

Summer November 2014

THE DISCOVERY AND STUDY OF FLUVIRUCIN B1 POLYKETIDE SYNTHASE

Tsung-Yi Lin
University of Massachusetts - Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

 Part of the [Biological Engineering Commons](#), and the [Organic Chemistry Commons](#)

Recommended Citation

Lin, Tsung-Yi, "THE DISCOVERY AND STUDY OF FLUVIRUCIN B1 POLYKETIDE SYNTHASE" (2014).
Doctoral Dissertations. 238.
<https://doi.org/10.7275/ehez-7370> https://scholarworks.umass.edu/dissertations_2/238

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

THE DISCOVERY AND STUDY OF FLUVIRUCIN B₁ POLYKETIDE SYNTHASE

A Dissertation Presented

by

TSUNG-YI LIN

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2014

Department of Chemistry

© Copyright by Tsung-Yi Lin 2014

All Rights Reserved

THE DISCOVERY AND STUDY OF FLUVIRUCIN B₁ POLYKETIDE SYNTHASE

A Dissertation Presented

by

TSUNG-YI LIN

Approved as to style and content by:

Nathan A. Schnarr, Chair

Lila M. Gierasch, Member

James Chambers, Member

Margaret A. Riley, Member

Craig T. Martin, Department Head

Department of Chemistry

DEDICATION

To my Mother Su-Chu Huang, who had a near-death experience for giving birth of me, for supporting and taking care of me in every aspect of my life and make me a better person than I thought I could be. To my Father Fang-Yen Lin who taught me about Nature and science, cultivated my curiosity since I was a little kid. To my sister Tsu-Ying Lin for standing by my side all the time. And to my love and friends, who support me in my selfish pursuit of PhD. Without you none of this would have been possible.

ACKNOWLEDGMENTS

I would like to thank my advisor, Nathan Schnarr for his knowledge, guidance, very strong support to this thesis and my career development. I would also like to thank my committee for their valuable insight and contributions, and the University of Massachusetts for supporting me and my work, and to all of my teachers.

I would also like to thank my group members who volunteered their support on these projects. Especially to S. Lawrence Borketey and Gitanjali Prasad for all of the hard work and wisdom they contributed, and to Jon Amoroso and Adam Gann for all the intelligent conversation about organic chemistry.

ABSTRACT

THE DISCOVERY AND STUDY OF FLUVIRUCIN B₁ POLYKETIDE SYNTHASE

SEPTEMBER 2014

TSUNG-YI LIN, B.S., NATIONAL TAIWAN UNIVERSITY

M.S., ANATOMY AND CELL BIOLOGY, NATIONAL TAIWAN UNIVERSITY

Ph.D., CHEMISTRY, UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Nathan A. Schnarr

Rapidly decreasing numbers of viable therapeutic leads in the pharmaceutical pipeline demand new, sustainable methods for improved drug discovery and development. Despite vast improvements in *de novo* drug design and target recognition, Nature remains the richest source of small molecule therapeutics. Among many natural products, polyketides are not only the most promising ones for developing new antibiotic leads, but also exhibit unusually high therapeutic value ranging from clinical use as anticancer, antiviral, and immunosuppressant drugs.

Modular polyketide synthases (PKSs) are dedicated nano-machinery that can be manipulated to produce a structurally diverse library for drug discovery programs. The ability to manipulate these natural systems to produce novel metabolites rests largely on increased mechanistic understanding of how these molecules are generated and how these processes can be manipulated. As impressive as their pharmaceutical properties are, the biosynthetic engineering potential of these compounds continues to draw widespread attention from the research community. Although some success has been realized in terms of polyketide structure diversification, severe limitations in engineered product output continue to impede efforts toward practical combinatorial biosynthesis. This thesis is focused on understanding and exploiting a new biosynthetic enzyme assembly and overcoming the engineering hurdles for making novel polyketide metabolites.

Fluvirucin B₁, produced by *Actinomadura vulgaris*, is a 14-membered macrolactam active against a variety of infectious fungi as well as influenza A. Despite considerable interest from the synthetic community, very little information is available regarding the biosynthetic origins of the fluvirucins. Herein, we report the identification and initial characterization of the fluvirucin B₁ polyketide synthase and related enzymes.

The cluster consists of five extender modules flanked by an N-terminal acyl carrier protein and C-terminal thioesterases domain. All but one of the synthase modules contain the full complement of tailoring domains (ketoreductase, dehydratase, and enoyl reductase) as determined by sequence homology with known polyketide synthases. Active site analyses of several key components of the cluster are performed to further verify that this gene cluster is associated with production of fluvirucin B₁. This work will both open doors toward a better understanding of macrolactam formation and provide an avenue to genetics based diversification of fluvirucin structure.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS.....	v
ABSTRACT.....	vi
LIST OF TABLES.....	xii
LIST OF FIGURES.....	xiii
CHAPTER	
1. INTRODUCTION OF POLYKETIDES, POLYKETIDE SYNTHASE AND ENGINEERING OF	
POLYKETIDE SYNTHASE.....	1
1.1 Background.....	1
1.2 Polyketides.....	1
1.3 Polyketide Synthases.....	3
1.3.1 DEBS for Elucidating the Mechanisms of Polyketide Synthase.....	4
1.4 Heterologous Expression and Genetic Engineering of Polyketide Synthase	
Genes.....	6
1.5 Polyketide Synthase Engineering Hurdles.....	9
1.6 Hypothesis.....	9
1.7 Reference.....	12
2. THE NEW TOOL FOR PKS ENGINEERING: FLUVIRUCIN B ₁ PKS AND GENOMIC	
INFORMATION ANALYSIS.....	16
2.1 Introduction.....	16
2.2 Fluvirucin B ₁ PKS.....	17
2.3 Results and Discussions.....	18
2.3.1 Hypothesis.....	18

2.3.2 Genomic Sequencing and Bioinformatic Analysis.....	19
2.3.3 Fluvirucin B ₁ PKS Genes	21
2.3.4 AT Selectivities	23
2.3.5 KR Domains	25
2.4 Experimental detail.....	26
2.4.1 Bacterial strains, culture conditions and DNA purification.	26
2.4.2 Genomic sequencing and bioinformatic analysis	27
2.4.3 Genomic PCR Procedure	28
2.5 Reference	29
3. THE CLONING AND HETEROLOGOUS EXPRESSION OF FLUVIRUCIN B ₁ PKS.....	33
3.1 Introduction	33
3.2 Result and Discussion.....	38
3.2.1 Cloning of Module 1 of Fluvirucin B ₁ PKS and PCR Optimization	38
3.2.2 Cloning of Module 3 and 5 of Fluvirucin B ₁ PKS.....	44
3.2.3 Cloning of Module 2 of Fluvirucin B ₁ PKS and the Study of PCR Additives	47
3.2.4 Cloning of Module 4 of Fluvirucin B ₁ PKS	51
3.2.5 Protein Expression	54
3.3 Experimental Procedures.....	55
3.3.1 Bacterial Strains, Culture Conditions and DNA Purification.	55
3.3.2 General Procedure for PCR Experiment	55
3.3.3 Cloning of Module 1 of Fluvirucin B ₁ Polyketide Synthase	56
3.3.4 Cloning of Module 3 of Fluvirucin B ₁ Polyketide Synthase	57
3.3.5 Cloning of Module 5 + TE of Fluvirucin B ₁ Polyketide Synthase	57

3.3.6 PCR Organic Solvent Study.....	58
3.3.7 Cloning of Module 2 of Fluvirucin B ₁ Polyketide Synthase	58
3.3.8 Cloning of Module 4 of Fluvirucin B ₁ Polyketide Synthase	59
3.3.9 General Procedure for Protein Expression and Isolation	60
3.3.10 Materials	61
3.4 Supplement Information for Cloning of Module 4	62
3.5 Reference	67
4. EXAMING SUSTRATE SELECTIVITY OF DOMAINS IN FLUVIRUCIN B₁ PKS SYSTEM AND	
BIOINFORMATIC ANALYSIS OF KETOSYNTHASE ON SUBSTRATE SPECIFICITY	70
4.1 Introduction	70
4.2 Results and Discussions	74
4.2.1 Cloning and Protein Expression of KSATs and ACPs	74
4.2.2 Examine the AT Selectivities	75
4.2.3 FluKS Substrate Selectivity	77
4.3 Experimental Procedures.....	80
4.3.1 Cloning of KSAT and ACP Domains of Fluvirucin B ₁ PKS	80
4.3.2 General Method for Preparation of All <i>N</i> -Acetylcysteamine (SNAc)	
Thioester Derivatives	82
4.3.3 Synthesis of malonyl and substituted malonyl SNAc thioesters.....	84
4.3.4 ACP Acylation by AT	86
4.3.5 KSAT3 Loading.....	87
4.3.6 Fluorescence Transfer Assay.....	87
4.3.7 Chromophore Attachment	88
4.3.8 Gel Assay	88

4.4 Reference	89
5. SUMMARY AND FUTURE DIRECTIONS	91
5.1 Discussion	91
5.2 Conclusions	95
5.3 Future Directions	95
5.3.1 Native Ketosynthase Selectivities in the Fluvirucin PKS	95
5.3.2 Altered Ketosynthase Selectivities via Mutation of Active Site Pocket Residues	98
5.3.3 Macrolactam Production from Engineered Fluvirucin PKS.....	103
5.3.4 Engineering FAS-like PKS into Iterative PKS.....	114
5.4 Summary	116
5.5 Reference	118
BIBLIOGRAPHY	122

LIST OF TABLES

Table	Page
Table 1. Expected AT Specificities : Results from SEARCHPKS.....	24
Table 2. The primer list for module 1	40
Table 3. The primer list for module 3	44
Table 4. The primer list for module 5	44
Table 5. The primer list for module 2 (the first round)	47
Table 6. The primer list for module 2 (the second round)	50
Table 7. Sequence comparison results for Flu modules and DEBS module 4	93
Table 8. Sequence Comparison Results for DEBS Modules.....	94

LIST OF FIGURES

Figure	Page
Figure 1. Representative polyketide structures. Rifamycin, Erythromycin, and Tetracycline are antibiotics	2
Figure 2. Structure of 6-dEB.....	4
Figure 3. Schematic diagram of 6-deoxyerythronolide B synthase (DEBS)	5
Figure 4. Schematic diagram of genetic engineering strategy for producing novel compounds from DEBS1 modules.....	7
Figure 5. Schematic diagram of a mutational inactivation strategy for production of novel compounds from modules containing the full compliment of β -processing domains.....	10
Figure 6. Schematic diagram of a mutational inactivation strategy for production of novel compounds form fatty acid-like PKS modules.....	17
Figure 7. Structure of fluvirucin B ₁ aglycone and fluvirucin B ₁	17
Figure 8. Predicted tailoring domains and AT selectivities for each fluvirucin B ₁ synthase module based on the fluvirucin B ₁ aglycone structure	19
Figure 9. Putative gene cluster of fluvirucin PKS	20
Figure 10. Gene organization for the fluvirucin B1 PKS cluster.....	21
Figure 11. Protein homology analysis of Flu-KR Sequence.....	26

Figure 12. Schematic diagram of polyketide engineering strategies	34
Figure 13. Schematic diagram for the putative fluvirucin B ₁ PKS.....	37
Figure 14. The PCR result for cloning module 1-a.	41
Figure 15. The PCR result for cloning module 1-b.	41
Figure 16. The PCR result for cloning module 1 with 5% DMSO.	42
Figure 17. The plasmid map of Flu-module1 (pTL-A01)	43
Figure 18. The plasmid map of Flu-module3 (pTL-A03)	45
Figure 19. The plasmid map of Flu-module5 (pTL-A05)	46
Figure 20. The gel analysis of PCR additives examination based on module 1 PCR setup.....	49
Figure 21. The plasmid map of Flu-module2 (pTL-A02).	51
Figure 22. The plasmid map of Flu-module4 (pTL-A04)	53
Figure 23. PAGE analysis of fluvirucin modules overexpressed in <i>E. coli</i> following Ni-NTA affinity purification.....	54
Figure 24. Schematic diagram of intermodular transfer in modular polyketide synthases.....	70
Figure 25. Schematic diagram of a mechanism-based, fluorescence transfer assay for KS- selectivity.....	73
Figure 26. SDS PAGE analysis of fluvirucin B ₁ PKS domains overexpressed in <i>E. coli</i> following Ni-NTA affinity purification.....	75

Figure 27. Schematic diagram of and observed results for the AT substrate selectivity studies of modules 1, 3, and 5.	76
Figure 28. LC–MS data for module 3 KS-acylation with (R)-3-hydroxybutyryl-SNAc (top) and (S)-3-hydroxybutyryl-SNAc (bottom).....	78
Figure 29. Mechanism based fluorescence transfer assay for Flu KSAT 3 substrate specificity towards (R)-and (S)-3-hydroxy butyryl SNAc.....	79
Figure 30. A panel of substrate analogs for KS selectivity.....	80
Figure 31. Panel of SNAc thioesters for determination of KS substrate selectivities in the fluvirucin PKS.....	97
Figure 32. Crystal structure of the DEBS module 3 KS active site pocket (PDB: 2QO3)	98
Figure 33. Consensus sequences for each of the four KS active site pocket residues of interest (see Figure 32)	99
Figure 34. The plasmid map of pTL-F01b.....	105
Figure 35. The plasmid map of pTL-F02.....	106
Figure 36. The plasmid map of pTL-F03p.....	107
Figure 37. Schematic diagram of our domain deletion-based engineering strategy for fluvirucin analog production	111
Figure 38. Structures of the thirteen fluvirucin analogs possible from single domain inactivations	113

CHAPTER 1.
INTRODUCTION OF POLYKETIDES, POLYKETIDE SYNTHASE
AND ENGINEERING OF POLYKETIDE SYNTHASE

1.1 Background

The steady rise of drug-resistant bacteria poses one of the most serious health risks in the 21st century. Deaths due to drug-resistant infections has raised the attention of most public health programs in the world, not only since deaths could have impact on human societal resources but also treating drug-resistant infections are far more costly for economic reasons.¹ Rapidly decreasing numbers of viable antibiotic leads in the pharmaceutical pipeline demand new, sustainable methods for improved drug discovery and development. Despite vast improvements in *de novo* drug design and target recognition, Nature remains the richest source of small-molecule therapeutics. Among many natural products, polyketides are among the most promising for developing new antibiotic leads.

1.2 Polyketides

Polyketides are a broad class of structurally diverse and biologically interactive natural products which are produced by bacteria, fungi, and plants. They exhibit a wide range of pharmacologically properties including antibiotic (erythromycin and its derivatives), anticancer (epothilone B), and immunosuppressant (rapamycin) among others (**Figure 1**).²

Since their discovery, polyketides have received enormous attention from the scientific community. It is estimated that there are about 10,000 known natural polyketides, and 1% of them are potent compounds, a significantly higher hit rate than is generally observed for other natural product classes (<0.001%). Given that polyketide-derived pharmaceuticals comprise 20%

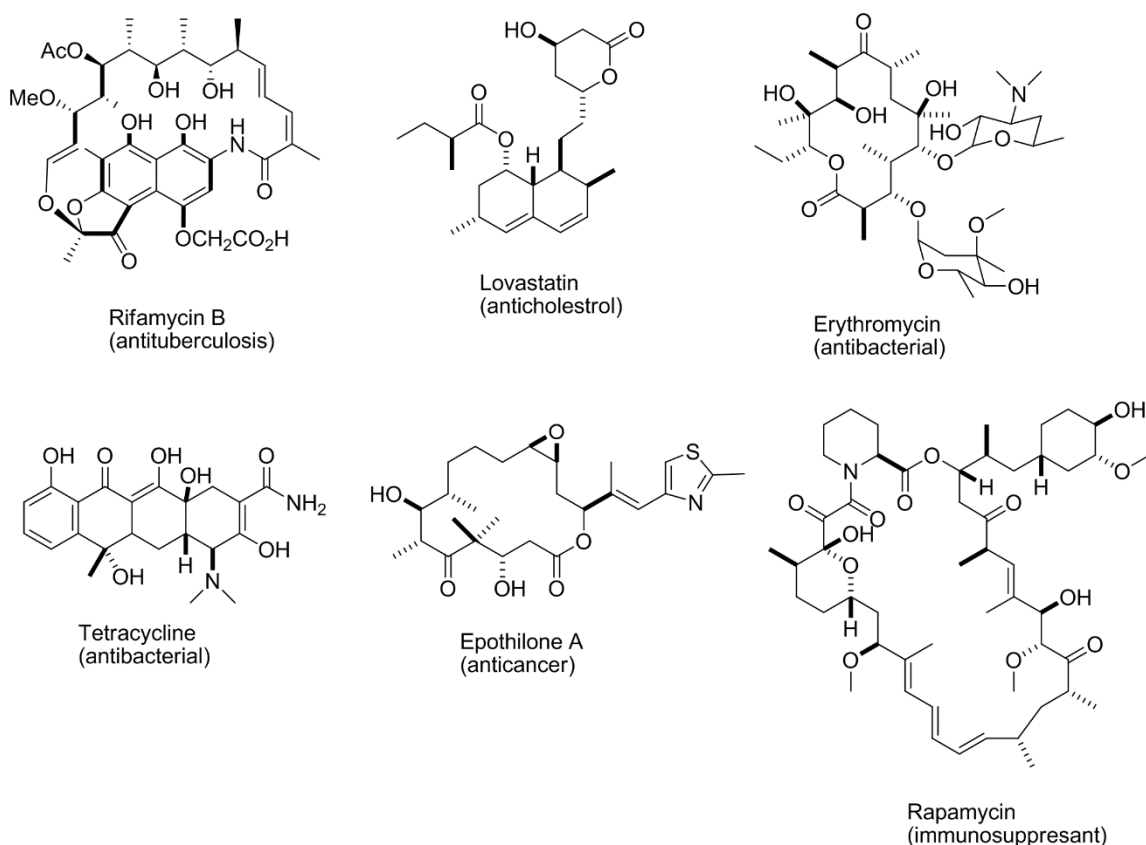


Figure 1. Representative polyketide structures. Rifamycin, Erythromycin, and Tetracycline are antibiotics. Lovastatin is a HMG-CoA reductase inhibitor (cholesterol lowering). Epothilone is an antitumor compound. Rapamycin is an immunosuppressant.

of the top-selling drugs in the world, it is very reasonable to develop viable therapeutic leads based on these compounds. However, due to the structural complexity of polyketides, they are generally difficult to be synthesized chemically and also troublesome to be produced naturally since the producing organisms are often hard to culture.^{3,4}

Despite all the difficulties with making polyketides, an alternative methodology has been proposed once after the natural machinery for making polyketides was discovered. Previous research has shown that polyketides are biosynthesized by repeated condensations from simple acyl-derived metabolites followed by varying degrees of reductive modification to render their structural and pharmacological diversity. Therefore, understanding the mechanisms involved in

polyketide biosynthesis can provide the necessary knowledge for manipulating these natural enzyme systems to produce novel metabolites. The enzymes responsible for making polyketides are known as polyketide synthases, and the mechanism of polyketide biosynthesis is very similar to the biosynthesis of fatty acids.

1.3 Polyketide Synthases

Polyketide synthases (PKSs) can be categorized into three different types based on the protein architecture and biosynthetic schemes. Type I PKSs, also called modular PKSs, are composed of multiple catalytic domains that are covalently linked together and they are organized as modules to produce macrolides with highly varied structures. Type II PKSs are multienzyme complexes assembled by stand-alone monofunctional enzymes, which utilize the same catalytic center iteratively to produce a polyketone chain which can be further processed into highly aromatized or oxidized products. Type III PKSs, also known as chalcone synthase-like PKSs, are homodimeric enzymes that can catalyze the condensation reaction iteratively to produce small aromatic molecules without acyl-carrier domains. This dissertation is mainly concerned with components from type I PKSs.

Modular PKSs, whose configurations are reminiscent of an industrial assembly line, contain various modules which are composed of similar domains, and the precise composition and arrangement determine the molecular make-up of the polyketide products. Each module represents a workstation in the assembly line and each domain functions as a worker at that workstation. Domains in a PKS module are responsible for catalyzing individual reactions that not only dictate the chain length of the polyketide but also control the functional groups present on the polyketide backbone. Hence, the linear setup of modular PKS systems provides the alternative

approach, which is to manipulate the modular PKS systems to produce novel metabolites like constructing an industrial assembly line for producing new product.

The ability to manipulate natural systems to produce novel metabolites rests largely on increased mechanistic understanding of how these molecules are generated in the first place. Over the past 20 years, a number of research groups have focused on genetic, biochemical, and structural properties of PKSs.^{5,6} It is well established that modular PKSs utilize simple extender units derived from malonyl-coenzyme A to assemble polyketides in a linear fashion. Among several characterized PKSs, the biosynthesis of the polyketide core of erythromycin A, 6-deoxyerythronolide B (6-dEB) (Error! Reference source not found.), has provided the paradigm for understanding the structure and function of the PKSs that are responsible for assembling complex polyketides.⁷

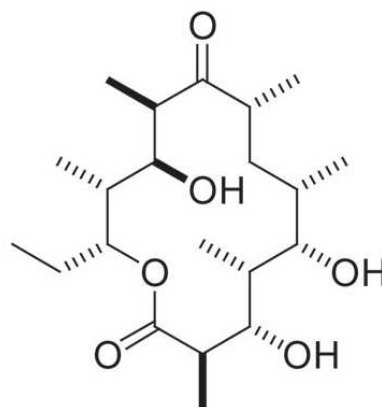


Figure 2. Structure of 6-dEB.

1.3.1 DEBS for Elucidating the Mechanisms of Polyketide Synthase

The 6-deoxyerythronolide B synthase (DEBS), which catalyzes the formation of 6-dEB, consists of three large subunits, DEBS1, DEBS2 and DEBS3, each containing two modules greater than 300 kD in size. A module is defined as the collection of enzymes required for elongation and processing of each intermediate structure. There are 2 domains in the N-terminal loading module, responsible for priming the synthase with a propionate starter unit, and 26 domains in the six extender modules. Each extender module contains at least three essential domains: a ketosynthase (KS), an acyl transferase (AT), and an acyl carrier protein (ACP). The AT domain selects the appropriate carbon extender unit and transfers the units from an acyl-CoA to the phosphopantetheine (Ppant) arm of ACP. The KS domain accepts the polyketide chain from the

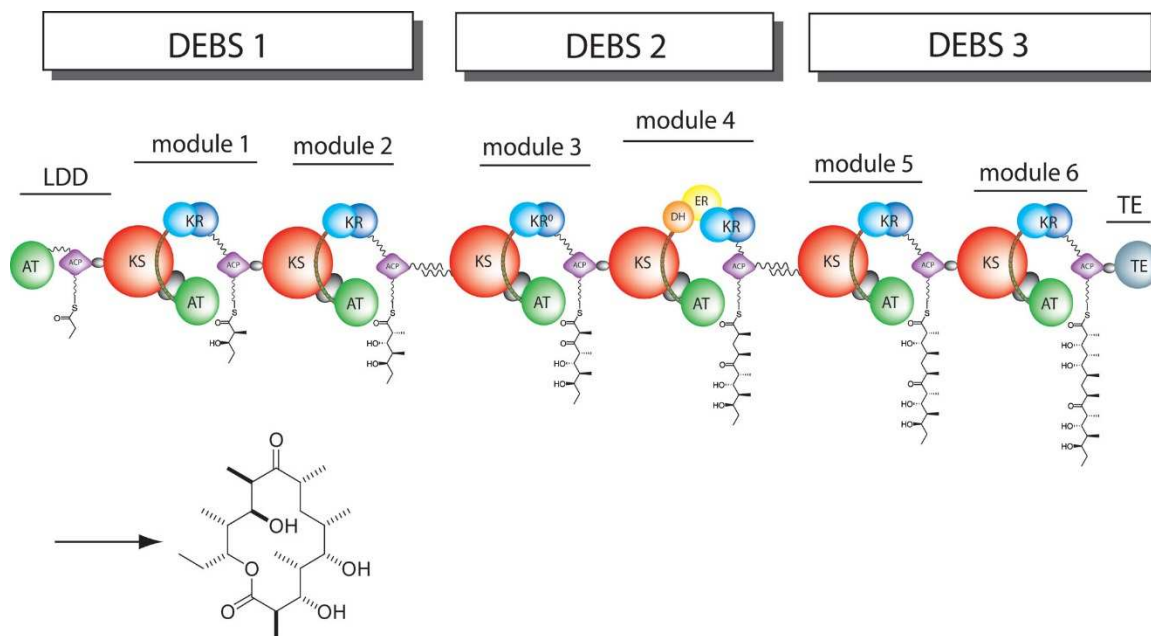


Figure 3. Schematic diagram of 6-deoxyerythronolide B synthase (DEBS). DEBS contains three polypeptide units each consisting of two modules which add two carbons to the polyketide backbone, as illustrated. A loading didomain (LDD) delivers the propionate starter unit to module 1. The thioesterase (TE) domain is responsible for cyclization/release of the final product. KS = ketosynthase, AT = acyl transferase, ACP = acyl carrier protein, KR = ketoreductase, DH = dehydratase, ER = enoyl reductase.

previous module and catalyzes chain elongation reaction by adding an ACP-bound extender unit through decarboxylative condensation (**Figure 3**).

After the extender unit is added, it can be further processed by optional tailoring domains, including ketoreductases (KRs), dehydratases (DHs), and enoyl reductases (ERs), to yield a hydroxyl, enoyl, or methylene group at the β -position. Finally, the thioesterase (TE) domain that is located at the C-terminus of DEBS module 6 promotes the macrocyclization event which releases the final product, 6-dEB. From this mechanistic insight, it is clear that the order of catalytic domains in a PKS module determines the structure and processing of the extender unit, while the number of PKS modules dictates the size of the polyketide chain. The mechanistic logic behind polyketide biosynthesis bodes well for rational manipulation of product structure. As a

result, several groups have shown that genetic engineering of polyketide biosynthetic pathways provides an attractive means of producing new polyketides.

1.4 Heterologous Expression and Genetic Engineering of Polyketide Synthase Genes

As mentioned above, modular PKS systems like DEBS provide ways to modify any portion of the polyketide through altering the corresponding domains or modules on the genetic level. However, several aspects need to be taken into account to make new polyketides. Firstly, one needs to demonstrate that the expression of large multi-domain proteins is not hindered by the host cell machinery, and secondly, that the post-translational modification of ACP can be properly loaded by the phosphopantetheinyl transferase in the host organism. Thirdly, some precursor biosyntheses are not universally available (methyl- and ethyl- malonyl CoA) and they may hinder the polyketide production. Lastly, one needs to consider the transportation of the precursor and product recovery.

The DEBS genes, isolated from the bacterium *Sacchapolyspora erythraea*, were cloned independently by two groups in the early 1990s,^{8,9} and were the first set of polyketide genes expressed in a heterologous host. The DEBS system was expressed in *Streptomyces coelicolor* CH999, a strain containing a chromosomal deletion of the entire actinorhodin biosynthetic gene cluster, and was used as the host strain to prove the idea of heterologous production.¹⁰ Later, 6-dEB was further produced in an engineered *E. coli* strain, BAP1, which was engineered from an *E. coli*-strain BL21 to encode the *sfp* gene, a promiscuous phosphopantetheinyl transferase from *B. subtilis*. The discovery of the *sfp* gene solves the problem of the posttranslational modification required for the ACP domain.³ Since then BAP1 is the primary *E. coli* strain serving as the heterologous host for expressing polyketide synthase genes.

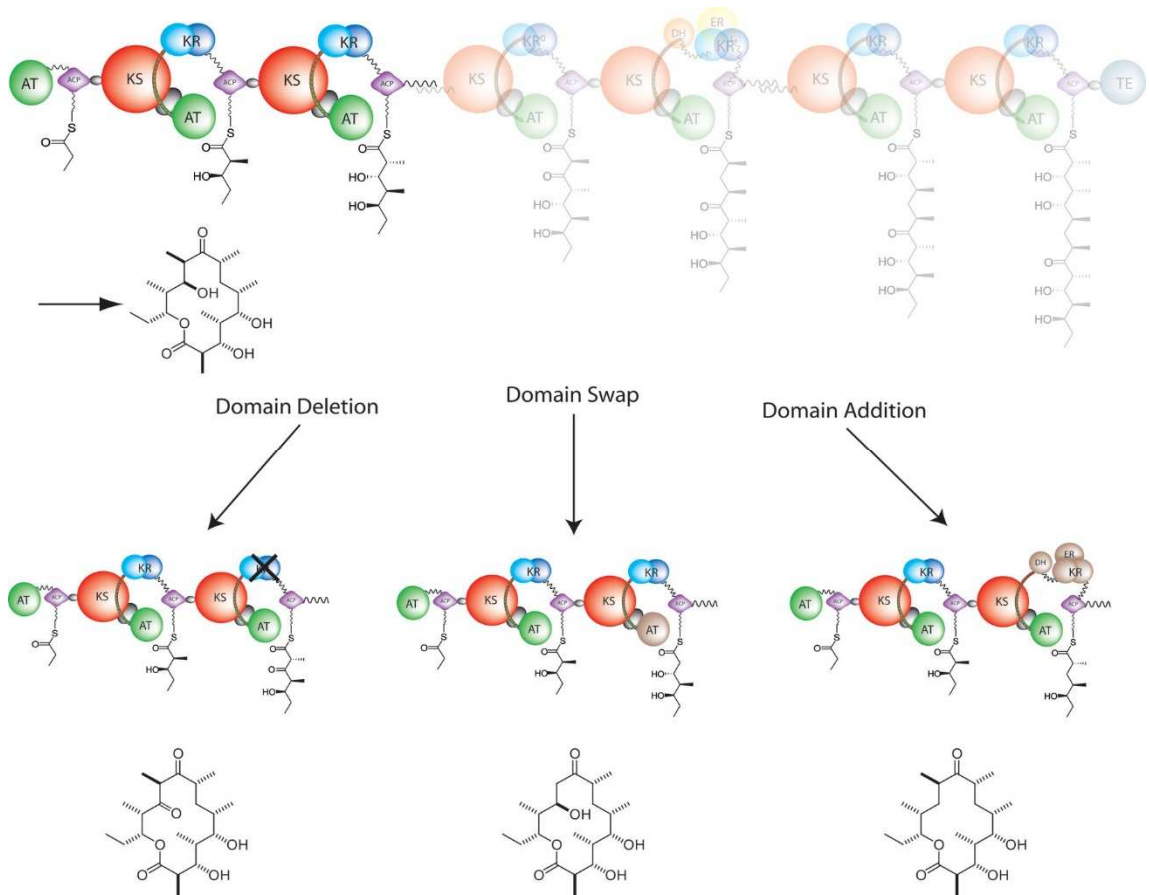


Figure 4. Schematic diagram of genetic engineering strategy for producing novel compounds from DEBS1 modules

To date, a number of PKSs have been cloned, analyzed, and engineered by recombinant DNA technology to make novel “unnatural” products. In DEBS, inactivation of specific domains has been accomplished through site-specific mutagenesis. Additionally, domain deletions, additions, and swaps have been achieved via genetic engineering efforts (**Figure 4**).¹¹ Because the polyketides synthesized by these modular PKSs resemble each other, strategies have been developed whereby modules of one PKS would recognize and process the structurally related intermediates from other PKSs. To elaborate, they can be applied into:

- 1. Manipulation of polyketide chain length:** For reducing chain length, it was reported that unnatural triketide and hexaketide lactones could be biosynthesized by a

deletion mutant comprising DEBS modules 1 and 2 + TE and modules 1 to 5 + TE respectively, indicating that TE can act on a variety of chain lengths (C6–C15)¹². For extending chain length, four different novel octaketides were biosynthesized by adding rapamycin module 2 or 5 between DEBS modules 1 and 2.¹³

- 2. Manipulation of building blocks:** As previously mentioned, incorporation of the carbon extender unit is determined by the AT domain. As a result AT is the most common domain engineered for domain replacement, in which the native AT domain is replaced with another AT with different starter or extender unit specificity. For example, nearly all six of the methylmalonyl-specific AT domains in DEBS have been successfully replaced by malonyl-specific AT domains from the rapamycin PKS to generate 6-dEB or erythromycin analogs lacking the corresponding methyl branches.¹⁴⁻¹⁷
- 3. Manipulation of the β -carbon processing activities:** In this category, all of the strategies mentioned above have been applied in DEBS manipulations. More specifically, inactivation of the ER domain in module 4 can generate the olefin at C-6-7 of the erythromycin derivative.¹⁸ Deletion of the KR domain in module 5 generates a C-5 keto derivative of 6-dEB¹⁹ and the swap of the KR domains in modules 2, 5, and 6 with rapamycin DH+KR produces C-10-11, C-4-5, and C-2-3 olefin derivatives of 6-dEB.²⁰ Similarly, the replacement of KR domain in modules 2 with a DH+ER+KR unit results in C-10 deoxy 6-dEB analogs.¹⁶

Furthermore, combining heterologous PKS domains into DEBS can create functional hybrid PKSs and result in multiple modifications of 6-deoxyerythronolide B as well as generating over 100 novel analogs.^{16, 21} The approaches further indicate that the modules of DEBS can be

altered separately, expressed on different plasmids and combined together to produce unnatural polyketide products.

