

ABSTRACT

CROWETIPTON, JARED MATTHEW. Controlling the substrate selectivity of polyketide synthases through semi-rational design. (Under the direction of Gavin Williams).

Polyketides are a diverse group of natural products that display a range of biological activities. Many polyketides are used extensively in clinical settings, such as the antibiotic erythromycin. However over the past decades antibiotic resistance has become a major hurdle for commonly used antibiotics. Many polyketides are produced by multi-modular megasynthases called polyketide synthases (PKSs). Type I PKSs work in an assembly line fashion that usually involves the specific installation of an extender unit building block by each module.

Here we employ a modeling-directed semi-rational design strategy to manipulate the substrate selectivity of the pikromycin polyketide synthase. This approach probes the roles of residues in the active site of the acyltransferase domain, the domain responsible for extender unit selection, to identify the residues responsible for extender unit selection. We demonstrate the identification of mutations that improve non-native extender unit incorporation over the wild-type enzyme.

Further we develop a cell extract based dual enzyme assay to test, *in vitro*, the effects of mutations on other modules in the PKS. This approach is one of the first demonstrations of non-natural substrate acceptance in a PKS that was not from the first or final modules. This dual enzyme assay also demonstrates evidence for substrate selectivity beyond the AT domain. Here we describe several mutations that begin to create a small list of potential sites to control the substrate selectivity of the PKS. Additionally, we are able to demonstrate a single AT mutation that inverts the selectivity of module five of the pikromycin PKS for propargylmalonyl-CoA.

© Copyright 2018 Jared Matthew CroweTipton

All Rights Reserved

Controlling the Substrate Selectivity of Polyketide Synthases Through Semi-Rational Design

by
Jared Matthew CroweTipton

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Chemistry

Raleigh, North Carolina

2018

APPROVED BY:

Gavin Williams
Committee Chair

Joshua Pierce

Chase Beisel

Wei-chen Chang

DEDICATION

I would like to dedicate this thesis to my fiancée Marybeth Garrett.

BIOGRAPHY

Jared Matthew CroweTipton was born in Waco, Texas to Vaughn and Beverly CroweTipton on May 8, 1993. He has a younger sister, Abigail CroweTipton. Jared grew up in Greenville, SC but spent time in Waco (TX), Auburn (AL), Ardmore (OK), and Macon (GA). He attended Riverside High school where he spent most of his time running or preparing for chemistry labs. Jared attended Furman University from 2011-2015 and graduated with a Bachelors of Science in chemistry. While at Furman he researched prebiotic chemistry under the instruction of Dr. Greg Springsteen. There he discovered his love for research and the understanding of biological processes. After graduation from Furman, Jared moved to Raleigh, NC where he began graduate studies at North Carolina State University under the direction of Dr. Gavin Williams. There he continued studying biological processes with the goal of utilizing nature to produce novel bioactive compounds. The work he has done at NC State has encouraged him to continue his research on the production of natural products. Upon completion of the Master of Science degree in chemistry, Jared is looking forward to moving to an industrial setting to continue his research.

ACKNOWLEDGMENTS

I would like to acknowledge Marybeth Garrett and my parents, Vaughn and Beverly. Without their support and encouragement this document and work would never have been completed.

TABLE OF CONTENTS

List of Tables	vii
List of Figures.....	viii
CHAPTER 1	1
1.1 Introduction to Engineering Polyketide Biosynthesis	1
1.2 Approaches to Engineering the Biosynthesis of Unnatural Polyketides.....	4
1.2.1 Loading Domain Engineering.....	4
1.2.2 Engineering the Polyketide Backbone	5
1.3 Engineering PKS Substrate Scope	5
1.3.1 Domain and Module Swapping	7
1.3.2 <i>Trans</i> Acyltransferase Complementation	8
1.3.3 Acyltransferase Chimeras	8
1.3.4 Rational Design.....	9
1.4 Pikromycin.....	10
1.4.1 The Pikromycin PKS	11
1.5 Scope of this thesis.....	13
CHAPTER 2	14
2.1 Description of the Overall Strategy.....	14
2.2 Characterization of the Extender Unit Specificity of Wild-Type PikAIIIITE	17
2.2.1 Selection and Design of PikAIIIITE Mutants	18
2.2.2 PikAIII AT Docking Studies	20
2.2.3 Development of a crude extract assay	23
2.2.4 Characterization of the Extender Unit Specificity of PikAIIIITE Mutants	24

2.3 Coupling PikAIII with PikAIV	28
2.3.1 Characterization of the Activity of PikAIII/PikAIV	28
2.3.2 Dissecting the Role of PikAIII and PikAIV in Determining Substrate Selectivity	32
2.3.3 Characterization of PikAIII and PikAIV Mutants	35
2.11 Conclusions	37
Methods.....	40
General	40
Mutagenesis of PikAIII and PikAIV	40
Protein Production	41
Protein Purification	41
Lysate Preparation	42
MatB Reactions and Acyl-CoA preparation	42
Pentaketide Assay	42
CHAPTER 3	44
3.1 The Future of Pikromycin Engineering.....	44
3.2 <i>In Vivo</i> engineeringPolyketide Engineering	45
REFERENCES.....	49
APPENDICES	58

LIST OF TABLES

Table 1: Product distributions of PikAIIIITE pentaketide assays. All results were run in duplicate and are the average of the 2 reactions. 1 is the expected methyl extension product and 2b is the expected propargyl extension product	25
Table 2: Product percentages of the extracted ion chromatograms of PikAIII + PikAIV reactions. Entries with no error are indicative of error less than 0.005. First number is the area of the EIC normalized to Entry 1 product 3 set as 100. Product percentages are in parentheses; 10-DML (*) products are considered separately from narbonolide products (**).	33
Table 3: 10-DML and narbonolide product distributions of PikAIII + PikAIV competition assays. First number is the area of the EIC normalized to 3b of the wild type enzyme run with each mutant set to 100. Product percentages are in parentheses; 10-DML (*) products are considered separately from narbonolide products (**). Error bars of 0 indicate error less than 0.005	36

LIST OF FIGURES

Figure 1: Representative polyketides and their industrially-relevant bioactivities	2
Figure 2: The erythromycin biosynthetic pathway (Top) Domains that make up a module, the minimal module is in bold (Bottom).....	3
Figure 3: Mechanism of thioester exchanges from CoA to the AT then from the AT to the ACP.....	6
Figure 4: PKS AT engineering overview. Each scheme represents a different technique for PKS engineering. From left to right: Native, AT mutagenesis, AT swapping, AT hybridization, domain swapping, trans AT complementation. Pink domains represent non-native domains and the red circles show the site of change.....	7
Figure 5: The pikromycin PKS. Pikromycin is produced through a 6 module PKS that includes 2 TE domains. This PKS is unique in that it can make 2 different macrolactones corresponding to cyclizations from the 5 th or 6 th extension. Post-PKS reactions hydroxylate and glycosylate either structure to tailor it into its final product	12
Figure 6: Scheme of the pentaketide assay with PikAIIITE as a competition. Here methylmalonyl-CoA, and propargylmalonyl-CoA are provided along with the pentaketide substrate. Not shown is a NADPH regeneration system that recycles NADPH through glucose-6-phosphate (G6P) and G6P dehydrogenase. The 2 expected products are detected by MS	15
Figure 7: (A) Scheme of CoA linked extender unit generation using MatB. (B) Representative HPLC trace of MatB reactions. Residual ATP and AMP are found around 3min, any left over CoA could be detected at 11min, and the malonyl-CoA derivatives would elute beyond the 11min mark	16
Figure 8: The substrate promiscuity of PikAIIITE (Blue) compared to the substrate promiscuity of Ery6TE (Green). PikAIIITE products were detected by MS. Ery6TE products were detected by a combination of HPLC-UV and	

MS. MS traces of the dehydration products of 1, 2b, and (2c). All products detected are less than ± 5 Δ ppm from the calculated masses	17
Figure 9: PikAIIITE Acyltransferase homology model. This model was constructed using the I-tasser server and refinement methods ⁵³ . Methylmalonyl-CoA is docked relative to where it would need to be for a reaction to occur. Residues critical for the activity or selection of the enzyme and substrate are in gold	19
Figure 10: PikAIII AT homology model showing the sites selected for mutation. (A) CoA on view, as CoA would enter the active site. (B) Left side on view, Arg680 is the back of the active site. Gold residues are critical for enzyme activity and if mutated result in inactivity. Pink residues have been tested in the literature previously. Green residues are novel sites highlighted by rational design or docking.....	21
Figure 11: Comparison of the product distribution of purified enzyme vs. clarified cell lysate	24
Figure 12: Activities of PikAIIITE and mutants relative to the wild-type enzyme	27
Figure 13: Scheme of the pentaketide assay with PikAIII coupled to PikAIV and the expected cyclization products	28
Figure 14: Representative chromatogram of an LC-MS sample of PikAIII with PikAIV. 10-DML products, 1 and 2b, are detected first at 8 and 9 min respectively. Narbonolide products, 3-5, are detected after and each product shows 2 peaks, each with the same mass. Chromatogram is not indicative of product ratios.	28
Figure 15: Product distribution of PikAIII + PikAIV reactions. Methyl vs. X. The 2 peaks from the LC are combined as narbonolide 3 as they are representative of the same product. Narbonolide 3 has been set to 100 and all other peaks normalized accordingly. Activity relative to 3 is labeled on top of each peak	30

Figure 16: Overview of the products from the dual enzyme assay. PikAIII accepts the Tp-pentaketide substrate and selects between the two possible extension units, methylmalonyl-CoA and propargylmalonyl-CoA in ratio W:X respectively. The methyl or propargyl hexaketide product is then cyclized to 10-DML (not shown) or accepted by PikAIV, which also selects between methyl and propargyl in ratio Y:Z respectively. After extension by PikAIV the cyclized narbonolide products 1, 2b, 3, 4b, and 5b are detected. The percentages of the narbonolide products are shown. 31

Figure 17: An overview of CRISPR-Cas9 for genome editing. *cas9* and plasmid pgRNA are transformed into a bacterial cell and expressed. Plasmid pgRNA transcribes the gRNA that serves as a genomic targeting strand and complexes with Cas9. Cas9 cuts the genome at the target site, if there is no repair the cell will die. A gene of interest (GOI) is provided along with homology arms matching areas flanking the target site. Homologous recombination can occur to repair the genome and incorporates the gene of interest. 47

CHAPTER 1

1.1 Introduction to Engineering Polyketide Biosynthesis

Polyketides are a diverse class of natural products that exhibit a wide range of bioactivities. Many have clinical applications and have been targeted for industrial production (**Figure 1**). From a synthetic standpoint, most polyketides are structurally complex and functionalized requiring long, intricate syntheses. For example, Woodward et al. first synthesized erythromycin A in 50 steps with a 0.2% yield.¹⁻³ A convergent strategy for a similar class of polyketide, the ketolides, has recently been developed and offers an elegant solution in only 14 steps with a 16% yield.⁴ Even with the recent advances in synthetic chemistry, most clinically-relevant polyketides on the market are produced through biology.⁵ Understanding the biological processes that produce these bioactive products will help to unlock the potential for derivatization of the products inside their natural hosts. Modification of the polyketide could tailor its bioactivity to increase its potency, avoid antibiotic resistance, or develop new activity.

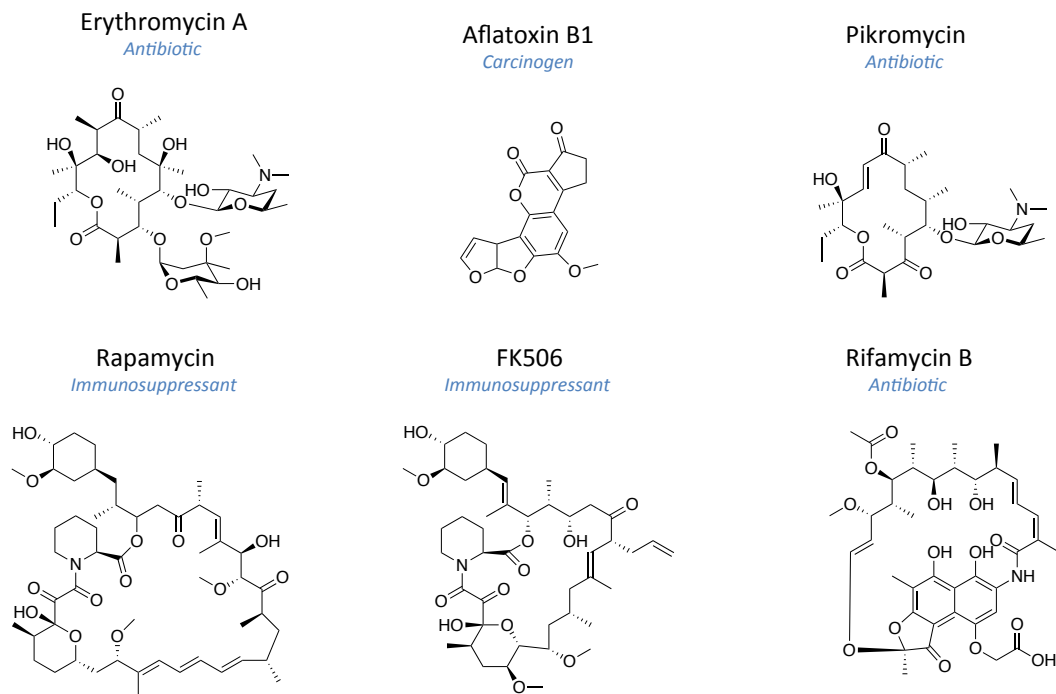


Figure 1: Representative polyketides and their industrially-relevant bioactivities.

Polyketides are produced naturally by massive enzymatic assembly lines named polyketide synthases (PKSs).⁶ Type I PKSs consist of groups of modular enzymes that work as an assembly line.⁷ This class of PKS is commonly found in bacteria such as *Actinomyces*.

