.7 Measurement of acyl CoA loading and hydrolysis by monitoring free CoA formation

By monitoring the rates of protein dependent free CoA formation with different acyl CoA substrates (acetyl, propionyl, isobutyryl) $k_{cat}$ and $K_m$ values were determined for the apo form of DEBS AT$_{L}$-ACP$_{L}$. Free CoA formation was monitored via a fluorescence assay using the commercially available thioglo reagent from Covalent Associates Inc (Corvallis, OR) (Figure 18). Thioglo is a commercially
available reagent, which derivitizes all free thiol containing compounds and fluoresces when excited at 379nm.

Determination of $k_{cat}$ and $K_m$ values were completed in 3 sets of duplicate, values averaged and standard error calculated. A mixture containing of 1.5 µM protein was incubated at room temperature with varying concentrations of each individual acyl CoA substrate (acetyl, propionyl, butyryl, isobutyryl) in a buffer containing 100 mM NaH$_2$PO$_4$, 0.2 mM TCEP, 1 mM EDTA, 20% glycerol, at pH 7.4. 25 µL aliquots were removed after 30 minutes of incubation and placed in individual wells of a 96 well plate pre-incubated on ice. The reactions were quenched by the addition of 75 µL of DMSO. After all samples were collected and quenched, 100µL of a 200 µM Thioglo stock solution (in DMSO) was added and the plate allowed to incubate at room temperature with gentle agitation for approximately 25 minutes. Alongside each set of experiments a standard curve was generated by preparing 0, 25, 50, 100, 200, and 400 µM free CoA in the same assay buffer followed by the same work up procedure as described above. Excitation at 379 nm and emission monitoring between 400 and 540 nm was completed with a Gemini EM Microplate Spectrofluorometer (Molecular Devices), equipped with Softmax Pro software. Maximum absorbance signal values were observed at 480 nm and were thus used to correlate to a molar value of free CoA generated for each individual data point.
Figure 18. Selective fluorescent properties of Thioglo melamide reagents. Only after selective free thiol addition to the melamide ring does the conjugated aromatic ring system exhibit fluorescence.

3.4 Results and Discussion

3.4.1 Substitution of Plm DD for intrapolyptide linkers in DEBS1

Fermentations of BB138 harboring pBK1, pBK1*, pBK1**, and pBK1* scar (Figure 21) were completed in at least two sets of triplicate to determine relative quantities of triketide lactone (TKL) production. TKL production was confirmed and quantified by GC-MS analysis. TKLs derived from acetate, propionate, and isobutyrate were identified by their expected retention times and m/z values 157 (7.9 min), 171 (8.3 min), and 185 (8.5 min) respectively, corresponding to published data for the expression of pBK1 in S. venezuelae (76) (Figure 19). In addition to each individual M⁺ ion, TKLs were also identified by characteristic fragment ions with m/z
values of 73 and 115 (Figure 20 a-c). Relative ratios of acetate, propionate, and isobutyrate derived TKLs were determined by integration of the extracted ion chromatogram GC peaks for each of the possible TKLs. Total amounts of TKLs produced by the expression of the two separated constructs pBK1* and pBK1** were 45±15 and 40±4 % respectively when compared to the observed production by the expression of pBK1 in BB138 (Figure 21). This approximate 50% reduction in total production was comparable to the production level decrease observed in the dissected pikromycin system described in chapter 2. This evidence has provided further evidence that multimodular PKS polypeptides can be separated with DDs without a significant drop in overall production.
Figure 19. GC-MS total ion chromatogram of an extract of *S. venezuelae* (BB138) carrying pBK1. Expected TKL lactone products derived from acetate, propionate, and isobutyrate elute at 7.9, 8.3 and 8.5 minutes respectively.
Figure 20a. Ions associated with acetate-derived TKL. Labeled m/z values of 157, represent the acetate-derived TKL molecular ion, while 73 and 115 represent characteristic fragment ions.
Figure 20b. Ions associated with propionate-derived TKL. Labeled m/z values of 171 represent the propionate-derived TKL molecular ion, while 73 and 115 represent characteristic fragment ions.
Figure 20c. Ions associated with isobutyrate-derived TKL. Labeled m/z values of 185 represent the isobutyrate-derived TKL molecular ion, while 73 and 115 represent characteristic fragment ions.
Figure 21. Triketide lactone production levels for DEBS1-TE expression constructs pBK1, pBK1*, pBK1** and pBK1* scar expressed in BB138.

Percentage values for each of the observed acetate (A), propionate (P), or isobutyrate (IB) derived TKL products were determined by integration of GC peaks from the extracted ion chromatogram of each product. Total TKL percentage values for fermentations of BB138 carrying pBK1*, PBK1** and pBK1* scar are relative values compared to that produced by intact DEBS1-TE (pBK1).
3.4.2 Expression of pBK1* in BB138 results in a shift in TKL ratios

Published reports have shown that the loading domain of DEBS possess some relaxed substrate specificity; however in these reports the preferred starter unit has consistently remained propionate \((54, 80, 84, 97)\). Based on this previous work, the dramatic shift from the propionate-derived product in favor of the isobutyrate derived TKL in fermentations of BB138 harboring pBK1* (Figure 22) is an intriguing unexpected observation.

![Abundance](image)

**Figure 22.** GC-MS total ion chromatogram of an extract of *S. venezuelae* (BB138) carrying pBK1*. Expected TKL lactone products derived from acetate, propionate, and isobutyrate elute at 7.9, 8.3 and 8.5 minutes respectively.
Published evidence has shown that strains of *Saccharopolyspora erythraea* where DEBS ATₐ₋ACPₐ has been deleted still retain the ability to produce erythromycin, albeit at a much diminished level (98). This evidence of direct KS domain loading made it essential to determine if the observed shift in TKL ratios arose through the change in expression of DEBS ATₐ₋ACPₐ as a stand-alone protein and not simply the direct loading of DEBS KS1 with isobutyrate. This was achieved by constructing an ATₐ₋ACPₐ deletion variant of pBK1, pBK1* scar (analogous to pJY7* scar described in chapter 2) with the C-terminal Plm DD fused to the N-terminal of KS1. Fermentations of BB138 caring pBK1* scar, which expresses DEBS1-TE without the ATₐ₋ACPₐ retained the ability to produce TKLs at approximately 2% of that observed for intact DEBS1-TE. Furthermore, nearly 90% of the observed TKL products were propionate-derived (Figure 23). This observation confirms previous published data that mutant strains lacking the ATₐ₋ACPₐ can still function and produce polyketide products, however with a significant drop in overall production (98). These observations indicate two additional points: a) that the KS of DEBS module 1 has the ability to be directly acylated by acyl CoAs in the absence of the naturally interacting ACP domain, and b) the KS domain has a strong preference for accepting and processing propionate into the final product. This observation confirms that the majority of the isobutyrate-derived TKL observed in fermentations of BB138/pBK1* can be attributed to the activity of the separated DEBS ATₐ₋ACPₐ, presenting the isobutyrate starter unit to the KS of module 1 for TKL production.
Figure 23. Extracted ion chromatogram of BB138 fermentation carrying expression construct pBK1* scar. Ions with m/z values of 157, 171, and 185 corresponding to acetate, propionate and isobutyrate derived TKLs were selectively extracted from the total ion chromatogram. Propionate derived TKL with the expected retention time of 8.3 minutes was the only major product produced by the expression of this AT_L-ACP_L deletion construct. Due to the low levels of TKL production with the expression of pBK1* scar, the ions associated with each TKL were only detectable under selected ion monitoring conditions.
3.4.3 Impact on TKL production for the module1-module 2 separation