1.5 Polyketide Synthase Engineering Hurdles

However, PKS engineering is not always successful. For example, the target compounds of domain addition strategies have only about 3–5% production efficiency compared to the normal polyketide products. In fact, the parent “unextended” polyketide is the major product because the extension module is bypassed within the DEBS multienzyme system. In an attempt to alter AT domains, McDaniel and coworkers failed to engineer a malonyl-AT domain in DEBS module 4 by using several different domain junctions.²² In the multiple modification experiments, the yields of the modified products dropped significantly when the number of combined mutations increased and, discouragingly, only 5 of the 21 triple mutant combinations reached detectable levels of polyketides, which may mean that multiple modifications on the polyketide structure are generally disfavored with this system.

Previous efforts have pointed out that the substrate selectivity and protein-protein interactions within a module or between modules play critical roles in polyketide biosynthesis. Because the strategies described above often require making changes in module composition to achieve the desired polyketide structure modifications, it is likely that the diminishing yields come, at least in part, from the disruption of 3-dimensional protein architecture.

1.6 Hypothesis

Successful biosynthetic engineering efforts will undoubtedly require methods which do not disrupt native protein architectures. The idea was bolstered by comparison of the inactivation and deletion of the KR6 domain in DEBS. A point mutation at the catalytic serine residue gave only the targeted analog, whereas the deletion of the whole domain affected the specificity of the

adjacent AT domain and led to unexpected products.²³ *To achieve this, techniques must be implemented where maximum product diversity can be achieved through manipulation of existing enzyme activities without introducing heterologous components.* However, many fashionable assemblies lack the elaborate modular compositions necessary to access the full range of functionalities. For instance, DEBS module 4 is the only module in that system containing all three tailoring domains (KR, DH, and ER) similar to mammalian fatty-acid synthase (FAS).²⁴⁻²⁸ All others possess, at most, a single KR which limits the possible chemical outcomes to either the naturally-formed hydroxyl moiety, or a ketone functionality if the KR is inactivated. In contrast, module 4 is capable of producing methylene, olefin, hydroxyl, and ketone groups simply by mutating active-site residues while keeping the overall 3-dimensional structure intact (**Figure 5**). With this in mind,

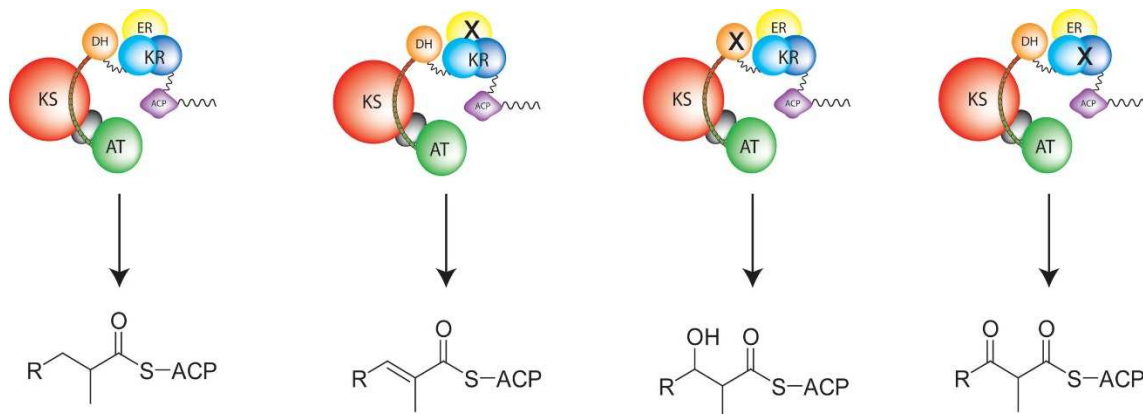


Figure 5. Schematic diagram of a mutational inactivation strategy for production of novel compounds from modules containing the full complement of β -processing domains. Alterations in chemical structure are designed at the genetic level

our strategy relies on discovery and characterization of PKS systems harboring KR, DH, and ER domains in the majority, if not all, of the active modules to maximize product diversification. As these synthases naturally produce less pharmaceutically-desired, hydrophobic molecules, they have generally received little attention so the PKS modules with the full array of tailoring domains (KR, DH, and ER) are relatively rare amongst PKSs studied to date. The content of this dissertation

will establish these biosynthetic machines as a powerful means to generate increased product diversity and, ultimately, improved small-molecule drugs.

Recognizing that the flow of intermediates along a given PKS requires, in large part, effective communication between neighboring enzymes, many groups have focused on determining the associated recognition motifs that govern these interactions.²⁹⁻³¹ One of the current research thrusts in the field involves redesigning individual modules by swapping primary sequence information, including these recognition motifs, in an effort to improve enzyme kinetics. Although successful implementations of this approach are beginning to emerge, we thought that there must be a way of obtaining product diversity without having to reinvent what Nature has already provided. To do so, a strategy where product diversification is solely the result of domain inactivation must be implemented. By leaving all recognition interfaces in their natural forms, the problem of producing and shuttling intermediates along the synthase is reduced to one of substrate selectivity. *In addition to delivering improved yields of engineered products, such a system will, for the first time, provide a systematic means of assessing the relative contribution of substrate selectivity to the flow of structures from one module to the next in a full-length PKS.*

In summary, the detail of this dissertation provides two vital missing pieces to the polyketide engineering puzzle: 1) A unique PKS where a wide variety of polyketide structures can be generated without introducing heterologous domains or linkers and 2) a facile method for examining ketosynthase substrate tolerance using simple, and readily obtainable, thioester substrates. The former takes advantage of the fact that the FAS-like PKS modules produce β -keto, β -hydroxyl, and β -olefin functionalities on the way to the final β -methylene product. Therefore, to access these desired intermediate structures, one must simply “turn-off” the appropriate tailoring domain within the module via site-directed mutagenesis. Most importantly, all protein-protein interactions and native linkers remain intact, thus allowing for direct determination of the

relative impact of chain transfer (in other words, KS substrate selectivity) on engineered polyketide production.

1.7 Reference

1. Roberts, R. R., Hota B., Ahmad I., Scott RD II, Foster S.D., Abbasi F., Schabowski S., Kampe L. M., Ciavarella G. G., Supino M., Naples J., Cordell R., Levy S. B., Weinstein, R. A., (2009) Hospital and societal costs of antimicrobial-resistant infections in a Chicago teaching hospital: implications for antibiotic stewardship. *Clin. Infect. Dis.* 49, 1175-1184.
2. Wawrik B, Kerkhof L, Zylstra G. J., Kukor J. J. (2005) Identification of Unique Type II Polyketide Synthase Genes in Soil. *Applied and Environmental Microbiology*, 71, 2232-38.
3. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E-coli*, *Science*, 291, 1790-1792.
4. Rohr J., (2000), A New Role for Polyketides, *Angew Chem Int Ed*, 39 (16), 2847-2849.
5. Weissman K. J, Leadlay P. F. (2005) Combinatorial Biosynthesis of Reduced Polyketides, *Nature Reviews Microbiology*, 3, 925-936.
6. Khosla, C. (2009) Structures and Mechanisms of Polyketide Synthases, *J. Org. Chem.*, 74 (17), 6416–6420
7. Khosla, C., Tang, Y., Chen A. Y., Schnarr N. A., and Cane D. E. (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase, *Annu. Rev. Biochem.* 76, 195-221.
8. Donadio, S., Katz, L. (1992) Organization of the Enzymatic Domains in the Multifunctional Polyketide Synthase Involved in Erythromycin Formation in *Saccharopolyspora erythraea*, *Gene*, 111(1): 51 – 60.

9. Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F. (1990), An Unusually Large Multifunctional Polypeptide in the Erythromycin-producing Polyketide Synthase of *Saccharopolyspora erythraea*, *Nature*, 348, 176-178
10. Kao, C. M., Katz, L., and Khosla, C. (1994) Engineered biosynthesis of a complete macrolactone in a heterologous host, *Science*, 265, 509-512.
11. Park SR, Han A.R, Ban Y.H, Yoo Y.J., Kim E.J., Yoon Y.J. (2010) Genetic engineering of macrolide biosynthesis: past advances, current state, and future prospects, *Appl Microbiol Biotechnol*, 85, 1227–1239
12. Kao CM, Luo G, Katz L, Cane DE, Khosla C. (1996) Engineered biosynthesis of structurally diverse tetraketides by a trimodular polyketide synthase. *J Am Chem Soc*, 118, 9184–9185
13. Rowe CJ, Böhm IU, Thomas IP, Wilkinson B, Rudd BA, Foster G, Blackaby AP, Sidebottom PJ, Roddis Y, Buss AD, Staunton J, Leadlay PF (2001) Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem Biol*, 8, 475–485
14. Katz L, McDaniel R. (1999) Novel macrolides through genetic engineering, *Medicinal Research Reviews*, 19, 543-558.,
15. Ruan X, Pereda A, Stassi DL, Zeidner D, Summers RG, Jackson M, Shivakumar A, Kakavas S, Staver MJ, Donadio S, Katz L. (1997) Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J. Bacteriol*, 179, 6416-6425.
16. Liu L, Thamchaipenet A, Fu H, Betlach M, and Ashley M., (1997) Biosynthesis of 2-Nor-6-deoxyerythronolide B by Rationally Designed Domain Substitution, *J. Am. Chem. Soc.*, 119, 10553-10554.

17. Petkovic H, Lill RE, Sheridan RM, Wilkinson B, McCormick EL, McArthur HA, Staunton J, Leadlay PF, Kendrew SG. (2003) A novel erythromycin, 6-desmethyl erythromycin D, made by substituting an acyltransferase domain of the erythromycin polyketide synthase. *J Antibiot (Tokyo)*, 56, 543-51.
18. Donadio S., McAlpine J.B., Sheldon P.J., Jackson M., and Katz L. (1993) An Erythromycin Analog Produced by Reprogramming of Polyketide Synthesis, *Proc. Natl. Acad. Sci. USA*, 90 (15), 7119-7123
19. Donadio S., Staver M.J, McAlpine J.B., Swanson S.J., and Katz L. (1991) Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science*, 252(5006), 675-679.
20. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. (1999) Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel "Unnatural" Natural Products, *Proc. Natl. Acad. Sci. USA*, 96, 1846-1851
21. Xue, Q., Ashley, G., Hutchinson, C., and Santi, D. (1999) A multiplasmid approach to preparing large libraries of polyketides, *Proc. Natl. Acad. Sci. USA*, 96, 11740-11745
22. Reeves, C.D., Murli, S., Ashley, G.W., Piagentini, M., Hutchinson, C.R., McDaniel. R. (2001) Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry*, 40, 15464-470.
23. Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R.; McDaniel, R. (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases, *Biochemistry*, 42, 72-79.
24. Smith, S.; Tsai, S-C. (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.*, 24, 1041-1072.

25. Smith, S. (1994) The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J.* 8, 1248-1259.
26. Asturias, F. J.; Chadick, J. Z.; Cheung, I. K.; Stark, H.; Witkowski, A.; Joshi, A. K.; Smith, S. (2005) Structure and molecular organization of mammalian fatty acid synthase. *Nat. Struct. Mol. Biol.* 12, 225-232.
27. Smith, S. (2006) Architectural options for a fatty acid synthase. *Science*, 311, 1251-1252
28. Maier, T.; Jenni, S.; Ban, N. (2006) Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science*, 311, 1258-1262.
29. Kapur, S., Chen, A.Y., Cane, D.E., Khosla, C. (2010) Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. USA*, 107, 22066-22071.
30. Tran, L., Broadhurst, R. W., Tosin, M., Cavalli, A., and Weissman, K. J. (2010) Insights into protein-protein and enzyme-substrate interactions in modular polyketide synthases. *Chem. Biol.* 17, 705-716.
31. Chandran, S.S., Menzella, H.G., Carney, J.R., Santi, D.V. (2006) Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chem. Biol.*, 13, 469-474.

CHAPTER 2.

THE NEW TOOL FOR PKS ENGINEERING: FLUVIRUCIN B₁ PKS

AND GENOMIC INFORMATION ANALYSIS

2.1 Introduction

Over the past several decades, researchers have devised numerous methods for manipulating PKS processes in an effort to diversify product structure and, ultimately, biological activity¹⁻⁵. Although some success has been realized toward this end, decreased yields of engineered products have limited the scope and efficacy of these methods. At this point it is clear that the two primary factors leading to low yields are substrate selectivity of downstream enzymes and disruption of protein-protein interactions when heterologous enzymes are introduced.⁶⁻¹¹ To circumvent the latter, strategies are needed where genetic alterations can be introduced without disrupting the native three-dimensional PKS architecture.

For modular polyketide synthases this means unearthing assemblies which bear the full complement of tailoring domains (KR, DH, and ER), similar to mammalian FAS, in most, if not all, active modules. Carefully executed mutagenesis of key active site residues should result in all possible β -functionalities while leaving the native protein-protein interactions intact (**Figure 6**).

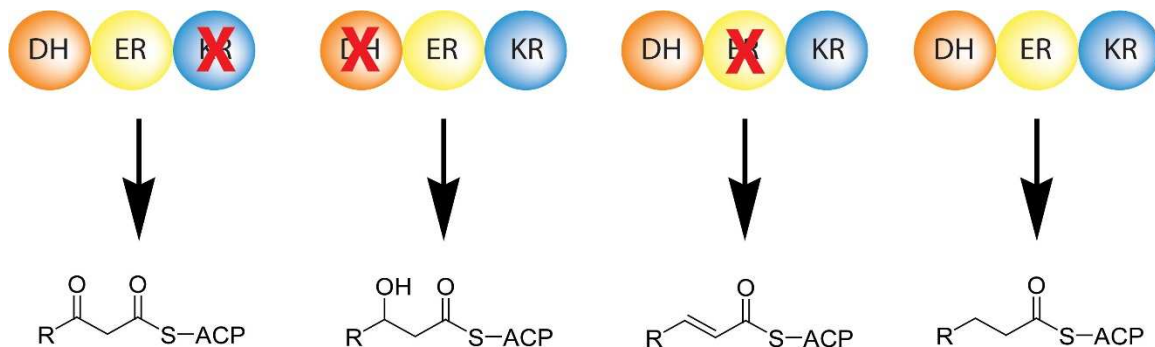


Figure 6. Schematic diagram of a mutational inactivation strategy for production of novel compounds from fatty acid-like PKS modules. Alterations in chemical structure are designed at the genetic level.

Therefore, PKSs that produce largely unfunctionalized polyketides (i.e. methylenes at most β -positions) may provide optimal engineering potential.

2.2 Fluvirucin B₁ PKS

Fluvirucin B₁ is 14-membered macrolactam produced by *Actinomadura vulgaris* with moderate to good antifungal and antiviral

activities (**Figure 7**).¹²⁻¹⁴ Following assembly

of the core macrocycle, a single 3-amino-

3,6-dideoxy-L-talopyranose is appended to

the sole exocyclic hydroxyl group. The lack

of additional ring functionalities peaked our

interest as we hypothesized that nearly all

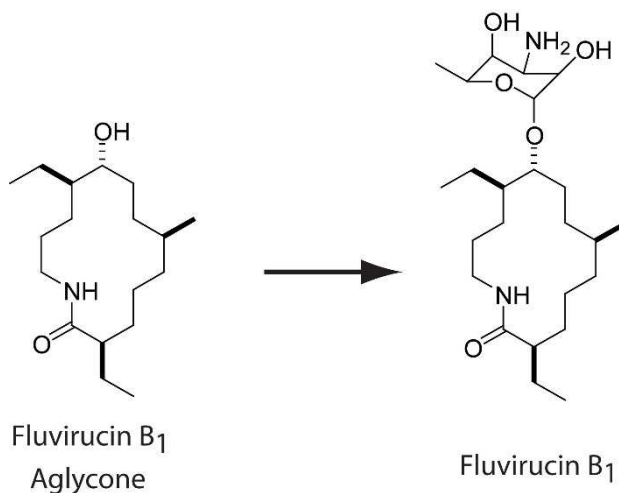


Figure 7. Structure of fluvirucin B₁ aglycone and

fluvirucin B₁

active modules should contain the FAS-like domain organization where KR, DH, and ER are all present. If one were to consider PKS systems as evolutionarily connected to FAS, the fluvirucin synthase may represent an early link between the two. As a result, we were motivated to explore the biosynthetic origin of fluvirucin B₁ with the ultimate goal of providing a platform for polyketide engineering that circumvented the need for incorporation of heterologous domains to achieve maximal product diversity.

2.3 Results and Discussions

2.3.1 Hypothesis

From the core structure of fluvirucin B₁, we hypothesized that the producing PKS would consist of five extender modules assuming that a β -alanine derivative is used as a starter unit in the process. The sole hydroxyl group would therefore arise from a module harboring only a KR domain while all other extender modules would contain the full compliment of KR, DH, and ER domains (**Figure 8**). Based on the positions of macrolactam ring substituents, we expected that: (1) the first and last modules contained ethylmalonyl-specific AT domains, (2) the second and fourth modules incorporated malonyl groups, and (3) the third module utilized methylmalonate (**Figure 8**). Finally, ring closure was most likely achieved via a C-terminal thioesterase (TE) domain as is the case with similar macrocyclic polyketides.¹⁵⁻¹⁷ To test these hypotheses and determine

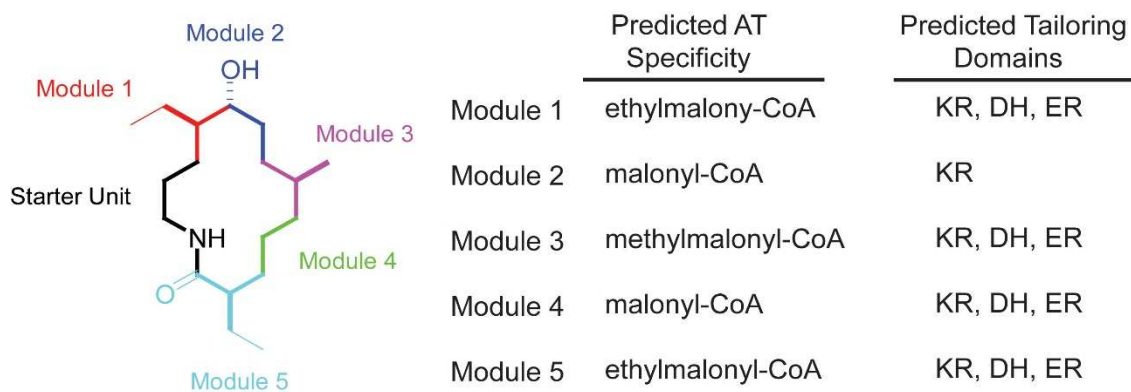


Figure 8. Predicted tailoring domains and AT selectivities for each fluvirucin B₁ synthase module based on the fluvirucin B₁ aglycone structure

the precise arrangement of enzymes within the assembly, we set out to identify and characterize the fluvirucin B₁ biosynthetic gene cluster.

2.3.2 Genomic Sequencing and Bioinformatic Analysis

To do so, the producing organism, *Actinomadura vulgaris*, was cultured following published procedures.¹²⁻¹⁴ Genomic DNA was isolated and sequenced (Beckman Coulter Genomics) affording 436,311 overlapping sequence fragments. These sequences were partially assembled resulting in 444 consensus sequences ranging in size from 5000 to 170,000 base pairs. We then utilized bioinformatic analyses to identify open reading frames that containing PKS proteins or domains *in silico* translation. The relatively large size of PKS constructs allowed us to quickly identify potential hits by searching each assembled sequence for open reading frames of

at least 4000 base pairs. Our search identified several PKS gene clusters, one of which contained the expected size and module composition of the proposed fluvirucin PKS.

For the proposed fluvirucin PKS, we initially found 5 contigs (33, 60, 1912, 3721, 3776) partially covering the putative PKS gene cluster (**Figure 9**). As it shows in **Figure 9**, the putative

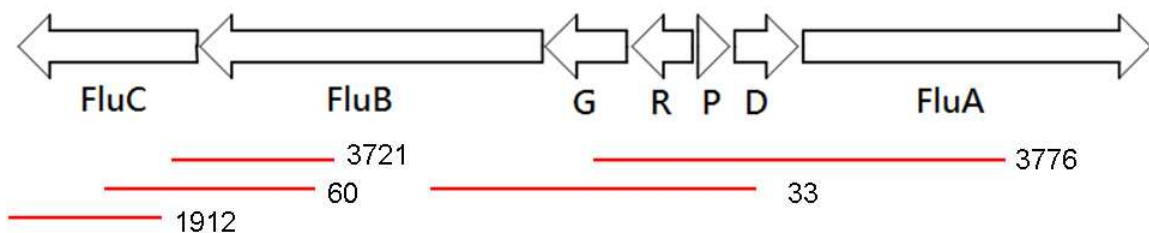
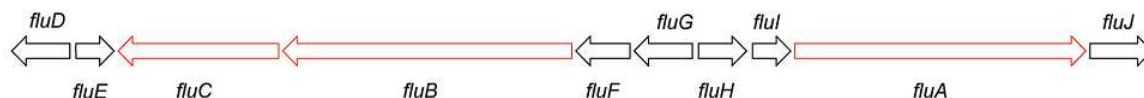


Figure 9. Putative gene cluster of fluvirucin PKS. FluA: module 1 and 2; FluB: module 3 and 4; FluC: module 5; D: Decarboxylase; G: Glycosyl tranferase; R: PKS regulator; P: proline iminopeptidase

gene map, it consists of two uncertain areas to be further investigated. For the first gap between the contig 3721 and 33, we performed a genomic PCR experiment to obtain the missing sequence. Two primers were designed for this purpose: forward primer (ATCTCGACTGCTACGCATCC) is at the putative DH domain sequence of module 3; reverse primer: (CCCGAAGTTGGATTTGACC) at the putative KS domain in module 4. The PCR experiment was successful, and we isolated an amplicon (4524 bp) and then further sequenced it to confirm that the missing gap was indeed flanking by the contig 3721 and 33. For the second gap, we conducted a less stringent search within the contigs we established, and we found that the contig 112 containing a putative crotonyl-coa reductase and some similar sequence to module 2 within contig 3776. We then manually adjusted

the sequence information based on the frequency and the degree of disagreement, and we proposed the new putative gene cluster of fluvirucin B₁ PKS (**Figure 10**).



Summary of the <i>flu</i> Genes and Comparison with Database					
Protein	Amino Acids	Proposed function	Sequence similarity (Protein, Origin)	Identity/Similarity	Accession No.
FluD	506	MFS-type efflux pump	Efflux pump/drug resistance transporter, <i>Streptomyces</i> sp. MP39-85	49%/67%	ACO94494.1
FluE	299	TetR family transcriptional regulator	Transcriptional regulator of TetR family, <i>Stactebrantdia nassauensis</i>	37%/50%	YP_003510265.1
FluF	425	Glycosyltransferase	Glycosyltransferase, <i>Streptomyces</i> sp. Tu21	46%/67%	ABO28821.1
FluG	351	LysR family transcriptional regulator	Putative transcriptional regulator of LysR family, <i>Streptomyces himastatinicus</i> ATCC 53653	37%/51%	ZP_07292079.1
FluH	317	proline iminopeptidase	Prolyl aminopeptidase, <i>Streptomyces halstedii</i>	67%/82%	ZP_07299775.1
FluI	420	decarboxylase	Decarboxylase, <i>Streptomyces halstedii</i>	62%/75%	BAD08372.1
FluJ	448	crotonyl-CoA reductase	Crotonyl-CoA reductase <i>Streptomyces violaceusniger</i> Tu 4113	74%/85%	ZP_07607278.1
FluA	3969	PKS modules 1 and 2	lasalocid modular polyketide synthase, <i>Streptomyces lasaliensis</i>	53%/65%	CAQ64687.1
FluB	4375	PKS modules 3 and 4	lasalocid modular polyketide synthase, <i>Streptomyces lasaliensis</i>	55%/67%	CAQ64687.1
FluC	2396	PKS module 5	Polyketide synthase, <i>Streptomyces graminofaciens</i>	53%/66%	BAJ16468.1

Figure 10. Gene organization for the fluvirucin B₁ PKS cluster. The table summarizes the gene cluster organization and gene products identified along with comparison to known homologs.

2.3.3 Fluvirucin B₁ PKS Genes

Three modular PKS genes, *flu A-C*, were found to contain an arrangement and composition of domains consistent with the expected fluvirucin PKS assembly (**Figure 10**).

In addition, several PKS-associated genes were uncovered within the cluster including a

pair of transcriptional regulators (*fluE*, *fluG*), a glycosyl transferase (*fluF*), a decarboxylase (*fluI*), and a drug transporter (*fluD*) (**Figure 10**). The *fluJ* sequence shows high similarity to crotonyl-CoA reductases and likely plays a role in ethylmalonate generation as is required for fluvirucin B₁ biosynthesis.

FluA contains modules 1 and 2 of the fluvirucin B₁ assembly. As expected, module 1 has the full complement of tailoring domains (KR, DH, and ER) while module 2 possesses only a KR domain putatively leading to the sole hydroxyl group on the macrolactam ring. A single loading ACP is found at the N-terminus of FluA similar to the vicenistatin PKS which utilizes an α -methyl- β -alanine starter unit.¹⁸

FluB contains modules 3 and 4 of the fluvirucin B₁ PKS. Both modules contain KR, DH, and ER domains as would be predicted from the fluvirucin core structure. Finally, FluC consists of module 5 and a C-terminal TE domain. Module 5 again has all three tailoring domains consistent with the lack of functionality at the corresponding macrolactam ring position. As described in detail below, modules 1, 3, and 5 are extremely similar to each other in terms of primary structure as are modules 2 and 4 but similarities drop significantly when comparisons are made between these two groups (Group1-module 1, 3, 5; Group2-module 2, 4). It is therefore tempting to speculate that

the fluvirucin B₁ synthase is composed of modules arising from two separate evolutionary ancestors; one leading to modules 1, 3, and 5 and another leading to modules 2 and 4.

2.3.4 AT Selectivities

Extender unit selectivity for AT domains provides compelling evidence for any link between a given PKS assembly and its associated polyketide product. Using the SEARCHPKS program developed by Mohanty and coworkers, probable coenzyme A substrates were determined for each of the five putative fluvirucin B₁ synthase AT domains.¹⁹ To our delight, all of the predicted AT domain specificities were consistent with the fluvirucin B₁ core structure (**Table 1**). Specifically, modules 2 and 4 showed high sequence similarity with malonyl-specific AT domains while module 3 was predicted to utilize methylmalonyl-CoA. Modules 1 and 5 returned a single hit for ethylmalonate specificity amidst several methylmalonyl-specific AT domains. This is most likely due to the considerably lower abundance of ethylmalonyl-specific ATs relative to methylmalonyl-specific ones rather than a true indication of relaxed specificity. These results, together with the observed tailoring domain patterns, provided the necessary evidence to link this polyketide synthase with the biosynthesis of fluvirucin B₁.

Table 1. Expected AT Specificities : Results from SEARCHPKS

Mod 1

Matching all known AT domains with your sequence

Query is the potential *AT* in your sequence

DOMAIN	E-VALUE	%IDENTITY	%POSITIVES	Substrate	ActivesiteMotif
Your_AT_sequence					QQGHSQGRSHTNV
nidda mod04 AT	7e-85	53	65	Methylmalonate	QQGHSQGRSHTNV
ampho mod01 AT	1e-81	49	62	Methylmalonate	QQGHSQGRSHTSV
nidda mod05 AT	3e-81	48	63	Ethylmalonate	QQGHSQGRGHTNV
rifam mod07 AT	7e-81	50	62	Methylmalonate	QQGHSQGRSHTNV
rifam mod04 AT	2e-80	50	63	Methylmalonate	QQGHSQGRSHTNV
ampho mod11 AT	1e-79	48	63	Methylmalonate	QQGHSQGRSHTSV
pimar mod07 AT	3e-79	50	62	Methylmalonate	QQGHSQGRSHTNV
averm mod01 AT	4e-79	50	63	Methylmalonate	QQGHSQGRSHTNV
averm mod07 AT	4e-79	50	63	Methylmalonate	QQGHSQGRSHTNV
nysta mod01 AT	4e-78	49	61	Methylmalonate	QQGHSQGRSHTNV

Mod2

Matching all known AT domains with your sequence

Query is the potential *AT* in your sequence

DOMAIN	E-VALUE	%IDENTITY	%POSITIVES	Substrate	ActivesiteMotif
Your_AT_sequence					QQGHSLGRFHNV
averm mod08 AT	4e-87	56	63	Malonate	QQGHSLGRFHAQV
averm mod04 AT	9e-87	56	63	Malonate	QQGHSLGRFHAQV
averm mod02 AT	1e-86	57	64	Malonate	QQGHSLGRFHAQV
averm mod03 AT	1e-86	56	63	Malonate	QQGHSLGRFHAQV
averm mod05 AT	3e-86	56	63	Malonate	QQGHSLGRFHAQV
ampho mod13 AT	3e-82	49	65	Malonate	QQGHSIGRFHTV
nysta mod13 AT	6e-79	48	63	Malonate	QQGHSIGRFHTV
ampho mod03 AT	8e-79	47	63	Malonate	QQGHSIGRFHNV
nysta mod04 AT	3e-78	50	65	Malonate	QQGHSIGRFHNV
nysta mod03 AT	9e-78	47	63	Malonate	QQGHSIGRFHNV

Mod3

Matching all known AT domains with your sequence

Query is the potential *AT* in your sequence

DOMAIN	E-VALUE	%IDENTITY	%POSITIVES	Substrate	ActivesiteMotif
Your_AT_sequence					QQGHSQGRSHTNV
pimar mod07 AT	e-104	61	73	Methylmalonate	QQGHSQGRSHTNV
ampho mod11 AT	e-101	58	72	Methylmalonate	QQGHSQGRSHTSV
ampho mod01 AT	1e-98	57	70	Methylmalonate	QQGHSQGRSHTSV
nysta mod01 AT	4e-98	58	71	Methylmalonate	QQGHSQGRSHTNV
nysta mod11 AT	7e-98	58	71	Methylmalonate	QQGHSQGRSHTNV
nidda mod04 AT	9e-98	58	68	Methylmalonate	QQGHSQGRSHTNV
sorap mod07 AT	3e-93	55	68	Glycerate	QQGHSQGRSHTNV
olean mod03 AT	1e-92	55	66	Methylmalonate	QQGHSQGRSHTNV
rifam mod03 AT	2e-92	54	69	Methylmalonate	QQGHSQGRSHTNV
olean mod01 AT	3e-92	55	66	Methylmalonate	QQGHSQGRSHTNV

Mod4

Matching all known AT domains with your sequence

Query is the potential *AT* in your sequence

DOMAIN	E-VALUE	%IDENTITY	%POSITIVES	Substrate	ActivesiteMotif
Your_AT_sequence					QQGHSLGR-RTQV
averm mod08 AT	2e-74	52	58	Malonate	QQGHSLGRFHAQV
averm mod04 AT	3e-74	52	58	Malonate	QQGHSLGRFHAQV
averm mod02 AT	7e-74	52	58	Malonate	QQGHSLGRFHAQV
averm mod03 AT	7e-74	52	58	Malonate	QQGHSLGRFHAQV
averm mod05 AT	2e-73	52	58	Malonate	QQGHSLGRFHAQV
ampho mod13 AT	2e-68	45	60	Malonate	QQGHSIGRFHTHV
nysta mod04 AT	5e-66	46	60	Malonate	QQGHSIGRFHNV
ampho mod03 AT	2e-65	43	58	Malonate	QQGHSIGRFHNV
nysta mod03 AT	7e-65	43	57	Malonate	QQGHSIGRFHNV
pimar mod05 AT	2e-63	44	57	Malonate	QQGHSIGRFHGHV

Mod5

Matching all known AT domains with your sequence

Query is the potential *AT* in your sequence

DOMAIN	E-VALUE	%IDENTITY	%POSITIVES	Substrate	ActivesiteMotif
Your_AT_sequence					QQGHSQGRSHTNV
nidida mod04 AT	7e-85	53	65	Methylmalonate	QQGHSQGRSHTNV
ampho mod01 AT	1e-81	49	62	Methylmalonate	QQGHSQGRSHTSV
nidida mod05 AT	3e-81	48	63	Ethylmalonate	QQGHSQGRGHTNV
rifam mod04 AT	3e-81	50	63	Methylmalonate	QQGHSQGRSHTNV
rifam mod07 AT	7e-81	50	62	Methylmalonate	QQGHSQGRSHTNV
pimar mod07 AT	5e-80	50	62	Methylmalonate	QQGHSQGRSHTNV
ampho mod11 AT	1e-79	48	63	Methylmalonate	QQGHSQGRSHTSV
averm mod01 AT	3e-79	50	63	Methylmalonate	QQGHSQGRSHTNV
averm mod07 AT	3e-79	50	63	Methylmalonate	QQGHSQGRSHTNV
nysta mod01 AT	4e-78	49	61	Methylmalonate	QQGHSQGRSHTNV

2.3.5 KR Domains

The Keatinge-Clay lab previously uncovered primary sequence patterns associated with different types of ketoreductase domains commonly found in modular PKS systems.²⁰ Examination of the fluvirucin KR sequences within this context revealed that all five align with the B1-type KR sequence pattern (**Figure 11**). This KR family is generally observed when the KR works in concert with other tailoring domains such as DH and ER, which is again consistent with the

<u>Module</u>	<u>Loop</u>	<u>Catalytic region</u>	<u>Lid</u>
Ave1	HTAGILDDAT-L	SSAAATFGAPGQANYAAANA	WGTWQNG--LA--DSDKARAYLDRRG
Ty11	HTAGILDDAV-I	SSVTGTWGNAGQGAYAAANA	WGLWGG--GM--AAGAGEESLRRG
Asc8	HTAATLDDGI-L	SSAAAVLGSPPGQNYAAANA	WGMWHTT-STLTGQLDDADRDRIRRGG dh
Ave7	HAAGVLDDAT-I	SSAAGILGSAGQNYAAANA	WGLWEEA-SGMTGHLACTDHRRIRRSG dh
Ave9	HAAGVLDDAT-I	SSAAGILGSAGQNYAAANA	WGLWEEA-SGMTGHLACTDHRRIRRSG DH
Rap10	HTAGVLDDGV-V	SSAAGVLGSAGQNYAVANA	WGLWEDA-SGLTAKLTGTDHDRIRRSG DH
Flu1	HTAGVLDDGV-V	SSSAATLDSAGQNYSAANA	WGLWHQP-SGMSAHLTTTDITRIEQAG
Flu2	HAAGVGHDDVLV	SSGAAVWGSQASNYAAANA	WGWGG--GM--MEDAGE-QLGRSG
Flu3	HTAGVLDDGV-V	SSSAATLDSAGQNYSAANA	WGLWAPETGGMTNQLSELDLRLSRSG
Flu4	HAAGVVDGIV-I	SSSAATLDSAGQNYSAANA	WGVWAPETGGMTRQLGIDLRLSRSG
Flu5	HTAGVLDDGV-V	SSSAGTLDSAGQNYSAANA	WGLWHQP-SGMSAHLTTTDITRIEQAG
	:. .:* :	** :. *_* .*:***	** * : : *

Figure 11. Protein homology analysis of Flu-KR Sequence. KR domain comparison between Fluvirucin modules and other B1 type KR domains. The homology analyses are done by Clustal Omega method of EMBL-EBI website, compared with the results in the reference. (Note: DH, functional; dh, nonfunctional PKS abbreviations: Ave: Avermectin, Tyl: Tylosin, Asc: Ascomycin, Rap: Rapamycin, Flu: Fluvirucin B₁)

expected enzyme composition for fluvirucin PKS modules 1, 3, 4, and 5. Interestingly, the module 2 KR domain also shows B1-type sequence character despite the absence of DH and ER domains from that module. This observation may further highlight the fascinating evolution of the fluvirucin synthase as discussed below. These results, together with the observed tailoring domain patterns and AT selectivities, provided the necessary evidence to link this polyketide synthase with the biosynthesis of fluvirucin B₁.