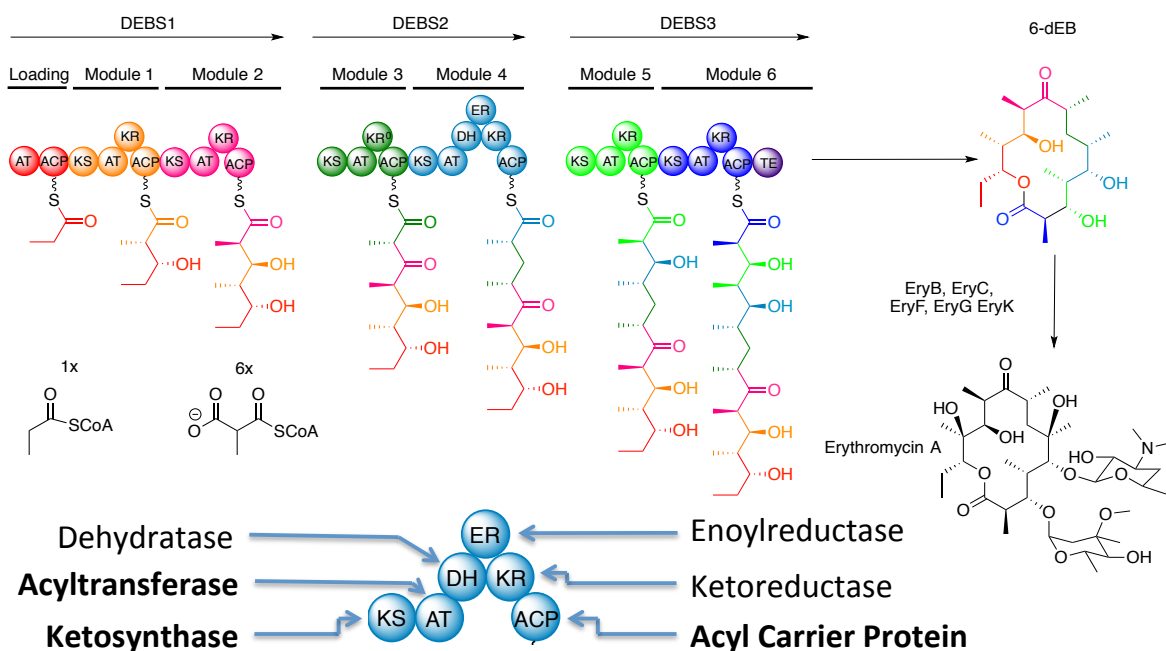


Figure 2: The erythromycin biosynthetic pathway (Top) Domains that make up a module, the minimal module is in bold (Bottom).

Type I PKSs are organized as a set of linear modules. The modules can be further subdivided into domains wherein each domain plays a singular role.⁸ A minimal module consists of at least three domains (**Figure 2**): the acyltransferase (AT), the ketosynthase (KS), and the acyl carrier protein (ACP). The type I PKS, 6-deoxyerythronolide B synthase (DEBS; **Figure 2**), produces 6-dEB, a precursor to the commonly-used antibiotic erythromycin A.⁹ There are three genes that code for six extension modules, a loading module, and a thioesterase (TE).¹⁰ The loading module accepts propionyl-coenzyme A (CoA) and then passes the propionate through each module where it receives one 3-carbon extension unit from the corresponding AT domain. The AT domain functions as a “gatekeeper” for substrate selection. Each AT allows for only an acyl-CoA extender unit to be loaded onto its active site from the pool of acyl-CoAs available in the environment.¹¹

The ACP acts as a shuttle for the chain as it moves from one domain (or module) to the next using a phosphopantetheine arm. First, the ACP moves the extender unit from the AT and into the KS domain where the previously elongated chain resides from the preceding

module. The KS catalyzes a decarboxylative Claisen condensation that adds the extender unit to the growing chain. Post chain extension, there are other in-line domains that can act on the chain such as methyltransferases, ketoreductases (KRs), dehydratases (DHs), and enoylreductases (ERs). Once the chain has gone through every part of a module, the ACP domain passes the chain to the KS domain of the next module. After the six extensions, the chain is hydrolyzed and cyclized via a thioesterase (TE) domain resulting in the product 6-dEB. Once 6-dEB is produced, the macrolactone undergoes two glycosylations, one methylation, and two hydroxylations to produce the fully decorated erythromycin A product.

1.2 Approaches to Engineering the Biosynthesis of Unnatural Polyketides

Precursor-directed biosynthesis was one of the earliest strategies attempted to produce unnaturally modified polyketides.¹² This strategy relies on the natural promiscuity of the native enzymes towards non-native substrates. Key to this approach, a non-native precursor is introduced into the wild-type polyketide producing strain.¹³ The PKS will then select between the native and non-native precursors to produce a distribution of products that use the two precursors. To demonstrate this concept, it has been shown that the inactivation of a KS domain will allow for DEBS module 6 to accept *N*-acetylcysteamine (SNAC) based substrates.¹⁴ This KS knockout was further used *in vivo* to successfully create an erythromycin analogue that contained an aromatic starter unit in place of the natural propionyl group.¹⁵ While this strategy provides a relatively simple and easy way to make a novel product, it is limited by the efficiency with which the native enzyme will tolerate the new precursor. To address this issue, PKSs can be engineered to tailor their enzymatic specificity toward the development of novel products.¹⁵ The engineering approach can take many forms; however, it generally works on the idea of developing a mutated enzyme to perform a new set of reactions or react with novel substrates to make a non-natural product.

1.2.1 Loading Domain Engineering

One notable engineering effort has been to modify the first part of the synthase, the loading domain (LD). This domain is responsible for priming the assembly line with an acyl-CoA derived starting unit that becomes the first unit of the polyketide chain. Many

polyketides have been shown to be amenable to changes in the loading domain.^{16,17} The avermectin PKS (Ave) loading domain is known to accept over 40 substrates including some with branched chains, which are uncommon in type I PKSs.¹⁸ Promiscuity of the Ave LD was utilized for the production of erythromycin derivatives by replacing the native DEBS LD with the Ave LD. The Ave LD/DEBS hybrid was able to produce both isopropyl and *sec*-butyl erythromycin derivatives.¹⁸ Another example is the derivatization of the insecticide spinosyn by replacing the LD of the erythromycin PKS with the LD of the spinosyn PKS and the combination of the avermectin PKS LD with the LD of the spinosyn PKS. These LD hybrids could produce spinosyn derivatives containing cyclobutyl, cyclopropyl, cyanomethyl, *sec*-butyl, methylthiomethyl, isopropyl, *N*-propyl, 2-furanyl, and 2-methylcyclopropyl groups at the loading domain position on the PKS backbone.¹⁹

1.2.2 Engineering the Polyketide Backbone

The engineering of the reductive domains can significantly alter the structure of polyketides produced by PKSs. KR domains are important for setting the stereochemistry of the polyketide backbone.¹² Inactivation is a powerful tool in the reductive loop, as the activity of these modules controls the oxidation of the backbone and the physical structure of the later product. It is also possible to swap the reductive loop from one PKS to another, resulting in new products.²⁰

Controlling the structure of the polyketide can also be done by mutating or swapping the TE domain. The TE provides the environment for any final macro-cyclization that occurs in many polyketides. The significant diversity in the pool of natural polyketides provides many options for TEs that give different cyclization patterns.^{20–22} After cyclization, many polyketides undergo post-PKS modifications. Common reactions found in this category are glycosylations, hydroxylations, and methylations. Each tailoring reaction on the PKS backbone can change its bioactivity.¹²

1.3 Engineering PKS Substrate Scope

One of the most difficult parts of a polyketide to diversify is the 2-position of each malonyl-CoA subunit. The AT domain controls which extender unit is chosen from the pool

of malonyl-CoA derivatives available to the cell.²³ The AT domain is a homodimer, and the monomer consists of a small subunit and a large subunit. The junction of the large subunit and the small subunit form the AT active site. AT domains use a ping-pong bi-bi mechanism similar to a serine hydrolase enzyme (**Figure 3**).²⁴

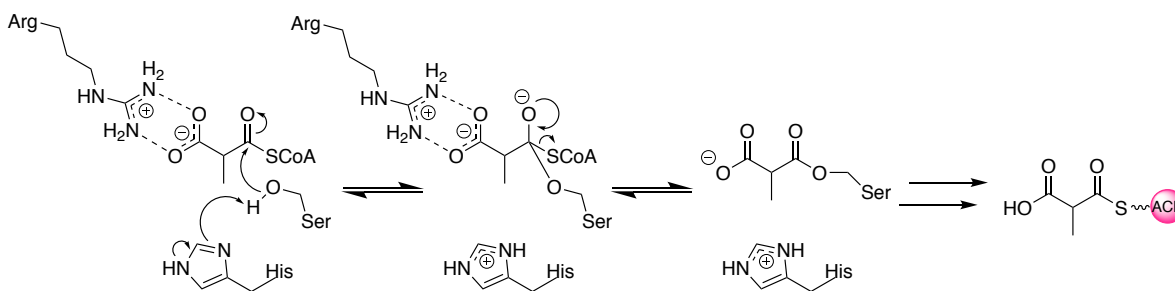


Figure 3: Mechanism of thioester exchanges from CoA to the AT then from the AT to the ACP.

The catalytic Ser and His are on the large and small subunits, respectively (**Figure 11**).²⁵ Also of importance is an Arg residue towards the back of the active site, which is critical for positioning the carboxylic acid of the malonyl-CoA derivative in the active site. Once in position, a thioester exchange reaction occurs that transfers the acyl group to the AT active site Ser residue.²⁴ After the free CoA leaves, a second exchange occurs between the AT-linked extender unit and the phosphopantetheine arm of the ACP.

Some ATs are more specific than others, but generally, they select between malonyl-CoA and methylmalonyl-CoA.²⁶ Many approaches to engineering AT domains have been explored; some have seen more successes than others. Domain swapping and module swapping were the first techniques used to engineer the AT module.

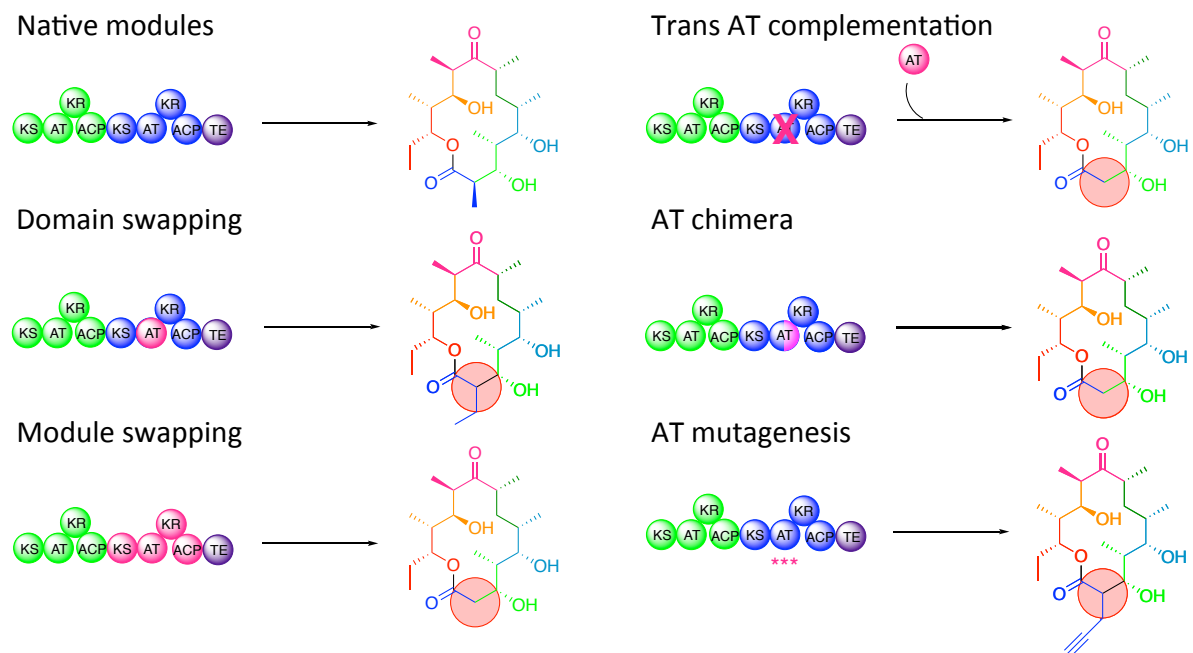


Figure 4: PKS AT engineering overview. Each scheme represents a different technique for PKS engineering. From left to right: Native, AT mutagenesis, AT swapping, AT hybridization, domain swapping, trans AT complementation. Pink domains represent non-native domains and the red circles show the site of change.

1.3.1 Domain and Module Swapping

Loading domain swapping was one of the first examples of domain or module swapping, but it only allows access to the substrate scope of the LD (**Figure 2**). AT domain swapping of the extender modules has allowed for the modification of more positions on the polyketide backbone. In order to swap the AT6 of DEBS for the AT4 of epothilone PKS, experimental testing of various AT boundaries, guided by sequence alignment, led to a DEBS variant that retained 70% activity compared to the native system.²⁷ Other examples of AT swaps include the swap of DEBS AT6 with the rapamycin AT2, rifamycin AT6 with the rapamycin AT2, DEBS AT4 with the niddamycin AT5, and the geldanamycin AT1 with rapamycin AT2 to produce derivatives with alternate specificity from the native PKSs.^{28,29}

In contrast to swapping the AT-domain only, the benefit of entire module swaps is the transfer of the AT specificity and the corresponding reductive loop of the non-native module. Over 154 bimodular combinations have been created from the swapping of modules from 8

different macrolide PKSs to produce triketide lactone products.³⁰ However, in both domain and module swapping, penalties are usually incurred in the form of decreased total activity due to disrupted protein-protein interactions. This decrease is especially highlighted by the bimodular combinations wherein ~50% of the combinations did not produce any detectable cyclized product.³¹ In the assembly line-like PKSs, communication between proteins is extremely important and disruptions to the system can significantly decrease activity.

1.3.2 *Trans* Acyltransferase Complementation

One strategy to get around the decrease in activity is to use an AT that works in *trans*. In some PKSs, the AT domain is not a part of the connected modular system, and one or more freestanding AT enzymes are able to service the PKS. In engineered systems, if a suitable *trans*-AT is provided to the PKS it can begin to compete with the native AT for substrate selection.³² To be most effective, the native AT is inactivated through a catalytic Ser→Ala mutation and the *trans*-AT is overexpressed in the microbial host. For example, DEBS module 6 has been complemented with a *trans*-AT, DSZS, to produce a malonyl-CoA extension in place of a methylmalonyl-CoA—with a small reduction in activity.³³ While the negative effects on protein activity are not observed, the pool of *trans*-ATs described in the literature is specific for only a few extender units *in vivo*.²³ Recent *in vitro* studies have identified the *trans*-AT KirCII as able to utilize the non-natural extender units allylmalonyl-CoA and propargylmalonyl-CoA.³²

1.3.3 Acyltransferase Chimeras

AT chimeras or hybrids are created by blending the DNA of several PKSs via DNA shuffling or homologous recombination to make a final hybrid AT, usually as a small or modest library (e.g. 200 variants).³⁴ While this technique can result in a blended AT that has shared characteristics of both parent domains, the vast majority of the library members are inactive or have not undergone much change from their parent gene. Hybridization of the pikromycin PKS modules 5 and 6 along with DEBS module 6 revealed that most of the active hybrids were only changes in the flexible linker regions between domains rather than in the active site of the AT where they could affect activity.³⁵

1.3.4 Rational Design

A more minimalist approach to AT engineering, rational design, utilizes knowledge about the AT's structure and mechanism to direct mutations that change the specificity of the enzyme. Additionally, the AT can be engineered to accept unusual extender units—inaccessible by other means.²⁸ The minimal perturbations of the AT active site could feasibly lead to engineering multiple ATs in a PKS to have different selectivities.