In order to thoroughly analyze the impact on polyketide production when the DEBS1-TE multimodular polypeptide is separated with matched docking domains, this polypeptide was also separated at the module 1-module 2 junction yielding expression construct pBK1**. This construct expresses DEBS1-TE with the Plm DD pair at the module1-module 2 junction and was generated in an analogous fashion to pJY7 described in chapter 2. As expected, expression of pBK1** in BB138 resulted in 40±4% total TKL production compared with pBK1 as observed by GC-MS analysis. Expression of DEBS1-TE now separated with the PLM DD pair between modules 1 and 2 resulted in a predominance in the propionate derived TKL (59%) comparable to the relative ratios observed with the expression of intact DEBS1-TE (67%) (Figure 24). The percentage of the isobutyrate derived TKL was 26% of the total TKLs for the expression of pBK1, lower than the 67% observed with the expression of pBK1* but higher than the 11% observed with pBK1 expression.
Figure 24. GC-MS total ion chromatogram of an extract of *S. venezuelae* (BB138) carrying pBK1**. Expected TKL lactone products derived from acetate, propionate, and isobutyrate elute at 7.9, 8.3 and 8.5 minutes respectively. Although the peaks associated with the propionate (8.3 minutes) and isobutyrate (8.5 minutes) derived TKL appear equivalent, selected ion monitoring and integration for m/z values of 171 and 185 for the respective TKLs showed at least a two fold larger production of propionate versus isobutyrate-derived TKL.
3.4.4 Expression of DEBS1-TE constructs in AX912

Expression of pBK1 and the above described variants (pBK1* and pBK1**) in *S. venezuelae* (BB138) results in the efficient and predictable production of TKL products. However the dramatic shift in TKL production ratios for the expression of pBK1*, when the AT$_L$-ACP$_L$ is expressed as a stand-alone protein, separate from the first and second extension module was not expected. An initial hypothesis to explain this shift, involved the activity of the pikromycin thioesterase II (Pik TE II), which is encoded with the other Pik PKS genes in *S. venezuelae* (BB138). Previously published data on the Pik TEII enzyme has assigned its probable role as a housekeeping enzyme. This work provided *in vitro* evidence that Pik TEII is capable of cleaving specific acyl groups from both DEBS AT$_L$-ACP$_L$ and PikAIII (Figure 5a) with varying efficiencies depending on the attached acyl group (94). A delayed transfer of the starter unit from the ACP$_L$ to KS1 of DEBS1-TE might permit preferential removal of a propionyl group from ACP$_L$, giving rise to the increase in isobutyrate-derived TKLs. To test this hypothesis, the above-described DEBS1-TE expression constructs were transformed and expressed in AX912, a Pik TEI and TEII deletion strain of *S. venezuelae* (65). TKL production was again screened for as before with fermentations of BB138 using GC-MS analysis. Again expression of pBK1 resulted in a predominance of the propionate derived TKL (61% of total TKLs) (Figure 25), while expression of pBK1* resulted in a nearly an equal ratio of the three observed TKLs (Figure 26). A summary of triketide lactone production for the expression constructs pBK1, pBK1* and pBK1** are summarized in Figure 27. As
the isobutyrate-derived TKLs still predominate, a factor other than Pik TEII must account for the shift in TKL ratios.

Figure 25. GC-MS total ion chromatogram of an extract of *S. venezuelae* (AX912) carrying pBK1. Expected TKL lactone products derived from acetate, propionate, and isobutyrate elute at 7.9, 8.3 and 8.5 minutes, respectively.
Figure 26. GC-MS total ion chromatogram of an extract of *S. venezuelae* (AX912) carrying pBK1*. Expected TKL lactone products derived from acetate, propionate, and isobutyrate elute at 7.9, 8.3 and 8.5 minutes, respectively.
Figure 27. Triketide lactone production levels for DEBS1-TE expression constructs pBK1, pBK1*, and pBK1** expressed in AX912. Percentage values for each of the observed TKL products acetate (A), propionate (P), or isobutyrate (IB) derived were determined by integration of GC peaks from the extracted ion chromatogram of each product. Total TKL percentage values for fermentations of BB138 carrying pBK1* and pBK1** are relative values compared to intact DEBS1-TE (pBK1).
3.4.5 PikAIII-PikAIV DD: PJY105 expression in YJ004

In addition to the Plm DD pair that has been utilized to separate multimodular PKS polypeptides, an additional pair that is naturally present between PikAIII and PikAIV of the Pik PKS was also used to separate DEBS1-TE. Construct pJY105 was designed to express DEBS1-TE in the same exact manner as pBK1*, except that the DD pair from between PikAIII and PikAIV substituted for the intrapolypeptide linker between the loading domain and module 1. Initially expression of this construct was completed, as before in the PikA deletion strain BB138, however fermentations generated no detectable TKL products. One possible explanation, was that there maybe an unintended docking interaction between the PikAIII-PikAIV DD pair now present between the loading domain and module 1 of DEBS 1-TE and modules 5 and 6 of the Pik PKS (Figure 5a), since the encoding region for PikAIII and PikAIV are present on the chromosomal DNA of BB138. Potential docking interactions between DEBS AT\(_L\)-ACP\(_L\) and module 6 of Pik and module 5 of Pik with DEBS module 1, might lead to the loss of TKL production. Given this possibility pJY105 was alternatively expressed in a PikAIII-PikAIV deletion strain of *S. venezuelae*, YJ004 (65), alongside pBK1 as the positive control. The expression of pJY105 in YJ004 resulted in trace levels of TKL production, which were approximately 1% of the wild type control experiment (expression of intact DEBS1-TE in the same strain of *S. venezuelae*).

This result using the PikAIII-AIV DD pair is in stark contrast to the approximately 50% production that is retained when the Plm DD pair were utilized. The drastic drop in production for the expression of pJY105 compared to pBK1* and
pBK1** supports the conclusions from chapter 2 (lack of production for the
expression of pJY9, where no docking domains are present) that docking domain are
important but also infers that the specific docking domains are necessary to retain the
ability of dissected PKS polypeptides to produce polyketide products. The basis for
the ability of one DD pair to provide the necessary protein-protein interactions for the
passage of polyketide intermediates over another is not clearly evident. Based on
published data specifically describing the PikA3 and PikA4 docking interactions (40),
it has been shown that the binding specificity which docking domains display for one
another can be increased dramatically by the presence of the extended downstream
protein sequences, in this case the entirety of Pik module 6. This data supports the
notion that docking domains are important for the proper protein-protein interactions
required for polyketide biosynthesis; however there are additional intrinsic
specificities which direct a module to properly interact with its cognate partner in the
biosynthetic process which lie in the downstream KS-AT region that has yet to be
determined (99, 100). Given these data, there are two possible explanations for the
lack of production when this set of DDs were used to dissect DEBS1-TE. One
possible explanation is that this DD pair perturbs the natural ACP-KS interactions,
which play key roles in the proper module-module interactions. A second possible
explanation is that the N-terminal DD which has been fused upstream of the KS
domain does not facilitate the adequate expression and proper folding of a functional
protein for polyketide production.
3.4.6 Incorporation of cyclic unnatural starter units

Previous feeding experiments with the DEBS PKS have shown that this system has the ability to not only incorporate a wide variety of naturally available starter units but is also able to utilize certain unnatural cyclic acyl groups when fermentation cultures are supplemented with the corresponding carboxylic acids (56). This observation was the result of a control experiment for a modified DEBS producing strain of *S. coelicolor*, where the loading domain for the production of avermectin from *S. avermitilis* was exchanged with the DEBS loading domain. It was understood that the avermectin loading domain could process a wide range of starter units including cyclic and branched chain acyl groups (86). The goal of the loading domain exchange was to utilize this loading flexibility to generate novel erythromycin analogs. While their experiment with this hybrid system was successful in generating novel erythromycin analogs, when cyclopropane and cyclobutane carboxylic acids were feed to the control experiment (wild type DEBS system) the same cyclic starter unit derived erythromycin analogs were unexpectedly observed.

Guided by the observation that the DEBS PKS is able to utilize some cyclic starter units, it was examined if feeding to fermentations of *S. venezuelae* expressing the separated DEBS AT_1-ACP_1 from module 1 would result in a higher ratio of the novel cyclic acyl group derived TKL products than the intact DEBS1-TE protein. This possibility was explored because the expression of the dissected DEBS1-TE construct (expression of pBK1*) resulted in unexpectedly higher levels of the isobutyrate-derived TKL compared to intact DEBS1-TE. Fermentations of BB138 harboring either pBK1 or pBK1*, supplemented with either cyclopropane or
cyclobutane carboxylic acid were completed in triplicate. Fermentations were analyzed by GC-MS analysis alongside selected ion monitoring for the expected m/z values of the M⁺ masses of the expected compounds, 183 and 197 respectively for cyclopropane and cyclobutane derived TKLs (Figure 28). Masses with the predicted m/z values of 183 and 197 were observed in both fermentations of BB138/(pBK1 or pBK1*) at 9.5 and 10.1 minutes respectively. Production levels of the unnatural cyclic carboxylic acid derived TKLs were not significantly different for the expression of pBK1* compared to pBK1.