2.4 Experimental detail

2.4.1 Bacterial strains, culture conditions and DNA purification.

Actinomadura vulgaris was purchased through American Type Culture Collection (ATCC) by the accession number ATCC 53715 and used as the source of DNA for shot-gun sequencing

service and the cloning of Fluvirucin B1 polyketide synthase. The strain was cultivated at ambient temperature in the liquid medium of ATCC Medium 172 (N-Z Amine with Soluble Starch and Glucose), which contains 1% glucose, 2% soluble starch, 0.5% yeast extract, 0.5% N-Z amine type A (Sigma C0626), 0.1% CaCO₃. The growth of *A. vulgaris* in ambient temperature can be observed after 3 days of culture. For genomic DNA extraction purpose, *A. vulgaris* was cultured for 6 days, and then genomic DNA was extracted by using the MasterPure gram-positive DNA purification kit (Epicenter, Madison, WI)

2.4.2 Genomic sequencing and bioinformatic analysis

The genomic DNA of *Actinomadura vulgaris* was collected and sent to Beckman Coulter Genomics for the genomic services as the requested concentration (20mg total). Later, the 436,311 sequence fragments of genomic DNA were sent back and subsequently imported into Geneious software (<http://www.geneious.com/download>). We then did the assembling function to harvest 444 contigs to be further analyzed. We chose 4000bp as a cut-off value for open reading frames and did in silico translation for those to obtain putative protein products. The putative proteins were then analysis by protien BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) for screening domains related to PKS. We found several gene clusters resemble to PKS function but only 5 contigs fit into what we hypothesized for fluvirucin B₁ PKS.

The protein sequences of AT domains were defined by BLAST, and subject to the analysis tool of SEARCHPKS (<http://www.nii.res.in/searchpks.html>). And the KR sequence were also defined by BLAST and the homology study were done by using ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>), a general purpose multiple sequence alignment program for DNA or proteins.

2.4.3 Genomic PCR Procedure

The PCR reaction for the gap between module 3 and module 4 was performed in a 50 ul reaction mixture containing: 1X Phusion GC buffer, 0.2 mM dNTP, 0.3 mM MgCl₂, 3%DMSO, 1 uM of each primer (forward primer: ATCTCGACTGCTACGCATCC; reverse primer: CCCGAAGTTGGATTGACC), 1U Phusion Hot Start II DNA polymerase (Thermo scientific, USA) and approximately 450ng of genomic DNA was added as template.

The thermal cycler (Mastercycler ep gradient, Eppendorf) was programmed according to the following “2-steps” amplification profile: 3 min denaturation at 98°C, then 10 initial cycles of 10 s denaturation at 98°C, 5 min annealing and elongation at 72°C, followed by 27 cycles of 10 sec denaturation at 98°C, 5 min + (5 sec/cycle) elongation at 72°C, and a final extension step at 72°C for 5 min. The amplified DNA fragments (4524bp) were then subjected to 0.8% agarose gel and single bands were excised and purified using QIAquick Gel Extraction Kit (QIAGEN, Germany)

according to the instructions from the manufacturer. The amplified DNA fragments were then purified, and directly inserted into the plasmid vector PCR-Blunt II Topo (Zero blunt TOPO PCR cloning kit, Invitrogen). Finally the resulting plasmid was sequenced under regular procedures to obtain the genomic information.

2.5 Reference

1. Schnarr, N. A. and Khosla, C. (2006) *In Chemical Biology: From Small Molecules to Systems Biology and Drug Design*, Wiley-VCH, Weinheim
2. Donadio, S. and Sosio, M. (2003) Strategies for combinatorial biosynthesis with modular polyketide synthases. *Comb. Chem. High Throughput Screening*, 6, 489-500
3. Walsh, C. T. (2002) Combinatorial biosynthesis of antibiotics: Challenges and opportunities. *ChemBioChem*, 3, 124-134
4. Staunton, J. and Wilkinson, B. (2001) Combinatorial biosynthesis of polyketides and nonribosomal peptides, *Curr. Opin. Chem. Biol.*, 5, 159– 164
5. Cane, D. E., Walsh, C. T., and Khosla, C. (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science*, 282, 63– 68

6. McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U.S.A.*, 96, 1846– 1851
7. Xue, Q., Ashley, G., Hutchinson, C., and Santi, D. (1999) A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. U.S.A.*, 96, 11740– 11745
8. Ruan X, Pereda A, Stassi DL, Zeidner D, Summers RG, Jackson M, Shivakumar A, Kakavas S, Staver MJ, Donadio S, Katz L. (1997) Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J. Bacteriol*, 179, 6416-6425.
9. Kapur, S., Chen, A.Y., Cane, D.E., Khosla, C. (2010) Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. USA*, 107, 22066-22071.
10. Tran, L., Broadhurst, R. W., Tosin, M., Cavalli, A., and Weissman, K. J. (2010) Insights into protein-protein and enzyme-substrate interactions in modular polyketide synthases. *Chem. Biol.* 17, 705-716.
11. Chandran, S.S., Menzella, H.G., Carney, J.R., Santi, D.V. (2006) Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chem. Biol.*, 13, 469-474.

12. Naruse, N., Tenmkyo, O., Kawano, K., Tomita, K., Ohgusa, N., Miyaki, T., Konishi, M., and Oki, T. (1991) Fluvirucins A₁, A₂, B₁, B₂, B₃, B₄ and B₅, new antibiotics active against influenza A virus. I. Production, isolation, chemical properties and biological activities. *J. Antibiot.* 44, 733– 740
13. Naruse, N., Tsuno, T., Sawada, Y., Konishi, M., and Oki, T. (1991) Fluvirucins A₁, A₂, B₁, B₂, B₃, B₄ and B₅, new antibiotics active against influenza A virus. II. Structure determination. *J. Antibiot.* 44, 741– 755
14. Naruse, N., Konishi, M., and Oki, T. (1991) Fluvirucins A₁, A₂, B₁, B₂, B₃, B₄ and B₅, new antibiotics active against influenza A virus. III. The stereochemistry and absolute configuration of fluvirucin A₁. *J. Antibiot.* 44, 756– 761
15. Kohli, R. M. and Walsh, C. T. (2003) Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem. Commun.*, 3, 297– 307
16. Parenty, A., Moreau, X., and Campagne, J. M. (2006) Macrolactonizations in the total synthesis of natural products. *Chem. Rev.* 106, 911– 939
17. Moutevelis-Minakakis, P., Neokosmidi, A., Filippakou, M., Stephens, D., Dennis, E. A., and Kokotos, G. (2007) Synthesis of lipophilic 2-oxoamides based on γ -aminobutyric and δ -aminovaleric analogues and their activity against phospholipase A2. *J. Pept. Sci.* 13, 634– 641

18. Ogasawara, Y., Katayama, K., Minami, A., Otsuka, M., Eguchi, T., and Kakinuma, K. (2004) Cloning, Sequencing, and Functional Analysis of the Biosynthetic Gene Cluster of Macrolactam Antibiotic Vicenistatin in *Streptomyces halstedii*. *Chem. Biol.* 11, 79– 86
19. Yadav, G., Gokhale, R. S., and Mohanty, D. (2003) SEARCHPKS: a program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res.* 31, 3654– 3658
20. Keatinge-Clay, A. T. (2007) A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* 14, 898– 908

CHAPTER 3.

THE CLONING AND HETEROLOGOUS EXPRESSION OF FLUVIRUCIN B₁ PKS

3.1 Introduction

For the past decade, the biosynthesis community has struggled to capture the true engineering potential of polyketide synthases despite a vast array of tools for separating, fusing, and rearranging the component enzymes. Of the many techniques employed in reengineering PKSs, three stand out as the primary means of genetic alteration: (1) Domain Swapping, (2) Domain Addition, and (3) Domain Deletion/Inactivation (**Figure 12**).

Domain Swapping: This technique involves physically replacing an individual PKS enzyme with one from another synthase. Most commonly used for AT domains, the key to effective domain swapping is a comprehensive knowledge of domain boundaries such that the surrounding enzymes are unaffected. The primary downside to this procedure is that heterologous domains often behave differently in their new environment. For instance, relaxed substrate selectivity has been observed for swapped AT domains leading to diminished yields of the desired polyketide products.¹⁻²

Domain Addition: This technique is generally employed when an increase in enzyme activity is desired. Many attempts have been made to incorporate KR, DH, and ER domains into

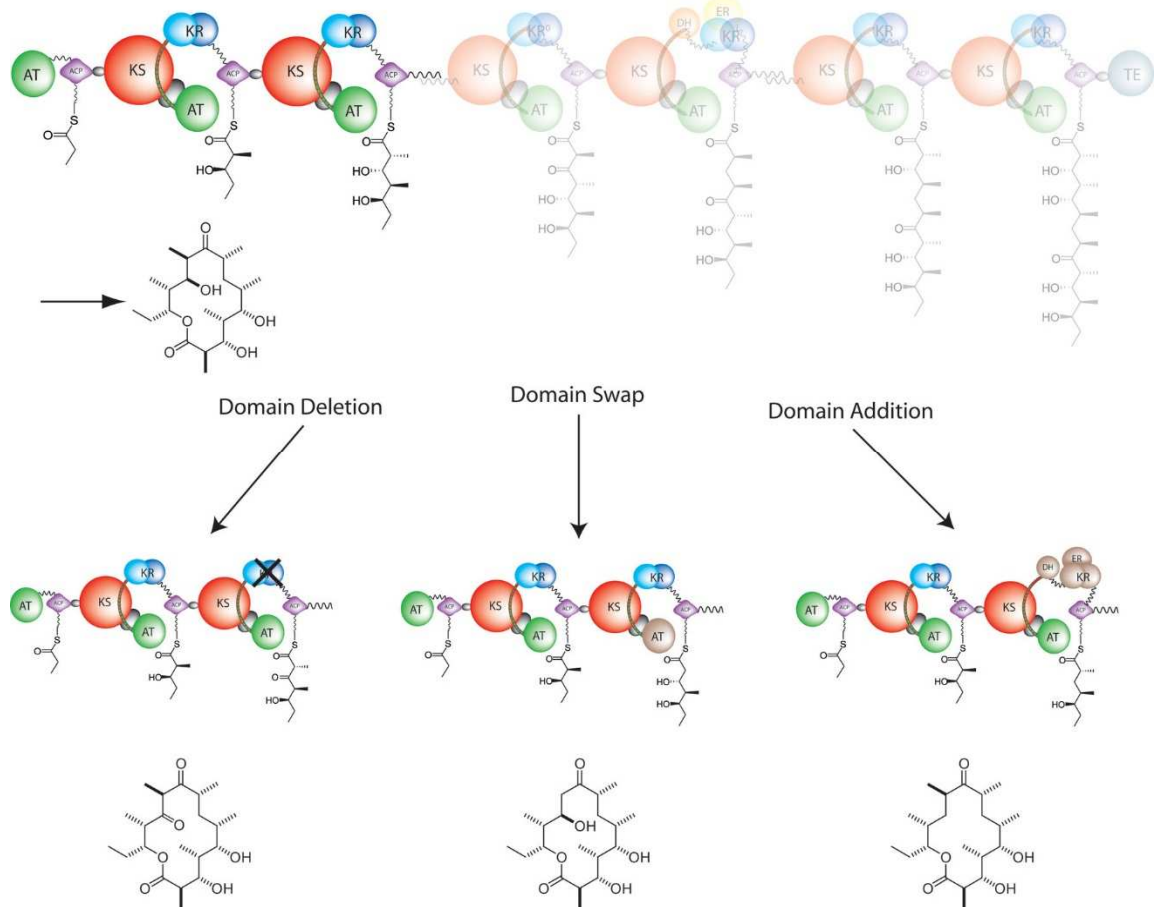


Figure 12. Schematic diagram of polyketide engineering strategies. Left to right: Domain deletion – KR knockout in DEBS module 2 results in a ketone functionality in place of the native hydroxyl group. Domain swap – substitution of a malonyl-specific AT for the native methylmalonyl-specific AT results in loss of a methyl group. Domain addition – replacement of the native KR domain with a DH/ER/KR cassette results in loss of a hydroxyl group from the final product. The latter two strategies require introduction of heterologous domains leading to disruption of native protein-protein interactions. We seek a PKS where domain deletion alone can provide maximal product diversity.

modules where they are absent, but the most effective strategy for domain addition involves replacing the entire module with a heterologous one bearing the appropriate tailoring domains.

In both cases, interaction issues between native and heterologous components often impede the flow of intermediates within and between modules, resulting in significantly decreased product yields.

Domain Deletion: As the simplest technique of the three, domain deletion/inactivation is functionally the opposite of domain addition. The activity of any module containing at least one tailoring domain (KR, DH, or ER) can be altered via mutation of active site residues or removal of the entire enzyme. It has been shown, however, that mutation (inactivation) provides significantly higher product yields compared to removal (deletion), owing largely to the fact that the former method maintains the native three-dimensional structure of the engineered module.³ This observation provided much of the initial motivation for this proposal and taking together with the domain addition and swapping data, it is clear that disruption of PKS structure and recognition motifs plays a significant role in diminished product yields.

To date, the most successful illustration of polyketide structure-diversification, executed by McDaniel and coworkers, resulted in nearly 100 novel analogs of 6-deoxyerythronolide B.^{4,5} In short, specific alterations to module components were genetically combined on unique plasmids to produce unnatural polyketide products. They observed yields of mutant products ranging from 1-70% of the wild-type 6-deoxyerythronolide B. It should be noted that in all but 2 single mutants, the relative yields were less than 50% and most of these fell into the realm of 20-30%. As a result,

only about one-quarter of the triple mutant combinations afforded detectable levels of polyketides. At this rate, access to the many thousands of conceivable engineered analogs is beyond reach, thus underscoring the urgent need for strategies that address the primary contributors to low product yields.

It is clear that decreased chemical output is a consequence of two primary barriers: (1) substrate selectivity of downstream enzymes and (2) disruption of protein-protein interactions. *Engineering strategies that circumvent either or both of these issues will afford much needed information regarding their relative contributions to low yields and ultimately lead to vastly improved product diversity and output.* However, tackling these engineering obstacles will first require an in-depth understanding of their origins.

Therefore, we proposed a new strategy to overcome this hurdles, and we planned to unearth a unique PKS system where a wide variety of polyketide structures can be generated without introducing heterologous domains or linkers. The strategy takes advantage of the fact that the FAS-like PKS modules produce β -keto, β -hydroxyl, and β -olefin functionalities on the way to the final β -methylene product. Therefore, to access these desired intermediate structures, one must simply “turn-off” the appropriate tailoring domain within the module via site-directed mutagenesis. Most importantly, all protein-protein interactions and native linkers remain intact,

thus allowing for direct determination of the relative impact of chain transfer (in other words, KS substrate selectivity) on engineered polyketide production.

Upon our previous identification of the gene cluster for making fluvirucin B₁, a fatty-acid-synthase-like PKS system, in *Actinomadura vulgaris*, it was soon appreciated as an interesting PKS system with many unique biosynthesis attributes (**Figure 13**).⁶ It contains 5 modules of which 4 contain the full set of PKS tailoring domains. According to our hypothesis, it is an ideal system for

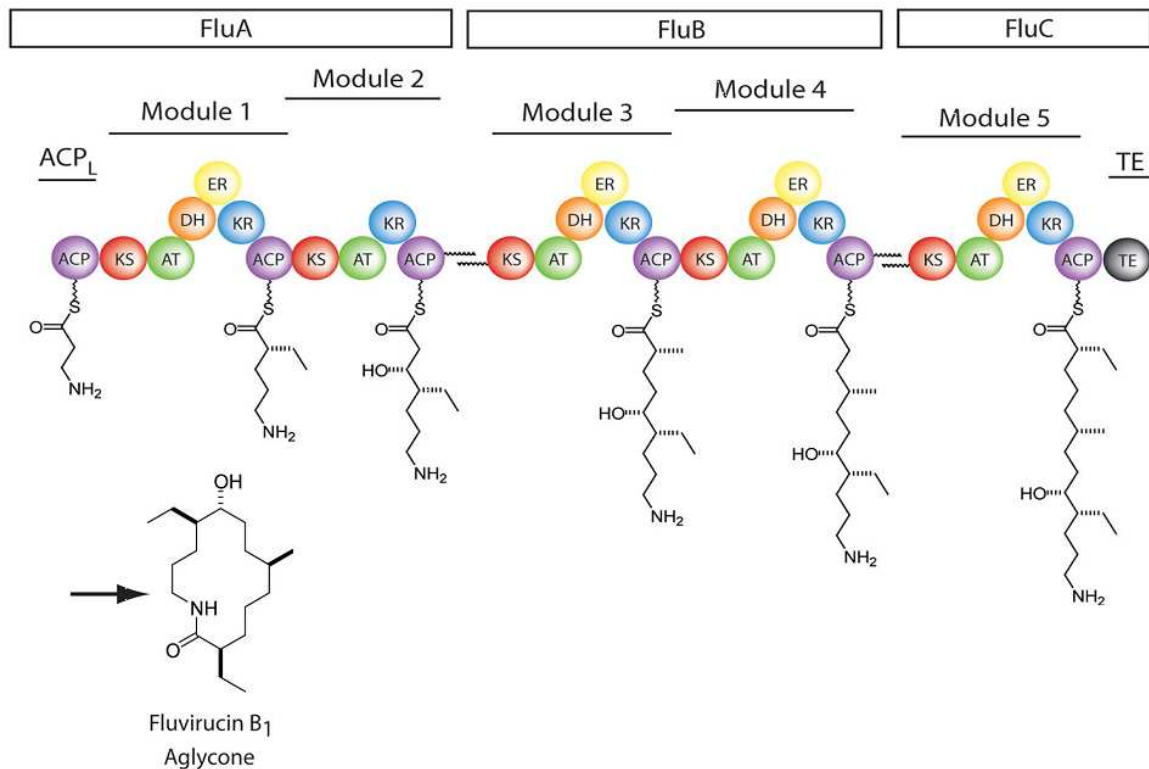


Figure 13. Schematic diagram for the putative fluvirucin B₁ PKS. The assembly consists of five extender modules flanked by an N-terminal loading ACP and C-terminal thioesterase (TE) domain. FluA consists of the loading ACP, module 1 and module 2. FluB consists of modules 3 and 4. FluC consists of module 5 and the TE domain.

engineering strategy of single domain inactivation via site-directed mutagenesis. As a result, we were motivated to explore the biosynthetic origin of fluvirucin B₁ with the ultimate goal of providing a platform for polyketide engineering that circumvented the need for incorporation of heterologous domains to achieve maximal product diversity. Herein, we describe our effort to unearth the assembly line of fluvirucin B₁ PKS.

3.2 Result and Discussion

3.2.1 Cloning of Module 1 of Fluvirucin B₁ PKS and PCR Optimization

To confirm the sequences obtained from partial assembly of the *A. vulagaris* genome and with the ultimate goal of reconstituting the entire assembly in *E. coli*, we turned our attention to cloning each module individually from genomic DNA. We firstly setup a genomic library (~800 colonies isolated) to screen for the genes that either contributed to or are associated with fluvirucin B₁ PKS. Unfortunately, all screening efforts came back with the indication that the genomic library of *A. vulagaris* was contaminated by the genome of *E. coli*, and we suspect that it was contaminated during the preparing processes of the genomic library.

After the failure of our genomic library, we began seeking alternative ways to clone the gene of fluvirucin B₁ PKS and one of which is genomic polymerase chain reaction (PCR). For replicating the gene of interest from genomic DNA, it is not only hard to get single product but

also difficult to amplify a long PCR product (>2kb) with high GC-content because GC-rich sequence can easily to form secondary structure and render PCR useless. According to the genomic sequencing information, the genome size of *A. vulagaris* is estimated about 12 million base pair (bp), and it is relatively larger than the genome size of *Streptomyces* strain. For instance, the complete genome of *S. coelicolor* strain A3(2) was decoded and published in 2002⁷, and the genome was sequenced about 8 million bp with a GC-content of 72.1%. We expect the GC content of *A. vulagaris* could be as high as *S. coelicolor* since they are close relatives under the order of Actinomycetales, and indeed the sequence analysis showed FluA (11910bp) consists GC-content of 71.3%, FluB (13128 bp) and FluC (7191bp) are 71.5% and 71.9% respectively. Considering all these difficulties, we began our quest for cloning fluvirucin B₁ PKS. Even though there is no literature documented for long PCR protocol applied to GC-rich sequence, there are some aspects of PCR can be optimized for our purpose. For instance, there are PCR methodologies for temperature manipulations, different organic solvents as PCR additives, and primer-design techniques, etc. We used all of PCR techniques mentioned above for cloning the assembly line of fluvirucin B₁ PKS.

Based on alignment with known PKS constructs and the in-silico analysis of restriction enzyme sites, we were able to determine the effective sequence boundaries for each fluvirucin

synthase module, and we began designing the PCR experiment and optimizing the PCR method for each module.

For starter, we would like to describe our effort for cloning of module 1. We used the sequence of module 1 as our amplification segment (6713bp, from the first ATG to the NotI site as the boundary between module 2) and designed the PCR primers from the upstream and downstream sequences within 1,000~1,500 bp to the boundaries of module 1. We then took the arbitrarily defined sequence to Primer-BLAST on NCBI website to perform the computer-aided primer-design in the default setting, and we obtained 20 sets of primers for cloning module 1. Subsequently, we screened all the primers obtained above with NetPrimer (<http://www.premierbiosoft.com/netprimer/>) software, and selected the primers scored over 90 in NetPrimer. So, our first round primers are listed in **Table 2**:

As soon as we obtained the oligonucleotides in the list for cloning module 1, we performed several general PCR reactions with Phusion Hot Start II DNA polymerase. Initially, the primer sets of Avul-Mod1-F1&R1 and Avul-Mod1-F2&R2 did not yield any meaningful products (**Figure 14**);

Table 2. The primer list for module 1

Avul-Mod1-F1: GGACATCGGCGGGCTTCG
Avul-Mod1-R1: GCGGGCGGCCAGGACCATCTT
Avul-Mod1-F2: CCGACTGCCTCCTGCGGGTG
Avul-Mod1-R2: GGACGTGGACGCGGCTCGGA
Avul-Mod1-F3: GCCACCACCAGCATCCGCAT
Avul-Mod1-R3: GTTCCGCCGATGAAGACGCC
Avul-Mod1-F4: GCCCTCCCAGTTCGGCTTCG
Avul-Mod1-R4: GGACTTCACGGTGGCCAGCC
Avul-Mod1-F5: ATCAACACGCTGGGCGGCAT
Avul-Mod1-R5: GGTTGGGGAGCACGGTCACG
Avul-Mod1-F6: TCCGTACCCGATCTCCGCC
Avul-Mod1-R6: CCGCCCAGACATGACCGAACTG

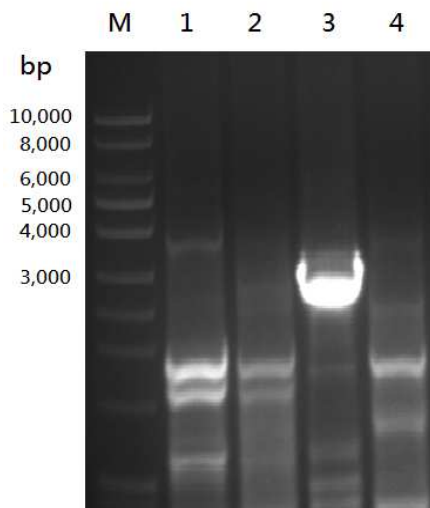


Figure 14. The PCR result for cloning module 1-a. Lane 1-2: F1+R1, R2; Lane 3-4: F2+ R1, R2. No promising band observed above 6,000bp.

however, in the following PCR experiment with for the permutation of primer matches (F3~F6 matched with R3~R6), we observed that Avul-Mod1-F3, F5, F6 are promising forward primers, R6 is a very promising reverse primer, and R5 only effective when combined with F5 (**Figure 15**). Those positive results of our first attempt were

subjected for the following cloning of TOPO-Blunt method in order to identify the sequence of the PCR product, but unfortunately the sequencing result did not suggest any full-length sequence of module 1.

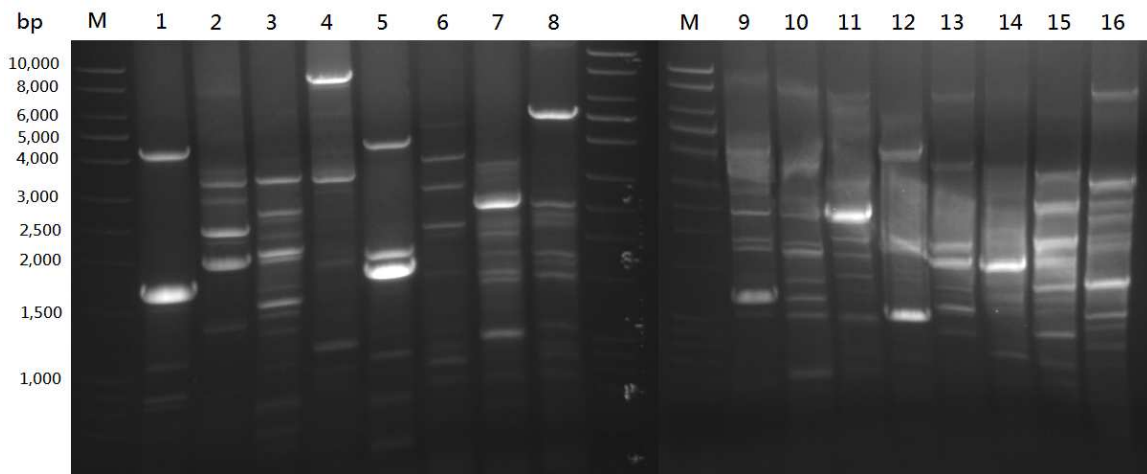


Figure 15. The PCR result for cloning module 1-b. Lane 1-4: F3+R3, R4, R5, R6; Lane 5-8: F4+ R3, R4, R5, R6; Lane 9-12: F5+R3, R4, R5, R6; Lane 13-16: F6+R3, R4, R5, R6. Lane 4, 12, 16 indicated R6 is a good reverse primer; Lane 9, 10, and 12 showed F5 is promising. The same conclusion can be deduced on F3 and F6.

With the data that some desired amplification of module 1 appeared in the DNA gel analysis, we hypothesized that the low yield/non-full length amplification of the PCR experiment could be caused by the GC-rich sequence within module 1. It has been suggested that DNA amplification by PCR is frequently complicated with low yield and specificity by the GC content of the target sequence, and the addition of some organic solvents like DMSO or betaine to PCR could improve the amplification of DNA by reducing the secondary structures of GC-rich regions in the DNA template or the primers.⁸ Therefore, we setup several PCR experiments with the primers previously identified promising in amplifying the DNA of module 1, and we titrated the concentration of DMSO from 0% to 8% in these experiment. To our delight, we got very promising

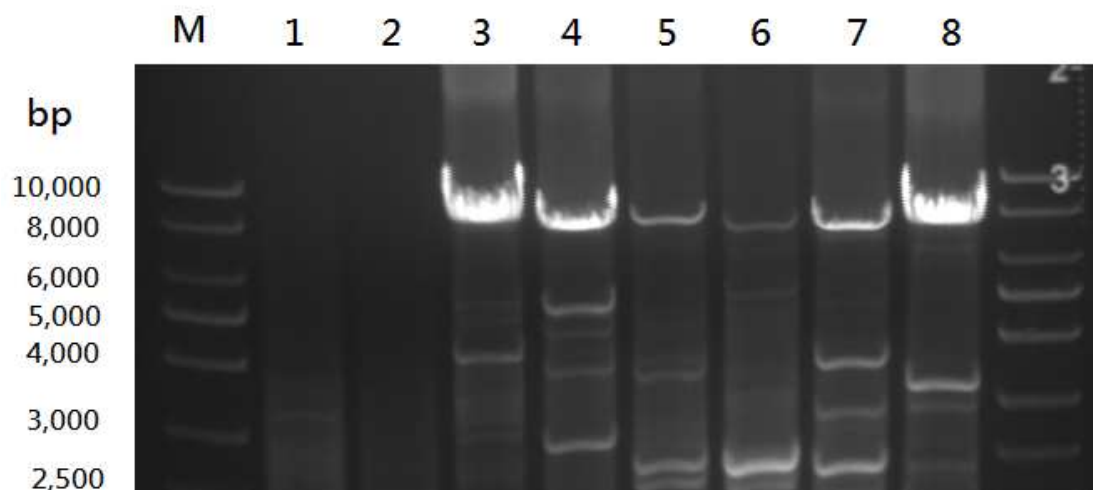


Figure 16. The PCR result for cloning module 1 with 5% DMSO. Lane 1: F1+R1; Lane 2: F1+ R6; Lane 3: F3+R6; Lane 4: F4+R6; Lane 5: F5+R4; Lane 6: F5+R5; Lane 7: F5+R6; Lane 8: F6+R6. The bands around 8,000bp indicated in Lane 3 to 8 are positive result, and the gel analysis confirmed that DMSO indeed improves the amplification of DNA in our experiment.

results in the PCR experiment with 5% DMSO (**Figure 16**). All PCR products with the predicted length were excised, and then later confirmed the PCR results from Avul-Mod1-F3 + R6, F5 + R6 and F6 + R6 are true replications of module 1. Subsequently, we further improved our cloning strategy not only by “touch-down” PCR methodology⁹ but also by designing a forward primer with a NheI restriction enzyme site (a non-cutter for module 1), AMod1-NheI-F1 (see experiment procedures), and we used in combination with Avul-Mod1-R6 in a similar setup of PCR to obtain the pre-module 1 DNA fragment that can be purified, digested, and directly ligated into pET21b without further steps of sub-cloning (**Figure 17**).