Substrate selection in AT is highly impacted by a conserved motif amongst many AT domains. This motif changes depending on what malonyl-CoA derivative is being selected for and was discovered by comparing conserved sequences across multiple AT domains of varying specificity.³⁶ In most methylmalonyl-CoA specific AT domains, this motif consists of the amino acids YASH, while other versions include HAFH and VASH for malonyl- and ethylmalonyl-CoA respectively (**Figure 9** left coil, insert).^{36–38} Considering the YASH motif, the His is critical for catalytic activity and is conserved across all domains, independent of substrate selectivity. The Tyr and Ser residues in the motif use their size and hydrogen bonding capabilities to form a screen for methylmalonyl-CoA that hinders other substrates from binding in the active site.¹¹ With ATs that have specificity for the larger ethylmalonyl-CoA, the Tyr→Val substitution likely shrinks the size of the screen for the larger substrate.³⁸ ATs with specificity for smaller substrates, containing HAFH, use the larger size of the F to screen out all residues but malonyl-CoA.³⁶ Across all specificities there is a general trend of increasing flexibility in the motif as larger substrates are selected.¹¹

In the DEBS PKS, several attempts have been made at controlling substrate specificity through rational design. To understand the initial substrate promiscuity, DEBS module 6 TE was used in an *in vitro* assay in which a library of extender units were provided in competition with the native substrate. It was found that the enzyme could create a triketide lactone product with methyl, ethyl, allyl, propargyl, butyl, hydroxymethyl, azidoethyl, and benzylmalonyl-CoAs.³⁹ Further, *in vivo* feeding experiments found that an engineered strain of *Saccharopolyspora erythraea* would incorporate a propargylmalonyl-SNAC into erythromycin at low yields.⁴⁰

The first rational design-based engineering efforts chose mutations according to AT sequence homology across multiple ATs.³⁷ Many of the first mutations were at residues that were required for activity and helped to establish a sequence-function relationship of sites inside the AT. Coupling the sequence homology to crystal structures implicated residues R222, Y297, and S299 as important for substrate positioning or binding. Of particular interest is a Val (V295) residue that when mutated to an Ala residue allowed for the increased use of propargylmalonyl-SNAC in DEBS AT6.³⁷

Molecular dynamics of the Val→Ala mutant demonstrated a wider and more flexible active site because of the mutation. Further, modeling bulky substrates into the wild-type AT active site showed disordered interactions with the substrate as it could not fit inside the active site. To increase the size and flexibility of the active site, a DEBS AT6 triple mutant was developed that caused a 3.79-fold increase in the amount of propargyl incorporation.¹¹

Continued use of this design principle by the William's Lab has identified another mutation, Tyr→Arg, in DEBS AT6 that flipped the substrate selectivity from methylmalonyl-CoA to propargylmalonyl-CoA *in vitro*. This mutation gave a greater total activity with the non-natural substrate than the natural enzyme demonstrates with methylmalonyl-CoA. A double mutant combining the Val→Ala mutation with the Tyr→Arg completely abolished methylmalonyl-CoA activity while retaining low activity with the propargyl substrate.⁴¹ This groundbreaking work shows the power of (semi-) rational design in the erythromycin PKS. To further demonstrate the rational design technique and to develop novel antibiotics beyond erythromycin, the mutations and lessons learned from this PKS need to be applied to other PKSs.

1.4 Pikromycin

Pikromycin is a polyketide produced by *Streptomyces venezuelae* and was the first macrolide antibiotic discovered.⁴² Though discovered in the 1950s, the macrolide was not chemically synthesized until 2011.⁴³ Pikromycin exhibits weak antibiotic activity when compared to other erythromycin, with a MIC₉₉ (minimum inhibitory concentration for 99% of the population) in *S. aureus* of 100 μM while erythromycin demonstrates a MIC₉₉ of 0.68

μM .^{13,44} Both compounds have the same mechanism of action, binding to the protein exit tunnel of the bacterial ribosome to inhibit protein synthesis.⁴⁵ Unlike erythromycin, pikromycin is a ketolide, indicating that it lacks a cladinose sugar and retains a keto group in its place (**Figure 5**). Novel ketolides are being developed as third or fourth generation antibiotics, which makes engineering the pikromycin PKS an interesting target.^{13,46} The pikromycin PKS serves as a comparison to the commonly studied PKS of erythromycin.¹³ Unlocking control of this PKS could allow for the addition of further chemical handles or reactive moieties that would be difficult to add by traditional chemical synthesis.

1.4.1 The Pikromycin PKS

The pikromycin PKS consists of five genes encoding a total of six modules and a freestanding thioesterase. Every module except for the second accepts a methylmalonyl-CoA extender unit, while the second accepts malonyl-CoA. The PKS starts with the loading of a propionyl-CoA starter unit that then undergoes five or six chain extensions. After the fifth or sixth extensions, the chain can be cyclized by the TE domain. This PKS is unique for the creation of two macrolactone products, one 14-membered and one 16-membered.⁴⁷

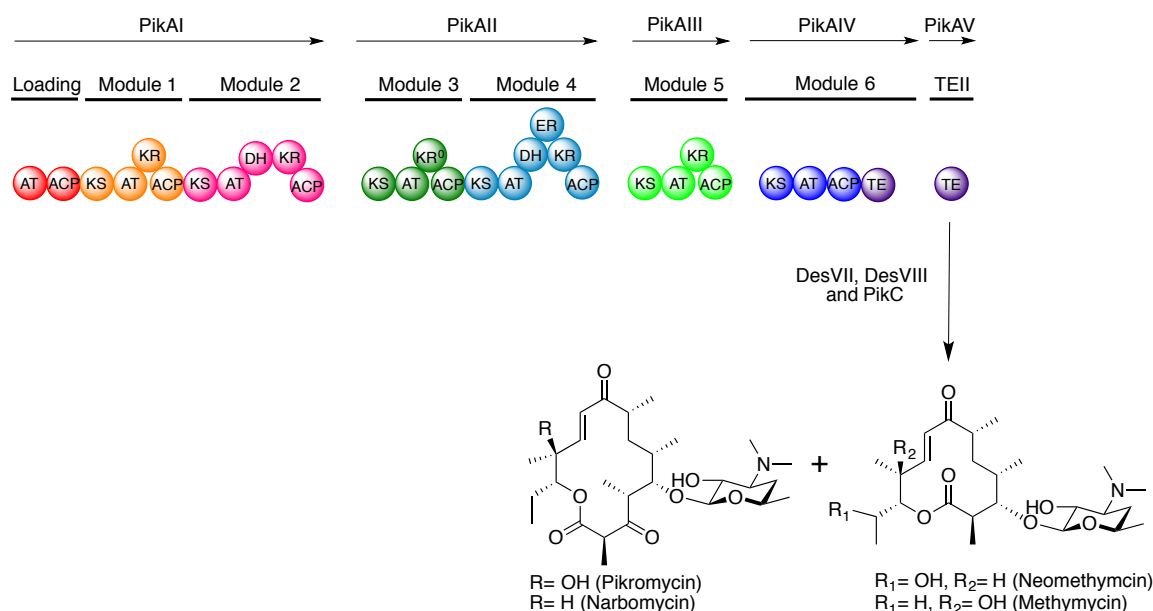


Figure 5: The pikromycin PKS. Pikromycin is produced through a 6 module PKS that includes 2 TE domains. This PKS is unique in that it can make 2 different macrolactones corresponding to cyclizations from the 5th or 6th extension. Post-PKS reactions hydroxylate and glycosylate either structure to tailor it into its final product.

The cyclization of the two compounds depends on when the thioesterase acts upon the chain. 10-deoxyymethynolide (10-DML, **1**) is formed after five extensions and narbonolide is formed after six extensions. These two compounds then become one of four possible products depending on the sugar and alcohol additions from post-PKS enzymes (**Figure 5**). Additionally, the final two modules are standalone, making them prime targets for enzyme engineering efforts. Standalone modules are considered easier to engineer because they have evolved to function more independently than covalently connected modules.

Preliminary results from Sarah Schultheis in the Williams group demonstrated that the PikAIII AT is highly selective for the methyl extender unit while comparing methylmalonyl-CoA and propargylmalonyl-CoA and that it will produce both **1** and the propargyl derivative **2b**.⁴⁸ The preliminary results of the Val753Ala mutation, first found in DEBS by Sundermann et al., in PikAIII AT indicated an increase in the amount of the propargyl product, but at a smaller percentage than observed in DEBS.³⁷ In PikAIII AT, the

Tyr755Arg mutation, found by Koryakina et al., which gave a complete inversion of substrate selectivity in DEBS module 6, causes complete inactivation.⁴¹

1.5 Scope of This Thesis

This thesis will cover our efforts toward the rational design of the AT domains of modules 5 and 6 of the pikromycin PKS to accept non-natural extender units. Herein, we develop and use a high-resolution mass spectrometry-based assay in clarified cell lysate to determine the structure to function relationship between selected mutations and the new substrates. We also examine whether the DEBS mutations described earlier can impact substrate specificity of the pikromycin PKS. These mutations will then be employed in bimodular experiments to determine the effects of the mutations on protein communication and product formation. Further, this thesis will lay the groundwork for a platform to genomically incorporate and edit PKS genes *in vivo* with multiple hosts.

CHAPTER 2

2.1 Description of the Overall Strategy

Of the final two pikromycin PKS modules, PikAIII shares the higher homology with DEBS module 6 (50%) and was thus chosen for further study. The MS-based pentaketide assay developed by Hansen, et al. and used as a competition assay by Koryakina, et al. was found to be the easiest way to study this system.⁴¹ In this reaction, the KS domain accepts the pentaketide as if it were the incoming chain from PikAII. The AT then screens the various malonyl CoAs present and, if accepted, allows for the selected unit to be passed on to the ACP which moves it to the KS where the Claisen condensation occurs. Once the extension occurs, the TE cyclizes the now-hexaketide into **1**. The thiophenyl-linked pentaketide substrate which serves as an exact chain mimic for the KS of PikAIII.⁴⁹ The assay can be used with only a single extender unit to test PikAIII's activity with a non-natural substrate. As the overarching goal of this project is to derivatize macrolides, propargyl incorporation is one of the more valuable targets. The installation of an alkyne in the backbone of a macrolide will change the macrolide's bioactive properties and also will add a functional handle for further semisynthetic approaches via click chemistry.³² Additionally, the mutations that increase propargyl incorporation can be applied to other large substrates.

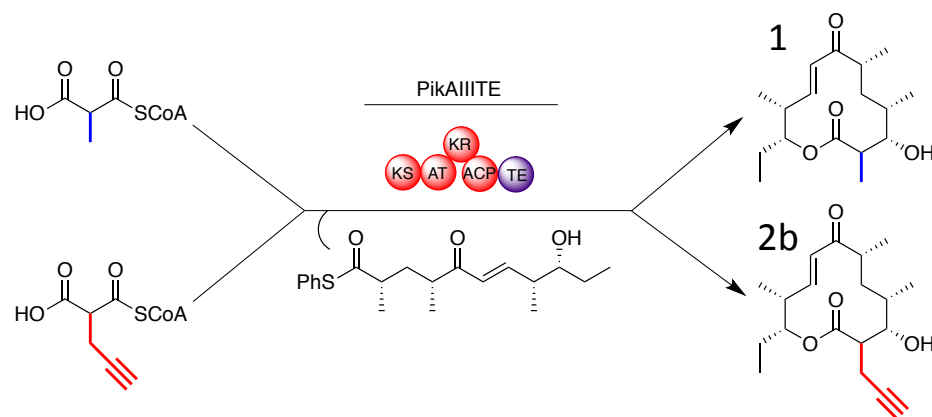


Figure 6: Scheme of the pentaketide assay with PikAIIIITE as a competition. Here methylmalonyl-CoA, and propargylmalonyl-CoA are provided along with the pentaketide substrate. Not shown is a NADPH regeneration system that recycles NADPH through glucose-6-phosphate (G6P) and G6P dehydrogenase. The 2 expected products are detected by MS.

PikAIII AT naturally accepts methylmalonyl-CoA. PikAIII alone will not produce a detectable amount of **1** so it was fused to the TE of PikAIV by Mortison, et al.⁵⁰ This fusion removed the C-terminal linker from PikAIII and replaced it with the TE and the TE linker. With a chain mimic, the module and TE fusion will catalyze one extension prior to cyclization to form product **1** (**Figure 6** blue).

Various extender units can also be provided in competition to see which extender(s) the enzyme favors (**Figure 6**). Polyketide host strains naturally provide the native extender unit for a native PKS. Therefore, in an effort to mimic the cellular environment, and to provide a measure of the selectivity of various Pik PKS mutants, both the natural methylmalonyl-CoA and non-native propargylmalonyl-CoA were provided in competition.