![Figure 28. Production of TKLs derived from cyclic starter units.](image)

Extracted ion chromatogram for a) cyclopropane-derived TKL, m/z 183 at 9.5 min and b) cyclobutane-derived TKL, m/z 197 at 10.1 min.
3.4.7 Hypothesis for the shift in production to favor isobutyrate-derived TKL with the dissection of the DEBS AT$_L$-ACP$_L$

The above described data where the expression of pBK1* scar (Figure 23) results in a drastic drop in TKL production but a majority of the production is derived from the propionate starter unit, indicates two important factors: one is that the AT$_L$-ACP$_L$ is required for efficient polyketide production and the second is that the shift to favor the isobutyrate-derived TKL product is the result of the DEBS AT$_L$-ACP$_L$ domain presenting the KS of module 1 with the isobutyrate starter unit. *In vitro* observations in the Reynolds lab has shown that a variety of PKS proteins, when incubated with acyl CoA substrates can result in an increase of free CoA concentration over time and the rate of free CoA formation is dependent on the acyl CoA substrate. Given these observations it has been hypothesized that once an acyl group has been loaded onto the AT domain of the loading module, the group can be hydrolyzed. Thus the AT domain acts as a thioesterase, consuming acyl CoA substrates.

Given this information, one hypothesis to explain the shift in TKL ratios for the expression of pBK1* was that there are differences in the rates of hydrolysis for different acyl groups bound to DEBS AT$_L$-ACP$_L$. In this model, propionate and isobutyrate load onto DEBS AT$_L$-ACP$_L$ with comparable rates as an intact protein, however propionate would be hydrolyzed much faster, leaving isobutyrate to be passed onto the KS of module 1 for polyketide processing. In this hypothesis, the assumption was that the normal polyketide biosynthetic process is relatively fast, where intermediates are efficiently passed from one protein active site to the next and that the hydrolytic activity of the AT domain is thus not significant. However when
the protein biosynthetic machinery is separated as in the constructs described in this thesis (pJY7, pJY12, pBK1*, pBK1**) the resultant production levels are approximately 50%, which is presumably due to the fact that PKS proteins separated with DDs result in a decreased efficiency in the passage of biosynthetic intermediates. A decrease in passage efficiency of intermediates would presumably result in covalently bound acyl groups to exist on the protein and would result in a greater role of hydrolytic process determining TKL product ratios.

In order to examine the dynamics of hydrolysis for different acyl groups from DEBS ATL-ACP L, this protein was expressed and purified from E. coli for further in vitro studies. DEBS ATL-ACP L was purified as a C-terminal His tagged protein from E. coli harboring the previously described expression construct, pBK12 (94). Initial purification of this protein (in both the apo and holo) using Ni-NTA resin yielded large quantities of soluble protein, however the high degree of impurities required further purification, which was accomplished via size exclusion chromatography (Figure 29).
Figure 29. Expression and purification of DEBS AT$_L$-ACP$_L$ (apo and holo). a) SDS PAGE gel of Ni-NTA resin purified protein (M-molecular weight marker, L-material loaded on Ni-NTA resin, FT-flow through, W1-W2 – washes, E1-E5 – elution fractions. b) Size exclusion FPLC chromatogram of Ni-NTA purified protein, DEBS AT$_L$-ACP$_L$ eluted in fractions 13 and 14. c) SDS PAGE gel of fractions containing partially purified protein after size exclusion chromatography.
3.4.8 Kinetic analysis of DEBS AT<sub>L</sub>-ACP<sub>L</sub> with varying acyl CoA substrates

Kinetic parameters ($k_{cat}$ and $K_m$) for the *apo* form of DEBS AT<sub>L</sub>-ACP<sub>L</sub> were determined by monitoring the protein dependent formation of free CoA with three different acyl CoA substrates (acetyl, propionyl, and isobutyryl) under pre-optimized linear conditions (Figure 30). *Apo* DEBS AT<sub>L</sub>-ACP<sub>L</sub> $k_{cat}$ and $K_m$ values with each substrate are summarized in Table 1.

![Kinetic analysis of apo DEBS AT<sub>L</sub>-ACP<sub>L</sub> with a) acetyl CoA, b) propionyl CoA, and c) isobutyryl CoA](image)

*Figure 30. Kinetic analysis of *apo* DEBS AT<sub>L</sub>-ACP<sub>L</sub> with a) acetyl CoA, b) propionyl CoA, and c) isobutyryl CoA*
Table 1. Kinetic parameters for DEBS AT$_{L}$-ACP$_{L}$

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_m$ (µM)</th>
<th>$k_{cat}/K_m$ (mM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetyl CoA</td>
<td>1.3 ± 0.1</td>
<td>812 ± 102</td>
<td>1.6</td>
</tr>
<tr>
<td>Propionyl CoA</td>
<td>1.5 ± 0.1</td>
<td>360 ± 54</td>
<td>4.2</td>
</tr>
<tr>
<td>Isobutyryl CoA</td>
<td>2.1 ± 0.1</td>
<td>486 ± 92</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Kinetic parameters were determined using the *apo* form of the protein, which allowed for the acylation and subsequent hydrolysis off of the AT domain to be specifically examined. Since these kinetic parameters were determined by the increase in free CoA concentration, which is dependent on both the rate constants of acylation ($k_A$) and hydrolysis ($k_H$) the observed $k_{cat}$ values are dependent on both of these processes. The $k_{cat}$ value can be described an extension of the standard Michaelis-Menten model by the equation $k_{cat}=(k_A*k_H)/(k_A+k_H)$ (101).

The approximately equivalent catalytic efficiency of DEBS AT$_{L}$-ACP$_{L}$ for the processing of propionyl and isobutyryl CoA, as determined by $k_{cat}/K_m$ values of 4.2 and 4.4 mM$^{-1}$ s$^{-1}$ respectively do not reflect any significant difference in hydrolytic rates. The observed kinetic data for DEBS AT$_{L}$-ACP$_{L}$, demonstrate that there is not a slower rate of hydrolysis for isobutyrate versus propionate from the AT domain and that there must be another basis to explain the dramatic shift in TKL production for the expression pBK1* when compared to pBK1.
3.4.9 Radioactive acylation of DEBS AT\textsubscript{L}-ACP\textsubscript{L}

The preceding experiments demonstrate that the shift in TKL ratios favoring the isobutyrate-derived product can not be explained by a slower hydrolytic rate for isobutyrate versus propionate or a preferential TEII catalyzed removal of propionate from ACP\textsubscript{L}. Thus an alternative explanation was sought. The DEBS AT\textsubscript{L}-ACP\textsubscript{L} recombinant protein was examined under a series of \textit{in vitro} \(^{14}\text{C}\) radioactive acyl CoA labeling experiments. The goal of this work was to determine the rate of AT domain acylation and subsequent transfer to the ACP domain for propionate and isobutyrate when only one substrate was present. The extent of \(^{14}\text{C}\) labeling for the \textit{apo} protein was used to determine the rate of AT domain acylation, while experiments with the \textit{holo} protein resulted in information about the transfer of acyl groups from the AT to the ACP domain. In addition, competition experiments where both substrates were present was used to track changes in protein labeling from the different acyl CoA substrates as a function of time.