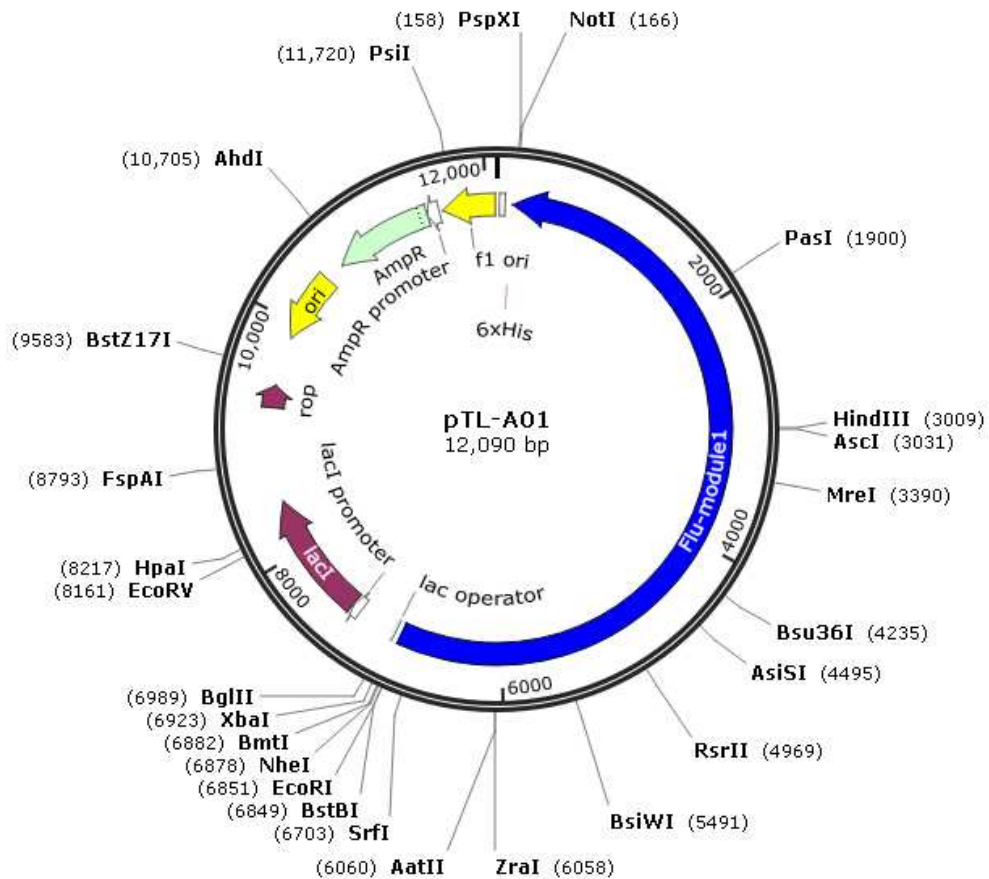


Figure 17. The plasmid map of Flu-module1 (pTL-A01)

3.2.2 Cloning of Module 3 and 5 of Fluvirucin B₁ PKS

With our successful experiments of replicating module 1 directly from genomic DNA, we moved forward to a new and fast protocol to clone module 3 and 5: design proper primers with restriction enzyme sites at the boundaries of each module. For cloning module 3, we designed the primers in **Table 3**:

Table 3. The primer list for module 3
AMod3-EcoRI-F1: AAAAAAGAATT <u>CG</u> ATGGCCACTGACGACAAGTTC
AMod3-R1: TTTTTTGACGTGGACGCGGCTCGGAC
AMod3-EcoRI-F2: AAAAAAGAATT <u>CG</u> ATGGCCACTGACGACAAGTTCCGG
AMod3-R2: TTTTTGTGGACGTGGACGCGGCTCGGAC
(Note: restriction enzyme sites were underlined and the additional “in-frame” bases were emphasized in boldface.)

We did similar permutation of primer matches and the same titration of DMSO from 0% to 8 % in the PCR for cloning module 3, and we successfully cloned out pre-module 3 DNA with 7% DMSO by the primer combination of AMod3-EcoRI-F2 and AMod3-R2. Later the PCR products were digested and incorporated into pET-21b to obtain pTL-A03 (**Figure 18**). Next, we applied the simple designing principle on primers for cloning module 5 (including the TE domain), we tested all the primers listed in **Table 4**:

Table 4. The primer list for module 5
AMod5-NheI-F1: TTTTTT <u>GCTAGCAT</u> GGCTGACGAA
AMod5-HindIII-R1: AAAAAAAGCTTT <u>CACGCGCCGTT</u>
AMod5-NheI-F2: AAAAAAGCTAGCATGGCTGACGAAGAGAAG
AMod5-HindIII-R2: TTTTTTAAGCTT <u>CGCGCCGTT</u> CGA
AMod5-NheI-F3: AAAAAAGCTAGCATGGCTGACGAAGAGAAGCTCCTC
AMod5-HindIII-R3: TTTTTTAAGCTT <u>CGCGCCGTT</u> CGAACGCC

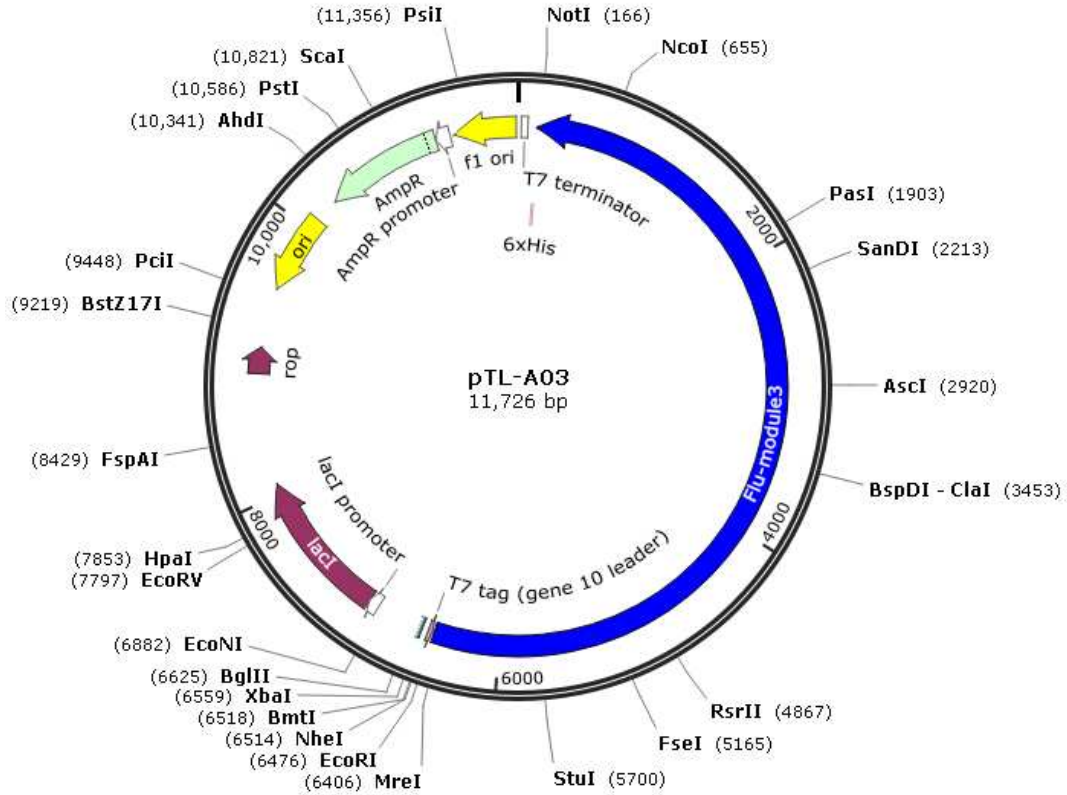


Figure 18. The plasmid map of Flu-module3 (pTL-A03)

The permutation of primer matches were again deployed in order to find the best primer set to clone the DNA of pre-module 5, and we observed the promising gel bands by the combination of AMod5-NheI-F1 + R2, F2 + R3, F3 + R2, and F3 + R3 (data not shown). However, the concentration of DMSO is slightly different than the previous titration for module 1 and 3, because the PCR products of pre-module 5 were not amplified until the concentration of DMSO was increased to 10%. This suggested that genomic PCR can be done in a relatively high concentration of DMSO even if the high concentration DMSO can slightly inhibit the activity of

DNA polymerase. In the end, pTL-A05 (Figure 19) was cloned out similarly as previously described for module 1 and 3.

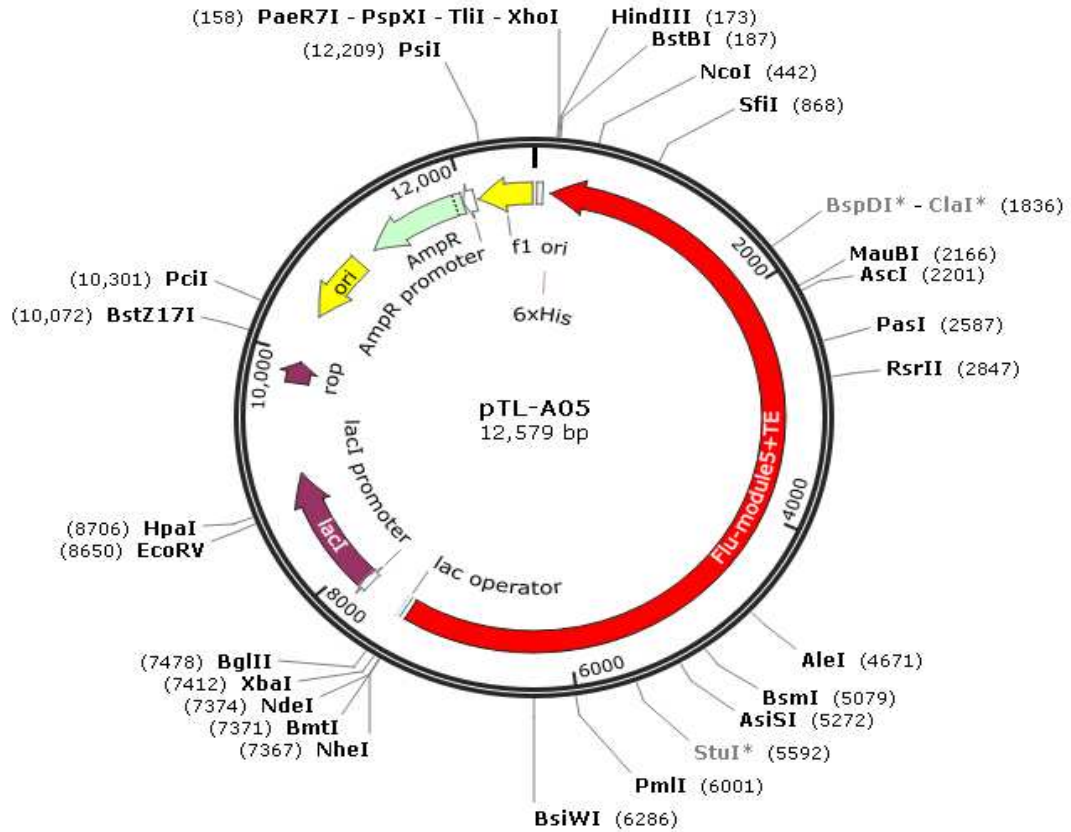


Figure 19. The plasmid map of Flu-module5 (pTL-A05)

As a summary for section 3.2.1 and 3.2.2, we were fortunately able to clone out Flu-module 1, 3, 5 directly with properly designed primers and various concentration of DMSO as additives in a typical genomic PCR experiment. All three PCR products are amplified correctly and verified by BLAST with the method provided in experimental procedures.

3.2.3 Cloning of Module 2 of Fluvirucin B₁ PKS and the Study of PCR Additives

Based on the strategy we developed, we applied to the same rules for designing primers (Table 5) for module 2; however, we failed to obtain any substantial PCR products with every possible match of forward and reverse primers and with different concentration of DMSO from 0% to 10%. We then hypothesized that the sequence within high GC content in module 2 (5160bp with 70.6% of GC content) or the interactions between primers and DNA template may be the cause for getting no/non-specific PCR products, and DMSO is not good enough to resolve the problem. Therefore, we designed a series tests on PCR additives to overcome this problem. We used the PCR condition for cloning module 1 with 8 different organic solvents as PCR additives in the concentration range that previously suggested (Figure 20).^{10,11}

Table 5. The primer list for module 2 (the first round)

Avul-Mod2-F1	CCACACCGGCAAGATCATCCTC
Avul-Mod2-F2	CCGGCAAGATCATCCTCACCATG
Avul-Mod2-F3	CCTCACCCCGACAAGATCGA
Avul-Mod2-F4	TTCGTCCTTACTCCTCCTCGG
Avul-Mod2-F5	CCTCACCACCACCGACATCAC
Avul-Mod2-F6	CGACATCACCCGCATCGAACA
Avul-Mod2-F7	ACCTCGCCACCGAACACCAGAC
Avul-Mod2-F8	CCTCCCGCCGAGGAGACCAAG
Avul-Mod2-F9	CGAGGAGACCAAGCCCCCGA
Avul-Mod2-F10	GCCATCCACCGCCACACCAC
Avul-Mod2-F11	ACCAACCCAGCGGCATGAGC
Avul-Mod2-F12	AAGCGGGACGCCCGTTCAAG
Avul-Mod2-F13	GCGACGCTGGTGTTCGACCA
Avul-Mod2-F14	AGGACAGGCTTCTGGACGGTT
Avul-Mod2-F15	GCCACCGAACACCAGACCCG
Avul-Mod2-F16	CCATCCACCGCCACACCACC
Avul-Mod2-F17	TCTGCTCGGGGTGGAGGACG

```

Avul-Mod2-F18 GCCTCTCTGCTCGGGGTGGA
Avul-Mod2-F19 CGGGATGCGGTGCTGGTGTT
Avul-Mod2-F20 TCGCCATCCACCGCCACACCACC (same as F10 but longer)
-----
Avul-Mod2-R1 GTACTGCTCGGACAGGAGGAC
Avul-Mod2-R2 GATCGGGAGGTTCTGAGCTG
Avul-Mod2-R3 CGAACATGTCCACGTCGTCGT
Avul-Mod2-R4 GTAGTTGATGCTGCTCGCCAT
Avul-Mod2-R5 CGTGTGCATCGGCTCGAAGAT
Avul-Mod2-R6 GTAGTTGGTCTCGTACCCCCAGA
Avul-Mod2-R7 GCCATCCACAGGTAGCGGTTG
Avul-Mod2-R8 TAGACGGCGGACAGGGTG
Avul-Mod2-R9 GCCCCGAACCGCTTCCACTC
Avul-Mod2-R10 TCGGCCTTGTGGTCGCTGGAGA
Avul-Mod2-R11 ACCTTGTCACCGGCCTGCCA
Avul-Mod2-R12 GGGCCGGTCTTCAGCACGAC
Avul-Mod2-R13 TCGGCGTCGTA CTGCTCGGA
Avul-Mod2-R14 GTCAGCCCGACCACACCGTG
Avul-Mod2-R15 TGTTGAACGCCCCGTCCAGC
Avul-Mod2-R16 CAGTACCCCGACCTTGCCGCT
Avul-Mod2-R17 CGGCGTCGTA CTGCTCGGAC
Avul-Mod2-R18 GATCGGCGTTCTCCAGGGCG
Avul-Mod2-R19 CGGCGATCTCCTCGGGTTGC
Avul-Mod2-R20 GCCATCCACAGGTAGCGGTTGTCG (same as R7 but longer)

```

We observed that sulfolane offers more specific amplification whereas glycerol provides with the best yield among all. After we found these phenomena/characteristics of organic solvents in genomic PCR are superior to DMSO alone, we made a new PCR recipe consists of 15% glycerol and 0.6M sulfolane as a new PCR additive mix for cloning module 2 with the primers previously designed.

Followed by series of PCR experiments, nevertheless, we observed some improvements in terms of specificity (data not shown), but we could not harvest enough DNA of pre-module 2

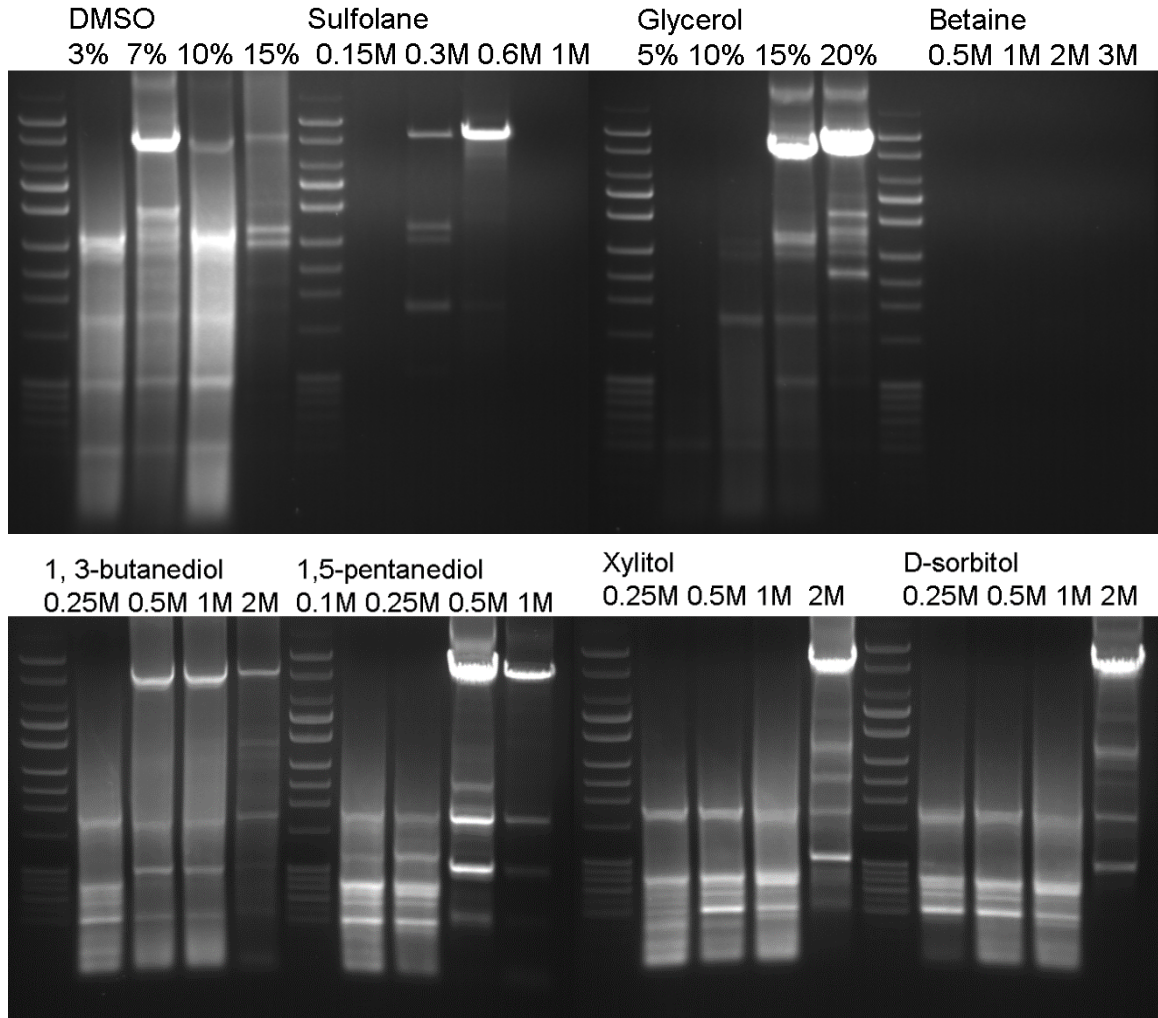


Figure 20. The gel analysis of PCR additives examination based on module 1 PCR setup. All panels contain 4 different concentrations and 1 lane of molecular marker

to be subjected into expression vector. Later, we designed the second round of primers for cloning module 2 (**Table 6**), and we successfully cloned out the DNA of pre-module 2 with the primer combination of Avul-Mod2-pF6+R7 and pF6+R17. The DNA of pre-module 2 was finally sub-cloned into pET21b to construct the expression vector of Flu-module2, pTL-A02 (**Figure 21**).

Table 6. The primer list for module 2 (the second round)

Avul-Mod2-pF1: CCATCAACACCCACACCCTCAG
Avul-Mod2-pR1: CCATCCACAGGTAGCGGTTGTC
Avul-Mod2-pF2: GCGTTCGTCCTCTACTCCTCCT (mis-primed at the end of mod4)
Avul-Mod2-pR2: AGGTAGCGGTTGTCGTAAGTGGTG (overlap with R7)
Avul-Mod2-pF3: TCATCCACATGTCCGTTATCGCAG
Avul-Mod2-pR3: ACCGTGTTGTAGTTGATGCTGCTC
Avul-Mod2-pF4: AACTGCTGGACCTGTTGCCT
Avul-Mod2-pR4: CTCATCCTCCGCAAGCTAGTTGTC
Avul-Mod2-pF5: GGATGTGGTGCTCAACTCGCTG
Avul-Mod2-pR5: TGAGGCCGTAGAAGACGCTG
Avul-Mod2-pF6: CCCATCAACACCCACACCCT
Avul-Mod2-pR6: TCCGCAAGCTAGTTGTCCGA
Avul-Mod2-pF7: ATCACCCGCATCGAACAAGCC
Avul-Mod2-pR7: TTGAGGCCGTAGAAGACGCTG
Avul-Mod2-pF8: TGGATGTGGTGCTCAACTCGCT
Avul-Mod2-pR8: GTGTTGTAGTTGATGCTGCTCGCC
Avul-Mod2-pF9: AGATGCTCGCCGAAGTCTGCTG
Avul-Mod2-pR9: TTGATCTGGTTCGATGTGGCGG
Avul-Mod2-pF10: CACACCGGCAAGATCATCCTCAC
Avul-Mod2-pR10: ACGTCGCCGACCTTGATCTG
Avul-Mod2-pF11: AGGCAACCAACTCCTCGACACCGCCCT
Avul-Mod2-pR11: AACTCCGTGACCGCTGACCGCACCA
Avul-Mod2-pF12: ACGCCGGACACGACCGTATTGGGCAGA
Avul-Mod2-pR12: TCCCGTTCACGCCGCCGACGTTGTT
Avul-Mod2-pF13: TCCAAGCGGGACGCCCGTTCAAGGA
Avul-Mod2-pR13: TGGTCGATGTGGCGGGCACGGAAT

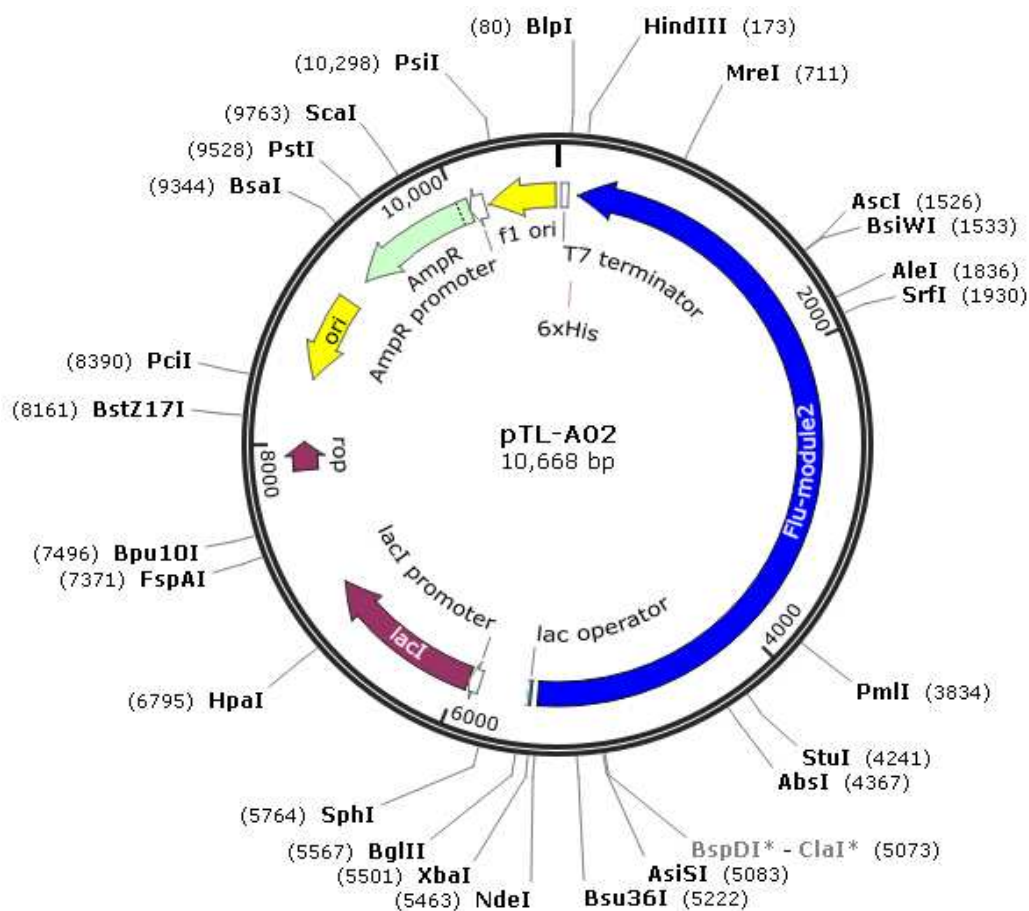


Figure 21. The plasmid map of Flu-module2 (pTL-A02).

3.2.4 Cloning of Module 4 of Fluvirucin B₁ PKS

As our effort for cloning module 1, 3, 5 and 2 being made, we simultaneously conducted similar experiments for cloning module 4. We designed more than 70 primers (see **3.4 supplement section** for the complete list) and tested in PCR experiments with DMSO or with glycerol/sulfolane mix, but we could not flank any desired products after more than 2,000 PCR run and ~200 gel analyses were casted. Then we reached to a conclusion that it may not be realistic to directly clone module 4 as one piece in genomic PCR, we hypothesized that module 2

and 4 may have the same evolution origin which makes them hard to be distinguished in the amplification of genomic PCR. As a result, we began to seek an alternative way to grasp module 4 out of the sea of genomic DNA, so we adopt the concept of “divide and conquer” by replicating module 4 DNA from two separate pieces, if not, as less pieces as possible, and merge each piece as the original module 4 DNA by the molecular cloning techniques.

By performing *in-silico* analysis for the restriction enzyme site of module 4 DNA, we quickly discovered a single cutter, MluI, located in the middle of the sequence, and we were convinced that it can be served as a stitch point for ligation reaction. Therefore, we planned to amplify two separate fragments of DNA, M4-1f and M4-2f, in genomic PCR and subsequently to stitch them together. We then designed the proper primers for cloning each DNA fragment, and we successfully replicated M1-4f with AMod4-HindIII-F1 + AMod4h-R6 and M4-2f with AMod4h-F7 + AMod4-XhoI-R7 (primers).

After the DNA information of the two fragments were sequenced by primer walking, however, we were surprised that M4-1f fragments are about 282bp shorter than we expected! Later, we conducted a thoroughly sequence comparison between the short and the expected sequence, and we found a repeated sequence, CCACCACCCA, located at the edge of the snipped sequence. We later recognized that the region is occupied with high content of C or G, which could contribute to secondary structure or G-quadruplex.¹²

In order to solve the problem, we lowered the concentration of Mg²⁺ in PCR reaction and extended the length of our primers, and we finally obtain the full length of M4-1f with AMod4-HindIII-F8 + AMod4h-R10. Subsequently, M4-1f and M4-2f were ligated together to harvest preM4 DNA, and the preM4 DNA was digested by NotI/XhoI and integrated with pET-21b to obtain pTL-A04 (Figure 22).

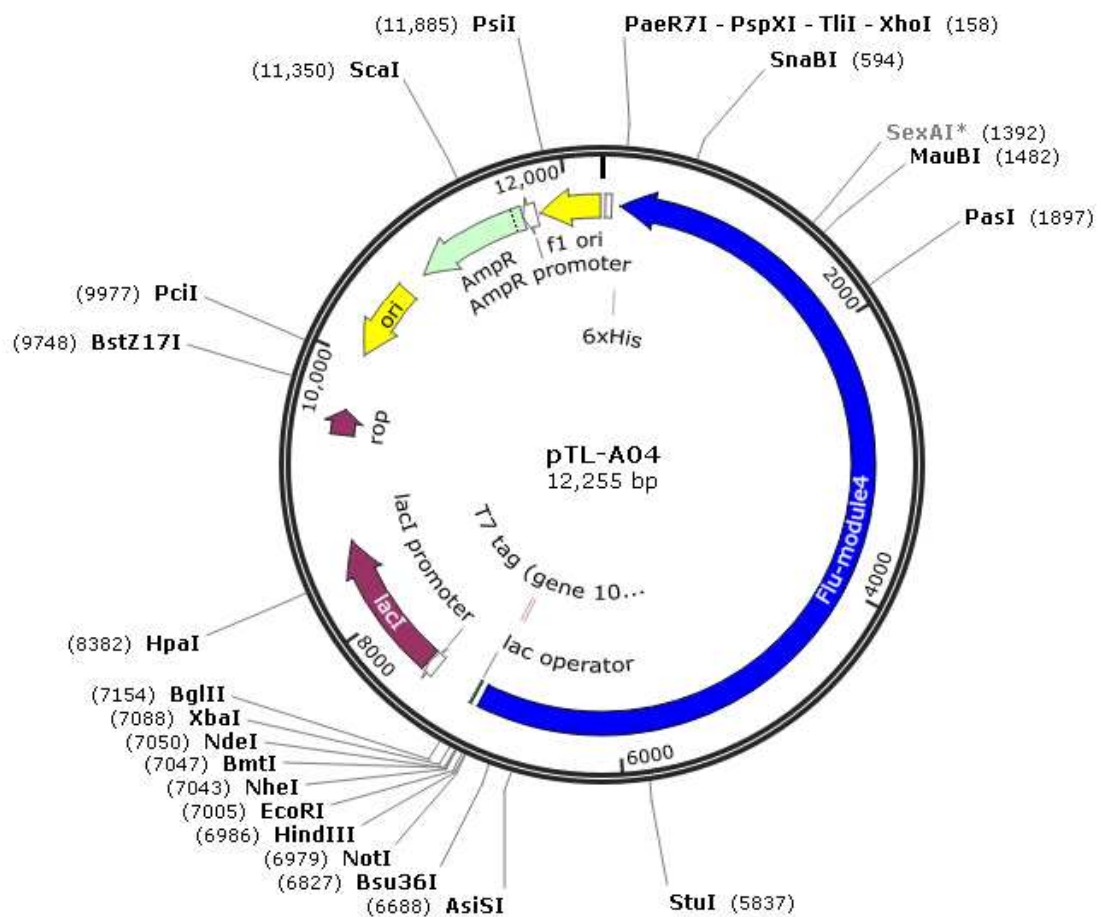


Figure 22. The plasmid map of Flu-module4 (pTL-A04)

3.2.5 Protein Expression

In conclusion, all five modules were cloned separately in pET vectors for expression in *E. coli*. Overexpression of each module was observed in BL21(DE3) cells, and gel migration patterns were consistent with calculated protein masses (**Figure 23**). The fact that *E. coli* seems to respond well to these foreign genes bodes well for our future efforts aimed at generating fluvirucin-derived structures in this heterologous host. In the near term, the ability to reliably produce usable quantities of each module will greatly facilitate studies concerning the substrate specificities and enzyme kinetics that govern fluvirucin B₁ biosynthesis.

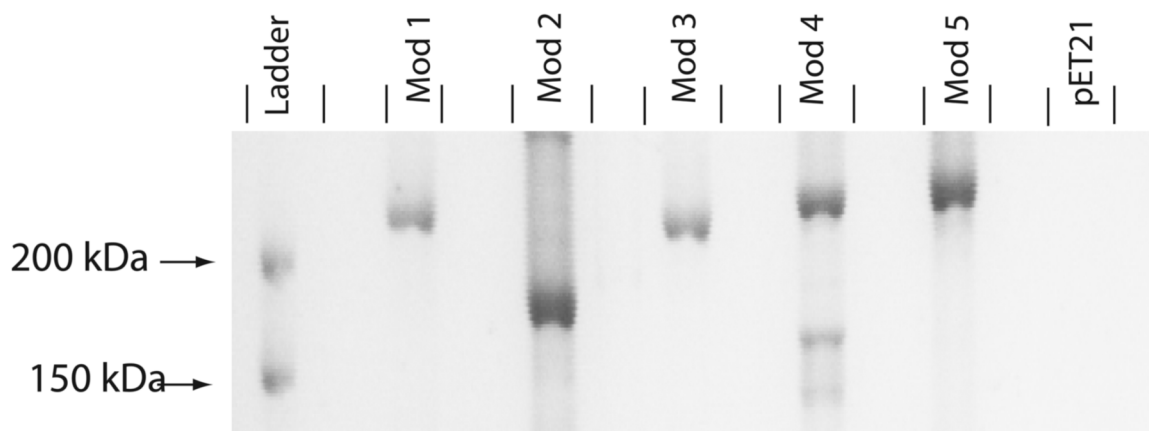


Figure 23. PAGE analysis of fluvirucin modules overexpressed in *E. coli* following Ni-NTA affinity purification. Lanes are marked with the corresponding protein or blank pET21 vector. Mod = Module. Approximate protein molecular weights: module 1 = 230 kDa, module 2 = 185 kDa, module 3 = 220 kDa, module 4 = 240 kDa, module 5 = 254 kDa. % acrylamide = 7.0.

3.3 Experimental Procedures

3.3.1 Bacterial Strains, Culture Conditions and DNA Purification.

Actinomadura vulgaris was purchased through American Type Culture Collection (ATCC) by the accession number ATCC 53715 and used as the source of DNA for shot-gun sequencing service and the cloning of Fluvirucin B1 polyketide synthase. The strain was cultivated at ambient temperature in the liquid medium of ATCC Medium 172 (N-Z Amine with Soluble Starch and Glucose), which contains 1% glucose, 2% soluble starch, 0.5% yeast extract, 0.5% N-Z amine type A (Sigma C0626), 0.1% CaCO₃. The growth of *A. vulgaris* at ambient temperature can be observed after 3 days of culture. For genomic DNA extraction purposes, *A. vulgaris* was cultured for 6 days and then genomic DNA was extracted by using the MasterPure gram-positive DNA purification kit (Epicenter, Madison, WI)

3.3.2 General Procedure for PCR Experiment

PCR reactions were performed in a 50µl reaction mixture containing: 1X Phusion GC buffer, 0.2mM dNTP, 0.3mM MgCl₂, 0~3%DMSO, 1M of each primer, 1U Phusion Hot Start II DNA polymerase (Thermo scientific, USA) and approximately 300~450ng of genomic DNA was added as template.