The preliminary results of PikAIII found by Sarah Schultheis demonstrated the enzyme is highly selective for the methyl extender unit while comparing methylmalonyl-CoA and propargylmalonyl-CoA.⁴⁸ Her results comparing the mutagenesis in DEBS module 6 AT with PikAIII AT indicated the mutations either had less of an effect or inactivated the PikAIII enzyme. In order to compare the AT substrate scope of DEBS module 6 and PikAIII

a library of extender units, similar to those used by Koryakina, et al., was produced (**Figure 7**).³⁹

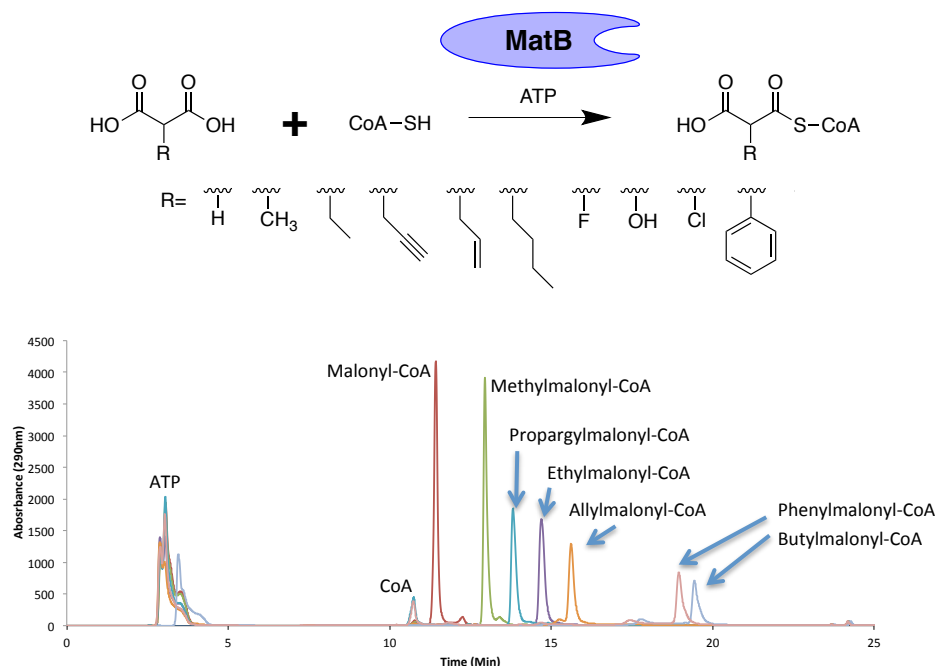


Figure 7 (A) Scheme of CoA linked extender unit generation using MatB. (B) Representative HPLC trace of MatB reactions. Residual ATP and AMP are found around 3min, any left over CoA could be detected at 11min, and the malonyl-CoA derivatives would elute beyond the 11min mark.

The enzyme MatB, a coenzyme A (CoA) ligase, and an engineered version were used to generate methylmalonyl-CoA and propargylmalonyl-CoA *in vitro* for the assay. MatB had previously been engineered by Dr. Koryakina to accept a large panel of substrates allowing for the *in vitro* generation of many different malonyl-CoA derivatives.⁵¹ These malonyl-CoA variants are necessary substrates for PKS specificity engineering and prior to engineering MatB, many of the derivatives were inaccessible by other means. A library of malonyl-CoA derivatives was generated by incubating MatB or the relevant MatB mutant with the respective malonic acid, ATP, and limiting CoA. Reactions proceeded overnight to completion and assayed by HPLC. All extender units used in pentaketide assays with pikromycin had undergone at least 95% conversion, and all extender units used in competition assays underwent greater than 99% conversion.

2.2 Characterization of the Extender Unit Specificity of Wild-Type PikAIIITE

To first characterize the promiscuity of the wild-type PikAIIITE enzyme, a library of extender units including the methyl, ethyl, butyl, allyl, benzyl, phenyl, fluoro, chloro, and malonyl-CoAs were generated using MatB wild-type, single mutant T207A, and double mutant T207G/M306I. Koryakina, et al. and Chang, et al. have previously demonstrated that DEBSmod6TE will accept all of the extender units except for malonyl, phenyl, hydroxy, and chloro (**Figure 8**; green box).^{41,52}

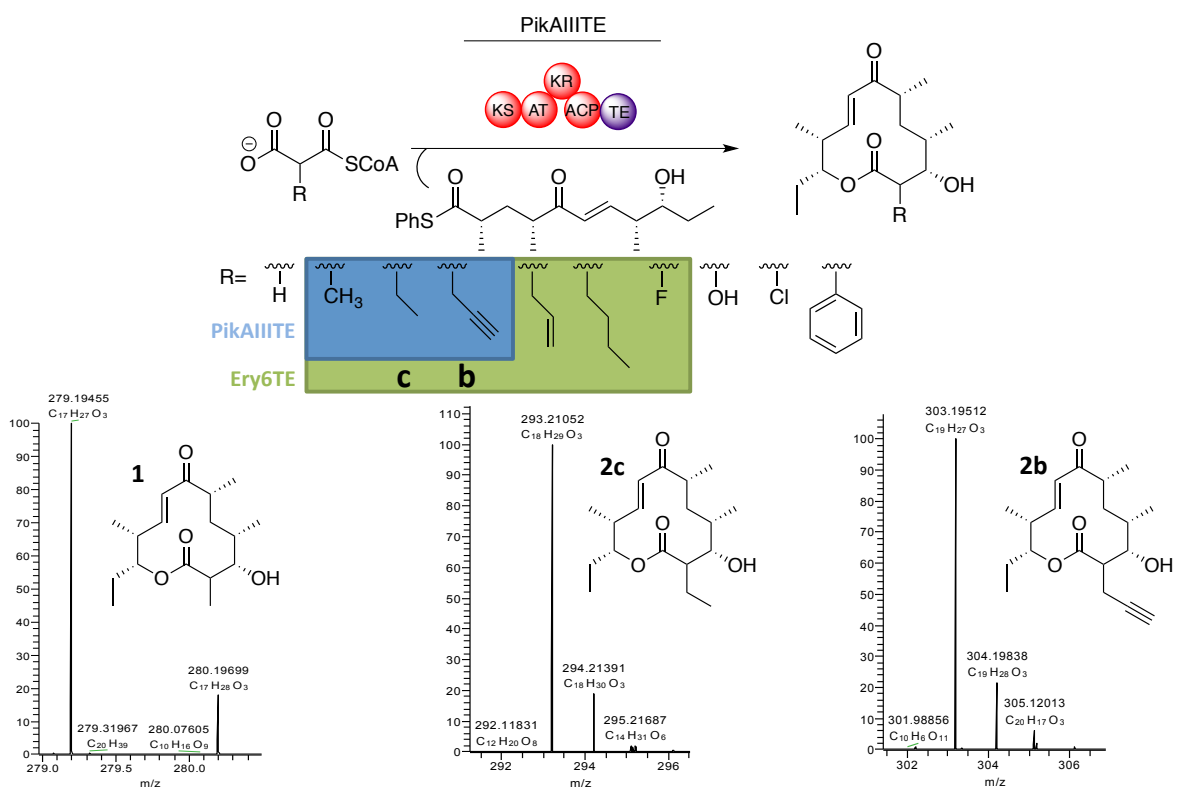


Figure 8 The substrate promiscuity of PikAIIITE (Blue) compared to the substrate promiscuity of Ery6TE (Green). PikAIIITE products were detected by MS. Ery6TE products were detected by a combination of HPLC-UV and MS. MS traces of the dehydration products of 1, 2b, and (2c). All products detected are less than ± 5 Appm from the calculated masses.

Notably, wild-type PikAIIIITE supported the production of the corresponding 10-DML analogue when methylmalonyl-, propargylmalonyl-, and ethylmalonyl-CoA extender units were provided (**Figure 8**; blue box and MS spectra). Methylmalonyl-CoA is incorporated 100-fold more efficiently than ethyl- and propargylmalonyl-CoA, as judged by the area of the extracted ion chromatogram. This result, in contrast to the previous work on DEBSmod6TE, suggests that the active site of the PikAIIIITE AT is more substrate-selective than its erythromycin PKS counterpart. While the AT is likely the major determinant of substrate selectivity, other domains, notably the KS and TE, could be affecting the substrate selectivity. The KS requires the non-natural substrate to fit and react inside its active site. The non-natural TE fusion could be affecting the entire structure of the module, thus forcing the module into a conformation different from its natural state. Also, cyclization in the TE requires the chain to have some degree of rotational freedom in the active site and the non-natural substrate could be removing that freedom. Preliminary studies (**Figure 8**) of mutants in the PikAIII AT do suggest that mutations to the AT active site are able to shift the product ratios and demonstrate that the AT is an important selectivity filter.

2.2.1 Selection and Design of PikAIIIITE Mutants

The observation that PikAIII AT has a more selective active site than DEBS module 6 AT led to the hypothesis that the PikAIII AT active site is tighter and more constricted than DEBS module 6 AT. This hypothesis can explain why the previously tested Tyr755Arg mutation caused inactivation of the enzyme and the Val753Ala mutation did not see as much improvement of non-natural substrate incorporation. Both the Val753 and the Tyr755 point inward toward the active site. Considering the Val→Ala mutation in DEBS module 6, the exchange to the smaller and more flexible residue opens up more space for a larger extender unit by allowing for rotation of the YASH motif.¹¹

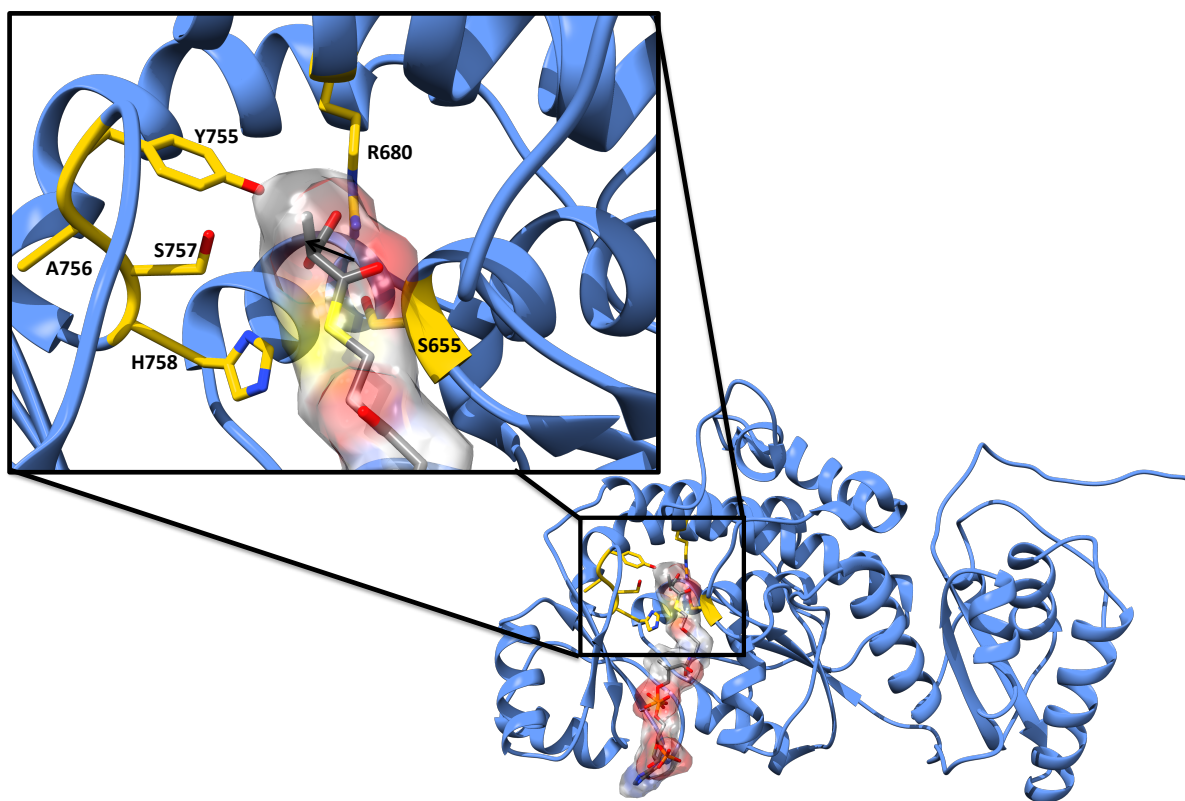


Figure 9 PikAIIIITE Acyltransferase homology model. This model was constructed using the I-tasser server and refinement methods⁵³. Methylmalonyl-CoA is docked relative to where it would need to be for a reaction to occur. Residues critical for the activity or selection of the enzyme and substrate are in gold.

By that same logic, there is a similar effect in PikAIII AT, but because the active site is tighter to begin with, the change in the active site flexibility is not as pronounced. The Tyr755Arg mutation replaces a large polar residue with a large, flexible, and positively charged residue. Arg residues are often found near solvent exposed regions on the surface of many enzymes but are not common in active sites. The flexible nature of the residue, coupled with its propensity for forming multiple hydrogen bonds has significant impact for protein structure.⁵⁴ Inclusion of an Arg residue in the active site of an enzyme will have strong impacts on the folding of the active site. The DEBS module 6 AT active site is robust enough to allow the mutation, and the structural changes caused by the residue open up the active site to allow for the formation of 92% propargyl 10-DML products. In PikAIII AT however, the

active site is too compact and the flexible charged residue destroys all activity by disrupting the active site completely.

2.2.2 PikAIII AT Docking Studies

To further understand the active site of PikAIII AT and its relationship to the AT of DEBS module 6, a molecular docking study was performed. A homology model was constructed based on published AT crystal structures PDB: 2QO3 and PDB: 2HG4.^{55,56} The model has a C score of 1.26 and a TM score of 0.89, consistent with literature values for high quality homology models.^{53,57} The substrate, methylmalonyl-CoA was docked into the model with AutoDockVina.⁵⁸ Due to the flexible nature of CoA, many conformations were unreasonable and screened out if they did not interact with the catalytic Ser655 residue. For the conformations that were more realistic, it was gratifyingly found that there are several previously reported residues that interact with the methylmalonyl-CoA sidechain (Arg680, Val753, and His758). Further, several other residues were found to be positioned near the active site but were not previously known to be important for substrate selection (**Figure 10**). Each of the new residues potentially influences the topography of the enzyme active site and were therefore considered important new targets for mutagenesis.