Incubations of the DEBS AT\textsubscript{L}-ACP\textsubscript{L} were carried out under 4 separate sets of conditions, utilizing propionyl and isobutyryl CoA labeled material mixed with non radioactive material; a) \(^{14}\text{C}\) propionyl CoA + unlabeled propionyl CoA (hot/cold P), b) \(^{14}\text{C}\) propionyl CoA + unlabeled isobutyryl CoA (hot P/cold IB), c) \(^{14}\text{C}\) isobutyryl CoA + unlabeled isobutyryl CoA (hot IB/cold P), d) \(^{14}\text{C}\) isobutyryl CoA + unlabeled isobutyryl CoA (hot/cold IB). As with the above described kinetic analysis, the \textit{apo} form of the protein was initially utilized, this allowed the dynamics of the AT acylation and hydrolysis to be examined specifically without the additional
complication of the dynamic AT-ACP transfer step. Aliquots of each individual reaction were removed over time and the protein was then precipitated and washed to remove any excess radioactive materials before liquid scintillation counting (LSC). LSC measurements in disintegrations per minute (DPM) were converted to mole values by use of the specific activity of the radioactive material in each experiment to determine the concentration of acylated protein at each time point. Initial work with the apo protein under the above described conditions with either propionyl or isobutyryl CoA (equal molar mixture of $^{14}$C labeled and unlabeled) alone resulted in an unexpected result (Figure 31a). Throughout the time course of this experiment the amount of recovered labeled protein was significantly higher in the presence of isobutyryl CoA (approximately 1.5 $\mu$M / 37% of maximum) compared to propionyl CoA (approximately 0.4 $\mu$M / 10% of maximum). These observed differences in the levels of labeling can be rationalized by either: a) propionate loading onto the AT domain and being subsequently hydrolyzed at a much faster rate than isobutyrate, or b) that isobutyrate loads onto the AT domain of the protein at much faster rate than propionate.

Given that the overall rate of hydrolysis for each substrate are known, the levels of acylated protein in these experiments permitted a calculation of the specific acylation and hydrolytic reaction rates to be modeled (Chris Hazzard, personal communication). This model assigned specific values for the rate constants for both AT domain acylation ($k_A$) and hydrolysis ($k_H$) for both substrates, propionate and isobutyrate based on the experimental observation with the apo protein and single acyl CoA substrates (Figure 32) (Chris Hazzard, unpublished data). In this model the
kinetic parameters and assigned $k_A$ and $k_H$ values were fitted to a remarkable and accurate description of the experimentally observed labeling data for both the apo and holo protein (Figure 31c-d). These kinetic values (Figure 32) show that isobutyrate loads faster than propionate but is hydrolyzed from the AT domain at a slower rate. As such the level of labeled protein is higher with isobutyryl CoA than with propionyl CoA as the substrate. Given that the overall rates of hydrolysis for isobutyrate and propionate are similar, the observed kinetics for the apo protein do not provide a sufficient explanation for the observed shift in TKL production when the DEBS AT$_L$-ACP$_L$ is separated from the first and second extension modules of DEBS1.

Figure 31. Experimentally determined and kinetic model predicted levels of DEBS AT$_L$-ACP$_L$ labeling with $^{14}$C propionyl or isobutyryl CoA.
Figure 32. Proposed dynamics of propionyl and isobutyryl CoA loading and off loading on DEBS AT\textsubscript{L}-ACP\textsubscript{L}. A faster acylation of the AT domain is observed with isobutyryl than propionyl CoA. However, formation of the propionyl-ACPL is favored over isobutyryl-ACPL as a result of differences in the AT-ACP transfer rates for the two different acyl groups.
A similar series of analyses were conducted with the holo form of DEBS AT_L-ACP_L. Examination of this form of the protein under the same conditions as the apo protein resulted in increases of recovered ^{14}C labeled protein with both propionyl and isobutyryl CoA substrates (Figure 31b). The concentration of labeled protein increased to approximately 2µM (25% of maximum) and 3 µM (37% of maximum) with propionyl and isobutyryl CoA respectively. This increase is logical since the ACP domain is able to bind acyl substrates when post-translationally modified. The higher levels of ^{14}C labeling for the holo protein compared to the apo, provide important insights into the dynamics of acyl group loading onto the AT domain and the subsequent passage to the ACP domain. The increase of ^{14}C labeling for propionyl CoA with the holo protein versus the apo form signifies that the majority of the observed labeling arises from covalent attachment to the ACP domain and not the AT. Conversely the small increase in labeling with isobutyryl CoA from 2 to 3µM indicates that with isobutyryl CoA, the majority of the protein bound acyl group is on the AT and not the ACP domain. Furthermore, the increase in isobutyryl CoA labeling is not observed during the initial time points of the experiment, but is with propionyl CoA. These observations indicate that the transfer of isobutyrate from the AT to the ACP domain is slow relative to the same process for propionate.
3.4.10 Acyl CoA competition experiments

The next question that was explored involved examining the behavior of DEBS AT\textsubscript{L}-ACP\textsubscript{L} in the presence of competing acyl CoA substrates. Radioactive acyl CoA competition experiments with combinations of \textsuperscript{14}C propionyl CoA with unlabeled isobutyryl CoA (hot P/cold IB) and \textsuperscript{14}C isobutyryl CoA with unlabeled propionyl CoA (cold P/hot IB) with both the apo and holo form were carried out and analyzed as previously described (Figure 33). The trends from the competition experiments with the apo protein displayed similar results to the above described assays when single substrates were present. Despite a drop in the maximum observed labeling, these results again show higher levels of AT labeling with \textsuperscript{14}C isobutyrate and lower levels of labeling with \textsuperscript{14}C propionate, despite the presence of a competing substrate. This result shows that even in the presence of competing substrates, isobutyrate is preferentially loaded on the AT domain at much higher levels than propionate (Figure 33a). These results agree with the hypothesis that isobutyrate loaded on the AT domain is the kinetically favored product versus propionate.
Figure 33. Experimentally observed levels of DEBS AT$_L$-ACP$_L$ labeling with competing acyl CoA substrates (propionyl and isobutyryl CoA) for a) apo and b) holo forms of the protein.

Experimental data from the competition experiments utilizing holo DEBS AT$_L$-ACP$_L$ provided insight regarding the placement of acyl groups, either on the AT or ACP domain of the protein. These experiments indicate that when $^{14}$C propionyl CoA and unlabeled isobutyryl CoA were incubated with the holo protein, the maximum level of labeling was observed at time=0, however over the time course of the experiment the levels of labeled protein was reduced by approximately 50% (Figure 33b). Conversely when $^{14}$C isobutyryl CoA and unlabeled propionyl CoA were incubated with the protein, the levels of labeled protein increase by approximately 40% over the time course of the experiment.

Acyl CoA competition experiments with the holo and apo data clearly indicates that when $^{14}$C propionate is present, the majority of the labeled protein is the result of attachment to the ACP domain specifically and not the AT. The observed
decrease in labeling over the time course of this experiment can be attributed to a combination of two possible processes. The first is the replacement of $^{14}$C propionate on the AT with unlabeled isobutyrate but can be discounted by the knowledge that these levels are low (see experiments with the apo protein and propionyl CoA). A second more reasonable explanation involves the dynamic equilibrium for ACP bound acyl groups, in this case propionate on the ACP domain (Figure 16c) can be passed back to the AT where hydrolysis of the acyl group can take place (Figure 16b). In this process the labeled propionate group can be replaced by unlabeled isobutyrate and the system can move from the kinetic product (propionyl ACP$_L$) to a thermodynamic equilibrium consistent with this hypothesis.

The observations with $^{14}$C isobutyryl and unlabeled propionyl CoA along with the apo data suggests that at time=0, the majority of the observed labeling in the competition experiment is the result isobutyrate on the AT domain. However over the duration of the experiment, an increase in the concentration of labeled protein was observed, which likely represents the transfer of isobutyrate from the AT to the ACP domain. The combined labeling data for both the apo and holo forms of DEBS AT$_L$-ACP$_L$ with either single or competing substrates has provided a model which explains the shift in TKL ratios for the separation of DEBS AT$_L$-ACP$_L$ from the first and second extension modules of DEBS1.

### 3.4.11 Correlation of in vitro and in vivo results

Based on the in vivo data for the expression of pBK1* scar, the source of the shift in starter unit preference away from propionate and towards isobutyrate has been
attributed to the DEBS AT\textsubscript{L}-ACP\textsubscript{L} di-domain presenting the KS of module 1 with a higher ratio of isobutyrate than propionate. This \textit{in vivo} change over in the acyl group that is bound to the loading domain ACP and being presented to the KS of module 1 is also observed \textit{in vitro}. The change which is observed in the \textit{holo} DEBS AT\textsubscript{L}-ACP\textsubscript{L} competition labeling experiments from an initial predominance in propionate to isobutyrate (Figure 33b) over the time confirms the \textit{in vivo} observation and supports the hypothesis that over time a stand alone protein, as in the \textit{in vivo} expression of pBK1*, results in higher levels of isobutyrate on the ACP domain.