The thermal cycler (Mastercycler ep gradient, Eppendorf) was programmed according to the following “3-step” amplification profile: 3min denaturation at 98°C, then 35~40 cycles of 10sec denaturation at 98°C, 30sec at annealing temp (depending by the primer set), Xmin (X=length of amplification/1000) elongation at 72°C and a final extension step at 72°C for 3~5min. The amplified DNA fragments were then subjected to 0.8%~1.1% agarose gel to verify the products. If a single band of desired length were revealed, it would be excised and purified using a QIAquick Gel Extraction Kit (QIAGEN, Germany) for any following analysis or cloning procedure.

3.3.3 Cloning of Module 1 of Fluvirucin B₁ Polyketide Synthase

The PCR reaction for module 1 was performed in a 50µl reaction mixture containing: 1X Phusion GC buffer, 0.2mM dNTP, 0.3mM MgCl₂, 5%DMSO, 1µM of each primer (AMod1-NheI-F1: AAAAAAGCTAGCATGAGCCAGTCCGGAAACAGCGAA; Avul-Mod1-R6: CCGCCCAGACATGACCG-AACTG), 1U Phusion Hot Start II DNA polymerase (Thermo scientific, USA) and approximately 450ng of genomic DNA was added as template.

The thermal cycler (Mastercycler ep gradient, Eppendorf) was programmed according to the following “2-step” amplification profile: 3min denaturation at 98°C, then 10 initial cycles of 10sec denaturation at 98°C, 5min annealing and elongation at 72°C, followed by 27 cycles of 10sec denaturation at 98°C, 5min + (5sec/cycle) elongation at 72°C, and a final extension step at 72°C

for 5min. The amplified DNA fragments (7221bp) were then subjected to 0.8% agarose gel and single bands were excised and purified using a QIAquick Gel Extraction Kit (QIAGEN, Germany) according to the instructions from the manufacturer. Subsequently, it was subjected to restriction enzyme digestion of NheI and NotI and the digested products were ligated to pre-digested pET-21b to obtain pTL-A01.

3.3.4 Cloning of Module 3 of Fluvirucin B₁ Polyketide Synthase

The PCR reaction for module 3 was performed in a 50µl reaction mixture containing: 1X Phusion GC buffer, 0.2mM dNTP, 0.3mM MgCl₂, 7%DMSO, 1µM of each primer (AMod3-EcoRI-F2: AAAAAAGAATTCGATGGCCACTGACGACAAGTTCCGG; AMod3-R2: TTTTTTGTGGACGTGGACGCGGCTCGGAC), 1U Phusion Hot Start II DNA polymerase (Thermo scientific, USA) and approximately 450ng of genomic DNA was added as template.

The thermal cycler was programmed as for module 1, and the amplified DNA fragments (6338bp) were purified, and directly digested by EcoRI and NotI. The digested products were ligated to pre-digested pET-21b to obtain pTL-A03.

3.3.5 Cloning of Module 5 + TE of Fluvirucin B₁ Polyketide Synthase

The PCR reaction for module 5 was performed in a 50µl reaction mixture containing: 1X Phusion GC buffer, 0.2mM dNTP, 0.3mM MgCl₂, 10%DMSO, 1µM of each primer (AMod5-NheI-

F3: AAAAAAGCTAGCATGGCTGACGAAGAGAAGCTCCTC; AMod5-HindIII-R2: TTTTAAAGCTTC-GCGCCGTTCGA), 1U Phusion Hot Start II DNA polymerase (Thermo scientific, USA) and approximately 450ng of genomic DNA was added as template.

The thermal cycler was programmed as for module 1, and the amplified DNA fragments (7196bp) were then extracted and subjected to NheI and HindIII double digestion. Finally, the digested products were ligated to pre-digested pET-21b to obtain pTL-A05.

3.3.6 PCR Organic Solvent Study

The PCR condition and the operation of thermocycler are the same with the method for cloning of module 1, except for the PCR additive was swap with the following organic solvent: 3% , 6%, 10%, 15% (v/v) of DMSO; 0.15M, 0.3M, 0.6M, 1M of sulfolane; 5% 10% 15% 20% of glycerol; 0.5M, 1M, 2M 3M of betaine; 0.25M, 0.5M, 1M, 2M Of 1, 3-butanediol; 0.1M, 0.25M, 0.5M, 1M of 1,5-pentanediol; 0.25M, 0.5M, 1M, 2M of xylitol; 0.25M, 0.5M, 1M, 2M of d-sorbitol, are used in the PCR reaction.

3.3.7 Cloning of Module 2 of Fluvirucin B₁ Polyketide Synthase

The PCR reaction for pre-module 2 was performed in a 50 µl reaction mixture containing: 1X Phusion GC buffer, 0.2mM dNTP, 0.3mM MgCl₂, 15% glycerol, 0.5M sulfolane, 1µM of each primer (Avul-Mod2-pF6: CCCATCAACACCCACACCCT; Avul-Mod2-R7: GCCATCCACAGGTAGCGG-

TTG), 1U Phusion Hot Start II DNA polymerase (Thermo scientific, USA) and approximately 450ng of genomic DNA was added as template.

The thermal cycler was programmed according to the following “stepdown” amplification profile: 3min denaturation at 98°C, then 10 initial cycles of 10sec denaturation at 98°C, 30sec annealing at 72-68°C, 6min elongation at 72°C where the annealing temperature was decreased by 0.4°C per cycle, followed by 27 cycles of 10sec denaturation at 98°C, 30sec annealing at 68°C, 6min + (5sec/cycle) elongation at 72°C, and a final extension step at 72°C for 5min.

The amplified DNA fragments (6760bp) were then purified, and directly inserted into the plasmid vector PCR-Blunt II Topo (Zero blunt TOPO PCR cloning kit, Invitrogen) to obtain pM2-44-4-3 (pTL-preM2). After obtaining pTL-preM2, the same PCR protocol was performed as stated above, except for using the following primers: AMod2-NdeI-F8: AAAAAACATATGACGCTGGTGTTCGACCAC; AMod2-HindIII-R1: TTTTTAAGCTTGGACGCGCCGAGCTGGTC. The DNA amplicons (5265bp) were digested by NdeI and HindIII, and then were ligated to pre-digested pET-21b to obtain pTL-A02.

3.3.8 Cloning of Module 4 of Fluvirucin B₁ Polyketide Synthase

The PCR protocol for pre-module4 is the same as for pre-module2, except for using the following primers: AMod4-HindIII-F8: AAAAAAAGCTTCGGCAAGATCATCCTGACCATGC and

AMod4h-R10: CGGGTACATGCCCAAGGAGTTGA are for M4-1f fragments (5863bp); AMod4h-F7: CAACGCACAAGACATCCAACA and AMod4-XhoI-R7: TTTTTTCTCGAGCAGGCCCTGGTCGATC-AGCGAGAAGAGC are for M4-2f fragments (4339bp). M4-1f fragments and M4-2f fragments are separately inserted into the plasmid vector PCR-Blunt II Topo to obtain pTL-M4-1f and pTL-M4-2f. Later, the pTL-M4-1f plasmids were digested by HindIII and MluI to obtain M4-1f fragments again to clone into pTL-M4-2f to harvest pTL-preM4.

Finally, the pTL-preM4 was digested by NotI and XhoI and ligated to pre-digested pET-21b to obtain pTL-A04.

3.3.9 General Procedure for Protein Expression and Isolation

E. coli (BL-21) bearing the appropriate plasmid was grown in 1 L shake cultures of LB-ampicillin media at 37 °C until the OD₆₀₀ was between 0.6 and 0.8. Overexpression was induced by adding 200 µL of 1 M IPTG at appropriate induction temperature (see below) for 16 h. After this point, all work was carried out at 4 °C. Cells were pelleted by spinning at 6000 rpm for 10 min and resuspended in 10 mL of lysis buffer (20 mM TrisHCl, 150 mM NaCl, 1 mM Na₂EDTA, 1 mM EGTA, 1% Triton, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM Na₃VO₄, 1 µg/mL leupeptin, pH 8). Cells were lysed for five 30 sec intervals with a 60 sec cool down period between each. Lysed cells were spun at 14,000 rpm for 60 min. The lysate supernatant was

equilibrated with 3 mL of Ni-NTA bead slurry for 60 min. The mixture was then poured into a 15 mL column, and the supernatant was eluted. The column was then washed with two 15 mL portions of wash buffer (50 mM phosphate, 300 mM NaCl, 50 mM imidazole, pH 8.0) and eluted with 3 mL of elution buffer (50 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0). The purified protein was loaded onto a 100 kDa cutoff centrifugal concentrator, diluted to 15 mL with storage buffer (100 mM Tris, 1 mM EDTA, 1 mM dithioerythritol, 10% glycerol, pH 8), and spun at 3000 rpm. Dilution and filtration was repeated a total of three times. Protein concentrations were determined by Bradford assay with an average concentration of approximately 500 μ M. Proteins were flash frozen and stored at -80 °C until use.

3.3.10 Materials

All biochemicals, chemicals, and media were obtained from Fisher Scientific, all restriction enzymes were obtained from New England Biolabs, and other molecular biological reagents were obtained from Fisher Scientific, New England Biolabs, or Invitrogen. All PCR primers were synthesized by Eurofins MWG Operon.

All DNA sequences were deposited in the GenBank database under the following accession numbers: JX308234 (*FluA*), JX915256 (*FluB*), and JX448408 (*FluC*)

3.4 Supplement Information for Cloning of Module 4

The Complete List of the Primers for Module 4

Part1: the following primers were designed for PCR when DMSO is applied.

Avul-Mod4-F1: CCGCACCTCAGCAACCGCC
Avul-Mod4-R1: CAGCCATGCCGTCCCGCCTTC
Avul-Mod4-F2: CCCGCACCTCAGCAACCGC
Avul-Mod4-R2: AGCCATGCCGTCCCGCCTTC
Avul-Mod4-F3: CGATTCTGCGTGGGCTGGCT
Avul-Mod4-R3: TGGATGGCGACGAGCGAGGA
Avul-Mod4-F4: CCCTCAGCAACCGCCACGAC
Avul-Mod4-R4: CCTCGGGCAGGGGTCCTCGG
Avul-Mod4-F5: CCGCTCAATCCCTCGCTGG
Avul-Mod4-R5: TCTCCTCGACCTCGCGCAAACGT
Avul-Mod4-F6: CGCTGGAACGCCGTCTGACC
Avul-Mod4-R6: CGCTCCTTCTCCTCGACCTCG
Avul-Mod4-F7: TAAAGCGCCTGGTGTGGCCTC
Avul-Mod4-R7: ATCTTGATCACACCGGCCACG
Avul-Mod4-F8: GTTCACCGCCATGGGTGCCA
Avul-Mod4-R8: CGTTCGGAGCCGTCAACCCAT
Avul-Mod4-F9: AGGAGACCAAGCCCCGAC
Avul-Mod4-R9: GACTGGCAGGCCATGTGGATG
Avul-Mod4-F10: CGATCCACCGCCACACCAC
Avul-Mod4-R10: CAACGACTACGAGGTGGGCCGA
Avul-Mod4-F11: GCCTGCCCCTCAATCCCTC
Avul-Mod4-R11: GAAGAACGCGGGGTCGAACTC
Avul-Mod4-F12: TCTGCTCGGGGTGGAGGACG
Avul-Mod4-F13: GCCTCTCTGCTCGGGGTGGA
Avul-Mod4-R12: CTCTTCGTCAGCCATGCCGTC
Avul-Mod4-R13: AAGGACCGTTCGGAGCCGTCAAC
Avul-Mod4-F14: ACGGGTGGCACAGGTGGTCT
Avul-Mod4-R14: ATCACCCGCTCCTGCGAAGGAC
Avul-Mod4-F15: CGATGGTGAGTTGGCGGCGT
Avul-Mod4-R15: GTGGATGGCGACGAGCGAGG

Avul-Mod4-F16: GGTGTTCCGCCGAGGTGGAGC
Avul-Mod4-R16: ACGCCGAACGACGACACACC
Avul-Mod4-F17: GTTCGCCTGGAACGGGGTGG
Avul-Mod4-R17: CCGACCGACTCCCACAACGC
Avul-Mod4-F18: GATCCACCGCCACACCACCG
Avul-Mod4-R18: GACCCCGACCGACTCCCACA
Avul-Mod4-F19: GATGGTGAGTTGGCGGCGTTGC
Avul-Mod4-R19: TGCCGTCCCGCCTTCCCTAA
Avul-Mod4-R20: CCGTTCAGTCCACCCCA
Avul-Mod4-R21: CCGTTCGGAGCCGTCAACCC
Avul-Mod4-R22: AGTCCCGATCCCGCCCGTAC

Part2: the following primers were primarily designed for simple cloning as a full module.

AMod4-F1: AAAAA AGCGTGTACCGGCAGCGGC
AMod4-XhoI-R1: TTTTTTCTCGAGCAGGCCCTGGTCGATCAGCGA
AMod4-F2: AAAAAA TCCAGCGTGTACCGGCAGCGGC
AMod4-XhoI-R2: TTTTTTCTCGAGCAGGCCCTGGTCGATCAGCGAGAA
AMod4-F3: ACCTGCTCCAGCGTGTACCG
AMod4-XhoI-R3: TTTTTTCTCGAGCAGGCCCTGGTCGAT (shorter and use HPSF purification when synthesized)
AMod4-F4: CAAGCGGGACGCCGTTCAAG
AMod4-XhoI-R4: TTTTTTCTCGAGCAGGCCCTGGTCGATCAG
AMod4-F5: ATGCCGGAGCCGTTCTCGTA
AMod4-XhoI-R5: AAAAAACTCGAGCAGGCCCTGGTCGATCAG
AMod4-F6: AGGCCACTCGCCACCTCGAT
AMod4-XhoI-R6: TTACTACTCGAGCAGGCCCTGGTCGATCAG
AMod4-F7: CCGCCCTCTTGACCGGGTA
AMod4-XhoI-R7: TTTTTTCTCGAGCAGGCCCTGGTCGATCAGCGAGAAGAGC
AMod4-F8: CCTCCCGCCGAGGAGACCAA
AMod4-XhoI-R8: TTTTTTCTCGAGCAGGCCCTGGTCGATCAGCGAG
AMod4-XhoI-R8-1: TTTTTTTTTTCTCGAGCAGGCCCTGGTCGATCAGCGAG
AMod4-XhoI-R8-2: TTTTTTTTTTTTTCTCGAGCAGGCCCTGGTCGATCAGCGAG
AMod4-XhoI-R8-3: TTTTTTTTTTTTTTCTCGAGCAGGCCCTGGTCGATCAGCGAG

AMod4-F9: GATTCTGCGTGGGCTGGCTCG

AMod4-F10: ATTCTGCGTGGGCTGGCT (like Avul-mod4-F3 but shorter and has lower GC% and use HPSF purification)

AMod4-F11: TTCGGGGTTCACCGCCAT (last 8bp have resemble site within the target sequence)

AMod4-F12: TCACGGCGGTCTTCCACA (last 8bp only present once in the margin of target sequence)

AMod4-F13: AAAAAACGCTCACGGCGGTCTTCCACA

AMod4-F14: AAAAAAGATTCTGCGTGGGCTGGCTCG

AMod4-F15: AAAAAACCTGCTCCAGCGTGTACCG

AMod4-F16: TGATCACGGGTGGCACAGGT

AMod4-F17: GCATCTGCACCAGGCCACTC

AMod4-F18: CGGGTGGCATGACGAACCAG

AMod4-F19: CGGAGCCGTTCTCGTACCT

AMod4-F20: CTCTGCCCATCAACACCCGCA

AMod4-F21: CGACGCTGGTGTTCGACCATC

AMod4-F22: GTGTTGACCATCCCACGCC

Avul-Mod4-aF1: TCAAGCGTGCTACCGCCGACCT

Avul-Mod4-aF2: GGGACTACCTCAAGCGTGCTAC

Avul-Mod4-aF3: CGTGCTACCGCCGACCTTCAG

Avul-Mod4-aF4: GCGAGGTCGAGGCGAAGGATC

Avul-Mod4-bF1: TGC GTGGCGACGAGGTGTTT

Avul-Mod4-bF2: AGGGTGTCTGGGTTGTCGCTGGT

Avul-Mod4-bF3: GGGAGGCGGTGCTGGTGTTC

Avul-Mod4-bF4: ATCACGGGGTGGGGAGACG

Avul-Mod4-bF5: AGGTGTTCTGCCGAGGTGGAGCT

Avul-Mod4-bF6: TGTCGTCGTTCTGGCGTCAGC

Avul-Mod4-bF7: GACTACGCCTCGCACTCCGC

Avul-Mod4-bR1: ATGCCGTCCCGCCTCCCTAAG

Avul-Mod4-bR2: GCAAACGTCTGGTTGGCCTGCTG

Avul-Mod4-bR3: ACCTCGTAGTCGTTGGCGAAGGC

Avul-Mod4-bR4: TCGGCCACCTCGTAGTCGTTG

Avul-Mod4-pF1: CTCGTACCTCTGCCCATCAACAC

Avul-Mod4-pR1: TTCTCTTCGTAGCCATGCCGT

Avul-Mod4-pF2: ATCAACACCCGCACCCTCAG

Avul-Mod4-pR2: TAAGTCGTCCCGCTGCTGTC

Avul-Mod4-pF3: CACACCGGCAAGATCATCCTGAC
Avul-Mod4-pR3: GGTTTCGAGTTGCTGTGGGAC
Avul-Mod4-pF4: ATGACGAACCAGTTGTCCGGAAC
Avul-Mod4-pR4: TAAGGGGTTTCGAGTTGCTGTG
Avul-Mod4-pF5: GACGAACCAGTTGTCCGGAACTC
Avul-Mod4-pR5: TTCCTAAGGGGTTTCGAGTTG
Avul-Mod4-pF6: TCCTTGACCTCTGTCGAGTTGC
Avul-Mod4-pR6: GGATCTCGGTGAAACATTCTC
Avul-Mod4-pF7: CCGTTCAAGGATCTCGGTTTC
Avul-Mod4-pR7: CCTTCCCTAAGGGGTTTCGAGT
Avul-Mod4-pF8: CAGTTGTCGGAACTCGACCTG
Avul-Mod4-pR8: CTCTCGATGGACTCCAGAGCA
Avul-Mod4-pF9: ACCAGTTGTCGGAACTCGACCT
Avul-Mod4-pR9: GCTGGAGATCCTCGGTCACTC
Avul-Mod4-pF10: CTCGCCTGGTAAGCAGCATC
Avul-Mod4-pR10: CGTGGTCACGTACAGGGTCTG
Avul-Mod4-pF11: CAACTCCTGGACGTCGCTCT
Avul-Mod4-pR11: AACCTGACCCTGCTCCCAAT
Avul-Mod4-sF1: ACTGACCTGGAGACGGCTGGTG
Avul-Mod4-sF2: GACGCCCGTTCAAGGATCTCG
Avul-Mod4-sF3: CCTGGACGTCGCTCTGACAATGC
Avul-Mod4-sF4: GGACGCCCGTTCAAGGATCTC
Avul-Mod4-sR1: GCGAGGAGCTTCTCTTCGTCAGC
Avul-Mod4-sR2: GTCCTCGGCCACCTCGTAGTC
Avul-Mod4-sR3: GCTGGAGATCCTCGGTCACTCG
Avul-Mod4-sR4: GTAAGTCGTCCCGCTGCTGTCC
Avul-Mod4-cF1: GTGCTGGTGTCCCTGGTCTGC
Avul-Mod4-cF2: CATCAACACCCGCACCCTCAG
Avul-Mod4-cF3: CGTTCCTCGTACCTCTGCCATC
Avul-Mod4-cF4: GGAGCCGTTCTCGTACCTCTGC
Avul-Mod4-cF5: GAGCTGTCGCCGATGGTGAG
Avul-Mod4-cR1: CTCCACGAACTCCTCCAGCACCT
Avul-Mod4-cR2: CCACGAACTCCTCCAGCACCTG
Avul-Mod4-cR3: GTTGTCCGCACTCACCTCCAACC
Avul-Mod4-cR4: CCACCTCGTAGTCGTTGGCGAAG
Avul-Mod4-cR5: CCTGCTGGAGATCCTCGGTCACT

Part3. the following primers were designed for cloning module 4 into two separated half-parts

DNA, when a Mlul site at 3343/3347 [GCGGCCGTGG A[▼]CGCG↓T CATGCCGCCG] of module 4 was identified.

3a. For M4-1f fragments:

AMod4-HindIII-F1: AAAAAA AAGCTT GGCAAGATCATCCTGACCA
AMod4-HindIII-F2: AAAAAA AAGCTT CGTTCAAGGATCTCGGTTTC
AMod4-HindIII-F3: AAAAAA AAGCTT GCGTTCGTGCTCTACTCCTC
AMod4-HindIII-F4: AAAAAA AAGCTT CAAGATCATCCTGACCATGC
AMod4-HindIII-F5: AAAAAA AAGCTT TATCGCAGCTTCGACCTGAT
AMod4-HindIII-F6: AAAAAA AAGCTT CAGCTTCGACCTGATGGATG
AMod4-HindIII-F7: AAAAAA AAGCTT CGCAGCTTCGACCTGATGGATG
AMod4-HindIII-F8: AAAAAA AAGCTT CGGCAAGATCATCCTGACCATGC

(The following were used with HindIII-forward)

AMod4h-R1: AGAGTCGCCCGTCTCCAGGGT
AMod4h-R2: AGCTCCCGCAGGATCAGCGA
AMod4h-R3: CTTGAGCCCCTGGAACGCCG
AMod4h-R4: GTGGCTCGACACGTCCACCG
AMod4h-R4: CCTGCTCGGGGTCGGACAGA
AMod4h-R5: GAGTCGTCGGAGAGTCGTTC
AMod4h-R6: GAAGGGCAGAGTGATCTTCG
AMod4h-R7: TCTTCGAGTCGTCGGAGAGT
AMod4h-R8: CACGTACTCCTCGACACCAC
AMod4h-R9: GGTACATGCCCAAGGAGTTGA
AMod4h-R10: CGGGTACATGCCCAAGGAGTTGA
AMod4h-R11: TTCGAGTCGTCGGAGAGTCGTTC

3b. For M4-2f fragments:

(The following were used with XhoI-reverse)

AMod4h-F1: ACCCATCTCACCGCCACCAA
AMod4h-F2: TCACATCGGCACGAGCAACG
AMod4h-F3: GCCCTCAAACCGCCCTCAC
AMod4h-F4: CGTTGTGGCTGGGGTCGGTC

AMod4h-F5: CCAAACCACCTACACCCAACC

AMod4h-F6: ACACATCACCCACCACTTCA

AMod4h-F7: CAACGCACAAGACATCCAACA

3.5 Reference

1. Lau, J., Fu, H., Cane, D. E., and Khosla, C. (1999) Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry*. 38, 1643-1651.
2. Ruan, X., Pereda, A., Stassi, D. L., Zeidner, D., Summers, R. G., Jackson, M., Shivakumar, A., Kakavas, S., Staver, M. J., and Donadio, S. (1997) Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J. Bacteriol.* 179, 6416-6425.
3. Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R.; McDaniel, R. (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry*, 42, 72–79.
4. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. (1999) Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel “Unnatural” Natural Products, *Proc. Natl. Acad. Sci. USA*, 96, 1846-1851

5. Xue, Q., Ashley, G., Hutchinson, C., and Santi, D. (1999) A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. U.S.A.*, 96, 11740– 11745
6. Lin T-Y, Borketey L.S, Prasad G, Waters S.A, Schnarr N.A. (2013) Sequence, Cloning, and Analysis of the Fluvirucin B1 Polyketide Synthase from *Actinomadura vulgaris*. *ACS Synth. Biol.*, 2, 635–642
7. Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabinowitsch, E.; Rajandream, M.-A.; Rutherford, K. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417 (6885), 141–147.
8. Rees WA¹, Yager TD, Korte J, von Hippel PH. (1993) Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry*, 32 (1), 137–44
9. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res*, 19 (14), 4008
10. Chakrabarti R. "Chapter 6. Novel PCR-Enhancing Compounds and Their Modes of Action" *PCR Technology: Current Innovations, Second Edition* (Edited by Hugh G. Griffin , Thomas Weissensteiner and Annette Griffin) CRC Press 2003, P.51–63

11. Chakrabarti R, Schutt CE. (2001) The enhancement of PCR amplification by low molecular-weight sulfones. *Gene*, 274 (1–2), 293–298.
12. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Research*, **34** (19), 5402–5415.

CHAPTER 4.

EXAMING SUSTRATE SELECTIVITY OF DOMAINS IN FLUVIRUCIN B₁ PKS SYSTEM AND BIOINFORMATIC ANALYSIS OF KETOSYNTASE ON SUBSTRATE SPECIFICITY

4.1 Introduction

In the past two decades, the effort of engineering polyketide synthase has focused on the domain manipulation of the enzyme systems, but the substrate specificities of critical domains obstruct the engineering effort and the development of producing “un-natural” natural product. A key step within the megasynthase is the substrate transferring of the polyketide intermediate between an upstream ACP and a downstream KS domain (**Figure 24**). This process is controlled

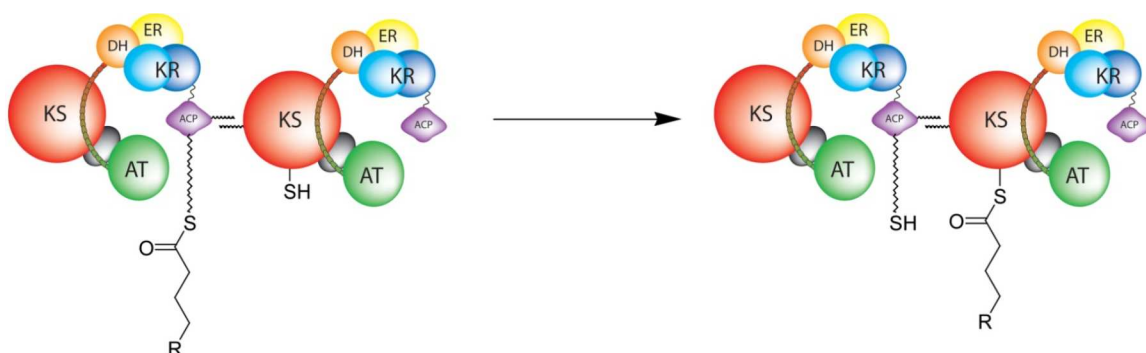


Figure 24. Schematic diagram of intermodular transfer in modular polyketide synthases.

A polyketide intermediate is passed from an upstream ACP to a downstream KS via the phosphopantetheine arm of holo-ACP. Often the substrate selectivity of the acceptor KS domain determines the efficiency of this process and controls the flow of intermediates down the assembly line. KS = ketosynthase, ACP = acyl carrier protein, AT = acyltransferase, DH = dehydratase, ER = enoyl reductase, KR = ketoreductase.

either via precise regulation of kinetics where ACP to KS transfer is slow compared to KR, DH, and ER activity, or strict inherent substrate selectivity of the KS domain. For kinetically-controlled transfer, changes in substrate functionality should not impede the overall production of engineered products. Conversely, those KS domains that exhibit tight substrate selectivities may detrimentally impact product yields through greatly reduced kinetics and off-target, shunt product formation. In these cases, a comprehensive understanding of substrate tolerance and the ability to alternative selectivities is required for polyketide diversification. Recent insights from KS structure determination have provided clues to potential substrate recognition motifs within the KS active site pocket. However, these principles remain largely untested and thorough analysis is needed to determine whether KS selectivities can be changed or broadened by proper mutation of residues within these motifs.

Early determinations of KS substrate selectivity relied heavily on radiolabeling of elongation products or, less commonly, the substrates themselves.¹⁻⁴ Small molecules were generally delivered to the KS active site via a compatible ACP or as an *N*-acetylcysteimine (SNAc) thioester, a more synthetically accessible mimic of the phosphopantetheine (Ppant) arm of *holo*-ACPs. Subsequent addition of radiolabeled extender units resulted in a single elongation event with the KS-bound substrate. The amount of labeled product excised from the protein provided an indirect measure of KS loading with each substrate. While very sensitive, the combination of

high reagent cost and limited substrate availability has severely restricted the use of this method for medium and high-throughput applications.

More recently, a number of groups have developed mass spectrometry-based techniques for examining KS active site acylation.⁵⁻⁷ Many of these strategies utilize tandem proteolysis/LC-MS to tease out differences in loading propensities for various substrates. Despite the many advantages compared to radiolabeling, MS-based methods still suffer from a number of important shortcomings. Firstly, the extremely large size of intact PKS modules generally prohibits their use in these analyses. Smaller subunits such as KS-AT didomains are often employed as substitutes for the full module and currently it is unclear what, if any, effect this has on substrate selectivity. In addition, the influence of various substrates on ionization potentials, especially when proteolysis is employed prior to analysis, has not been fully explored and may significantly impact data interpretation and reproducibility. Finally, when particular hydrophobic groups (e.g. octyl, naphthyl, cyclohexyl) are utilized as a substantial part of a given substrate, loss of signal for proteolysis fragments associated with KS domains is often observed, likely due to solubility issues.

In addition, our lab has developed a fluorescence transfer assay for facile determination of KS selectivities.⁸ The degree of KS acylation for a given SNAc substrate can be assayed by introducing a fluorescent ACP, prepared from Ppant acylation with a clickable β -lactam, after incubation of KS and substrate and monitoring ACP to KS fluorescence transfer via gel

electrophoresis. The amount of fluorescence transfer indicates to what extent the SNAc compound is accepted by the KS domain (**Figure 25**). High KS-loading blocks subsequent fluorescence transfer from ACP, and thus produces little to no fluorescent signal in the gel. Alternatively, insufficient KS-acylation with a given SNAc compound leaves most KS active sites free to accept the fluorescent substrate from ACP resulting in a bright band. This technique provides a direct measure of KS selectivity while circumventing the limitations of radiolabeling

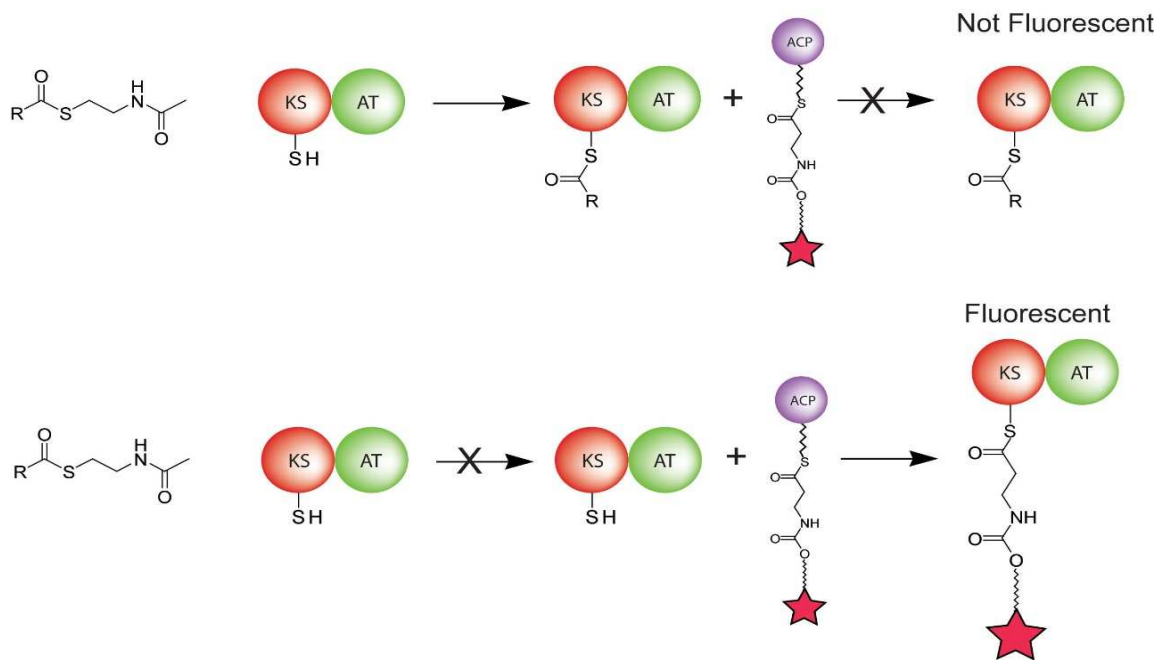


Figure 25. Schematic diagram of a mechanism-based, fluorescence transfer assay for KS-selectivity. A given KS domain is incubated with acyl-SNAc followed by addition of fluorescently labeled ACP. The amount of fluorescence transferred from the ACP to the KS active site provides a direct readout of the substrate tolerance for each acyl-SNAc examined. KS = ketosynthase, ACP = acyl carrier protein, AT = acyltransferase, SNAc = *N*-acetylcysteamine.

and mass spectrometry methods described above. Most importantly, the use of mechanism-based fluorescence transfer from acyl-ACPs ensures that any fluorescence observed on the acceptor module results solely from the KS active site and not from competing, nucleophilic surface residues within the complex protein system.