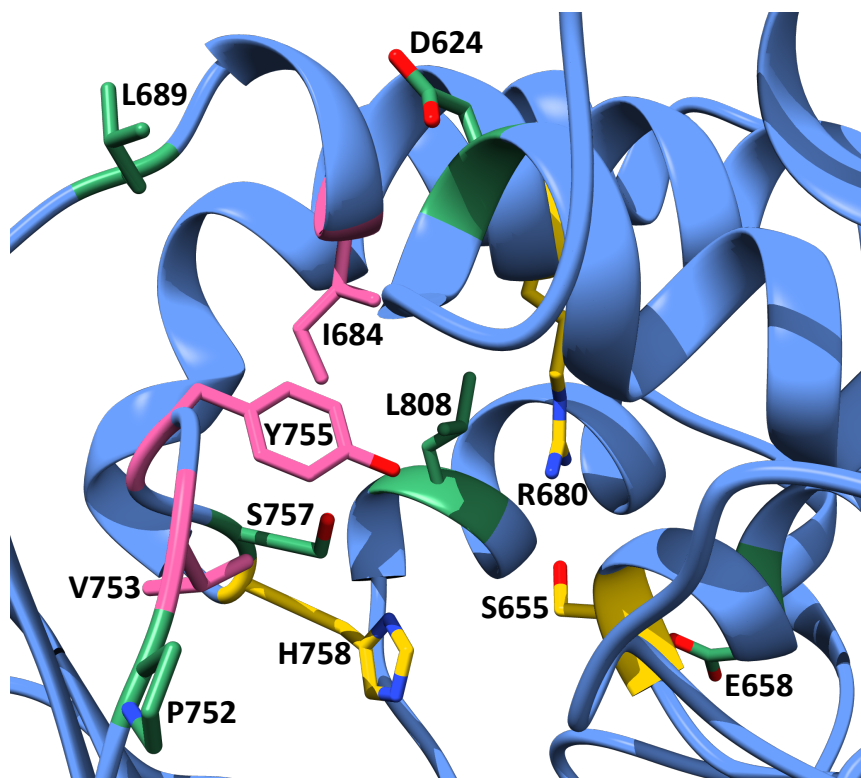


Figure 10 PikAIII AT homology model showing the sites selected for mutation. (A) CoA on view, as CoA would enter the active site. (B) Left side on view, Arg680 is the back of the active site. Gold residues are critical for enzyme activity and if mutated result in inactivity. Pink residues have been tested in the literature previously. Green residues are novel sites highlighted by rational design or docking.

Arg680 and is the residue responsible for positioning the methylmalonyl-CoA or derivative in the active site while the actual catalytic activity is between Ser655 and His758. Mutations at one of these sites will result in an inactive protein, and thus, they were not targeted. Sites Val753, Tyr755, and Ile685 have all been reported previously in other AT engineering literature. Val753 is the position of the Val→Ala mutation that has previously worked well in the DEBS module 6 AT and Tyr755Arg was the mutation that inverted the substrate selectivity for the DEBS AT, but found to cause PikAIII to be inactive. The position of ILE685 has been shown by Koryakina, et al. to be important for the topography of the active site and a mutation to a His residue abolished nearly all propargyl incorporation in DEBS module 6 AT.⁴¹

Other sites were selected based on their sequence alignment with the PikAIII AT and other ATs or by rational design. Leu808 in particular contributes to the lower back of the active site. Mutating Leu808 to a smaller residue will directly remove a barrier to the acceptance of the extender units with larger C2-sidechains (**Figure 10**). Leu689 is similar to Leu808, but its position more directly affects the top and back of the active site. Our hypothesis is that by mutating the larger residues to smaller residues, that have the same hydrophobic qualities, enzymatic activity will be conserved while substrate selectivity will change. One example of this hypothesis is the Val754Ala mutation, which has already been shown to effect substrate selectivity in DEBS without sacrificing activity.

There are several salt bridges that interconnect the helices of the active site Asp624: Asp625 and Glu658 are a part of two of them. Disrupting the salt bridge would change the relationship of the helix in respect to its position in the active site. Another structural consideration was the role of proline residues in the active site. Proline is known to assist in polypeptide backbone turning due to its constrained geometry. Removing or adding Pro residues should have large structural effects for the active site. Pro752 is four residues away from the YASH motif in an unstructured coil just after the final β sheet of the small subunit. Due to the geometric restrictions of the proline, changing it to a more flexible residue in this position may allow for larger substrates to be accepted by adding more flexibility to the loop of the YASH motif. Further, the area around Pro752 has no overall structure, therefore decreasing the rigidity of the area should not have too much of a destabilizing effect.

While the YASH motif of the AT has been known to help direct substrate selectivity, the effect of the flexibility of the Ser residue has not been explored. This residue is typically conserved across methylmalonyl (YASH) and ethylmalonyl (VASH) specific AT domains and is usually larger in malonyl (HAFH) specific domains. There are some ethylmalonyl specific AT domains that contain the motif VAGH; therefore, it was presumed that by shrinking the serine or adding more flexibility to the site would allow for larger substrates to be accepted. Additionally, a Ser→Gly mutation in DEBS module 6 AT was used as part of a

triple mutant that demonstrated increased propargylerythromycin A production than compared to the wild-type.¹¹

2.2.3 Development of a crude extract assay

A consideration for the pentaketide assay was how to provide the PKS enzyme. Purified enzyme, cell lysate, and reactions with intact cells were considered. While an *in vivo* reaction would give the best understanding of the cellular environment, it is cost prohibitive, due to the large quantities of pentaketide necessitated for feeding into *E. coli* or would require the use of the slow growing native host strain. Running the reaction as purified protein or cellular lysate is more cost effective and higher in throughput. Lysate is the most high-throughput of the three approaches considered, as it does not require lengthy purification or genomic editing of host strains. On the other hand, running the reaction in cellular lysate renders determination of protein concentrations more difficult. A potential downside to lysate reactions is the inclusion of cellular proteins that could lead to undesired substrate use, changing the concentration of substrates in the reaction, or protein inhibition.⁵⁹

Preliminary experiments revealed that PikAIIITE does not express well in the heterologous *E. coli* host and requires supplemental tRNAs provided by a secondary plasmid. The poor expression of PikAIIITE makes purification difficult as large culture volumes are needed to purify a reasonable amount of enzyme. However, reactions with purified enzyme are the standard for *in vitro* protein studies and is how the substrate specificity of DEBS has been quantified.⁴¹ Thus, it was necessary to compare the product distributions of the proteins in both lysate and as a purified enzyme to determine the equivalence of the two.

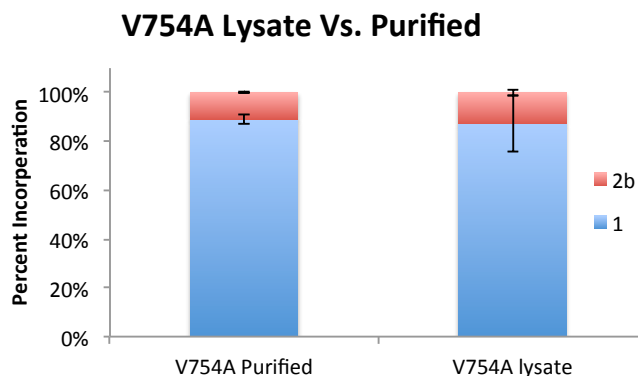


Figure 11 Comparison of the product distribution of purified enzyme vs. clarified cell lysate.

To test the product profiles supported by PikAIIIITE in crude extracts and as purified enzyme, a pentaketide competition assay was run with the PikAIIIITE Val753A mutant with methylmalonyl-CoA and propargylmalonyl-CoA in a 0.5:3 mM ratio. The mutant was used for this experiment because previous results indicate that it gives a reasonable quantity of 2b whereas the wild type enzyme yields <1% of 2b. There was very little difference in the distribution of the methyl and propargyl product when the crude lysate was used vs. the purified enzyme (**Figure 11**). Both methylmalonyl-CoA and propargylmalonyl-CoA are non-native to *E. coli*, thus it is unsurprising there are no reactions that would substantially differentiate the concentration of the two extender units. Additionally, once protein concentration in lysate reactions was quantified in the lysate reactions, on average, gave 1 fold higher activity than purified enzyme. The higher activity per microgram of enzyme is likely due to the lack of harsh purification conditions that could result in enzyme disruption. Due to the small change in the product distribution and the higher throughput of the lysate, cell lysates were used for experimentation unless otherwise noted.

2.2.4 Characterization of the Extender Unit Specificity of PikAIIIITE Mutants

When methylmalonyl-CoA and propargylmalonyl-CoA were run in equimolar competition at saturating concentrations with PikAIIIITE, propargylmalonyl 10-DML, **2b**, is formed as 0.4% of the total products (**Table 1**). The poor propargyl incorporation is unsurprising given the previous results where the total activity of the enzyme with the

propargyl extender unit demonstrated only 1% activity compared to the activity with methylmalonyl-CoA (**Figure 8**). The 0.4% served as a baseline threshold for the improvement of propargyl incorporation.

Table 1 Product distributions of PikAIIITE pentaketide assays. All results were run in duplicate and are the average of the 2 reactions. 1 is the expected methyl extension product and 2b is the expected propargyl extension product.

Percent Product Formation of PikAIIITE Mutants					
	1	2b		1	2b
WT	99.61	0.39±0.002			
WT	99.58	0.42±0.006	D624E	0.00	0.00
WT (apo)	0.00	0.00			
			E658D	0.00	0.00
V753A	98.35	1.65±0.043	E658V	100.00	0.00
V753I	100.00	0.00			
			S757P	99.73	0.27±0.376
I684L	100.00	0.00	S757A	99.83	0.17±0.014
I684V	99.68	0.32±0.006	S757G	100.00	0.00
I684G	98.84	1.16±0.031			
I684H	99.50	0.50	P752V	99.38	0.62
I684R	100.00	0.00			
			Y755V	68.23	31.77±0.534
L808A	98.02	1.98±1.857	Y755R	0.00	0.00
L808G	100.00	0.00			
L808S	99.81	0.19±0.049			
L808V	99.68	0.32±0.06			
L689V	99.34	0.66±0.314			

Most notably, the Tyr755Val mutation supported a 79-fold increase in the percent propargyl product, compared to wild-type PikAIIITE. Interestingly, Arg at this position in PikAIII resulted in zero conversion with either extender unit. In contrast, in DEBS module 6, an Arg at the equivalent position completely inverted extender unit selectivity. The site of this position in the PikAIII AT homology model (**Figure 10**) is directly in front of and pointed at the active site. Tyr755Val was tested specifically because of the commonly found sequences VASH and VAGH of ethyl specific ATs. While the entire YASH motif has

previously been implicated in substrate selectivity, mutations at the conserved Tyr residue clearly have a significant impact on the substrate selectivity.

Mutations at other residues resulted in less substantial changes in substrate selectivity. For example, introduction of Val754Ala in the PikAIIIITE led to an approximately 4-fold increase in the percentage of propargyl product produced. The effect of this equivalent mutation in DEBS was very similar.^{11,41} In contrast, the Val753Ile mutation completely abolished propargyl acceptance, presumably the larger volume of the Ile sidechain compared to that of Val cannot accommodate the unnatural extender unit. Additionally, Val753Ile causes the site to become slightly more rigid, while the Val753Ala mutation adds more flexibility.⁶⁰ While flexibility is a concern in these mutations, the Schultz group has tested the Val753Leu mutation in DEBS module 6 AT, adding similar flexibility to the Val→Ala mutation, and found no change in the amount of larger extender units accepted from the WT. Additionally, they have also tested a Val753Gly mutation, which did not give any products.³⁷ These data taken together indicates that at this site the size of the residue is critical. Smaller residues can allow for larger substrates to be accepted. Too much flexibility at the site leads to an inactive enzyme. Similarly, mutating Ile685, Leu808, and Leu689 to smaller residues increases the percentage of propargyl product produced. Interestingly, the Ile685Gly mutation caused nearly equivalent propargyl incorporation as Val753Ala, even though it is expected to provide flexibility, while Leu808Gly did not support propargyl incorporation at all. Determination of the overall activity of each enzyme revealed that the mutations that involved the most significant change in sidechain residue led to the largest decrease in enzyme activity (**Figure 12**).

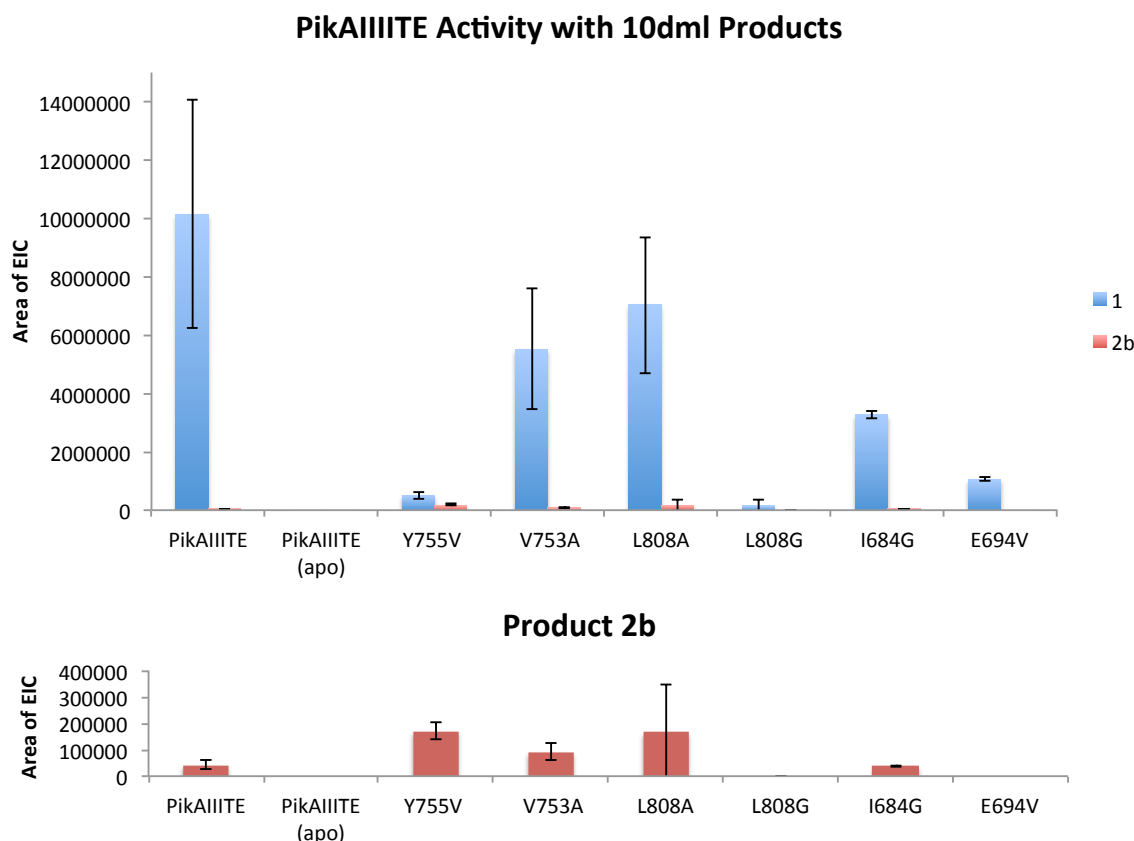


Figure 12 Activities of PikAIIITE and mutants relative to the wild-type enzyme.