3.5 Conclusion

The biosynthetic process for type I modular polyketide synthases involves the selection of a starter unit by the loading module AT, which is passed to the cognate ACP domain. Once a starter unit is on the ACP it is then passed to KS domain of the first extension module for the first round decarboxylative condensation with the pre-selected extender unit. In a multimodular system the passage of biosynthetic intermediates is a relatively fast and efficient process. However as was observed in the pikromycin PKS described in chapter 2 and the DEBS PKS in this chapter, when multimodular PKS polypeptides are separated with DD pairs there is a drop in the biosynthetic efficiency. One of the reasons for this drop can be attributed to the slower passage of intermediates. This decrease in biosynthetic efficiency would leave intermediates bound to the preceding module’s ACP and or AT domains susceptible to either thioesterase activity or hydrolysis and would allow for the build up of
intermediates that are not the normal kinetic product. In the case of the DEBS PKS, when the loading module is separated from the first and second extension modules, the slower passage of the starter unit from the ACP domain to the KS of module 1, has allowed for the accumulation and predominance of isobutyrate bound to the ACP in place of propionate.

The observed shift in starter unit incorporation by the DEBS PKS when the loading module has been separated from the rest of DEBS1 is quite remarkable and adds much insight into the dynamics of modular PKS enzymes. The shift towards isobutyrate TKLs along with the above described in vitro data clearly indicate that AT domain acylation and AT-ACP transfer are dynamic processes. By slowing down and reducing the efficiency at which the ACP of the loading module passes the starter unit to the KS of module 1, it has been revealed that loading module ACP domains can present an unexpected starter unit, which normally does not predominate. These insights into the dynamics of acyl group loading and incorporation into the polyketide product, add another level of consideration which must be accounted for when designing hybrid and genetically modified PKS enzyme systems.
4. *In vitro* investigation of the iterative mono module BorA5 from the borrelidin PKS from *Streptomyces parvulus* Tu4055

4.1 Summary

The PKS responsible for the biosynthesis of borrelidin consists of a loading domain and six extension modules, which are housed on six separate polypeptides. The loading domain along with each of the six extension modules are contained on their own polypeptide, with the exception of modules 2 and 3 which are contained on the dimodular single protein, BorA3. Despite the non-iterative nature of most type I modular PKS modules, BorA5 has been proposed to be unusual, acting iteratively by incorporating three molecules of methylmalonyl CoA into the final polyketide structure. *In vitro* examination of this module was accomplished by cloning, expressing and purifying the entire PKS module as well as specific domains using *E. coli* as an expression host. This work has shown that some of the catalytic domains of BorA5 can be expressed as separate proteins and can act in *trans* with one another. Specifically, this work has demonstrated by mass spectrometric (MS) analysis that when expressed as a stand-alone protein, the BorA5 ACP domain in the *holo* (with the 4’-phosphopantetheine group attached) form can be acylated with methylmalonyl CoA substrate in the presence of the entire BorA5 protein. This work has laid the groundwork and basis for additional *in vitro* studies of the BorA5 iterative module by monitoring specific reactions on the stand-alone BorA5 ACP domain.
4.2 Introduction

The structure of the macrocyclic polyketide natural product borrelidin (Figure 31) from *Streptomyces parvulus* Tu4055 was elucidated in 1967 (102). Initially borrelidin was discovered for its antimicrobial activity (103, 104), but has since also been reported to have antiviral (105) and antiangiogenesis (106) activities. The gene cluster responsible for borrelidin biosynthesis from *Streptomyces parvulus* Tu4055 was cloned, sequenced and reported in 2004 (107). The structure of borrelidin has two features not commonly observed in polyketide products; 1) the starter unit for biosynthesis is trans-cyclopentane (1R, 2R)-dicarboxylic acid and 2) a nitrile group at C12 of the borrelidin structure (Figure 34). Additionally the borrelidin PKS is unusual in that it does not follow the general rule of one extension step per module. Specifically, BorA5 has been proposed to function in an iterative fashion, catalyzing three rounds of polyketide chain elongation. The presence of an iterative module in an otherwise modular type I bacterial PKS systems has also been observed in the aureothin system from *Streptomyces thioluteus* (47). Iterative PKS activities are typically observed in type II and type III PKS systems described in chapter 1 but are quite rare in type I PKSs.
Figure 3.4. Modular organization of the borrelidin PKS

The borrelidin PKS modular organization (Figure 3.4) consists of a loading domain and six extension modules that have been proposed to carry a total of eight elongation cycles. Based on the structure of the final product and module organization of the borrelidin PKS, it has been proposed (46) that BorA5 functions iteratively, catalyzing three successive rounds of elongation using methylmalonyl CoA extender units with full reduction of C4-C9 to methylene moieties. If BorA5 is responsible for multiple condensation steps, it is logical to think that in between each round of
elongation that BorA5 catalyzes, the growing polyketide chain is transferred from the ACP domain back to the KS domain and that only after all three rounds of elongation and modification (keto reduction, dehydration, and enoyl reduction) is the growing chain passed to module 6 for further elongation and subsequent cyclization. Iterative PKS modules represent a presumed relaxed substrate specificity, since the KS domains of these modules are required to both accept and process polyketide chains of different carbon length. This presumed relaxed substrate processing capacity of BorA5 was the initial target of this study for its possible utilization in efficient hybrid PKS systems capable of producing a variety of compounds.

The goal of this work entailed cloning and expressing BorA5 in E. coli as a recombinant protein for in vitro studies. Initially the goal of this project involved examining the flexibility of intact BorA5 to load and process different starter units for subsequent polyketide chain elongation. In addition to this it was sought to examine if these starter units would undergo one, two or three rounds of chain elongation and to what extent there would be keto group reduction and dehydration. However, due to limitations of the amount of BorA5 in the holo form, which could be purified, the focus of this study shifted towards separating and expressing individual BorA5 domains. Examples of deconstructing an entire PKS module and expressing specific functional domains as separate discrete proteins have been reported for both DEBS module 3 from the erythromycin PKS (108) as well as PksA, which initiates biosynthesis of the environmental carcinogen aflatoxin B1 (109-111). This work has shown that multidomain PKS proteins can be functionally reconstituted as a combination of proteins while retaining their respective catalytic functions.
Deconstruction of BorA5 was attempted by cloning and expressing the individual catalytic domains or combinations of domains and reconstituting them in vitro with one another. Intact BorA5 in the apo form (without the phosphopantetheine arm attached to the ACP domain) along with the ACP domain were successfully expressed and purified from E. coli as N-terminal His\textsubscript{6} tagged proteins. It has been shown here that BorA5 ACP can be acylated with methylmalonyl CoA by the intact apo BorA5 module. This preliminary work has thus set the stage for additional studies regarding the reactions, which BorA5 can catalyze by monitoring modifications to BorA5 ACP.

4.3 Materials and methods

4.3.1 Commercially purchased materials

All chemicals and antibiotics were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Ipswich, MA). PCR primers were custom synthesized from Integrated DNA Technologies (Coralville, IA). PCR reactions were carried out with either the GC Rich PCR System from Roche (Basel, Switzerland) or LA PCR Kit version 2.1 from Takara Bio INC (Madison, WI).