As we unearthed a new PKS system, fluvirucin B₁ synthase, an ideal system for engineering un-natural polyketides, we were eager to figure out the substrate specificity within the domains of fluvirucin B₁ PKS in order to overcome the engineering hurdle that we describe above. Herein, we described our finding for FluAT selectivity and FluKS specificity, and also provided our bioinformatic analysis of FluKS for future engineering purposes (see section 5.3.1).

4.2 Results and Discussions

4.2.1 Cloning and Protein Expression of KSATs and ACPs

To determine the selectivities of the domains in this assembly line, appropriate primers were designed for the cloning of all KS-AT didomains within fluvirucin B₁ PKS. However, even though all five KSATs were cloned out without any difficulties, only KSAT of modules 1, 3 and 5 were successfully expressed in *E. coli* BL21 strain. On the contrary, ACPs of all the five modules were successfully cloned and expressed in *E. coli* BAP1 strain in order to harvest the fully

functional ACPs purified with the Ppant group. In addition, the terminal TE domain were also cloned and overexpressed in BL21 (**Figure 26**).

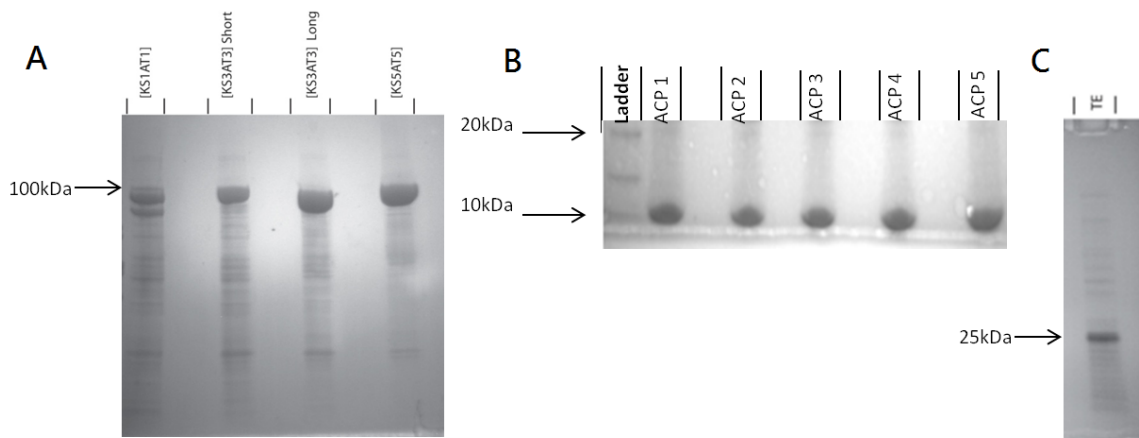


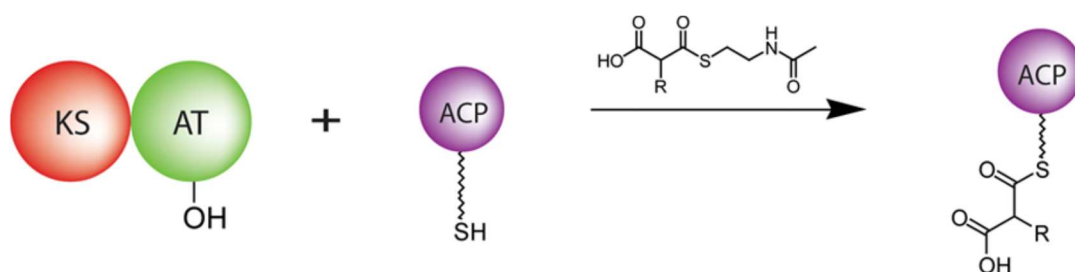
Figure 26. SDS PAGE analysis of fluvirucin B₁ PKS domains overexpressed in *E. coli* following Ni-NTA affinity purification. Lanes are marked with corresponding protein. Molecular weights are indicated for each gel. **(A)** Flu-KSAT didomains **(B)** Flu ACP domains **(C)** Flu TE domain.

4.2.2 Examine the AT Selectivities.

To examine the domain selectivities, we began synthesizing substrate analogs to probe FluATs. We are convinced that not only the information could be useful to confirm the result from our bioinformatic analysis, but also we could establish the methodology we deployed for other domains as well in terms of using LC-MS.

As we alluded in chapter 2, using the SEARCHPKS program developed by Mohanty and co-workers⁹, modules 2 and 4 showed high sequence similarity with malonyl-specific AT domains, while module 3 was predicted to utilize methylmalonyl-CoA. Modules 1 and 5 returned a single

hit for ethylmalonate specificity amidst several methylmalonylspecific AT domains. To experimentally verify the putative AT selectivities for modules 1, 3, and 5, ketosynthase-acyltransferase (KSAT) didomains were cloned from these modules along with the corresponding ACP domains. Each KSAT didomain was mixed with ACP from the same module followed by introduction of either malonyl-, methylmalonyl-, or ethylmalonyl-N-acetylcysteamine (SNAc)



	FluAT1	FluAT3	FluAT5
	X	X	X
	X	✓	X
	✓	X	✓

Figure 27. Schematic diagram of and observed results for the AT substrate selectivity studies of modules 1, 3, and 5. Red checks indicate that the substrate shown on the left is transferred from the indicated AT to the ACP domain, whereas a black X indicates that no substrate transfer was observed. See Supporting Information for raw LC-MS data. FluATX = fluvirucin AT domain from module X.

thioesters. Following 30 min of incubation with each substrate, samples were trypsinized, and the extent of AT to ACP transfer for each extender unit was analyzed by LC-MS (**Figure 27**). Gratifyingly, experimentally determined AT selectivities for modules 1, 3, and 5 were consistent with those suggested by sequence homology. It is important to note that no direct acylation of ACP was observed with any of the extender units, indicating that AT to ACP transfer is the sole mechanism for formation of the acylated ACP (data not shown).

4.2.3 FluKS Substrate Selectivity

As we summarized before, the capability of KS domain accepting different substrates could potentially increase product yields in polyketide engineering. On the way in characterizing our unique fluvirucin B₁ PKS, one of our objectives is to test the tolerance of KSs by applying the technique we developed and described above. We believed that gaining the information of the KS selectivity would provide the reference of altered β -functionalities and insight for better polyketide engineering. Based on our current bioinformatic analysis and prediction, FluKSs only have two different kind preferences of β -substituents: one is a methylene and the other a β -hydroxyl substituent, we will focus on Flu KS3 which only accepts a β -hydroxy acyl group of all the KS domains.

To examine the selectivity in FluKS3 for the hydroxyl group stereo-configuration and provide further evidence connecting this PKS with fluvirucin B₁ production, tandem proteolysis/LC-MS was again employed. SNAc thioesters of both enantiomers of 3-hydroxybutyrate were prepared and introduced separately to KSAT3. Following 1 h of incubation, samples were trypsinized, and KS-acylation was observed via LC-MS analysis. As predicted from the fluvirucin B1 structure, only the (S)-3-hydroxybutyryl-SNAc compound, which places the hydroxyl group in the same three-dimensional orientation as the ring hydroxyl of the final product, was accepted by the module 3 KS domain (**Figure 28**). Very little to no KS-acylation with the (R)-isomer was observed.

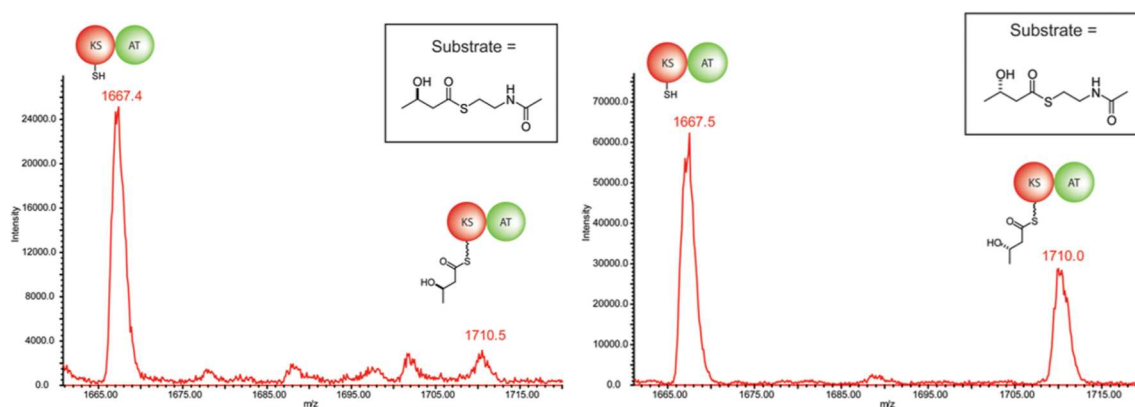


Figure 28. LC-MS data for module 3 KS-acylation with (R)-3-hydroxybutyryl-SNAc (top) and (S)-3-hydroxybutyryl-SNAc (bottom). Only the (S)-isomer is accepted by module 3 KS as is expected from the fluvirucin B1 structure. Peaks are labeled with the corresponding acylated or unacylated KSAT3 didomain. All peaks are $m/z = +2$.

The mechanism based fluorescence transfer assay that was developed in our lab was also used to examine the stereochemical preference of Flu KS3. The fluorescent ACP was prepared by incubating holo Flu-ACP3 with N-propargyloxycarbonyl- β -lactam for an hour. The acylated ACP was then “clicked” with rhodamine azide. The KSAT3 was preloaded with the substrates for an hour followed by addition of the fluorescent ACP for 30 minutes and SDS PAGE analysis of each reaction. The bright KSAT3 bands observed for the (R) - stereoisomer indicates that there was little to no KS acylation. The weak fluorescent band observed in the (S) - stereoisomer lane is an indication of efficient KS acylation and confirms the earlier LC-MS results obtained (**Figure 29**).

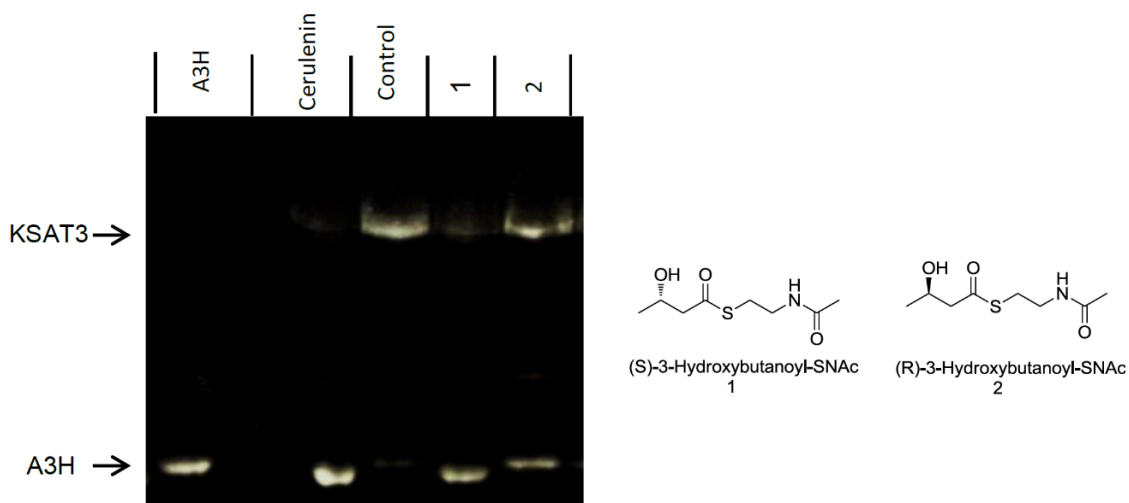


Figure 29. Mechanism based fluorescence transfer assay for Flu KSAT 3 substrate specificity towards (R)-and (S)-3-hydroxy butyryl SNAc

Subsequently, we prepared a small panel (**Figure 30**) of SNAc thioesters for a preliminary test on FluKS3. They incorporate different variations of hydroxyl, enoyl and fully reduced carbon at the beta position. This panel is a good starting point to investigate the preferences of KS for

variation of functionalities at the

beta position and be used further

as a framework for testing the

other four KSs in this PKS system.

By following the same procedure

for LC-MS methodology, we found

that FluKS3 prefer not only the (S)-

isomer but also the β -keto thioester probe, and that FluKS3 slightly interacts with the β -

methylene analog but does not accept any other substrates (data not shown). In sum, we suggest

these preliminary experiments will provide some insight into the extent and contrasts of tolerance

exhibited within other KSs.

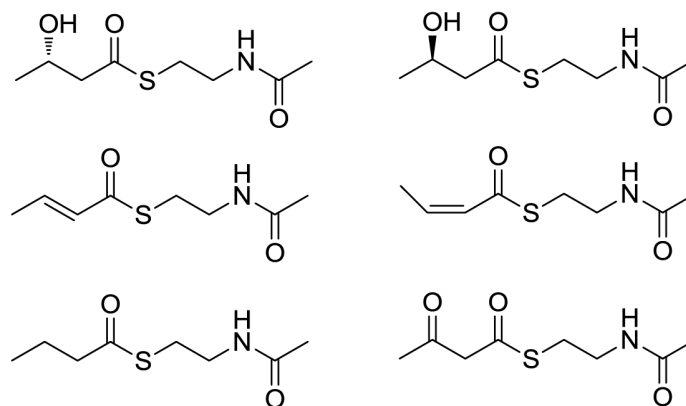


Figure 30. A panel of substrate analogs for KS selectivity

4.3 Experimental Procedures

4.3.1 Cloning of KSAT and ACP Domains of Fluvirucin B₁ PKS

The DNA sequence encoding Flu-KSAT1 was amplified from pTL-A01 by the PCR protocol described for pre-module 2. Flu-KSAT1 was constructed as an NheI-EcoRI fragment by using following primers, pTL-KSAT1-F: TTTTTTGCTAGCGAGCCCATCGCGATCGTC and pTL-KSAT1-R: AAAAAAGAATTCTGGTCCACGGCGGCCTGG. This NheI-EcoRI fragment was cloned into the pET21b

expression vector to yield plasmid pTL-KSAT1. The DNA sequence encoding Flu-KSAT3, Flu-KSAT5 and Flu-ACP1, Flu-ACP3, Flu-ACP5 were cloned similarly by the corresponding templates the corresponding primers as follows, pTL-KSAT2-F: TTTTTTGCTAGCGAGCCGATCGCGATCGTG

pTL-KSAT2-R: AAAAAAGAATTC**AG**GGAGGACAGGGCGGGAAG

pTL-KSAT3-F1: TTTTTTGCTAGCATGGCCACTGACGACAAG,

pTL-KSAT3-R1: AAAAAAGAATTCGGATCCACCCGGGCAGG;

pTL-KSAT4-F: TTTTTTGCTAGCATCGGGGAGCCGATCGCGATC,

pTL-KSAT4-R: AAAAAAGAATTC**TG**GTCCGCGGGGCCTGGC;

pTL-KSAT5-F1: TTTTTTGCTAGCATGGCTGACGAAGAGAAG,

pTL-KSAT5-R2: AAAAAAGAATTC**TG**ATCCACCCGAGCCTG;

pTL-ACP1-F: TTTTTTGCTAGCCTGACCGGACTACCGGCG,

pTL-ACP1-R: AAAAAAGAATTC**GT**GACACGCTGGAGCAG;

pTL-ACP2-F: TTTTTTGCTAGCCTCGCGGAGCTGCCGCCGGCC,

pTL-ACP2-R: AAAAAAGAATTC**GT**GAGCCTGTCTGGACGAAC;

pTL-ACP3-F: TTTTTTGCTAGCCTGACCGCCTGCCCGCG,

pTL-ACP3-R: AAAAAAGAATTC**GT**GACACGCTGGAGCAG;

pTL-ACP4-F: TTTTTTGCTAGCCTGGCCGGGCTGCCCCCG

pTL-ACP4-R: AAAAAAGAATTC**AG**GAGCTGGCTCCCCATGTA

pTL-ACP5-F: TTTTTTGCTAGCCTGGCCGGGCTGTCG,

pTL-ACP5-R:AAAAAAGAATTCGCGATCTCCTCCGCCAG

The resulting plasmids for Flu-KSATs (pTL-KSAT1 to 5) were transformed into BL21(DE3).

The plasmids constructed for Flu-ACPs (pTL-ACP1 to 5) were introduced into BAP1.

4.3.2 General Method for Preparation of All *N*-Acetylcysteamine (SNAc) Thioester Derivatives

To a solution of triethylamine (2.80 mmol) in dichloromethane (10 mL) was added the appropriate acid (1.40 mmol), (3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (1.40 mmol), 1-hydroxybenzotriazole (HOBt) (1.40 mmol) and *N*-acetylcysteamine (SNAc) (1.35 mmol) under argon. The reaction mixture was stirred overnight. The organic layer was washed with saturated NaHCO₃ solution, 0.1 N HCl solution and brine. It was then dried over anhydrous sodium sulfate, concentrated under vacuum, and purified by flash column to provide the final product in pure form.

Butyroyl SNAc

Using butyric acid, the general procedure was followed

¹H NMR (400 MHz, CDCl₃) δ ppm 0.92-0.96(t, J=, 3H), 1.67-1.69(m, 2H), 1.95(s, 3H), 2.52-2.56(t,2H),2.99-3.00(t, 2H), 3.40-3.42(q, 2H), 6.12(s, 1H).

^{13}C NMR (400 MHz, CDCl_3) δ ppm 11.54, 17.17, 21.27, 26.55, 37.64, 44.12, 168.64, 197.97.

IR: ν 3291.13, 2967.08, 1651.12, 1547.47, 1436.23, 1370.50, 1284.55, 1112.64, 988.76, 905.33,

723.31 cm^{-1}

LC-MS $[\text{M}+\text{H}]^+ = 190.0$.

Trans-2-butenoyl SNAc

Using crotonic acid, the general procedure was followed.

^1H NMR (400 MHz, CDCl_3) δ 1.86 (dd, 3 H), 1.93 (s, 3 H), 3.05 (t, 2 H), 3.44 (dt, 2 H), 6.11 (dq, 1H),

6.21 (s, 1 H), 6.90 (dq, 1 H);

^{13}C NMR (400 MHz, CDCl_3) δ 17.9, 23.1, 28.1, 39.8; 129.8, 141.7, 170.3;

LC-MS $[\text{M}+\text{H}] 188.2$

(R)-3-hydroxy butyroyl SNAc

Using (R)-hydroxy butyric acid, general procedure was followed

^1H NMR (400 MHz, $\text{CH}_3\text{Cl}-d$) δ ppm 1.33 (3 H, d, $J=6.82$ Hz), 1.98 (3 H, s), 2.65 - 2.76 (2 H, m), 2.78

- 2.82 (1 H, d), 3.04 - 3.10 (2 H, t), 3.27 - 3.31 (1 H, m), 3.42-3.46(3H, q), 6.36 (1 H, br. s.)

^{13}C NMR (400 MHz, $\text{CH}_3\text{Cl}-d$) δ ppm 19.76, 21.35, 26.95, 28.50, 34.57, 37.48, 49.19, 168.60, 195.73

LC-MS $[\text{M}+\text{H}]$ for $\text{C}_8\text{H}_{15}\text{NO}_3\text{S}$ calcd 206.1 found 206.1

(S)-3-hydroxy butyroyl SNAc

Using (S)-hydroxy butyric acid, general procedure was followed.

^1H NMR (400 MHz, $\text{CH}_3\text{Cl}-d$) δ ppm 1.33 (3 H, d, $J=6.82$ Hz), 1.98 (3 H, s), 2.65 - 2.76 (2 H, m), 2.78 - 2.82 (1 H, d), 3.04 - 3.10 (2 H, t), 3.27 - 3.31 (1 H, m), 3.42-3.46(3H, q), 6.36 (1 H, br. s.)

^{13}C NMR (101 MHz, $\text{CH}_3\text{Cl}-d$) δ ppm 19.76, 21.35, 26.95, 28.50, 34.57, 37.48, 49.19, 168.60, 195.73

LRMS [M+H] for $\text{C}_8\text{H}_{15}\text{NO}_3\text{S}$ calcd 206.1 found 206.1

β -keto butyroyl SNAc

To a solution of meldrum acid (10mmol) in dichloromethane at 0°C was added pyridine (21mmol). Acetyl chloride (12mmol) was added dropwise to the mixture, with the solution turning deep red. The reaction was stirred for an hour at 0°C and then warmed to room temperature overnight. Mixture was washed 3X with 0.1M HCl, dried with sodium sulfate and concentrated to yield the acetylated meldrum acid as a dark orange solid in 84% yield. 8mmol of this acetylated meldrum acid and 7.7mmol of N-acetylated cysteamine (SNAc) were refluxed in toluene for 5 hours. After the mixture was cooled to room temperature, it was concentrated to yield titled compound in 78% yield.

4.3.3 Synthesis of malonyl and substituted malonyl SNAc thioesters

These syntheses we carried out using established procedures²⁵. A general outline follows:
To a solution of appropriate malonic or substituted malonic acid (1eq) in dry THF (5mL), was added pyridine (2.2eq) and t-butanol (1.8eq). The solution was cooled to 0°C . Methanesulfonyl chloride

(1.05eq) was then added dropwise over a 10 minute period. The reaction mixture was warmed to room temperature and stirred for 3 hours. The mixture was filtered and the resulting filtrate diluted with water. The pH was adjusted to ~12 and washed 3X with dichloromethane. The aqueous layer was acidified (pH ~2), extracted 3X with dichloromethane and dried with sodium sulfate. The product was then coupled to SNAc via EDC coupling. In an RB flask, the acid (1.1eq) was dissolved in dichloromethane. EDCI (1.2eq), DMAP (0.02eq) and SNAc (1eq) were then added and reaction stirred at room temperature overnight. The mixture was diluted with water and dichloromethane. The organic layer was washed with NH₄Cl, NaHCO₃, and brine, and then dried with Na₂SO₄. Concentration *in vacuo* yielded the expected product. The product was dissolved in TFA at 0°C. After stirring at 0°C for 24 hours, TFA was evaporated. The product was diluted with diethyl ether and concentrated. This was repeated three times. The crude product was purified by chromatography to yield the titled compound.

Malonyl SNAc thioester (pale white solid) (70% yield)

¹H NMR (400 MHz, DMSO-*d*₆) δ ppm 1.79 (br. s., 3 H), 2.95 (br. s., 2 H), 3.18 (br. s., 2 H), 3.65 (br. s., 2 H), 8.06 (br. s., 1 H)

¹³C NMR (400 MHz, DMSO-*d*₆) δ ppm 21.04, 27.06, 36.57, 48.21, 165.97, 167.91, 190.24

LRMS [M+H] for C₇H₁₁NO₄S calcd 206.0 found 206.1

Methyl Malonyl SNAc thioester (pale white solid) (63% yield)

^1H NMR (400 MHz, CH_3Cl -*d*) δ ppm 1.44 (3 H, d, $J=7.07$ Hz), 1.99 (3 H, s), 3.06-3.17 (2 H, m), 3.40 - 3.52 (2 H, m), 3.63 - 3.77 (1 H, m), 6.71 (1 H, t, $J=5.56$ Hz), 10.35 (1 H, br. s.)

^{13}C NMR (400 MHz, CH_2Cl_2 -*d*) ppm 12.13, 20.74, 26.59, 37.64, 52.07, 170.24, 170.59, 194.96

LRMS [M+H] for $\text{C}_8\text{H}_{13}\text{NO}_4\text{S}$ calcd 220.1 found 220.1

Ethyl malonyl SNAc thioester (pale yellow solid) (51% yield)

^1H NMR (400 MHz, CH_3Cl -*d*) δ ppm 1.01 - 1.06 (2 H, m), 2.01 (3 H, s), 3.05 (1 H, dd, $J=13.26$, 6.69 Hz), 3.11 - 3.22 (1 H, m), 3.37 (1 H, t, $J=7.20$ Hz), 3.48 (2 H, q, $J=5.89$ Hz), 3.56 (1 H, t, $J=7.45$ Hz), 6.59 (1H, br. s.), 11.53 (1H, br. s.)

^{13}C NMR (101 MHz, CH_3Cl -*d*) δ ppm 9.83, 20.63, 20.98, 26.62, 37.73, 51.02, 59.44, 170.96, 172.31, 194.21

LRMS [M+H] for $\text{C}_9\text{H}_{15}\text{NO}_4\text{S}$ calcd 234.1 found 234.1

4.3.4 ACP Acylation by AT

To a mixture of ACP (50 μM) and KS-AT (2 μM) in 100mM pH 7.0 phosphate buffer (50 μL total volume) containing 2.5 mM TCEP at 4°C was added appropriate SNAc thioester substrate (500 μM). The mixture was incubated 4°C for 30 min to achieve AT assisted acylation of the ACP. Sequence grade modified trypsin was added to prepare samples with final ration trypsin:ACP to

be 1:10 (w/w). The mixture was incubated for 60 min at 37°C. Digestion was quenched by addition of equal volume of 10% formic acid. Digests were stored at -80°C until analysis.

4.3.5 KSAT3 Loading

To a mixture of Flu-KSAT3 (25 μ M) in 100 mM pH 7.0 phosphate buffer (50 μ L total volume) containing 2.5 mM TCEP at 4°C was added appropriate SNAc thioester substrate (5mM). The mixture was incubated 4°C for 60 min to achieve KS acylation. Sequence grade modified trypsin was added to prepare samples with final ratio trypsin:ACP to be 1:10 (w/w). The mixture was incubated for 60 min at 37°C. Digestion was quenched by addition of equal volume of 10% formic acid. Digests were stored at -80°C until analysis.

4.3.6 Fluorescence Transfer Assay

Unless otherwise stated, phosphate buffer refers to 100 mM, pH 7.0 phosphate. Flu-ACP3 (25 μ M total protein) was equilibrated at ambient temperature for 15 min in phosphate buffer containing 2.5 mM TCEP. Poc β -lactam (10 \times with respect to total ACP concentration) was added and the mixture was equilibrated at ambient temperature for 45 min. Rhodamine Azide (2 \times with respect to Poc β -lactam), sodium ascorbate (1 mM), and copper (II) sulfate (1 mM) were added to ACP samples which had been labeled with Poc β -lactam. The reaction was performed at ambient temperature for 45 min. To remove excess Poc β -lactam and Rhodamine Azide, the mixture was

loaded into a 3 kDa NMW concentrator and the volume reduced to 100 μ L by spinning in a centrifuge cooled to 4 °C. The mixture was diluted to 500 μ L with phosphate buffer, and then concentrated to 100 μ L again. This process was repeated a total of 5 times. Protein was removed from the concentrator by inverting it and spinning. For transfer of the fluorescent product, [Flu-KSAT3] was pre-treated with respective SNAc (2.5 mM) derivatives for 1 h. This mixture was introduced to filtered acyl-ACP3 and incubated for 30 min. The [Flu-KSAT3] control was executed under the same acylation and click conditions as in other samples but without the SNAc derivative and therefore represents full acyl transfer.

4.3.7 Chromophore Attachment

The reaction was carried out at 25 μ L, final concentrations reported. Rhodamine Azide, (2 \times with respect to Poc β -lactam), sodium ascorbate (1 mM), and copper (II) sulfate (1 mM) were added to ACP samples which had been labeled with Poc β -lactam. The reaction was performed at ambient temperature for 45 min.

4.3.8 Gel Assay

Labeled samples were diluted to 20 μ L with gel-loading buffer. Proteins were separated using a 4–20% gradient HEPES-PAGE gel (100 V, 50 mA, 70 min). Gels were developed in 10% acetic acid to visualize Rhodamine Azide. Labeled proteins were imaged on a UV-transilluminator.

Total protein was stained using Coomassie stain. trypsin was added to prepare samples with final ratio trypsin:ACP to be 1:10 (w/w). The mixture was incubated for 60 min at 37°C. Digestion was quenched by addition of equal volume of 10% formic acid. Digests were stored at -80°C until analysis.

4.4 Reference

1. Wu, N., Cane, D. E., and Khosla, C. (2002) Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* (N. Y.). 41, 5056-5066.
2. Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) Dissecting and exploiting intermodular communication in polyketide synthases, *Science*, 284, 482-485.
3. Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* (N. Y.). 40, 2326-2331.

4. Wu, N., Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc.* 123, 6465-6474.
5. Schnarr, N. A., Chen, A. Y., Cane, D. E., and Khosla, C. (2005) Analysis of covalently bound polyketide intermediates on 6-deoxyerythronolide B synthase by tandem proteolysis-mass spectrometry. *Biochemistry* (N. Y). 44, 11836-11842
6. Hong, H., Appleyard, A. N., Siskos, A. P., Garcia-Bernardo, J., Staunton, J., and Leadlay, P. F. (2005) Chain initiation on type I modular polyketide synthases revealed by limited proteolysis and ion-trap mass spectrometry. *FEBS Journal.* 272, 2373-2387.
7. Hicks, L. M., Mazur, M. T., Miller, L. M., Dorrestein, P. C., Schnarr, N. A., Khosla, C., and Kelleher, N. L. (2006) Investigating nonribosomal peptide and polyketide biosynthesis by direct detection of intermediates on >70 kDa polypeptides by using Fourier-Transform mass spectrometry. *ChemBioChem.* 7, 904-907.
8. Prasad G, Borketey LS, Lin T-Y, Schnarr NA. (2012) A Mechanism-Based Fluorescent Assay for Examining Ketosynthase Activity. *Org. Biomol. Chem.*, 10, 6717-23
9. Yadav, G., Gokhale, R. S., and Mohanty, D. (2003) SEARCHPKS: a program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res.* 31, 3654– 3658

CHAPTER 5.

SUMMARY AND FUTURE DIRECTIONS

5.1 Discussion

Fluvirucin B₁ is a relatively simple natural product stemming from a rather complex set of biosynthetic transformations. Despite the diminutive size of the PKS responsible for its production in *A. vulgaris*, each round of elongation and subsequent β -carbon tailoring requires extensive manipulation of functionality. Four of the five putative extender modules bear the full complement of tailoring domains, meaning that at each of these positions within the assembly, keto-, hydroxyl-, and olefin-containing intermediates are generated enroute to the fully saturated product, similar to mammalian fatty acid synthase. We have hypothesized that this type of module composition will afford the highest engineering potential as product diversification can be achieved without the need for incorporation of heterologous domains. In other words, one can potentially access each of the aforementioned functionalities by simple active site mutagenesis of KR, DH, and ER domains leaving the highly evolved protein–protein communication and recognition interfaces in their native states. This is in stark contrast to more popular assemblies such as 6-deoxyerythronolide B synthase (DEBS), where nearly all of the extender modules contain, at most,

a KR domain, where only ketone functionalities are accessible via similar active site mutagenesis strategies.¹ For this reason, we were eager to uncover the biosynthetic origins of fluvirucin B₁.

As predicted, the fluvirucin B₁ polyketide synthase consists of 5 extender modules flanked by an N-terminal loading ACP and C-terminal TE domain. All but one of the extender modules contains a KR, DH, and ER domain in addition to the required KS, AT, and ACP leading to the relatively un-functionalized nature of the macrolactam product. On the basis of this arrangement of composition of modules, β -alanine is expected to serve as the starter unit for fluvirucin B₁ biosynthesis. As strong evidence for this hypothesis, *flul*, which putatively encodes for a PLP-dependent decarboxylase, displays high homology with both *vinO* from the vicenistatin PKS cluster and *azicN* from the azicemicin PKS cluster.^{2,3} The former is responsible for decarboxylation of 3-methylaspartate, while the latter decarboxylates aspartic acid itself, leading to 3-methyl- β -alanine and β -alanine, respectively. While further studies are needed to confirm the starter unit identity for fluvirucin B₁ biosynthesis, this data strongly suggests a role for β -alanine in the early stages of macrolactam construction.

As alluded to above, thorough analysis of the protein sequences for each module reveals an intriguing trend with implications as to the evolutionary origins of these PKS components. Pairwise alignments between fluvirucin B₁ PKS modules 1, 3, and 5 yield protein sequence identities ranging from 75% to 81% (**Table 7**).

Table 7. Sequence comparison results for Flu modules and DEBS module 4						
Sequence comparison (Identities/Similarities)	FluMod1	FluMod2	FluMod3	FluMod4	FluMod5	DEBS Mod4
FluMod1	100%/100%	61%/72%	81%/86%	60%/70%	80%/86%	49%/61%
FluMod2	61%/72%	100%/100%	64%/75%	95%/96%	64%/75%	58%/72%
FluMod3	81%/86%	64%/75%	100%/100%	63%/73%	74%/82%	50%/63%
FluMod4	60%/70%	94%/96%	63%/73%	100%/100%	60%/70%	41%/56%
FluMod5	80%/86%	64%/75%	75%/82%	60%/70%	100%/100%	49%/62%
<p>*All sequence comparison done by BLAST of NCBI website. All modules listed in the column are used as query, and the modules listed in the row are used as subject.</p> <p>**In cases where the compared modules are of dramatically different sizes (i.e Flu Mod 1 vs. Flu Mod2) identities are related only to the KS-AT regions (regions of max identity) of the modules.</p>						

Similarly, modules 2 and 4 show 94% sequence identity. When analogous alignments are executed between these two groups (e.g., module 1 vs module 2), more typical identities ranging from 60% to 64% are observed. By comparison, sequence identities between modules from the well characterized DEBS assembly as well as between DEBS modules and fluvirucin PKS modules fall in the more modest 40–60% range (**Table 8**). The similarities between fluvirucin B₁ synthase modules might suggest independent ancestry for modules 1, 3, and 5 versus 2 and 4. It is important to note that the remarkable sequence identities observed within these two groups

occur despite the fact that each module both accepts and processes appreciably different polyketide intermediates.