For example, Leu808Ala and Val753Ala are quite minimal structural changes and are also the most active mutants. On the other hand, Leu808Gly or Ile685Gly, which involve more significant changes in side chain structure and flexibility, provided the least active enzymes. With the Ile685Gly mutation, there was an increase in the amount of propargyl relative to the methyl product; however, the total activity of the enzyme is low. In the case of the Ile685Gly mutation, it demonstrates that the increase in propargyl ratio is actually more due to a decrease in methyl activity.

Disruptions in salt bridges of the enzyme also have significant effects on the activity of the enzyme. As seen with the Asp624Glu and Glu695Asp mutations, while unsurprising, swapping one end of the bridge for a residue of the same charge but different lengths will result in disruption at the site, causing an inactive enzyme. More interestingly however,

removing the interaction with Glu658Val did not inactivate the enzyme. Rather, the activity was similar to that of the mutations that add more flexibility. However, there was no noticeable non-natural extender unit incorporation.

2.3 Coupling PikAIII with PikAIV

To further understand the effects of these mutations in the pikromycin pathway, PikAIII was coupled to PikAIV, the next enzyme in the pathway. The single module assay gives little information about how the mutations will affect the overall protein-protein communication between modules in the PKS. Additionally, the PikAIIITE system relies on a fusion of a TE, which could disrupt the native structure of the enzyme. Coupling PikAIII to PikAIV will show how the mutations affect the natural system and offer mechanistic insights into how the PKS is able to form two macrolactone products.

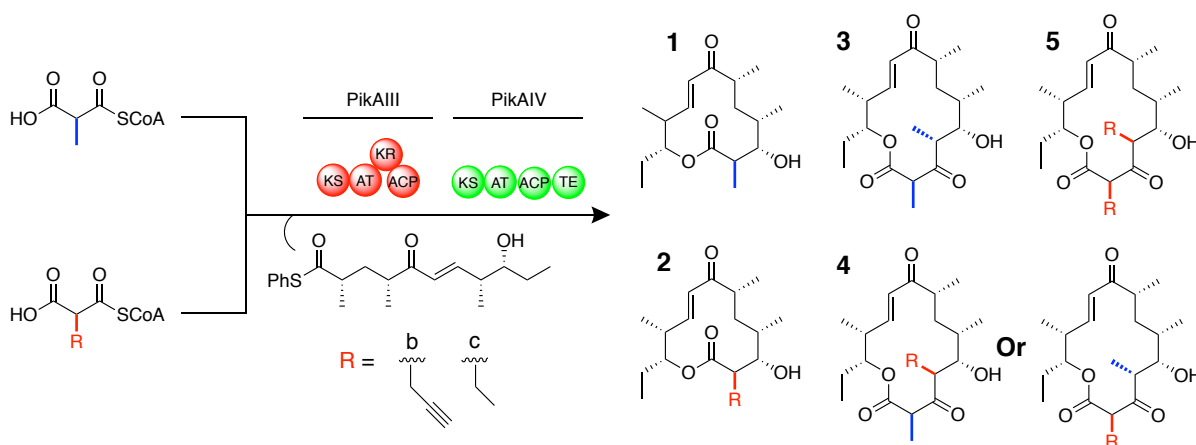


Figure 13 Scheme of the pentaketide assay with PikAIII coupled to PikAIV and the expected cyclization products.

2.3.1 Characterization of the Activity of PikAIII/PikAIV

Initial validation of the dual enzyme assay (**Figure 13**) confirmed that PikAIII without PikAIV will not produce detectable products **1** or **3**. This result was expected, as PikAIII does not have a TE domain and cannot cyclize the hexaketide chain. PikAIV alone will also not produce product when incubated with the pentaketide substrate, as the native PikAIV KS is selective for a hexaketide substrate. When run in competition, only six peaks

were initially expected; however, eight were found. The **1** and **2b** peaks retained the same elution time and profile as they did with PikAIIITE. The narbonolide products: **3**, **4b**, and **5b** each gave two peaks per product. The narbonolide products are mass degenerate and cannot be isolated. The two peaks of **3** were confirmed via a standard provided by Dr. Sherman and are thought to be different ring conformers of the same molecule (David Sherman, personal communication).

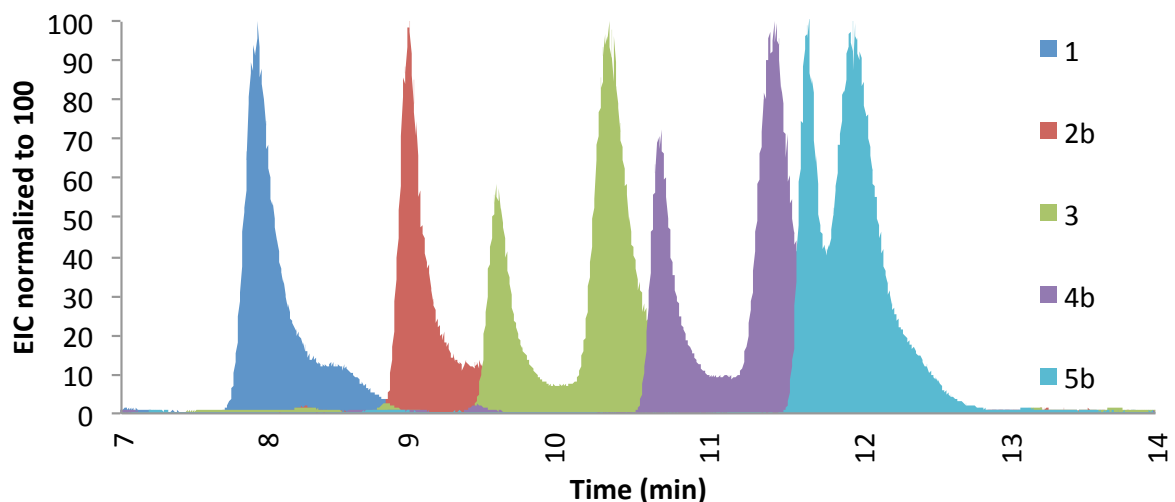


Figure 14 Representative chromatogram of an LC-MS sample of PikAIII with PikAIV. 10-DML products, **1** and **2b**, are detected first at 8 and 9 min respectively. Narbonolide products, **3-5**, are detected after and each product shows 2 peaks, each with the same mass. Chromatogram is not indicative of product ratios.

Next, the selectivity of the bimodular system was tested by including methylmalonyl-CoA in competition with one of two non-native extender units (**Figure 13**). The narbonolide products are produced in a higher quantity than the 10-DML products. While the products still significantly favor methyl products over the propargyl derivative, the relative amounts of propargyl derivative are higher than expected when compared to the amounts found with PikAIIITE alone.

The percentage of the propargyl 10-DML analogue **2b** is 0.1% of all the 10-DML and narbonolide products. When only the 10-DML products are considered the percentage rises to 1.0% (See **Figure 15** and **Table 2**). The same ratio holds for the competition with of

ethylmalonyl-CoA, which gave more ethyl products, 0.05% **2c**, 38% **4c**, and 1.31% **5c**, than the results from PikAIIIITE would have predicted. The large increase in **3**, **4b**, and **5b** likely indicates that PikAIV is more promiscuous than PikAIII.

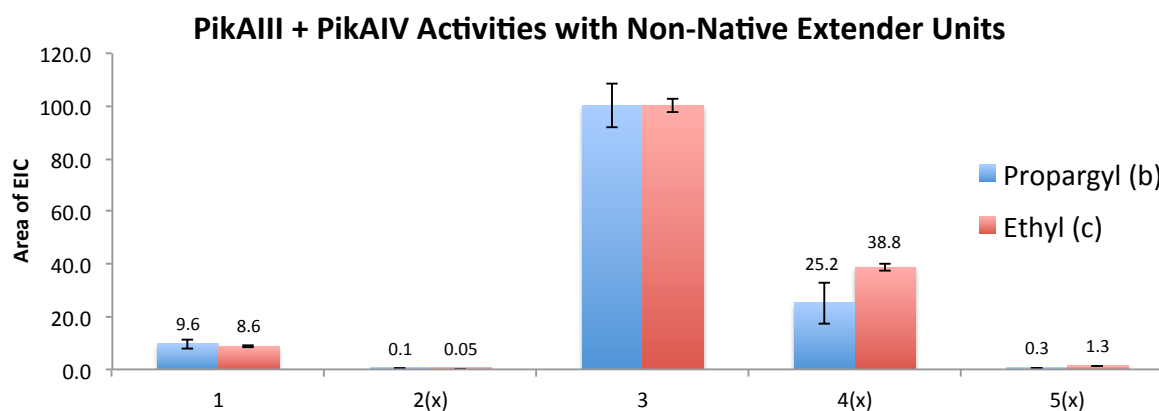


Figure 15 Product distribution of PikAIII + PikAIV reactions. Methyl vs. X. The 2 peaks from the LC are combined as narbonolide **3** as they are representative of the same product. Narbonolide **3** has been set to 100 and all other peaks normalized accordingly. Activity respective to **3** is labeled on top of each peak.

The PikAIII AT and PikAIV AT share around 85% homology and were originally expected to display similar promiscuity. By comparing the product distributions and assuming the formation of **2b** is indicative of PikAIII selectivity, it is possible to determine the ratios of the two possible extensions (**Figure 16**) by comparing the conditional probability of the two dependent events. The distribution found (**Figure 15**) is indicative that PikAIII is actually allowing for the formation of anywhere between 1%-10% propargyl products while PikAIV is allowing for the formation of 13%-22% propargyl products. These ranges are based on the assumptions that the AT is the only selective filter in either enzyme, and that PikAIV is more promiscuous than PikAIII. Therefore, the data from PikAIII shows that its lower bound is ~1% and its upper bound must be less than half of the total percentage of propargyl products.

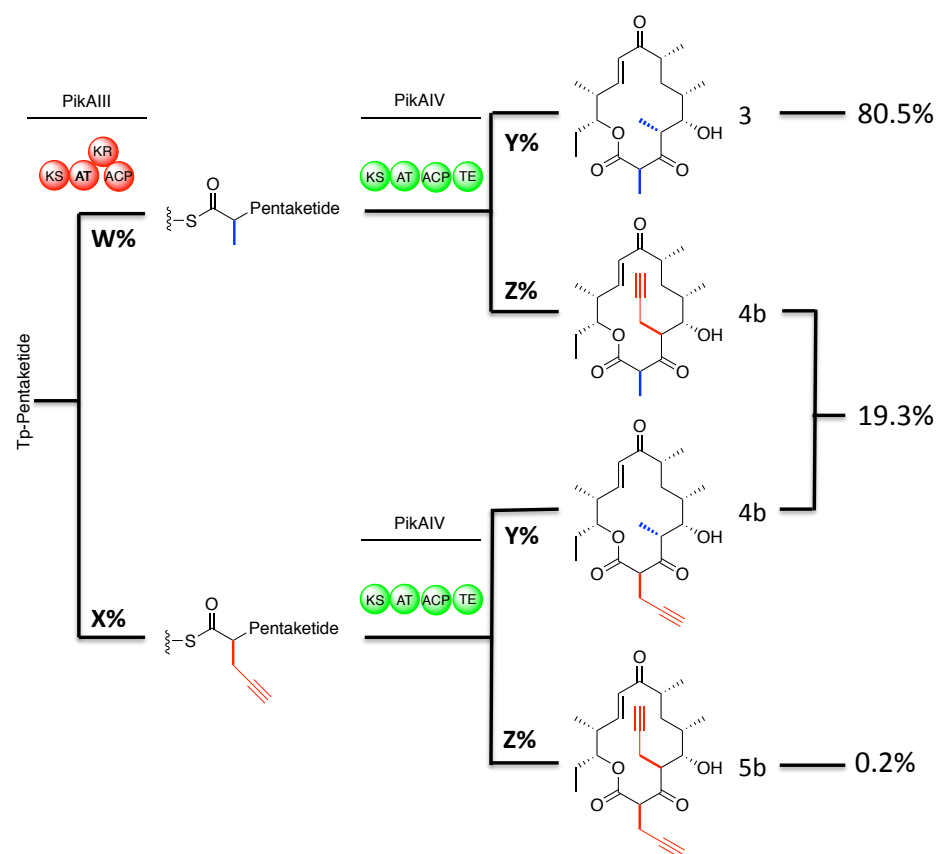


Figure 16 Overview of the products from the dual enzyme assay. **PikAIII** accepts the **Tp-pentaketide** substrate and selects between the two possible extension units, **methylmalonyl-CoA** and **propargylmalonyl-CoA** in ratio **W:X** respectively. The methyl or propargyl hexaketide product is then cyclized to **10-DML** (not shown) or accepted by **PikAIV**, which also selects between methyl and propargyl in ratio **Y:Z** respectively. After extension by **PikAIV** the cyclized narbonolide products **1**, **2b**, **3**, **4b**, and **5b** are detected. The percentages of the narbonolide products are shown.

2.3.2 Dissecting the Role of PikAIII and PikAIV in Determining Substrate Selectivity

To determine whether the production of propargyl products was dependent on both PikAIII and PikAIV, the ability to provide active (holo) and inactive (apo) forms of the PKS modules was leveraged. To this end, both holo and apo forms of each module were prepared as purified enzymes and incubated in various combinations in the presence of methylmalonyl-CoA and propargylmalonyl-CoA.