4.3.2 Construction of BorA5 E. coli expression constructs

Expression construct pJY50, which encodes BorA5 as an N-terminal His\textsubscript{6} tagged protein was generated via three separate PCR reactions using cosmid Bor19B9 (107) that contains the borrelidin PKS gene cluster as the PCR template (Figure 35).
Each individual PCR product was gel purified with use of Qiagen Gel Extraction Kit (Valencia, CA) and sub cloned into pCR 8/GW/TOPO TA cloning vector, Invitrogen (Carlsbad, CA) and subsequently transformed in One Shot Mach1 E. coli cells also from Invitrogen. The first PCR product was generated as an NdeI/FspI fragment with PCR primer pair BA5AF (5’- CAT ATG GTG ACC TCG GTC GGC CGC CG - 3’) + BA5AR (5’ – AGT ACG CCC TTG CGC AGG GC – 3’). The PCR program consisted of an initial hold for 2 minutes at 96 °C, followed by 30 cycles of 96 °C for 30 seconds, 58 °C for 30 seconds, and 72 °C for 1 minute. This was followed by a final extension period at 72 °C for 5 minutes. The second PCR product was generated with primer pair BA5BF (5’ – CAT ATG GCC CTG CGC AAG GGC GTA CT – 3’) + BA5BR (5’ – GGA TCC GAA ACC GAG CTG GT – 3’) as an NdeI/BamHI fragment, where the NdeI restriction site was built into the primer overlapping the natural FspI site of BorA5. The PCR program consisted of an initial hold for 2 minutes at 94 °C followed by 30 cycles of 94 °C for 30 seconds, 56 °C for 60 seconds, and 72 °C for 2.5 minutes. This was followed by a final extension period at 72 °C for 10 minutes. The third PCR product was generated with primer pair BA5CF (5’ – CAT ATG CTC GCC TCG CTG TCC GGC CCC – 3’) + BA5CR (5’ – AAG CTT TCA CAG CAG GCC GGC GAG GTG CCC – 3’) as an NdeI/HindIII fragment, where the NdeI restriction site was again built into the primer overlapping the natural BamHI site of BorA5. The reverse primer contains the natural stop codon of BoA5 plus a built in HindIII site for subsequent cloning steps. Each of the three inserts were verified by plasmid sequencing with manufacturer supplied M13F and M13R primers, yielding constructs pJY51A, pJY51B, and pJY51C respectively. The intact BorA5 coding
sequence was generated by ligating each of the three fragments together as follows; pJY51A was restricted with Ndel/FspI and the approximately 1.2 kb fragment was ligated into pJY51B after restriction with Ndel/FspI. This intermediate plasmid (pJY51AB) was transformed in *E. coli* and the plasmid purified as previously described. Purified intermediate plasmid pJY51AB was restricted with Ndel/BamHI and ligated into pJY51C after restriction with the same endonucleases to give pJY51ABC. The compiled 6.5 kb fragment was then subsequently cloned from pJY51ABC into pET 28 (Novagen) as an Ndel/HindIII fragment yielding pJY50.

Construct pJY55, which expresses just the ACP domain of BorA5 as an N-terminal His tagged protein was generated by PCR amplification followed by TOPO TA cloning and subsequent cloning into the pET 28 expression plasmid as an Ndel/HindIII fragment. The PCR product was generated with PCR primer pair BA5ACPF (5’ – CAT ATG CTC GCC TCG CTG TC  C GGC CCC – 3’) + BA5ACPR (5’ – AAG CTT TCA CAG CAG GCC GGC GAG GTG CCC – 3’). The PCR program consisted of an initial hold for 2 minutes at 94 °C, followed by 30 cycles of 94 °C for 30 seconds, 56 °C for 60 seconds, and 72 °C, for 1 minute. This was followed by a final extension period at 72 °C for 10 minutes.
Figure 35. PCR amplification and cloning of BorA5. Plasmids pJY51A/B/C were individually generated by PCR amplification from cosmid Bor19B9. These constructs were then subsequently restricted and ligated together. The entire 6.5 kb fragment encoding BorA5 was ligated into pET28 as an Ndel/HindIII fragment yielding pJY50.
4.3.3 Expression and purification conditions for BorA5 recombinant proteins

BorA5 was over expressed and purified from E. coli by expression of pJY50 as a N-terminal 6 His-tagged protein. Expression was carried out in either BL21 DE3 chemically competent cells (Invitrogen) or BAP1 cells. BAP1 E. coli cells are a derivative of BL21 DE3 cells, which contain a chromosomally integrated copy of the sfp gene encoding the requisite 4’-phosphopantetheinyl transferase, which allows for posttranslational modification of the ACP domain to the holo form. BAP1 E. coli cells were provided as a gift from Professor Chaitan Khosla’s laboratory, Stanford University. The stand alone BorA5 ACP was over expressed and purified from E. coli in the same manner by expression of pJY55 in BAP1 cells.

Overnight E. coli seed cultures harboring pJY50 or pJY55 were grown in LB medium supplemented with 100 g/mL kanamycin at 37 °C with agitation. 1000 mL LB medium cultures supplemented with 50 g/mL kanamycin were inoculated with 5% (v/v) of the overnight seed culture and grown at 37 °C with agitation to an O.D of approximately 0.6 (approximately 2.5 hours). Cultures were allowed to cool to room temperature and expression was induced by the addition of isopropyl thio-D-galactoside (IPTG) to a final concentration of 1.0 mM. Cultures for the expression of BorA5 were allowed to grow for an additional 16 hours at room temperature with agitation, while cultures for the expression of BorA5 ACP were allowed to grow for an additional 3 hours at 37 °C. E. coli cells with either of the over expressed proteins BorA5 or BorA5 ACP were harvested by centrifugation and either stored at -80°C or immediately lysed for protein purification. His6-tagged protein was purified following the protocol outlined in Qiagen’s QiAexpressionist handbook under native conditions.
with the following modifications. *E. coli* cells were re-suspended in 25 mL of BugBuster Protein Extraction Reagent (Novagen). Cells were disrupted by sonication and allowed to incubate on ice with slow agitation for at least 1 hour. The mixture was then subjected to centrifugation and the soluble fraction (supernatant) was incubated with 1 mL of Ni-NTA agarose (Qiagen) on ice with gentle agitation for at least 1 hour. The mixture was then transferred to a Polyprep chromatography column (BioRad) and allowed to equilibrate at 4°C. The cell free extract (CFE) was drained and the Ni-NTA resin washed with 10 column volumes of wash buffer 1 (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM imidazole, pH 8.0) followed by 10 column volumes of wash buffer 2 (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 20 mM imidazole, pH 8.0). His$_6$-tagged protein was eluted from the Ni-NTA resin with 3mL of elution buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 250 mM imidazole, pH 8.0). 5 separate elution fractions were collected each containing between 500-600 µL total volume. Each of the 5 fractions along with the initial load, column flow through and the two column washes were analyzed by SDS-PAGE gel electrophoresis using NuPAGE Novex 4-12% Bis-Tris gels (Invitrogen). Expression and purification of the entire BorA5 module was confirmed by the presence of a band over 200 kDa, the desired protein has a molecular weight of approximately 230 kDa. Similarly, expression and purification of BA5 ACP was confirmed by the presence of a band which was approximately 10 kDa, the expected molecular weight of the recombinant protein is approximately 12 kDa. Both desired proteins were further purified to homogeneity by size exclusion chromatography using an AKTA FPLC System (GE Health care, formally Amersham Biosciences) equipped with a Superdex GL 200 sizing column and a UV visible
detector set at 280 nm. Fractions containing the His$_6$-tagged protein were confirmed by SDS PAGE gel electrophoresis and concentrated with Amicon Ultra Centrifuge columns (Milipore) and stored at -80°C for future use.

### 4.3.4 In vitro acylation assay conditions

Due to the low quantity of recoverable BorA5 in the *holo* form (data not shown), the *apo* form was used for in vitro enzyme assays. *In vitro* acylation experiments consisted of three essential components a) acyl CoA substrates (malonyl, and methylmalonyl) b) the *apo* form of BorA5 (to be utilized as a catalyst) and c) the post translationally modified *holo* form of the stand alone ACP domain. The presence of these components allowed for any covalent modifications of the ACP protein to be monitored by subsequent mass spectrometric analysis.

Assays to probe whether or not the AT domain from an intact *apo* form of BorA5 could act in *trans* and acylate the stand-alone ACP domain were carried out under the following set of conditions. As negative controls the following reactions which would not be expected to yield a covalently modified ACP protein were carried out; a) ACP + malonyl CoA, b) ACP + methylmalonyl CoA, c) ACP + BorA5, d) ACP + BorA5 + malonyl CoA. Contrastingly, the experimental reaction which would be expected to result in the modification of BA5 ACP contained all of the following components, ACP + BorA5 + methylmalonyl CoA. Each of these assays was carried out with a 3:1 molar ratio of BorA5 ACP to BorA5 with 1mM of the respective acyl CoA substrate in a buffer consisting of 200 mM NaH$_2$PO$_4$, 1 mM TCEP, 1 mM EDTA, 20% glycerol, pH 7.4. Reactions were incubated at room temperature
(approximately 25 °C) and aliquots were removed at t=0, 2, 4, 8 and 18 hours.