Table 8. Sequence Comparison Results for DEBS Modules						
Sequence comparison (Identities/Similarities)	DEBS Mod1	DEBS Mod2	DEBS Mod3	DEBS Mod4	DEBS Mod5	DEBS Mod6
DEBS Mod1	100%/100%	49%/60%	48%/61%	56%/68%	49%/61%	59%/68%
DEBS Mod2	(49%/60%)	100%/100%	48%/59%	52%/64%	50%/61%	49%/61%
DEBS Mod3	(48%/61%)	(48%/59%)	100%/100%	56%/69%	48%/62%	51%/64%
DEBS Mod4	(56%/68%)	(52%/64%)	(56%/69%)	100%/100%	57%/68%	(52%/64%)
DEBS Mod5	(49%/61%)	(50%/61%)	(48%/62%)	57%/69%	100%/100%	(50%/61%)
DEBS Mod6	(59%/68%)	(49%/61%)	(51%/64%),	52%/64%	50%/61%	100%/100%

*Comparison results in parentheses are results from reverse query/subject pairs.

Another interesting aspect of the fluvirucin B₁ synthase involves the TE domain. Most macrocycle-forming thioesterases bear a conserved serine residue charged with accepting the fully mature, linear polyketide intermediate from an immediately upstream ACP followed by cyclization and product release.

The fluvirucin B₁ TE domain instead uses a cysteine active site for this task. This type of serine to cysteine substitution has been observed in other PKS systems, prompting speculation as

to possible divergent evolutionary origins between these two active site arrangements.⁴⁻⁶

Although beyond the scope of this thesis, this somewhat unique feature of the fluvirucin B₁ synthase should provide additional insights into any kinetic consequences of this switch and thus warrants further study.

5.2 Conclusions

In summary, we have identified and characterized the putative PKS genes associated with fluvirucin B₁ aglycone biosynthesis in *A. vulgaris*. The number and composition of modules as well as predicted AT specificities are consistent with the fluvirucin B₁ structure. The abundance of tailoring domains within the assembly is expected to provide increased engineering potential, through straightforward active site mutagenesis. Reconstitution of biosynthesis in a more workable host will greatly facilitate these studies and efforts to do so are currently underway in our laboratory.

5.3 Future Directions

5.3.1 Native Ketosynthase Selectivities in the Fluvirucin PKS

As alluded to above, our lab has recently developed a mechanism-based fluorescence transfer assay for examining KS substrate tolerance (**Figure 25**, chapter 4). This method stems from previous work using β -lactones and β -lactams as site-specific, electrophilic partners for the

Ppant thiol of holo-ACPs.⁷⁻⁸ The thioester-ACP resulting from incubation of holo-ACP with a propargyloxycarbonyl (Poc)-activated β -lactam was shown to be competitive with traditional SNAc thioesters for KS acylation. Based on these results, it occurred to us that fluorescent thioester-ACPs, derived from an alkyne-bearing β -lactam, coupled with acyl-SNAc compounds, could provide a straightforward means of examining KS selectivity in whole PKS modules.

We expect that substrate recognition in KS domains arises primarily from functionality near the thioester bond. Mutational inactivation of KR, DH, and ER domains will afford polyketide intermediates with modification at the β -position and thus, this position is the most logical place to begin our analysis of substrate recognition. A panel of SNAc thioesters bearing ketone, hydroxyl, olefin, and methylene β -functionalities along with the appropriate hydrogen, methyl, and ethyl α -functionalities will be prepared via established methods (Error! Reference source not found.).⁹ An added benefit of this analysis is that hydroxyl-group stereoselectivities, not immediately obvious from the products accepted by modules 2, 4, and 5, will be established for future studies utilizing the full fluvirucin PKS. KS domains from modules 2, 3, 4, and 5 will be incubated with SNAc compounds containing each β -functionality and the appropriate α -functionality, as determined by the immediately upstream AT domain, to establish a substrate scope for each. In the case of inconclusive or irregular observations in the data, traditional LC-MS techniques will be used for further confirmation.

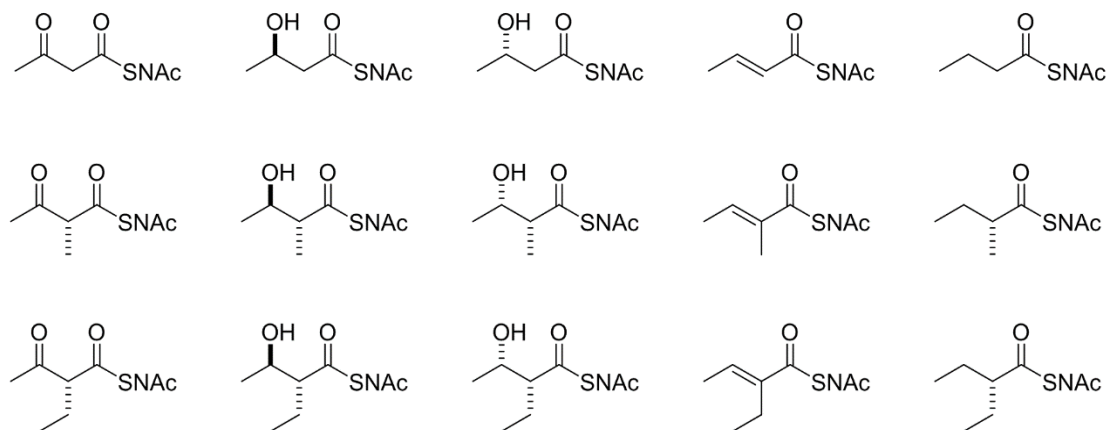


Figure 31. Panel of SNAc thioesters for determination of KS substrate selectivities in the fluvirucin PKS.

From these data, we expect two principal outcomes: 1) KS-domains utilizing purely kinetic-control for transfer of the proper intermediate will show little to no selection for a particular substrate and, 2) those evolved to only accept a certain substrate will exhibit the opposite response to non-native functionality. Domains that fall into the first category will require no alteration and can be used directly for fluvirucin analog production described in section 5.3.3. Conversely, for situations where strict substrate selectivities or a combination of both outcomes is observed, we will attempt to alter or broaden the substrate scope using targeted mutagenesis of residues located in a variable region of the KS active site pocket as detailed in section 5.3.2.

5.3.2 Altered Ketosynthase Selectivities via Mutation of Active Site Pocket Residues

Recent crystallographic data for the KS domains of DEBS modules 3 and 5 has provided new insights into substrate recognition interfaces surrounding the active site.¹⁰ Most of the deep KS active site pocket is highly conserved among modular PKSs. However, as Khosla and coworkers observed upon modeling a triketide and pentaketide substrate into the active site of DEBS module 3 and 5 KS domains, respectively, there exists a small variable region within the pocket that may allow the KS to distinguish certain substrate functionalities and sizes (**Figure 32**. Crystal structure

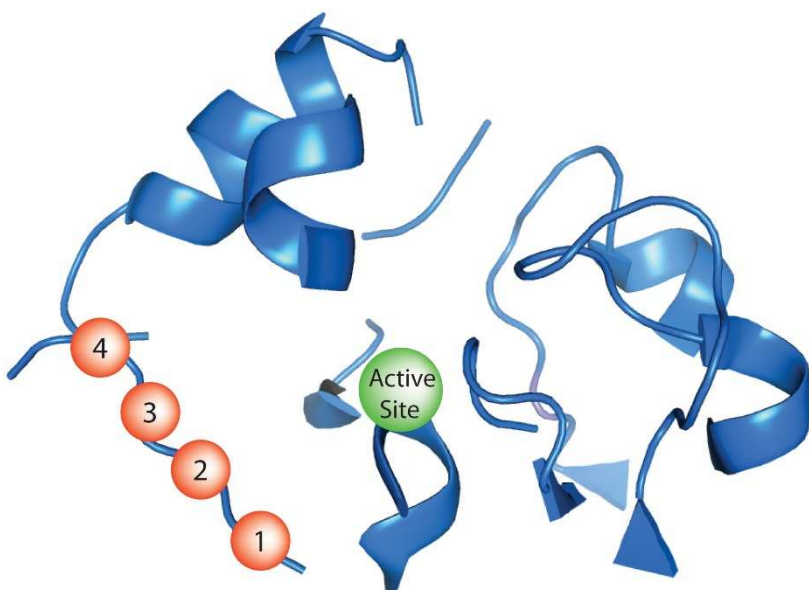


Figure 32. Crystal structure of the DEBS module 3 KS active site pocket (PDB: 2QO3). Active site cysteine indicated by a green sphere. The four loop residues thought to play a role in substrate recognition are labeled accordingly.

of the DEBS module 3 KS active site pocket (PDB: 2QO3). Active site cysteine indicated by a green sphere. The four loop residues thought to play a role in substrate recognition are labeled accordingly.). This hypothesis is yet untested

and the fluvirucin PKS KS domains exhibiting strict substrate selectivities from section 5.3.1 will offer the ideal platform for experimental investigation of the model.

Motivated by this observation, we were interested in whether any distinct sequence patterns could be detected within this region that might lead to selectivity for each common polyketide β -functionality.

Separate alignments were executed for KS domains falling into one of four categories: those that natively accept an intermediate bearing a 1) ketone, 2) hydroxyl, 3) olefin or, 4) methylene at the β -

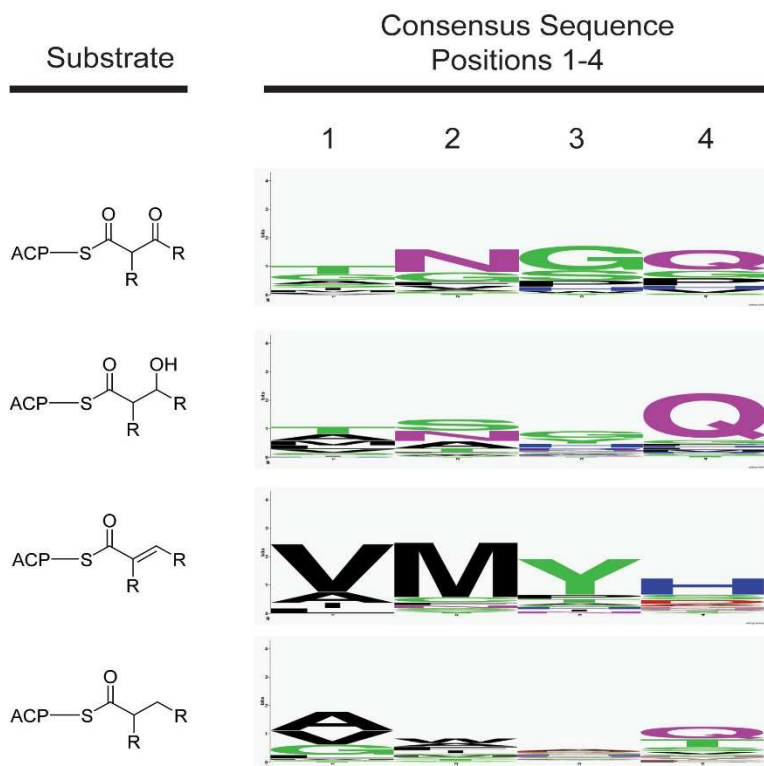


Figure 33. Consensus sequences for each of the four KS active site pocket residues of interest (see **Figure 32**. Crystal structure of the DEBS module 3 KS active site pocket (PDB: 2QO3). Active site cysteine indicated by a green sphere. The four loop residues thought to play a role in substrate recognition are labeled accordingly.). Molecules on the left indicate which β -functionality is accepted by the KS domain from an upstream ACP. Each consensus sequence was determined by alignment of at least 25 individual sequences. The more hydrophobic olefin and methylene functionalities generally coincide with more hydrophobic residues at positions 1 and 2 while ketone and hydroxyl groups tend to prefer hydrophilic residues at these positions.

position from an upstream ACP. Consensus sequences derived from this data revealed some intriguing trends (**Figure 33**. Consensus sequences for each of the four KS active site pocket residues of interest (see **Figure 32**. Crystal structure of the DEBS module 3 KS active site pocket (PDB: 2QO3). Active site cysteine indicated by a green sphere. The four loop residues thought to play a role in substrate recognition are labeled accordingly.). Molecules on the left indicate which β -functionality is accepted by the KS domain from an upstream ACP. Each consensus sequence was determined by alignment of at least 25 individual sequences. The more hydrophobic olefin and methylene functionalities generally coincide with more hydrophobic residues at positions 1 and 2 while ketone and hydroxyl groups tend to prefer hydrophilic residues at these positions). Firstly, it is not immediately obvious that this region of the active site pocket is able to distinguish between β -keto and β -hydroxyl functionalities. In contrast, positions 1 and 2 may play a critical role in selecting between hydrophobic and hydrophilic groups. For the more polar ketone and hydroxyl groups, these two positions are dominated by hydrophilic residues capable of hydrogen bonding while KS domains that natively accept olefin and methylene functionalities generally possess more hydrophobic amino acids at these positions. For example, close examination of these residues in fluvirucin module 3, which accepts a β -hydroxyl group, and module 5, which accepts a β -methylene, reveals a similar trend. Module 3 KS bears a glycine at position 1 and threonine at position 2 while module 5 KS contains an alanine at position 1 and a

phenylalanine at position 2. If either of these KS domains exhibit strict selection for its native substrate, introducing more hydrophobicity to module 3 and more hydrophilicity to module 5 may serve to reverse its substrate tolerances.

In a broader sense, the consensus sequences for each β -functionality will be introduced, via site-directed mutagenesis, to each fluvirucin KS domain that displays high substrate selectivity from section 5.3.1. The impact of these mutations will be readily observed using our fluorescence transfer assay. Any changes to the fluorescence pattern between the unmodified and mutated KS domains will be confirmed by tandem proteolysis-LCMS. Mutant KS domains with altered or relaxed substrate selectivity relative to wild-type will join the kinetically-controlled KS domains for production of novel fluvirucin analogs.

If grafting consensus sequences onto the selective KS domains results in little change to substrate tolerances, alternative strategies are in place to further explore the role of this pocket region in substrate recognition. The 20+ sequences used to generate each consensus sequence above come from a wide variety of PKS systems at vastly different positions along each assembly line. Therefore, it may be worth adopting narrower search criteria for modification of KS-selectivities. In lieu of consensus sequences, we will search within each subset of available sequences for KS domains that natively accept substrates with the highest similarities, in terms of length, α -functionality, and more distal functionality to the natural fluvirucin substrates. For

instance, the natural substrate of DEBS module 4 is a tetraketide containing α -methyl and β -keto functionalities, and a β -keto form of the fluvirucin module 4 substrate would bear significant resemblance to this. Therefore, the DEBS module 4 active site pocket sequence (MNGQ), which has little in common with the fluvirucin module 4 sequence (GTQT), is a potential alternative to the consensus sequence (TNGQ). Given the projected flexibility of the loop which harbors the target residues and sequence variability observed for each type of β -functionality, it is probable that even minor changes to this region will have a dramatic impact on KS-selectivities.

As a second alternative strategy, we will broaden our pursuit using saturation mutagenesis to explore full sequence variation at positions 1 and 2 of the target active site pocket region. The resulting mutant KS domains will be expressed and subsequently bound to Ni-NTA beads or functionalized plates. After thorough washing, the bound KS domain will be incubated with substrate(s) bearing the desired β -functionalities followed by addition of an untagged, fluorescent ACP. The mutant KS domains will then be eluted from the Ni-NTA and the degree of ACP to KS fluorescence transfer will be determined by PAGE analysis. Gels will then be stained to confirm that any lack of fluorescence is the result of active site acylation with substrate and not poor expression of the KS domain. As above, any mutant KS domains that show altered or broadened substrate tolerance will be utilized in section 5.3.3 for optimization of fluvirucin analog production. Together with the more targeted strategies described above, we are well positioned

to clearly define a role for these residues in substrate recognition which will aid in future exploration of KS active site pockets well beyond the fluvirucin system.

5.3.3 Macrolactam Production from Engineered Fluvirucin PKS

In order to reconstitute the biosynthesis of fluvirucin B₁ aglycone and further to produce its analogs, we will utilize the constructs and information generated from native KSATs selectivities (section 5.3.1) and bioinformatic analyses (section 5.3.2) to systematically examine the relative contributions of KS substrate selectivity and disruption of protein-protein interactions on fluvirucin analog production. We envision three possible outcomes from the work described in previous chapters. Firstly, it may be the case that all KS domains in the fluvirucin PKS are kinetically controlled, in which case broad substrate tolerance will be observed across the board. In the unlikely event that this is the case, we still plan to explore the role of active site pocket residues as described in section 5.3.2. Any changes to substrate selectivity arising from these studies will provide critical insight for other PKSs where KS substrate selectivity is more tightly regulated. Should we instead observe the opposite and equally unlikely case where all KS domains exhibit high selectivity for a particular substrate, we stand to gain maximal insight as to the role played by the aforementioned active site pocket residues in substrate recognition. While we are

prepared for either of these two extremes, the most likely scenario for chain transfer in the fluvirucin PKS involves a combination of these mechanisms.

In any case, we have reconstructed the full fluvirucin PKS on three separate plasmids with unique resistance markers. The first two plasmids contain fluA (modules 1 and 2, **Figure 34**) and fluB (modules 3 and 4, **Figure 35**) while the third harbors fluC (module 5, **Figure 36**) and the genes necessary for methylmalonyl-CoA production (**Figure 36**). Using these three constructs, we will attempt to reconstitute the fluvirucin PKS pathway in *E. coli*. This versatile host has been instrumental in much of the DEBS work and has proved quite tolerant of fluvirucin-related genes in our preliminary work with this system. Despite the many advantages of using *E. coli* for fluvirucin production, this host does have some limitations and a plan to tackle these potential issues is essential.

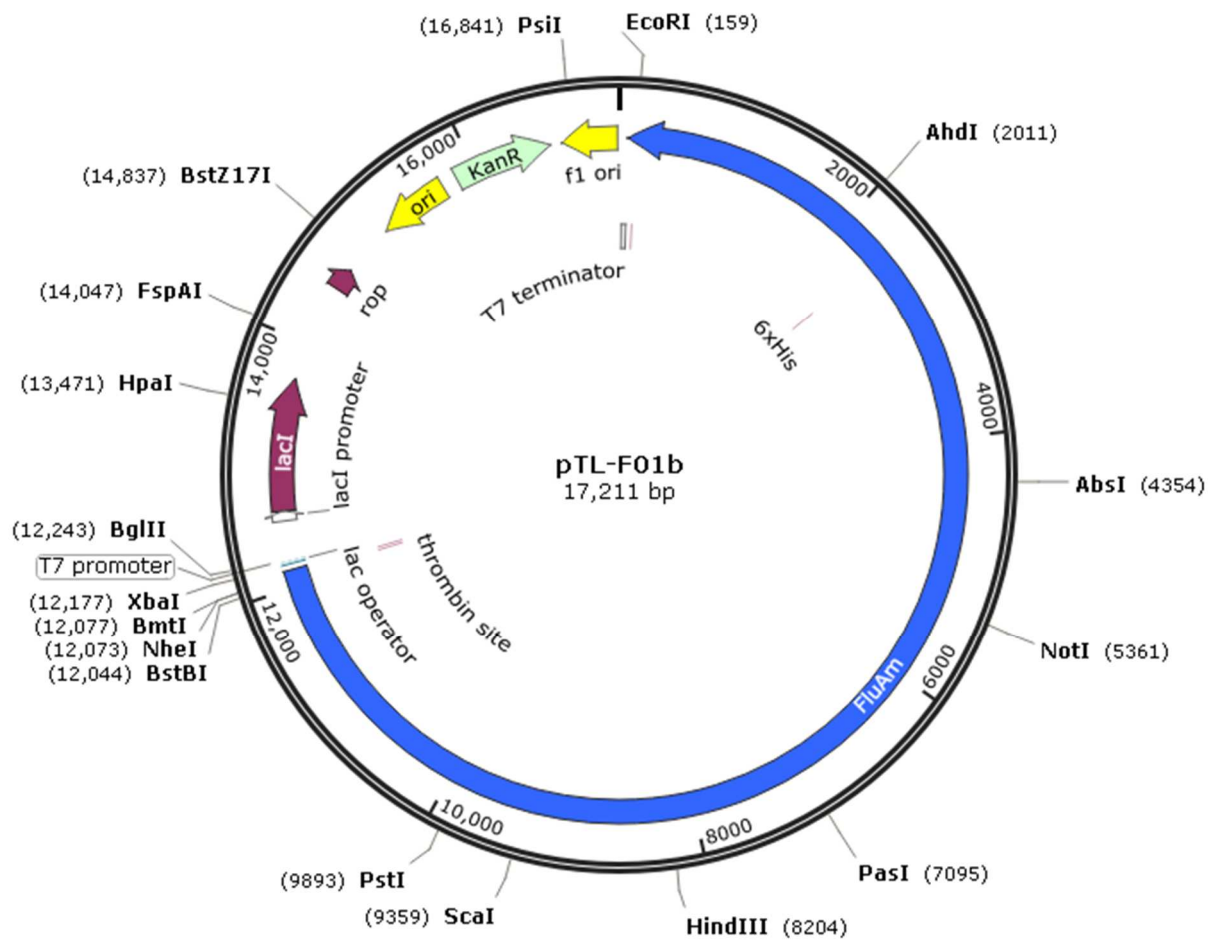


Figure 34. The plasmid map of pTL-F01b

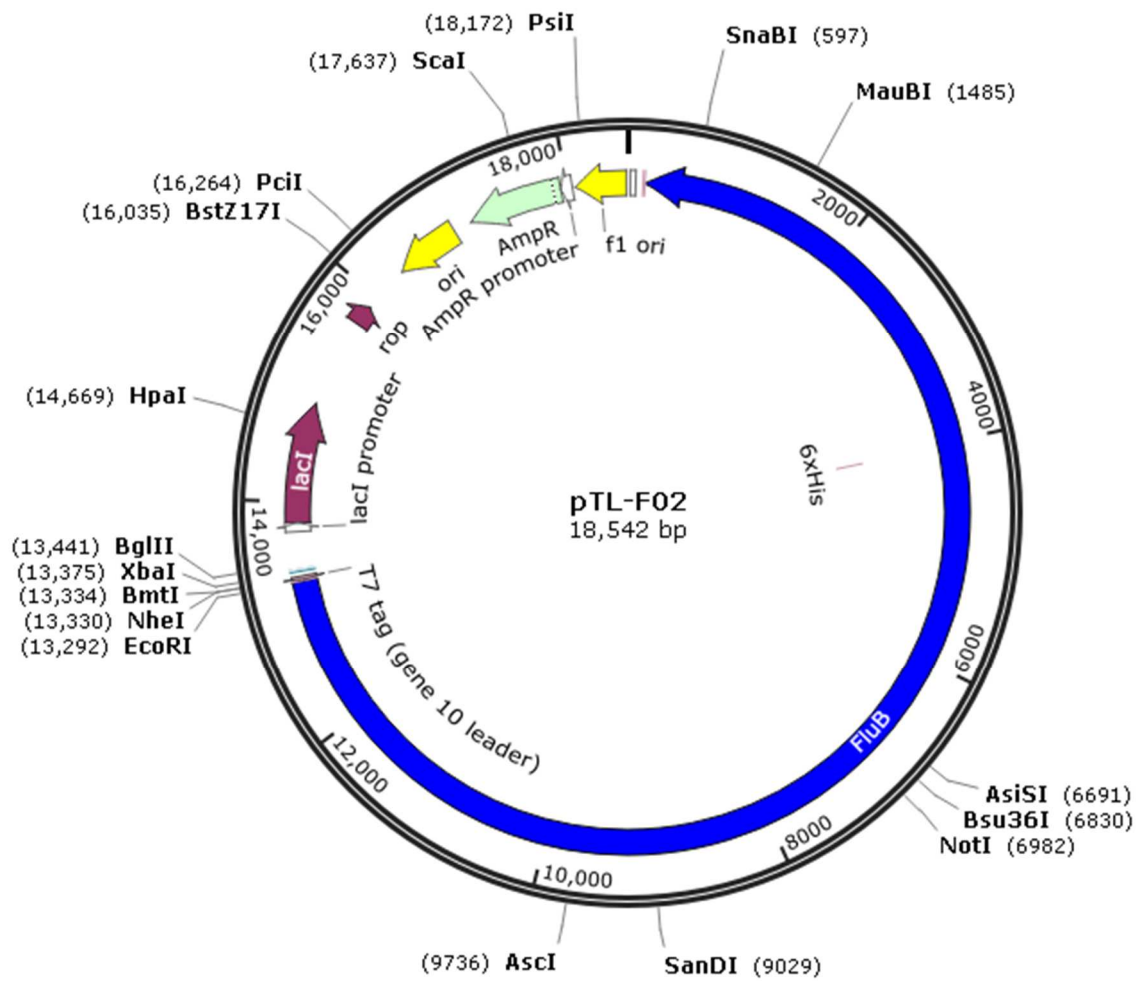


Figure 35. The plasmid map of pTL-F02

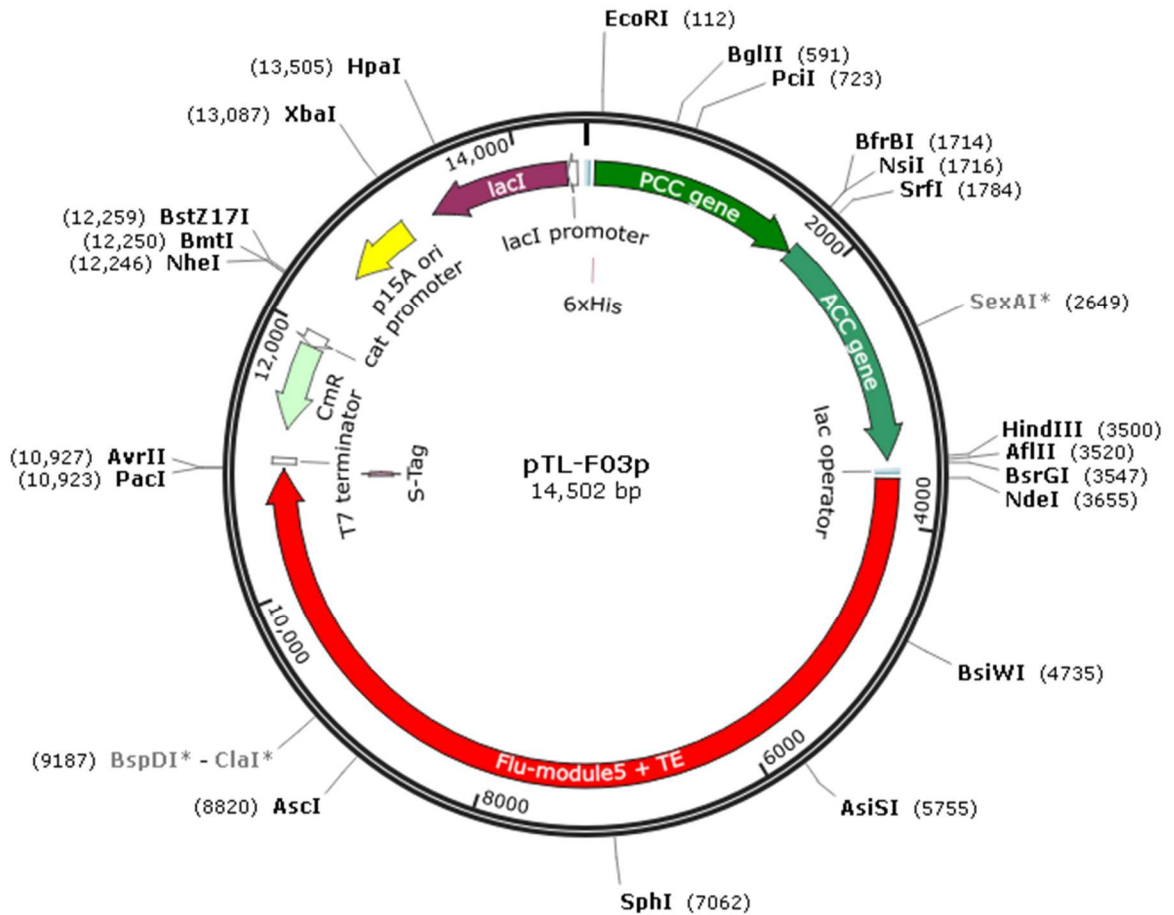


Figure 36. The plasmid map of pTL-F03p

5.3.3.1 Reconstituting Fluvirucin B₁ Aglycone Production in *E. Coli*

The most challenging part of using a heterologous host for fluvirucin production will be rendering the PKS functional. This system utilizes methylmalonyl- and ethylmalonyl-CoA as extender units, both of which are produced at very low levels, if at all, in *E. coli*. Luckily, the issue of methylmalonyl-CoA production has previously been solved for the DEBS system via

introduction of propionyl-CoA carboxylase (PCC) and related genes.¹¹⁻¹² We will rely on this technique for the fluvirucin PKS, but an equivalent means of ethylmalonyl-CoA generation will likely necessitate further engineering of the host. We have devised several complementary strategies for tackling this issue including (1) exploiting the relaxed substrate tolerance of PCC, and (2) introduction of crotonyl-CoA carboxylase/reductase (CCR).

Propionyl-CoA Carboxylase: As the most straightforward of these potential solutions, we may be able to simply use the same enzymes responsible for methylmalonyl-CoA production. It has been observed that PCC has moderately relaxed substrate selectivity and will catalyze, albeit with lower efficiency, carboxylation of butyryl-CoA.^{13,14} We will attempt to generate both methylmalonyl- and ethylmalonyl-CoA by addition of propionate and butyrate, respectively, to our engineered *E. coli* cultures. Since propionate is converted faster than butyrate, we plan to add the latter in excess. Analysis of lysates by LC-MS will help determine the extent to which each species is formed and modifications to the addition ratio will be made accordingly.

Crotonyl-CoA Carboxylase Reductase: Our backup plan if the PCC method proves unsuccessful will be to employ a natural enzyme involved in ethylmalonyl-CoA biosynthesis in *S. collinus*, CCR.^{15,16} This strategy has been used previously in *Streptomyces* to incorporate a single ethylmalonate into the backbone of 6-deoxyerythronolide B.¹⁷ By adding the appropriate precursors to an engineered *S. erythraea* strain overproducing CCR, they were able to

demonstrate incorporation of ethylmalonate. A similar strategy will be attempted in our engineered *E. coli* system to provide the correct extender units for fluvirucin biosynthesis. As an interesting side note, the same research team that produced the ethyl analog of 6-deoxyerythronolide B observed that the ethylmalonate-specific AT domain would accept methylmalonate under conditions where ethylmalonyl-CoA was unavailable.¹⁷ If the fluvirucin AT domains exhibit a similar tolerance for methylmalonyl-CoA, the methylated form of the final product will suffice for all downstream experiments.

With a functional strategy for extender unit construction in place, we will begin assessing fluvirucin output from our engineered *E. coli* strain. The N-acetylcysteamine thioester of β -alanine (β -Ala-SNAc) will serve as the starter unit for PKS priming. A series of 100 ml cultures of appropriately engineered BAP1, an *E. coli* variant with a phosphopantetheinyl transferase embedded in the genome, will be incubated with β -Ala-SNAc at concentrations ranging from 1 mM to 1 M to determine the optimal conditions for fluvirucin production. Small molecule output will be assayed using LC-MS following organic extraction of each lysate. Further optimization of fermentation parameters will be accomplished using additional variables such as culture volume, incubation temperature, and time. Ultimately, the maximal yield ascertained in these studies will provide the benchmark for comparison to all mutant forms. Although it is just the beginning of

our work with this system, successful production of fluvirucin B₁ will represent the first demonstration of heterologous macrolactam generation in *E. coli*.

We are fully aware that reconstitution of the fluvirucin PKS pathway in *E. coli* may not be possible. Although we are thoroughly committed to this host for its genetic simplicity, fast growth, and the countless tools available for its manipulation, alternate plans are in place if at any point it becomes clear that our efforts are unproductive. Other bacteria, including *S. coelicolor* and *S. lividans* have also been used to heterologously produce polyketides and may require less host-engineering than *E. coli* should this pose a problem.¹⁸ In addition, we will explore the possibility of a cell-free synthesis of fluvirucin using synthetically-derived starter and extender units. Although not equivalent to cell-based fluvirucin biosynthesis, this strategy may afford the desired information regarding the engineering obstacles with little background noise arising from unrelated cellular metabolites.