In the presence of holo-PikAIII (active) and apo-PikAIV (inactive), entry 3, the majority of the detected products were the 10-DML macrolactones **1** or **2b** (Table 2), which agrees with the previous data from PikAIIIITE. Some amount of narbonolide products were also detected, albeit at low levels, presumably as a result of the CoA-linked extender directly loading onto the KS of PikAIV. Iterative chain elongation, where one module is responsible for two or more elongations, is another possibility, however unlikely, for the formation of narbonolide products seen in entry 3. Evidence of iteration by PikAIII has been previously demonstrated through the production of triketide lactones when the enzyme is incubated with methylmalonyl-CoA alone.⁴² However, if iteration were occurring then narbonolide products would be detected in entry 7. Additionally, given that the ACP in apo-PikAIV is not functional, this result indicates that the ACP of PikAIV is not required for TE-catalyzed macrocyclization of the linear intermediates. While the ACP of PikAIII can pass the newly-formed hexaketide chain to the KS of PikAIV it is not clear how the chain will then be passed to the TE of PikAIV without the functional ACP. However, it has been proposed that PikAIV may adopt a conformation that allows for the transfer of the chain from PikAIII ACP to the TE directly.⁴⁷

Table 2 Product percentages of the extracted ion chromatograms of PikAIII + PikAIV reactions. Entries with no error are indicative of error less than 0.005. First number is the area of the EIC normalized to Entry 1 product 3 set as 100. Product percentages are in parentheses; 10-DML (*) products are considered separately from narbonolide products ().**

PikAIII + PikAIV Codependance and AT Swaps							
Entry	10dml		3	Narbonolide		Total Activity	
	1	2b*		4b**	5b**		
<div><div>PikAIII</div><div><div><div>KR</div><div>KS</div><div>AT</div><div>ACP</div></div></div></div> <div><div>PikAIV</div><div><div><div>KS</div><div>AT</div><div>ACP</div><div>TE</div></div></div></div>	1	9.05±0.09	0.16±0.07 (1.3)	100±0.03	17.28 (19.3)	0.2±0.18 (0.2)	126.69
<div><div>PikAIII</div><div><div><div>KR</div><div>KS</div><div>AT</div><div>ACP</div></div></div></div> <div><div>PikAIV</div><div><div><div>KS</div><div>AT</div><div>ACP</div><div>TE</div></div><div><div>X</div></div></div></div>	2	0.85±0.02	0.01±0.04 (1.1)	9.43±0.02	2.33 (19.8)	0.01±0.29 (0.1)	12.64
<div><div>PikAIII</div><div><div><div>KR</div><div>KS</div><div>AT</div><div>ACP</div></div><div><div>X</div></div></div></div> <div><div>PikAIV</div><div><div><div>KS</div><div>AT</div><div>ACP</div><div>TE</div></div></div></div>	3	32.66±0.02	0.37 (1.1)	0.55±0.05	0.13±0.03 (19.1)	0.00	33.71
<div><div>PikAIII</div><div><div><div>KR</div><div>KS</div><div>AT</div><div>ACP</div></div></div></div> <div><div><div>PikAIV</div><div><div><div>KS</div><div>AT</div><div>ACP</div><div>TE</div></div></div></div><div><div>2x</div></div></div>	4	8.70	0.1 (1.2)	95.16±0.04	23.3±0.05 (19.6)	0.25±0.03 (0.2)	127.52
<div><div>PikAIII</div><div><div><div>KR</div><div>KS</div><div>AT</div><div>ACP</div></div></div></div> <div><div>PikAIV</div><div><div><div>KS</div><div>AT</div><div>ACP</div><div>TE</div></div></div></div>	5	4.53±0.09	0.08±0.01 (1.8)	36.95±0.05	18.46±0.15 (33.2)	0.2±0.05 (0.4)	60.22
<div><div>PikAIII</div><div><div><div>KR</div><div>KS</div><div>AT</div><div>ACP</div></div></div></div> <div><div>PikAIV</div><div><div><div>KS</div><div>AT</div><div>ACP</div><div>TE</div></div></div></div>	6	1.19±0.03	0.06±0.21 (4.9)	8.67±0.08	1.41±0.09 (13.9)	0.05±0.12 (0.5)	11.39
<div><div>PikAIII</div><div><div><div>KR</div><div>KS</div><div>AT</div><div>ACP</div></div></div></div>	7	0.00	0.00	0.00	0.00	0.00	0.00
<div><div>PikAIV</div><div><div><div>KS</div><div>AT</div><div>ACP</div><div>TE</div></div></div></div>	8	0.00	0.00	0.00	0.00	0.00	0.00

When PikAIV holo is incubated with PikAIII apo, entry 2, a lower overall level of product is observed. Low activity is expected, as PikAIV will not take the pentaketide substrate and without PikAIII's ACP, the reaction is stalled, possibly, by the rate of PikAIII's extension, subsequent hydrolysis from the KS, and loading onto PikAIV's KS from solution.

Another possibility is the ACP of PikAIV may adopt a conformation where it can supplant the role of the missing ACP, reminiscent of the proposed transfer mechanism for 10dml formation.⁴⁷ Next, to examine whether PikAIV was limiting the incorporation of the unnatural extender unit, both holo enzymes were included but PikAIV was provided at twice the concentration of PikAIII, entry 4. The amount of narbonolide products (**3-5**) remained the same as when the modules are provided at 1:1, which shows that a higher concentration of PikAIV will not change the activity or product distribution under these conditions. The higher percentage of propargyl narbonolide products compared to propargyl 10-DML indicates that the specificity of PikAIV is different than that of PikAIII. To achieve the ratios seen in Table 2, while assuming PikAIII only accepts 1% propargyl, would require PikAIV to accept around 20% propargyl (**Figure 16**).

To further characterize the specificity of PikAIV, the AT domains were swapped from PikAIII and PikAIV using the boundaries defined by the Keasling group in the KS-AT linker and post AT linker.²⁷ When the PikAIV AT was put in place of the PikAIII AT (P3→P4AT, entry 5), there was only a very small increase in the percentage of **2** found in the product mixture. However, the increase is much lower than the 20% **4** seen from PikAIV WT, entry 1. The lower than the expected 20% **2b** formation suggests that something beyond the AT in PikAIII is acting as a means of selection against the non-natural product. Another point of interest from this reaction is the 13% increase in **4b** product but only a 0.2% increase in the amount of **5b**. It would be expected that if the amount of PikAIV controlled extensions with propargyl (**4b** and **5b**) increased overall, then we should see a similar increase in the amount of **5b** compared to the increase in **4b**. The lack of observed increase in **5b** indicates that something is also selecting against the dipropargyl derivative.

When the PikAIII AT was swapped into PikAIV (P4→P3AT, entry 6), a large decrease in total activity and relative production of **4b** was observed. The decrease in ratio is indicative that the PikAIII AT is less accepting of the propargyl substrate. However, there is a 3.7% increase in **2b** with this AT swap and only a 6% decrease in **4b** relative to the product distribution of entry 1. While the PikAIIIITE data suggested that PikAIII AT would only

allow the formation of 1% **2b**, the 13% of **4b** found here indicates the PikAIII AT will allow a similar product ratio to PikAIV AT when under similar conditions.

Together, these data demonstrate that PikAIV is the more promiscuous enzyme of the two in the PikAIII/PikAIV assay. It also sets the expected baseline for product ratios when comparing mutant enzymes. In the dual enzyme system, the data shows that PikAIII will produce 1-10% propargyl product while PikAIV will produce 13-22% propargyl product. The total propargyl acceptance of PikAIII's AT is likely higher than 1% under different conditions as demonstrated by the AT swap into PikAIV. The difference in PikAIII AT specificity when under different conditions indicates that another factor is affecting the enzyme's promiscuity beyond the AT domain. Further, the only 0.6% increase in PikAIV AT swap into PikAIII continues to support this conclusion. Several other domains, including a KS, KR, and TE, act on the chain in the PikAIII and PikAIV modules and could be acting as a means of selection. Another likely possibility is that PikAIV is the final module of the PKS; therefore, after extension, the chain will not interact with another KS domain. From an evolutionary perspective, the final extension would be the easiest to allow for differentiation, as the rest of the pathway would never see the difference.

2.3.3 Characterization of PikAIII and PikAIV Mutants

Given that PikAIV was found to be more tolerant of unnatural extender units than PikAIII, the most promising mutations described from the PikAIIIITE assays were introduced into PikAIII and tested in the context of the bimodular system. This would serve as a robust test of the ability of amino acid mutations to overcome the stringent specificity of the selectivity filter(s) in PikAIII. None of the mutant enzymes exhibited the same total activity as wild-type (**Table 3**). Notably though, several mutants did display a significant increase in the amount of propargyl products produced, as compared to the wild-type system.

Table 3 10-DML and narbonolide product distributions of PikAIII + PikAIV competition assays. First number is the area of the EIC normalized to 3b of the wild type enzyme run with each mutant set to 100. Product percentages are in parentheses; 10-DML (*) products are considered separately from narbonolide products (). Error bars of 0 indicate error less than 0.005.**

PikAIII + PikAIV Mutant Activity				
Mutant	Entry	10dml		Total Activity of all Products
		1	2b*	
PikAIII + PIV	1	9.65±1.74	0.12±0.03 (1.2)	135.29
PikAIII Y755V + PIV	2	4.69±0.4	8.53±0.19 (64.5)	111.13
PikAIII V753A + PIV	3	7.74±0.3	0.5±0.03 (6.1)	125.19
PikAIII P762V + PIV	4	3.27±0.01	0.03±0 (1)	42.95
PikAIII CPSH + PIV	5	4.12±0.07	0.04±0 (1)	49.20
PikAIII L808A + PIV	6	0.95±1.08	0.01±0.02 (1)	5.00
PikAIII L808V + PIV	7	1.56±1.02	0.01±0.01 (0.6)	19.56
PikAIII D624E + PIV	8	4.32±3.32	0.06±0.04 (1.4)	67.14
PikAIII E658D + PIV	9	0.04±0.01	0±0 (0)	0.56
PikAIII + PIV Y753V	10	3.68±0.29	0.11±0.01 (2.9)	11.75
PikAIII (apo) +PIV (apo)	11	0±0	0±0	0

Mutant	Entry	Narbonolide		
		3	4b**	5b**
PikAIII + PIV	1	100±8.2	25.23±7.86 (20.1)	0.29±0.14 (0.2)
PikAIII Y755V + PIV	2	16.45±0.22	63.17±2.04 (64.5)	18.29±0.96 (18.7)
PikAIII V753A + PIV	3	95.14±2.68	21.14±0.84 (18.1)	0.67±0.01 (0.6)
PikAIII P762V + PIV	4	29.31±0.87	10.26±0.24 (25.9)	0.08±0 (0.2)
PikAIII CPSH + PIV	5	34.76±0.47	10.18±0.21 (22.6)	0.1±0.01 (0.2)
PikAIII L808A + PIV	6	3.13±1.88	0.91±0.56 (22.5)	0±0 (0)
PikAIII L808V + PIV	7	13.27±9.19	4.68±3.25 (26)	0.04±0.03 (0.2)
PikAIII D624E + PIV	8	51.2±42.67	11.47±11.82 (18.3)	0.09±0.08 (0.1)
PikAIII E658D + PIV	9	0.42±0.16	0.1±0.05 (19.2)	0±0 (0)
PikAIII + PIV Y753V	10	1.98±0.09	5.89±0.07 (74)	0.09±0.01 (1.1)
PikAIII (apo) +PIV (apo)	11	0±0	0±0	0±0

All of the enzymes tested saw a minimum increase of 2-3 fold in propargyl incorporation in the coupled system compared to PikAIIIITE (**Table 3**). Mutant Tyr755Val (entry 2) in PikAIII demonstrates the largest shift in product distributions. For example, the fraction of **2b** is ~64% of the total 10-DML products, whereas the wild-type system only

produces 1% **2b**. Notably, the fraction of propargyl narbonolides supported by this mutant are also increased, such that 64% of the narbonolide product is mono-propargyl **4b** and 19% is dipropargyl **5b**. In summary, the product distribution of this mutant is completely inverted towards production of propargyl macrolactones. The Tyr755Val mutation was also applied to PikAIV (Tyr753Val, entry 10) where a 3.8-fold increase in the percentage of **4b** was observed. (**Table 3**) This increase agrees with the previous data and demonstrates that the mutation can, at a minimum, be applied to multiple modules within the same PKS.

In addition to determining the percentage distribution of products, the activity of each enzyme was also compared by plotting the total ion counts for each product. In this regard, PikAIII Tyr755Val clearly outshines wild-type and every other mutant. With this mutant, there is a 71-fold increase in the amount of **2b**, a 2.6-fold increase in **4**, and a 70.0-fold increase in **5b** production. Thus, the increased fraction of propargyl products is a result of improved conversion to the propargyl products rather than just a reduction in conversion to the natural methyl products. Remarkably, Tyr755Val produces the monopropargyl- and dipropargyl narbonolides in quantities that are 63% and 18% that of the narbonolide production by the wild-type enzyme.

Several mutants, YASH→CPSH (entry 5) and Asp624Glu (entry 8) in particular, are indicative of activity changes based on the removal of the TE from PikAIII and the addition of PikAIV. This result is a strong argument for testing mutations in as near to the natural system as possible. It is also important to note that mutations in unstructured coils of the protein had less effect on the overall activity of the enzyme. Mutations in β sheets or helices, however, typically had large decreases in activity such as the Leu808 and Glu658 mutations. Further, it appears that mutations toward the front of the active site exhibit the best ability to flip the substrate selectivity.

2.11 Conclusions

Rational design of PKS AT domains is a powerful technique that has demonstrated substantial control over AT substrate selectivity. Previously, the AT of DEBSmod6 has been mutated to accept the larger substrate propargylmalonyl-CoA over its native methylmalonyl-

CoA.⁴¹. In an attempt to translate the DEBS mutations, PikAIII AT was chosen for study as it shares the highest homology with DEBS module 6 AT. Interestingly, it was found that the PikAIII AT was more substrate-selective than its DEBS counterpart. Unlike DEBS, the PIKAIII AT allowed for fewer total substrates to be accepted and lower percentages of those it did accept. Understanding the selectivity of the PikAIII AT active site provided a rationale as to why the PikAIII AT Tyr755Arg mutation, translated from DEBS, did not work. As the PikAIII AT active site is more selective than DEBS module 6 AT, the PikAIII AT site must have less free space for extender units. Then the addition of a large, charged residue pointing into the active site is unreasonable as the residue would have too many disrupting interactions with the site for the active site to retain activity.

Docking studies were performed alongside sequence alignments to compare the ATs of different PKSs. Several sites were found that could potentially lead to opening up the active site for larger substrates. These sites were explored via site-directed mutagenesis and a pentaketide chain mimic assay. The protein fusion PikAIIITE was used as an initial screen for testing AT substrate selectivity. Mutations Leu808Ala, Val753Ala, and Tyr755Val were found to have the most significant control of the product ratios. Tyr755Val was of particular interest, as it is the only mutation that demonstrates a significant shift in the substrate selectivity. While Tyr755Val has the highest activity with propargyl, it also demonstrates the lowest total activity in PikAIIITE. The low total activity of Tyr755, however, may not be an issue when moving forward due to the higher selectivity and higher activity with propargyl compared to the wild-type enzyme.