4.3.5 LC-MS analyses of BorA5 ACP acylation

Prior to LC-MS analysis each sample was dialyzed using Slide-A-Lyzer mini dialysis units from Thermo Scientific (Rockford, IL) against 500 mL of a buffer consisting of 100 mM NaH$_2$PO$_4$ and 100 mM NaCl at pH 7.4 at 4 °C for approximately 2 hours. LC-MS analysis was performed using reverse phase HPLC on an Agilent 1100 series LC (liquid chromatography) system coupled to a Bruker MicroToF-Q mass spectrometer in the positive mode. LC separation was carried out at an elevated temperature of 50 °C with a 5µm Discovery Bio Wide Pore C8 reverse phase column (5 x 2.1 cm, Supelco), with a linear gradient system from solvent A (water – 0.01% formic acid) to solvent B (acetonitrile – 0.01% formic acid) over 33 minutes at a flow rate of 0.3 mL/minute. In addition to acylation assays, the mass of stand alone BorA5 ACP was also analyzed in this same manner.

4.4 Results

4.4.1 Expression and purification of BorA5 and BorA5 ACP

BorA5 as an N-terminal His$_6$-tagged protein was over expressed in BL21 DE3 containing the construct pJY50. Expression of this construct under the above described conditions presumably results in the generation of the apo form of the protein (without the 4’-phosphopantetheine post translational modification), since this *E. coli* strain lacks the sfp gene required for this modification. Protein expression was observed by SDS-PAGE analysis, which demonstrated a significant band with an
approximate molecular weight of 230 kDa, which is consistent with the expected molecular weight based on the sequence of the cloned BorA5 open reading frame (Figure 36). The Ni-NTA purified protein was further purified to homogeneity via FPLC size exclusion chromatography. Repeated attempts to express BorA5 using pJY50 in BAP1 cells (which contains a chromosomally integrated copy of the sfp gene) and purify the holo form of this protein resulted in high levels of insoluble proteins. All attempts to improve the solubility of the holo BorA5 by varying the media, induction temperature and other culture conditions failed to result in higher quantities of soluble protein.
Figure 36. Expression and purification of apo BorA5. a) SDS PAGE gel of Ni-NTA resin purified protein (M-molecular weight marker, L-material loaded on Ni-NTA resin, FT-flow through, W1-W2 – washes, E1-E5 – elution fractions. b) Size exclusion FPLC chromatogram of Ni-NTA purified protein, apo BorA5 eluted in fractions 7-10. c) SDS PAGE gel of fractions containing partially purified protein after size exclusion chromatography.
Stand alone BorA5 ACP in the *holo* form was expressed and purified in an analogous manner to the complete BorA5 module. BorA5 ACP as an N-terminal His$_6$-tagged protein was over expressed in BAP1 cells containing the construct pJY55. Expression of this protein was detected as a significant band in SDS PAGE analysis, with an approximate molecular weight of 12 kDa, this is consistent with the expected mass based on the sequence of the cloned BorA5 ACP open reading frame (Figure 37). Ni-NTA purified protein was further purified via FPLC size exclusion chromatography.
Figure 37. Expression and purification of holo BorA5 ACP. a) SDS PAGE gel of Ni-NTA resin purified protein (M-molecular weight marker, L-material loaded on Ni-NTA resin, FT-flow through, W1-W2 – washes, E1-E5 – elution fractions. b) Size exclusion FPLC chromatogram of Ni-NTA purified protein, holo BorA5 ACP eluted in factions 7-10. c) SDS PAGE gel of fractions containing partially purified protein after size exclusion chromatography.
4.4.2 LC-MS analysis of *holo* BorA5 ACP domain

Mass spectrometric analysis of BorA5 ACP domain was completed to confirm the identity of the purified 12 kDa protein (Figure 38).

Figure 38. Positive mode mass analysis of BorA5 ACP. a) Selected ion mass spectrum of the LC chromatogram elution at approximately 16 minutes. Each mass represents a different charge state of the multiply charged protein molecule. b) Charge deconvoluted mass spectrum of (a), BorA5 ACP is identified by a mass value of 11498.4 Da.

BorA5 ACP was expressed in BAP1 *E. coli* cells, in an attempt to generate the *holo* form (with the 4′-phosphopantetheine group covalently attached) of the protein, which is required for the covalent attachment of polyketide intermediates and acyl group extender units. The predicted mass of the recombinant protein with the N-terminal
His$_6$ tag in the *holo* form is 11497.6 Da and when analyzed in the positive mode, has a mass of 11498.6 Da. The observed mass of 11498.4 Da (Figure 37b) is within 10 ppm of the calculated theoretical mass. This result is consistent with the expression and purification of the target protein, *holo* BorA5 ACP.

**4.4.2 LC-MS analysis of *apo* BorA5 in *vitro* acylation activity**

PKS proteins which contain ACP domains requires posttranslational modification (with the phosphopantetheine arm) to be active and capable of accepting polyketide extender units. Because the ACP domain of the complete and intact *apo* BorA5 module lacks this modification it was sought to examine if the remainder of the domains within the protein retained activity. Initially, work began by assessing if the AT domain of *apo* BorA5 was capable of acylating the stand-alone *holo* BorA5 ACP protein when combined in *trans*. The two purified proteins were coincubated in the presence of the natural extender unit for this module, methylmalonyl CoA (Figure 39).
Figure 39. Proposed reaction between *apo* BorA5 and stand alone *holo* BorA5 ACP domain in trans with one another. The ACP domain of *apo* BorA5 does not possess the posttranslational modification required to accept the AT domain selected extender unit. *Apo* BorA5 along with the specific BorA5 AT domain extender unit (methylmalonyl CoA) was incubated along with the BorA5 ACP in the *holo* form to examine if the AT and ACP domains could facilitate acyl group transfer.

With the ability to identify the BorA5 ACP protein by LC-MS analysis (Figure 38), covalent modifications of the protein can be observed by the appropriate and predicted increases in mass for specific acyl groups or polyketide intermediates bound to the protein. Indeed, overnight incubations with *apo* BorA5, *holo* BorA5 ACP, and methylmalonyl CoA resulted in a new mass of 11598 Da, which is consistent with covalent modification of the stand-alone ACP domain with methylmalonate (Figure 40).
Figure 40. Positive mode mass analysis of an overnight incubation of *apo* BorA5 with *holo* BorA5 ACP and methylmalonyl CoA. a) Selected ion mass spectrum of the LC chromatogram elution at approximately 16 minutes. b) Charge deconvoluted mass spectrum of (a) resulted in two m/z values of interest. *Holο* BorA5 ACP is identified by an m/z value of 11498.5 while a new mass of 11598.5 correlates to the mass of the ACP domain acylated with methylmalonate.