5.3.3.2 Mutational Inactivation of Individual Domains for Increased Product Yield

During the biosynthesis of fluvirucin B₁, thirteen unique tailoring domains are involved in β -carbon processing. In contrast, DEBS relies on only seven of these enzymes for 6-deoxyerythronolide B production despite harboring an extra module compared to the fluvirucin PKS. Therefore, if our goal is to manipulate small molecule structure without disrupting native

protein-protein interactions, mutational inactivation of fluvirucin PKS domains nearly doubles the product landscape. It has been said that the key to successful combinatorial methods in natural product biosynthesis lies in the discovery of higher-yielding engineering strategies.¹⁹ This section of our research program will systematically illustrate the benefits of simple mutagenesis to control polyketide structure (**Figure 37**). In addition, the relative product yields from each single-point mutant will begin to uncover when, and to what extent, substrate selectivity plays a role in the

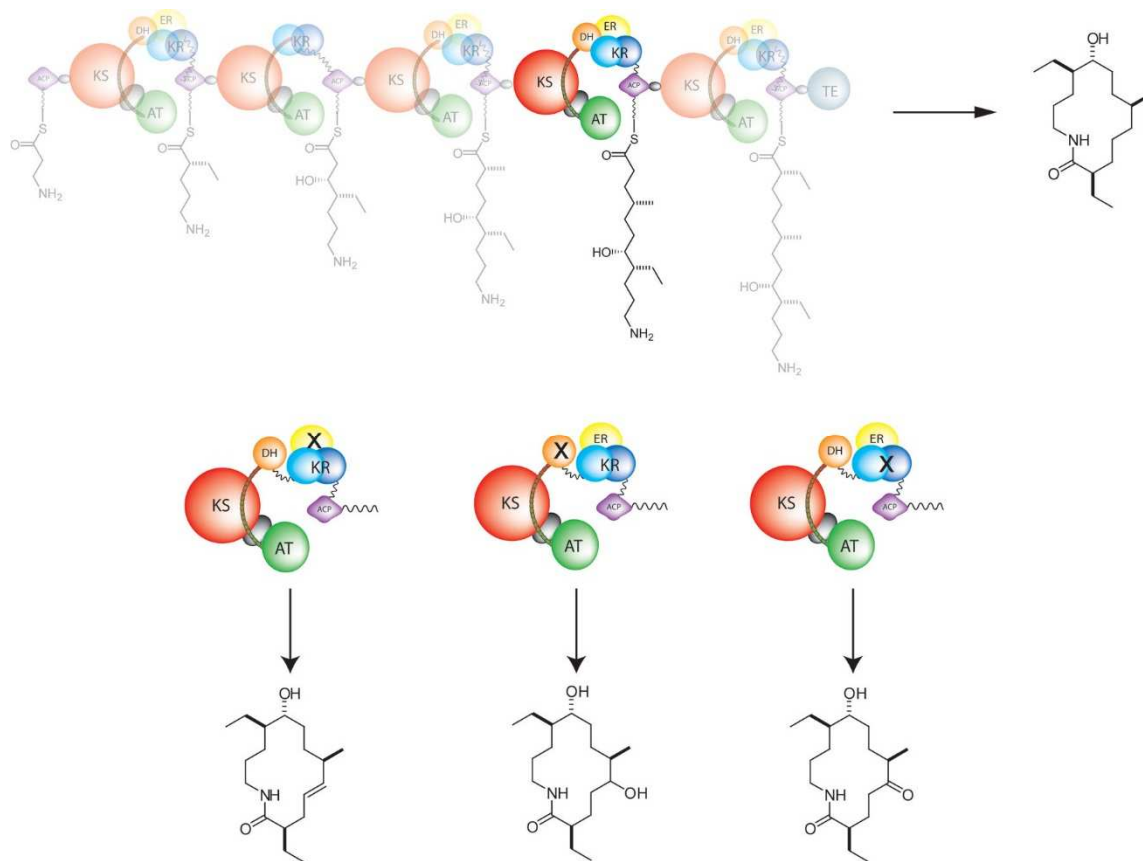


Figure 37. Schematic diagram of our domain deletion-based engineering strategy for fluvirucin analog production. Active site mutagenesis of KR, DH, and ER domains will result in site-specific incorporation of ketone, hydroxyl, and olefin functionalities respectively. This strategy eliminates the need for introduction of heterologous PKS components thus circumventing the disruption of protein-protein interactions, a major engineering hurdle.

flow of intermediates along the assembly line. It should be noted that our strategy for mutational activation is almost exclusively based on work with similar PKS assemblies. However, we are prepared to take a more exploratory route (i.e. random mutagenesis, alanine scanning, etc.) should the need arise.

KR domain inactivation: We will initially focus mutagenesis efforts on conserved NADPH binding residues for inactivation of ketoreductase domains as described in the preliminary studies section. Alternatively, most active KR domains harbor a highly conserved catalytic triad, Ser/Tyr/Lys, which serves to activate the electrophilic carbonyl carbon toward hydride attack. Any issues encountered with cofactor binding mutants may be resolved by additionally altering these residues. Disruption of the catalytic triad may also improve kinetics in these modified systems if KR-recognition of the ACP-bound intermediates slows intermodular transfer. Combinations of mutations to both the NADPH binding site and catalytic residues will be examined to determine optimal inactivation parameters.

DH domain inactivation: Although not yet rigorously studied in modular polyketide synthases, based on work with analogous fatty acid synthases, dehydratase activity likely relies heavily on a conserved histidine residue acting as a general base on the proton of the β -hydroxyl group during dehydration. We expect that mutating this residue to any hydrophobic amino acid such as alanine or leucine will abolish activity.

ER domain inactivation: Similar to KR domains, enoyl reductases require an NADPH cofactor for activity. Both single and multiple mutations within this binding site will be examined to determine the most favorable configuration as above.

Each of the thirteen mutants will be individually assayed using the optimized parameters for wild-type fluvirucin production elucidated in section 5.3.3.1 (**Figure 38**). Yields of the engineered products, relative to the wild-type fluvirucin, will be determined by LC-MS using an internal standard. Although we do not expect the ionization potentials of these products to be

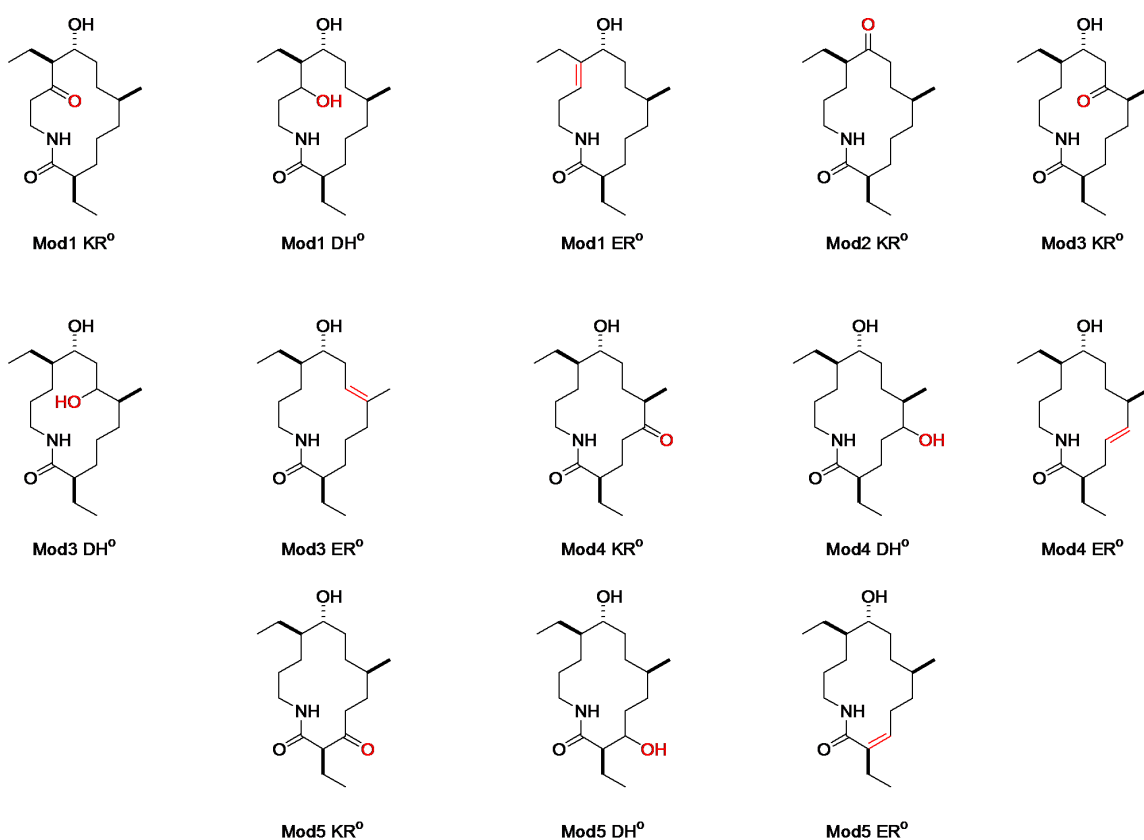


Figure 38. Structures of the thirteen fluvirucin analogs possible from single domain inactivations. Each mutation results in site specific modification of macrolactam structure. Relative yields of these products will provide vital information as to the relative contributions of the two primary obstacles in polyketide engineering.

vastly different, standard curves will be prepared for added assurance. Any polyketide products generated in high enough quantities upon scale-up will be analyzed by 1D and 2D NMR to uncover any stereochemical trends as they relate to intermodular shuttling of intermediates. We are aware that, in some cases, the thioesterase domain may not be able to efficiently cyclize the intended product, resulting in unexpectedly low yields. Every effort will be made to identify and quantitate hydrolysis products, and plans are in place to utilize alternative thioesterase domains should the need arise. Our overriding goal is to meet or exceed the highest yields obtained with the engineered DEBS system (~70% of wild type) for all single mutants of the fluvirucin PKS.^{19,20} At these levels, we can confidently state that we have overcome the primary polyketide engineering hurdles and that realization of functional combinatorial biosynthesis is well within our grasp.

5.3.4 Engineering FAS-like PKS into Iterative PKS

Similarities between FASs and modular PKS has been studied from the level of the individual reaction mechanisms to the structures of each catalytic domains and the architectural organization of the entire enzymatic complexes^{21,22}, and furthermore, the evolutionary links between FAS and PKS has been emphasized by comprehensive phylogenetic analyses to show they may have a common origin.²³ Therefore, it is very reasonable to explore the possibility to

engineer a PKS module into an iterative FAS system, and the successful result could be applied into metabolic engineering or biofuel production.

In 2012, Khosla and coworkers have “reprogrammed” module 3 of DEBS to catalyze two successive rounds of chain elongation.²⁴ However, they could not observe products more than two iterations, and they hypothesized that the active site of KS3 may not be able to accommodate substrate longer than triketide. Similarly, Mohanty and colleagues conducted bioinformatic analyses on 217 pure modular KS domains, 82 pure iterative domains, 19 enediyne, 43 trans-type and 34 KS domains from hybrid NRPS-PKS clusters,²⁵ and they have discovered two crucial positions may govern the chain length in different iterative PKSs. They also combined the structural information of FAS and PKS to deduce that the size of substrate cavity in KS hold the key to fully engineer a PKS into a FAS.

We believe that fluvirucin PKS system provides a unique opportunity and potential to be explored for this purpose. Not only four out of five modules possess the similarity to FAS, but also they are subdivided into two distinct group which could be useful in terms of bioinformatic analyses and in vitro experiments.

5.4 Summary

In this thesis, I seek to examine the relative contributions of the two primary causes of low yields in polyketide engineering, substrate selectivity and disruption of protein-protein interactions, issues that have severely limited drug discovery and diversification programs for the past several decades. In contrast to most traditional PKS systems used for metabolic engineering efforts, the fluvirucin PKS is composed primarily of fatty acid synthase-like modules which contain all three β -processing domains (KR, DH, and ER). These enzymes are solely responsible for the oxidation state of every other carbon on the polyketide chain and, therefore, provide enormous potential for controlling product variation through genetic engineering. Most importantly, alterations to polyketide structure can be achieved by simple mutagenesis of key amino acids which leaves the three-dimensional architecture and all native protein-protein recognition motifs intact. This is in stark contrast to other, well-studied systems such as DEBS, where obtaining the desired level of product diversity requires the introduction of heterologous domains and modules which are not necessarily evolved to communicate with the native components.

Our strategy with the fluvirucin PKS effectively eliminates the disruption of protein-protein interactions as an engineering hurdle. As a result, we are uniquely positioned to tackle the second hurdle, substrate selectivity, in isolation. This challenge requires a thorough understanding of chain transfer mechanisms between ACP and the downstream KS domain. Using

the fluorescence assay for KS substrate selectivity and guided by recent structural data, we will both determine the native substrate tolerance for each fluvirucin KS domain and attempt to alter or broaden selectivity in those domains exhibiting inherent, thermodynamically-controlled substrate recognition. By eliminating both of these engineering obstacles, we aspire to obtain near wild type production levels of fluvirucin B1 analogs. In cases where low yields persist, we aim to uncover alternative engineering obstacles and ultimately offer a set of universal guidelines for polyketide structure diversification.

Successful execution of the proposed research program will lead to improved understanding of chain transfer mechanisms, more clearly-defined approaches for domain deletion, and the necessary foundation for practical combinatorial biosynthesis. To illustrate the latter, future efforts will focus on design and implementation of experiments where high-yielding single KR, DH, and ER mutants are combined to produce usable quantities of novel macrolactams. Realization of this goal will permit greatly expanded exploration of polyketide functionality and amass an abundance of diverse macrolide scaffolds for building the next generation of small molecule therapeutics, and even elucidate the evolution relationship between FASs and PKSs.

5.5 Reference

1. Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* 76, 195–221.
2. Ogasawara, Y., Katayama, K., Minami, A., Otsuka, M., Eguchi, T., and Kakinuma, K. (2004) Cloning, Sequencing, and Functional Analysis of the Biosynthetic Gene Cluster of Macrolactam Antibiotic Vicenistatin in *Streptomyces halstedii*. *Chem. Biol.* 11, 79–86.
3. Ogasawara, Y., and Liu, H.-W. (2009) Biosynthetic studies of afaziridine formation in azicemicins. *J. Am. Chem. Soc.* 131, 18066–18068.
4. Irwin, D. M. (1998) Evolution of an active-site codon in serine proteases. *Nature* 336, 429–430.
5. Witzowski, A., Naggert, J., Witkowska, H. E., Randhawa, Z. I., and Smith, S. (1992) Utilization of an active serine 101-cysteine mutant to demonstrate the proximity of the catalytic serine 101 and histidine 237 residues in thioesterase II. *J. Biol. Chem.* 267, 18488–18492.
6. Witzowski, A., Witkowska, H. E., and Smith, S. (1994) Reengineering the specificity of a serine active-site enzyme. Two active-site mutations convert a hydrolase to a transferase. *J. Biol. Chem.* 269, 379–383.

7. Prasad, G.; Amoroso, J.W.; Borketey, L.S.; Schnarr, N. A. (2012) N-Activated β -Lactams as Versatile Reagents for Acyl Carrier Protein Labeling. *Org. Biomol. Chem.*, 10, 1992-2002
8. Amoroso, J.W., Borketey, L.S., Prasad, G., Schnarr, N.A. (2010) Direct Acylation of Carrier Proteins with Functionalized β -Lactones. *Org. Lett.*, 12, 2330-33.
9. Prasad G, Borketey LS, Lin T-Y, Schnarr NA. (2012) A Mechanism-Based Fluorescent Assay for Examining Ketosynthase Activity. *Org. Biomol. Chem.*, 10, 6717-23.
10. Tang, Y.; Chen, A.Y.; Kim, C-Y.; Cane, D.E.; Khosla, C. (2007) Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chem. Bio.*, 14, 931-943.
11. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E-coli*, *Science*, 291, 1790-1792.
12. Wang, Y.; Pfeifer, B. A. (2008) 6-Deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in *E. coli*. *Metabolic Engineering*, 10, 33-38.
13. Diacovich, L.; Peiru, S.; Kurth, D.; Rodriguez, E.; Podesta, F.; Khosla, C.; Gramajo, H. (2002) Kinetic and structural analysis of a new group of acyl-CoA carboxylases found in *Streptomyces coelicolor* A3(2). *J. Biol. Chem.*, 277, 31228-31236.

14. Rodriguez, E.; Banchio, C.; Diacovich, L.; Bibb, M. J.; Gramajo, H. (2001) Role of an essential acyl coenzyme A carboxylase in the primary and secondary metabolism of *Streptomyces coelicolor* A3(2). *Applied and Environmental Microbiology*, 67, 4166-4176.
15. Liu, H., Reynolds, K.A. (2001) Precursor Supply for Polyketide Biosynthesis: The Role of Crotonyl-CoA Reductase. *Metabolic Engineering*, 3, 40-48.
16. Wallace, K.K., Bao, Z., Dai, H., Digate, R., Schuler, G., Speedie, M.K., Reynolds, K.A. (1995) Purification of Crotonyl-CoA Reductase from *Streptomyces collinus* and Cloning, Sequencing and Expression of the Corresponding Gene in *Escherichia coli*. *Eur. J. Biochem.*, 233, 954-62.
17. Stassi, D.L.; Kakavas, S.J.; Reynolds, K.A.; Gunawardana, G.; Swanson, S.; Zeidner, D.; Jackson, M.; Liu, H.; Buko, A.; Katz, L. (1998) Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proc. Natl. Acad. Sci. USA*, 95, 7305-09.
18. Pfeifer, B.; Khosla, C. (2001) Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.*, 65, 106-118.
19. Xue, Q., Ashley, G., Hutchinson, C., and Santi, D. (1999) A multiplasmid approach to preparing large libraries of polyketides, *Proc. Natl. Acad. Sci. USA*, 96, 11740-11745

20. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. (1999) Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel "Unnatural" Natural Products, *Proc. Natl. Acad. Sci. USA*, *96*, 1846-1851.
21. Revill, WP.; Bibb, MJ.; Hopwood, DA. (1996) Relationships between fatty acid and polyketide synthases from *Streptomyces coelicolor* A3(2): characterization of the fatty acid synthase acyl carrier protein. *J. Bacteriol.*, *178*(19), 5660-5667
22. Smith, S.; Tsai, S-C. (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.*, *24*, 1041-1072.
23. Jenke-Kodama, H.; Sandmann, A.; Müller, R.; Dittmann, E. (2005) Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol.*, *22*(10), 2027-39
24. Kapur S, Lowry B, Yuzawa S, Kenthirapalan S, Chen AY, Cane DE, Khosla C. (2012) Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc Natl Acad Sci USA*, *109*, 4110-4115.
25. Yadav G.; Gokhale, RS.; Mohanty, D. (2009) Towards prediction of metabolic products of polyketide synthases: an in silico analysis. *PLoS Comput Biol.*, DOI: 10.1371/journal.pcbi.1000351

BIBLIOGRAPHY

- Amoroso, J.W., Borketey, L.S., Prasad, G., Schnarr, N.A. (2010) Direct Acylation of Carrier Proteins with Functionalized β -Lactones. *Org. Lett.*, 12, 2330-33.
- Asturias, F. J.; Chadick, J. Z.; Cheung, I. K.; Stark, H.; Witkowski, A.; Joshi, A. K.; Smith, S. (2005) Structure and molecular organization of mammalian fatty acid synthase. *Nat. Struct. Mol. Biol.* 12, 225-232.
- Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabinowitsch, E.; Rajandream, M.-A.; Rutherford, K. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417 (6885), 141–147.
- Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Research*, 34 (19), 5402–5415.
- Cane, D. E., Walsh, C. T., and Khosla, C. (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science*, 282, 63–68
- Chakrabarti R. "Chapter 6. Novel PCR-Enhancing Compounds and Their Modes of Action" *PCR Technology: Current Innovations, Second Edition* (Edited by Hugh G. Griffin , Thomas Weissensteiner and Annette Griffin) CRC Press 2003, P.51–63
- Chakrabarti R, Schutt CE. (2001) The enhancement of PCR amplification by low molecular-weight sulfones. *Gene*, 274 (1–2), 293–298.
- Chandran, S.S., Menzella, H.G., Carney, J.R., Santi, D.V. (2006) Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chem. Biol.*, 13, 469-474.

- Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F. (1990), An Unusually Large Multifunctional Polypeptide in the Erythromycin-producing Polyketide Synthase of *Saccharopolyspora erythraea*, *Nature*, 348, 176-178
- Diacovich, L.; Peiru, S.; Kurth, D.; Rodriguez, E.; Podesta, F.; Khosla, C.; Gramajo, H. (2002) Kinetic and structural analysis of a new group of acyl-CoA carboxylases found in *Streptomyces coelicolor* A3(2). *J. Biol. Chem.*, 277, 31228-31236.
- Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res*, 19 (14), 4008
- Donadio, S. and Sosio, M. (2003) Strategies for combinatorial biosynthesis with modular polyketide synthases. *Comb. Chem. High Throughput Screening*, 6, 489-500
- Donadio, S., Katz, L. (1992) Organization of the Enzymatic Domains in the Multifunctional Polyketide Synthase Involved in Erythromycin Formation in *Saccharopolyspora erythraea*, *Gene*, 111(1): 51 – 60.
- Donadio S., McAlpine J.B., Sheldon P.J., Jackson M., and Katz L. (1993) An Erythromycin Analog Produced by Reprogramming of Polyketide Synthesis, *Proc. Natl. Acad. Sci. USA*, 90 (15), 7119-7123
- Donadio S., Staver M.J, McAlpine J.B., Swanson S.J., and Katz L. (1991) Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science*, 252(5006), 675-679.
- Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) Dissecting and exploiting intermodular communication in polyketide synthases, *Science*, 284, 482-485.

- Hicks, L. M., Mazur, M. T., Miller, L. M., Dorrestein, P. C., Schnarr, N. A., Khosla, C., and Kelleher, N. L. (2006) Investigating nonribosomal peptide and polyketide biosynthesis by direct detection of intermediates on >70 kDa polypeptides by using Fourier-Transform mass spectrometry. *ChemBioChem*. 7, 904-907.
- Hong, H., Appleyard, A. N., Siskos, A. P., Garcia-Bernardo, J., Staunton, J., and Leadlay, P. F. (2005) Chain initiation on type I modular polyketide synthases revealed by limited proteolysis and ion-trap mass spectrometry. *FEBS Journal*. 272, 2373-2387.
- Irwin, D. M. (1998) Evolution of an active-site codon in serine proteases. *Nature* 336, 429-430.
- Jenke-Kodama, H.; Sandmann, A.; Müller, R.; Dittmann, E. (2005) Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol.*, 22(10), 2027-39
- Kao, C. M., Katz, L., and Khosla, C. (1994) Engineered biosynthesis of a complete macrolactone in a heterologous host, *Science*, 265, 509-512.
- Kao C.M., Luo G., Katz L., Cane D.E., Khosla C. (1996) Engineered biosynthesis of structurally diverse tetraketides by a trimodular polyketide synthase. *J Am Chem Soc*, 118, 9184-9185
- Kapur, S., Chen, A.Y., Cane, D.E., Khosla, C. (2010) Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. USA*, 107, 22066-22071.
- Kapur S, Lowry B, Yuzawa S, Kenthirapalan S, Chen AY, Cane DE, Khosla C. (2012) Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc Natl Acad Sci USA*, 109, 4110-4115.
- Katz L, McDaniel R. (1999) Novel macrolides through genetic engineering, *Medicinal Research Reviews*, 19, 543-558.,

- Keatinge-Clay, A. T. (2007) A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* 14, 898–908
- Khosla, C. (2009) Structures and Mechanisms of Polyketide Synthases, *J. Org. Chem.*, 74 (17), 6416–6420
- Khosla, C., Tang, Y., Chen A. Y., Schnarr N. A., and Cane D. E. (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase, *Annu. Rev. Biochem.* 76, 195-221.
- Kohli, R. M. and Walsh, C. T. (2003) Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem. Commun.*, 3, 297–307
- Lau, J., Fu, H., Cane, D. E., and Khosla, C. (1999) Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry.* 38, 1643-1651.
- Liu, H., Reynolds, K.A. (2001) Precursor Supply for Polyketide Biosynthesis: The Role of Crotonyl-CoA Reductase. *Metabolic Engineering*, 3, 40-48.
- Lin T-Y, Borketey L.S, Prasad G, Waters S.A, Schnarr N.A. (2013) Sequence, Cloning, and Analysis of the Fluvirucin B1 Polyketide Synthase from *Actinomadura vulgaris*. *ACS Synth. Biol.*, 2, 635–642
- Liu L, Thamchaipenet A, Fu H, Betlach M, and Ashley M., (1997) Biosynthesis of 2-Nor-6-deoxyerythronolide B by Rationally Designed Domain Substitution, *J. Am. Chem. Soc.*, 119, 10553-10554.
- Maier, T.; Jenni, S.; Ban, N. (2006) Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science*, 311, 1258-1262.

- McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. (1999) Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel "Unnatural" Natural Products, *Proc. Natl. Acad. Sci. USA*, *96*, 1846-1851
- Moutevelis-Minakakis, P., Neokosmidi, A., Filippakou, M., Stephens, D., Dennis, E. A., and Kokotos, G. (2007) Synthesis of lipophilic 2-oxoamides based on γ -aminobutyric and δ -aminovaleric analogues and their activity against phospholipase A2. *J. Pept. Sci.* *13*, 634– 641.
- Naruse, N., Konishi, M., and Oki, T. (1991) Fluvirucins A₁, A₂, B₁, B₂, B₃, B₄ and B₅, new antibiotics active against influenza A virus. III. The stereochemistry and absolute configuration of fluvirucin A₁. *J. Antibiot.* *44*, 756– 761
- Naruse, N., Tenmkyo, O., Kawano, K., Tomita, K., Ohgusa, N., Miyaki, T., Konishi, M., and Oki, T. (1991) Fluvirucins A₁, A₂, B₁, B₂, B₃, B₄ and B₅, new antibiotics active against influenza A virus. I. Production, isolation, chemical properties and biological activities. *J. Antibiot.* *44*, 733– 740
- Naruse, N., Tsuno, T., Sawada, Y., Konishi, M., and Oki, T. (1991) Fluvirucins A₁, A₂, B₁, B₂, B₃, B₄ and B₅, new antibiotics active against influenza A virus. II. Structure determination. *J. Antibiot.* *44*, 741– 755
- Ogasawara, Y., and Liu, H.-W. (2009) Biosynthetic studies of afaziridine formation in azicemicins. *J. Am. Chem. Soc.* *131*, 18066–18068.
- Ogasawara, Y., Katayama, K., Minami, A., Otsuka, M., Eguchi, T., and Kakinuma, K. (2004) Cloning, Sequencing, and Functional Analysis of the Biosynthetic Gene Cluster of Macrolactam Antibiotic Vicenistatin in *Streptomyces halstedii*. *Chem. Biol.* *11*, 79– 86
- Park SR, Han A.R, Ban Y.H, Yoo Y.J., Kim E.J., Yoon Y.J. (2010) Genetic engineering of macrolide biosynthesis: past advances, current state, and future prospects, *Appl Microbiol Biotechnol*, *85*, 1227–1239.

- Parenty, A., Moreau, X., and Campagne, J. M. (2006) Macrolactonizations in the total synthesis of natural products. *Chem. Rev.* 106, 911–939
- Petkovic H, Lill RE, Sheridan RM, Wilkinson B, McCormick EL, McArthur HA, Staunton J, Leadlay PF, Kendrew SG. (2003) A novel erythromycin, 6-desmethyl erythromycin D, made by substituting an acyltransferase domain of the erythromycin polyketide synthase. *J Antibiot (Tokyo)*, 56, 543-51.
- Pfeifer, B.; Khosla, C. (2001) Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.*, 65, 106-118.
- Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E-coli*, *Science*, 291, 1790-1792.
- Prasad, G.; Amoroso, J.W.; Borketey, L.S.; Schnarr, N. A. (2012) N-Activated β -Lactams as Versatile Reagents for Acyl Carrier Protein Labeling. *Org. Biomol. Chem.*, 10, 1992-2002
- Prasad G, Borketey LS, Lin T-Y, Schnarr NA. (2012) A Mechanism-Based Fluorescent Assay for Examining Ketosynthase Activity. *Org. Biomol. Chem.*, 10, 6717-23
- Rees WA¹, Yager TD, Korte J, von Hippel PH. (1993) Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry*, 32 (1), 137–44
- Reeves, C.D., Murli, S., Ashley, G.W., Piagentini, M., Hutchinson, C.R., McDaniel. R. (2001) Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry*, 40, 15464-470.
- Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R.; McDaniel, R. (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases, *Biochemistry*, 42, 72–79.

- Revill, WP.; Bibb, MJ.; Hopwood, DA. (1996) Relationships between fatty acid and polyketide synthases from *Streptomyces coelicolor* A3(2): characterization of the fatty acid synthase acyl carrier protein. *J. Bacteriol.*, 178(19), 5660-5667
- Roberts, R. R., Hota B., Ahmad I., Scott RD II, Foster S.D., Abbasi F., Schabowski S., Kampe L. M., Ciavarella G. G., Supino M., Naples J., Cordell R., Levy S. B., Weinstein, R. A., (2009) Hospital and societal costs of antimicrobial-resistant infections in a Chicago teaching hospital: implications for antibiotic stewardship. *Clin. Infect. Dis.* 49, 1175-1184.
- Rodriguez, E.; Banchio, C.; Diacovich, L.; Bibb, M. J.; Gramajo, H. (2001) Role of an essential acyl coenzyme A carboxylase in the primary and secondary metabolism of *Streptomyces coelicolor* A3(2). *Applied and Environmental Microbiology*, 67, 4166-4176.
- Rohr J., (2000), A New Role for Polyketides, *Angew Chem Int Ed*, 39 (16), 2847-2849.
- Rowe CJ, Böhm IU, Thomas IP, Wilkinson B, Rudd BA, Foster G, Blackaby AP, Sidebottom PJ, Roddis Y, Buss AD, Staunton J, Leadlay PF (2001) Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem Biol*, 8, 475–485
- Ruan X, Pereda A, Stassi DL, Zeidner D, Summers RG, Jackson M, Shivakumar A, Kakavas S, Staver MJ, Donadio S, Katz L. (1997) Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J. Bacteriol*, 179, 6416-6425.
- Stassi, D.L.; Kakavas, S.J.; Reynolds, K.A.; Gunawardana, G.; Swanson, S.; Zeidner, D.; Jackson, M.; Liu, H.; Buko, A.; Katz, L. (1998) Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proc. Natl. Acad. Sci. USA*, 95, 7305-09.
- Schnarr, N. A. and Khosla, C. (2006) *In Chemical Biology: From Small Molecules to Systems Biology and Drug Design*, Wiley-VCH, Weinheim

- Schnarr, N. A., Chen, A. Y., Cane, D. E., and Khosla, C. (2005) Analysis of covalently bound polyketide intermediates on 6-deoxyerythronolide B synthase by tandem proteolysis-mass spectrometry. *Biochemistry* (N. Y). 44, 11836-11842
- Smith, S. (2006) Architectural options for a fatty acid synthase. *Science*, 311, 1251-1252
- Smith, S. (1994) The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J.* 8, 1248-1259.
- Smith, S.; Tsai, S-C. (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.*, 24, 1041-1072.
- Staunton, J. and Wilkinson, B. (2001) Combinatorial biosynthesis of polyketides and nonribosomal peptides, *Curr. Opin. Chem. Biol.*, 5, 159– 164
- Tang, Y.; Chen, A.Y.; Kim, C-Y.; Cane, D.E.; Khosla, C. (2007) Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chem. Bio.*, 14, 931-943.
- Tran, L., Broadhurst, R. W., Tosin, M., Cavalli, A., and Weissman, K. J. (2010) Insights into protein-protein and enzyme-substrate interactions in modular polyketide synthases. *Chem. Biol.* 17, 705-716.
- Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* (N. Y.). 40, 2326-2331.
- Wallace, K.K., Bao, Z., Dai, H., Digate, R., Schuler, G., Speedie, M.K., Reynolds, K.A. (1995) Purification of Crotonyl-CoA Reductase from *Streptomyces collinus* and Cloning, Sequencing and Expression of the Corresponding Gene in *Escherichia coli*. *Eur. J. Biochem.*, 233, 954-62.

- Walsh, C. T. (2002) Combinatorial biosynthesis of antibiotics: Challenges and opportunities. *ChemBioChem*, 3, 124-134
- Wang, Y.; Pfeifer, B. A. (2008) 6-Deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in *E. coli*. *Metabolic Engineering*, 10, 33-38.
- Wawrik B, Kerkhof L, Zylstra G. J., Kukor J. J. (2005) Identification of Unique Type II Polyketide Synthase Genes in Soil. *Applied and Environmental Microbiology*, 71, 2232-38.
- Weissman K. J, Leadlay P. F. (2005) Combinatorial Biosynthesis of Reduced Polyketides, *Nature Reviews Microbiology*, 3, 925-936.
- Witzowski, A., Naggert, J., Witkowska, H. E., Randhawa, Z. I., and Smith, S. (1992) Utilization of an active serine 101-cysteine mutant to demonstrate the proximity of the catalytic serine 101 and histidine 237 residues in thioesterase II. *J. Biol. Chem.* 267, 18488–18492.
- Witzowski, A., Witkowska, H. E., and Smith, S. (1994) Reengineering the specificity of a serine active-site enzyme. Two active-site mutations convert a hydrolase to a transferase. *J. Biol. Chem.* 269, 379–383.
- Wu, N., Cane, D. E., and Khosla, C. (2002) Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry (N. Y.)*. 41, 5056-5066.
- Wu, N., Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc.* 123, 6465-6474.
- Xue, Q., Ashley, G., Hutchinson, C., and Santi, D. (1999) A multiplasmid approach to preparing large libraries of polyketides, *Proc. Natl. Acad. Sci. USA*, 96, 11740-11745

Yadav, G., Gokhale, R. S., and Mohanty, D. (2003) SEARCHPKS: a program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res.* 31, 3654– 3658

Yadav G.; Gokhale, R. S.; Mohanty, D. (2009) Towards prediction of metabolic products of polyketide synthases: an in silico analysis. *PLoS Comput Biol.*, DOI: 10.1371/journal.pcbi.1000351