Incubations devoid of either apo BorA5, BorA5 ACP or methylmalonyl CoA as well as incubations utilizing malonyl CoA (not the cognate substrate for BorA5 AT) resulted in no observed covalent modification of the *holο* BorA5 ACP domain.
4.5 Discussion

4.5.1 Trans in vitro acylation of stand alone holo BorA5 ACP

Based on published work with module 3 of the DEBS PKS (108) and PKSA (111) which is involved in the biosynthesis of the environmental carcinogen aflatoxin B1, the ability of the AT domain from the intact apo BorA5 module to acylate a separate ACP domain protein specifically with methylmalonyl CoA, is consistent with the notion that multi-domain PKS proteins can be reconstituted as separate proteins in vitro, while retaining catalytic activity. In both of these published examples of multi domain PKS proteins being reconstituted as multiple proteins in vitro while retaining enzymatic activity, each of the respective domains (such as KS, AT, ACP) were only present in one form, either as a single stand-alone protein or in combination with other catalytic domains. Contrastingly, the results shown here indicate that even when an intact module including the ACP domain in the apo form is combined in trans with a stand alone holo ACP domain, acyl group transfer from the AT domain of the intact module to the holo ACP domain is possible (Figure 40). The ability of the single domain holo ACP protein to accept an acyl group substrate from an AT that is part of an intact modular protein may signify the inherent specificity of an ACP domain to interact properly with its cognate AT, as well as the versatility of the phosphopantetheine arm of an ACP domain to accept the extender unit acyl group from the AT domain. The fact that no covalent modifications of BorA5 ACP were observed when malonyl CoA was utilized may reflect a reinforcement of the strict substrate specificity of extension module AT domains for specific extender units (79). Conversely, the inability to detect malonyl BorA5 ACP may reflect the acylation of
the AT domain with malonyl CoA, followed by the rapid hydrolysis of the malonate acyl group before transfer to the ACP can occur. The work in the DEBS PKS system described in chapter 3, demonstrates that there are significant acylation and hydrolysis reaction rates for AT domains and that transfer to an ACP can be a limiting factor. The intriguing observation of trans acylation of a stand-alone ACP domain protein by an AT domain contained within an entire intact module represents a potential key step towards understanding the BorA5 iterative module. In principle, the ability to monitor the reactions which BorA5 catalyzes by MS analysis of a stand alone ACP domain using either apo BorA5 or dissected components of the module may aid in elucidating the specific mechanism of iteration as well as provide insights into the potential substrate flexibility of this module. However this approach is not without some drawbacks. Under the optimized conditions, that were developed, the acylation of holo BorA5 ACP was only observed after overnight (> 16 hour) incubations and resulted in approximately 50% conversion to the acylated form. Attempts to increase the efficiency of acylation were attempted by adjusting the ratio of the apo BorA5 catalyst to the holo BorA5 ACP substrate, but did not result in any significant change in acylation efficiency. Other attempts to increase the efficiency of acyl group transfer from the AT to the ACP domain included expressing other forms of the BorA5 AT domain (KS-AT, AT, KS-AT-DH-ER-KR), however these expression constructs either resulted in poor expression or insoluble proteins.

In conclusion, it has been shown that in vitro reconstitution of the apo form of BorA5 from the borellidin PKS is capable of acylating its cognate ACP in trans when it is a stand-alone protein. Based on the preliminary observations of this study, this
trans acylation reaction of the separate ACP protein is not a rapid or efficient process. This highlights the innate specificity that cognate PKS domains have for one another. These observations suggests that future studies of BorA5 to better understand the iterative nature and the potential substrate flexibility of this module can be approached by monitoring modifications to the stand-alone BorA5 ACP domain when coincubated with either apo BorA5 or dissected components of the module.
5. Conclusions and recommendations for future work

The work presented in this thesis describes the utilization of a matched docking domain pair from the phoslactomycin PKS system to functionally dissect single multimodular polypeptides from both the pikromycin and erythromycin PKSs into separate proteins. Both of the examples from the pikromycin and erythromycin systems, where the natural intrapolypeptide linkers between neighboring modules was replaced with the matched docking domain pair resulted in 50% polyketide production compared to the intact un-separated intact control experiments. In addition to the significant retention of polyketide production which these dissections yielded, the results from the separation of DEBS1-TE has provided significant insights into the dynamics of AT domain acyl group loading and hydrolysis as well as new insights into the rate of transfer from the AT to the ACP domain for different acyl groups.

Conclusion 1

Experiments dissecting PikAI and DEBS1-TE, multimodular polypeptides from the pikromycin and erythromycin PKS have provided additional *in vivo* examples that multimodular PKS can be separated into individual modules by replacing the intrapolypeptide linker with a heterologous DD pair, in this work a docking domain pair from the phoslactomycin PKS. Additionally this work has specifically demonstrated that docking domains are required for module-to-module interaction and that simply separating multimodular PKS proteins without the proper docking interactions results in a tremendous drop in polyketide production. These
observations demonstrate for the first time that the separation of these complex multifunctional proteins can lead to only a modest decrease in the overall production of the final polyketide product compared to the intact protein.

**Conclusion 2**

In the case of PikAI from the pikromycin system, it has been shown that the efficiency of *in vivo* utilization of the NAC-thioester mimic, (of the natural diketidie intermediate) presented to the KS of module 2 can vary depending on the context of the feeding experiment. Depending on whether the preceding module was covalently linked, separated with docking domains, or removed entirely, resulted in significantly differing levels of incorporation into the final product. The results of this work demonstrate that accessing internally linked modules of multimodular PKS proteins with synthetic intermediates is very inefficient but can be improved significantly by removal of the preceding modules and expression of the internal module with an N-terminal docking domain. The ability to access internal PKS modules efficiently adds another tool to precursor-directed biosynthetic efforts, which can lead to the production of new antibiotics with useful properties.

**Conclusion 3**

Separation of the trimodular PKS polypeptide DEBS1-TE, at the module-module junctions with a matched docking domain pair from the phoslactomycin PKS results in approximately 50% triketide lactone production compared to intact DEBS1-TE. Dissection of DEBS1-TE at the loading domain-module 1 junction results in a
dramatic shift in starter unit incorporation, with a predominance in the isobutyrate-derived triketide lactone. *In vitro* data with the DEBS AT<sub>L</sub>-ACP<sub>L</sub> has shown isobutyrate being preferentially loaded onto the AT domain versus propionate. However propionate being kinetically favored in the subsequent passage step of acyl groups from the AT to the ACP domain. Under wild type fermentation conditions when DEBS1 is an intact protein, propionate is the kinetic product presented to the KS domain of module 1 by ACP<sub>L</sub>. However when the system is dissected at the loading domain module 1 junction, the efficiency of acyl group transfer between ACP<sub>L</sub> and the KS of module 1 diminishes, allowing the dynamics of acyl group loading, hydrolysis and domain transfer to equilibrate with isobutyrate predominating on the ACP domain. By slowing down the natural biosynthetic process between DEBS AT<sub>L</sub>-ACP<sub>L</sub> and the KS of module 1, insights into the dynamics of acyl group loading, hydrolysis and transfer between AT<sub>L</sub> and ACP<sub>L</sub> have been gained.

**Future work**

Given the complexity of the compiled data presented in this thesis, which indicates that isobutyrate loaded on the AT domain, is favored over propionate but that propionate is kinetically favored on the ACP domain. Future work will consist of mass spectrometric analysis of the specific peptides containing the active site serine residue of the AT domain, where acyl groups are covalently bound, as well as the serine residue of the ACP domain where the phosphopantetheine moiety is attached. Incubating the recombinant DEBS AT<sub>L</sub>-ACP<sub>L</sub> protein expressed and purified from *E. coli* with both propionyl and isobutyryl CoA competing substrates, followed by
proteolytic digestion and mass spec analysis, will yield information regarding which substrates are bound to the AT and ACP domains and in what proportion they exist relative to one another. Direct identification of which acyl group predominates on the AT and ACP domains under competition conditions will help support the current $^{14}$C labeling data.

**Conclusion 4**

*In vitro* work with the iterative BorA5 PKS module from the borrelidin PKS has shown that the AT domain an intact and complete module in the *apo* form, is capable acylating a stand alone form of its cognate ACP in trans with methylmalonyl CoA, when the ACP domain is in the *holo* form. This example of reconstituting enzymatic activity by expressing individual domains of PKS modules as discrete proteins is consistent with previously published work with other PKS systems. This preliminary work on BorA5 is the first observation that an AT domain from an intact and complete module can act in trans, acylating a separate stand alone ACP domain protein and has laid the groundwork for future work on this iterative PKS module.

**Future work**

Future work on the iterative BorA5 module should be focused on optimizing the acylation reaction of the stand alone *holo* BorA5 ACP protein. Perhaps the presence of the other domains (DH, ER, KR, *apo* ACP) may hinder the trans acylation reaction and it is possible that removal of these domains will increase the accessibility of the *holo* ACP phosphopantetheine arm to the acyl group bound to the AT domain.
Despite the difficulties in generating soluble proteins thus far from expression constructs aimed at truncating BorA5 perhaps alternate expression constructs or expression conditions may result in improved results.
References


Interactions Direct Channeling of Intermediates between Polyketide Synthase Modules, *Biochemistry* 40, 2326-2331.


Efficient Triketide Lactone Synthesis†, Biochemistry 41, 10827-10833.