To further test the effects of the mutations on an *in vitro* PKS, a coupled assay was developed from the original pentaketide assay. This lysate-based assay coupled PikAIII with PikAIV, the terminal module in the native PKS. While it was first thought that PikAIII and PikAIV would have the same substrate selectivity, this was quickly found to not be the case. Analysis of the product ratios revealed that while PikAIII was allowing for 1%-10% of non-natural extender units, and PikAIV was allowing 13%-22% non-natural extender units. This result suggests that either the ~15% difference in homology between PikAIII and PikAIV

contribute to a large specificity shift or indicates the final module of a PKS is likely the most promiscuous. Promiscuity of the final module is supported by the data from DEBS module 6 AT, as it is the final module of its PKS; however, there has not been full testing of the substrate selectivity of every module in either PKS.

The best performing mutants from the PikAIIIITE assay were translated to the PikAIII + PikAIV coupled assay. All of the non-natural product profiles increased when used in this assay. The increase seen here may be caused by a selectivity shift in PikAIII based on an unknown structural change. As PikAIII alone will not produce a detectable product with this assay, it is difficult to know if the change is from the removal of the TE or the addition of PikAIV.

The mutation Tyr755Val demonstrated the largest change in the product ratios. This mutant was found to favor the production of the non-natural propargyl product over the methyl. Beyond the product ratios, this mutant produced 2.6-fold more total **4b** than the wild-type. Interestingly, this mutation is at the same site as the Tyr→Arg mutation in DEBS module 6 AT that also flipped the substrate selectivity. Together this indicates that while the individual mutations may not translate from one system to another, the sites for controlling the substrate selectivity remain the same. Thus, Tyr755, Val753, and Leu808 constitute at least three AT residues that partially control selectivity for larger extender units.

Methods

General

Materials and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Isopropyl β -D-thiogalactoside (IPTG) was purchased from Calbiochem (Gibbstown, NJ). The pRARE plasmid, the tp-pentaketide, the PikAIII construct, the PikAIIITE construct, and the PikAIV construct were provided by Dr. David Sherman at the University of Michigan. The BAP1 *E. coli* strain was provided by Dr. Blaine Pfeifer at the University at Buffalo. Primers were purchased from Integrated DNA Technologies (Coralville, IA). All holo proteins were expressed in BAP1 cells and all apo proteins were expressed in BL21 *E. coli* cells.

Mutagenesis of PikAIII and PikAIV

According to the manufactures instructions, Phusion polymerase (New England Biolabs, NEB, Ipswich, MA) was used for round the horn mutagenesis of all PikAIII and PikAIV mutants. Primer sequences can be found in the appendix. AT swaps of PikAIII AT and PikAIV AT were constructed using Gibson assembly (NEB) and boundaries defined by the Keasling group, primers in appendix.²⁷ All DNA constructs were sequenced by Genewiz (Research Triangle Park, NC) using oligonucleotides, in appendix. Successful mutants were transformed into BAP1 competent cells along with the pRARE plasmid then plated onto LB agar plates along with 100 μ g/mL kanamycin and 50 μ g /ml spectinomycin. Plates were incubated overnight at 37°C then stored at 4°C. Individual colonies were inoculated into 4mL LB media along with 100 μ g/mL kanamycin and 50 μ g /ml spectinomycin and incubated with shaking at 37°C overnight. After growth a 0.5mL sample of the culture was taken and added to 0.5mL of 50 % glycerol solution, then stored as a glycerol stock at -80°C.

Protein Production

5mL of autoclaved LB media along with 5 μ L of the appropriate antibiotic (50 μ g/mL kanamycin for modules, 100 μ g/mL spectinomycin for pRARE) was inoculated from a glycerol stock then allowed to grow overnight at 37°C while shaking. The next day 4mL of the starter culture was added to a 1L flask containing 300ml of LB media along with 300 μ L of the appropriate antibiotic. The culture was grown at 37°C shaking at 250rpm until optical density at 600nm (OD600) reached 0.6. The culture was then induced with 300 μ L of 1M IPTG and allowed to express protein overnight at 16°C.

Protein Purification

After overnight expression, the culture was spun down at 4700rpm for 20min and the media discarded. The cells were re-suspended in ~10ml wash buffer (50mM monobasic sodium phosphate, 300mM sodium chloride, 20mM imidazole, pH 7.4) then sonicated using 51% amplitude, 10sec on, 20sec off for 10min. After sonication the lysed cells were spun down at high speed at 18000rpm for 1hr, then the supernatant was collected. MatB Proteins were then purified using the Bio-Rad Profinia purification system with a native IMAC method, 5ml cartridge and low flow rate. Module proteins were purified with 1mL HisTrap HP columns on a biologic duoflow FPLC (Bio-Rad, Hercules, CA) with wash buffer and elution buffer (same as wash buffer except 250mM imidazole). The gradient used 0% elution from 0-22min, then 0 \rightarrow 100% elution from 22-27min. 100% elution was held for 3 min then returned to 0% elution for 5min. 2mL fractions were collected throughout the time elution buffer was used. After purification, proteins were spun down in molecular weight cutoff filters (Millipore Corp. Billerica, MA) to concentrate and buffer exchange the proteins into storage buffer. Proteins were stored in MatB storage buffer (50mM Tris-hydrochloride, 100mM sodium chloride, 10% glycerol, pH 7.4) at -80°C.

Lysate Preparation

After overnight expression the culture was spun down at 4700rpm for 20min and the media discarded. The cells were re-suspended in ~1ml module storage buffer (100mM sodium phosphate, 1mM ethylenediaminetetraacetate (EDTA), 1mM tris(2-carboxyethyl)phosphine (TCEP), 20% glycerol, 0.1μL Benzonase (NEB), pH 7.4) and sonicated using 51% amplitude, 10sec on, 20sec off for 5min. After sonication, the lysed cells were spun down at high speed at 18000rpm for 1hr, then the supernatant was collected and stored at -80°C as 120μl aliquots.

MatB Reactions and Acyl-CoA preparation

MatB reactions were set up with 120μL of 100mM ATP, 80μl of 100mM CoA, 160μl of the respective malonic acid, 30μl of MatB enzyme (approximately 10-50μg), 100μl of 10x MatB buffer (1M sodium phosphate, 20mM magnesium chloride, pH 7.0), and 510μl of water. Reactions were run overnight, and then a 50μL sample was taken and quenched with 50μL ice-cold methanol. The sample clarified by centrifugation at max speed (13,300 rpm) for 10min, and then analyzed by HPLC (Varian Prostar, Walnut Creek, CA) at 254nm on a Phenomenex Kinetex 5u c18 100A 250x4.6mm column with 0.1% trifluoroacetic acid in water (solvent A) and methanol (solvent B). Method was run over 30min 0-3min 100% A, then from 3min- 20min an gradient was run to 60%B solvent B. After the 20min mark, 100% solvent B was run for 5 min, followed by 5 min of 0% solvent B. Reactions were stored, unquenched, at -20°C.

Pentaketide Assay

The pentaketide assay was set up to a total volume of 80μL with 17.6μl of water, 8μl of the 10x MatB buffer, 1.6μl of the (50mM stock) TP-pentaketide, 35μl of the 8mM malonyl-CoA derivative from a MatB reaction, 4μl of 100mM glucose 6 phosphate, 0.8μl of 50mM NADP⁺, 3.2μl of 0.2U/ml of glucose 6 phosphate dehydrogenase, and 25ng of

PikAIIIITE enzyme. Reactions proceeded at RT for 16-18hrs then quenched with an equal volume of ice-cold methanol. After quenching, all reactions were centrifuged at max speed (13,300 rpm) for 10min, and the supernatant was collected and centrifuged again. Once no more precipitate was seen in the supernatant, the supernatant was filtered through a nylon 0.2 micron filter. The filtrate was allowed to sit at RT for 1hr, then centrifuged again. The clarified reactions were then analyzed.

Analysis was carried out on a high resolution mass spectrometer – a Thermo Fisher Scientific Exactive Plus MS, a benchtop full-scan Orbitrap™ mass spectrometer – using Heated Electrospray Ionization (HESI). The sample was analyzed via LCMS injection into the mass spectrometer at a flow rate of 225 μ L/min. The mobile phase B was acetonitrile with 0.1% formic acid and mobile phase A was water with 0.1% formic acid, see appendix for gradient, scan parameters, and product ions. The mass spectrometer was operated in positive ion mode. The LC column was a Thermo Hypersil Gold 50 x 2.1 mm, 1.9 μ particle size.

The dual enzyme competition assay was run with the same substrate concentrations as the pentaketide assay except no water or purified enzyme was provided. The enzyme lysates for both PikAIII and PikAIV were provided in equal volumes up to a total reaction volume of 80 μ L. The concentration of protein in each reaction was obtained by assessing the total protein concentration of the lysate via a Bradford assay kit. Then the lysate was run on an SDS-PAGE gel (4% stacking, 12% running) where the gel image was processed used to determine the percent PikAIII protein from the total protein concentration. This assay produces six final products that can be seen as their $[M+H]^+$, $[M+H-H_2O]^+$, or $[M+Na]^+$ ions. For quantification purposes, only the $[M+H-H_2O]^+$ was followed as it gave the cleanest and most consistent peaks. Error was calculated as the standard deviation of the average of the extracted ion chromatogram of duplicate samples or as the standard deviation of the product percentages.

CHAPTER 3

3.1 The Future of Pikromycin Engineering

Rational design of the pikromycin PKS has proven to be a powerful tool for controlling the selectivity of singular PKS modules. The Tyr755V mutation has provided a foothold for the total swap of substrate selectivity from methylmalonyl-CoA to propargylmalonyl-CoA. Coupling this mutation with other sites, such as Val753 or Leu808, may provide a total flip in selectivity without sacrificing much activity. Additional testing of the AT active site may also reveal other sites that can similarly affect product distributions.

The sites and mutations found to control substrate selectivity in the pikromycin PKS also need to be applied to other PKSs. As was seen with the translation of DEBS mutants to PikAIII, not all mutations will continue to have the expected effect. Therefore, the Tyr or its cognate residue in other PKS's YASH motif should be tested via saturation mutagenesis. To fully control the selectivity imparted by this site, PKSs that are not methylmalonyl-CoA specific should be chosen for future testing.

As PikAIV appears significantly more promiscuous than PikAIII, testing the AT of PikAIV directly is of interest. The KS of PikAIV will recognize a hexaketide chain substrate. The substrate can be accessed through chemical synthesis or via chromatography of PikAIII pentaketide reactions. Testing PikAIV can also be accomplished with the pentaketide assay, as the high specificity of PikAIV will rule out the products from incorporation by PikAIII. The interesting result of PikAIV's promiscuity should also be further explored through mutagenesis, to further shift the selectivity.

When the AT of PikAIV was put into PikAIII, the increase in propargyl incorporation was not near the 13-22% incorporation that was found for PikAIV. This result indicates that there is something beyond the AT that is limiting the promiscuity. To further test the ATs of PikAIII and PikAIV should be swapped into another PKS module, unrelated to pikromycin biosynthesis, as a way to test the ATs in identical environments. Along the same lines,

replacing the TE of the pikromycin PKS with the TE of another PKS will determine if the TE is acting as a point of selection for the final macrolide beyond the promiscuity of the AT.

Each PKS module is a large enzyme with many residues of interest, so increasing the throughput of testing would be a valuable achievement. Biosensor systems, such as the repressor protein MphR coupled with green fluorescent protein, have begun to be developed for macrolide products such as erythromycin.^{61,62} The repressor proteins can be engineered to recognize a new substrate, which directs each biosensor system for the detection of the new substrate. Construction of a biosensor system to report the production of a untailored PKS product would allow for high throughput *in vivo* or *in vitro* screening.⁶³ In turn, high throughput screening would allow for further testing of rational PKS mutagenic libraries or directed evolution approaches to engineering.^{64,65}

3.2 *In Vivo* Polyketide Engineering

While *in vitro* assays are a good test of the enzymes in a PKS, they are not fully indicative of the cellular environment. The pikromycin PKS has not been fully reconstituted in *E. coli*. On the other hand, DEBS is fully reconstituted in *E. coli* and is produced across several host organisms.⁶⁶⁻⁶⁸ Thus, DEBS serves as the ideal progenitor for the development of an efficient, multi-host PKS editing platform. The *E. coli* expression of erythromycin A utilizes either a five plasmid system, with large plasmids of 25kB, or a two bacterial artificial chromosomes (BAC), each of around 50kB^{66,69}.

This system will need to install mutations on plasmids or BACs up to 50kB in size. Site-directed mutagenesis and round-the-horn mutagenesis require polymerases that can amplify the full plasmid. Unfortunately, no commercial polymerases can reliably amplify up to that size. There are several possible strategies to manipulate DNA constructs of large sizes. However, genomic manipulation of PKS genes is also of interest for *in vivo* production. Therefore, in the interest of having both genomic manipulation and plasmid manipulation available, recombination and CRISPR are some of the best tools available to reliably handle both.^{70,71}

CRISPR, or clustered regularly spaced palindromic repeats, is a bacterial immune system against invading DNA or RNA.⁷² Once foreign DNA is found several CRISPR associated proteins (Cas) assist in storing the DNA in repeat regions known as arrays. These arrays serve as a memory for the immune system. Next the array is expressed and tailored with tracer RNA, a short strand to help recruit Cas proteins, into functional CRISPR RNA (crRNA).⁷³ The crRNA and tracer RNA complexes with Cas nucleases and guides it to the foreign DNA which the nuclease then cuts. These complexes can target almost any strand of DNA, so long as the target is adjacent to a protospacer adjacent motif (PAM).⁷⁴ With no way of repair, the foreign DNA cannot survive and is rapidly degraded by the cell.

While native CRISPR systems are originally for targeting foreign DNA, the system can be modified to target selected DNA.⁷⁵ Retargeting the system only requires the provision of the nuclease and a guide RNA (gRNA) that mimics the targeting effect of functional crRNA. The gRNA must still target a sequence next to a PAM. Each CRISPR system utilizes a different PAM sequence.⁷⁶ One of the best-studied CRISPR systems is CRISPR-Cas9 from *S. pyogenes*.^{71,77,78} The spCas9 nuclease recognizes a NGG PAM sequence and has been developed to introduce DSBs in many heterologous hosts. The nuclease is especially effective in hosts that do not possess a means of repairing a DSB. Further, if a plasmid experiences a DSB, it will not be able to express the proteins on the plasmid. If one of those proteins were necessary for survival, e.g. an antibiotic resistance gene, then the cell would die without repair (**Figure 17**). As *Aeromicrobium erythreum*, *Saccharopolyspora erythraea*, and the plasmid platform in *E. coli* have high genomic GC content, the well-studied spCas9 nuclease is optimal for introducing the DSB event.

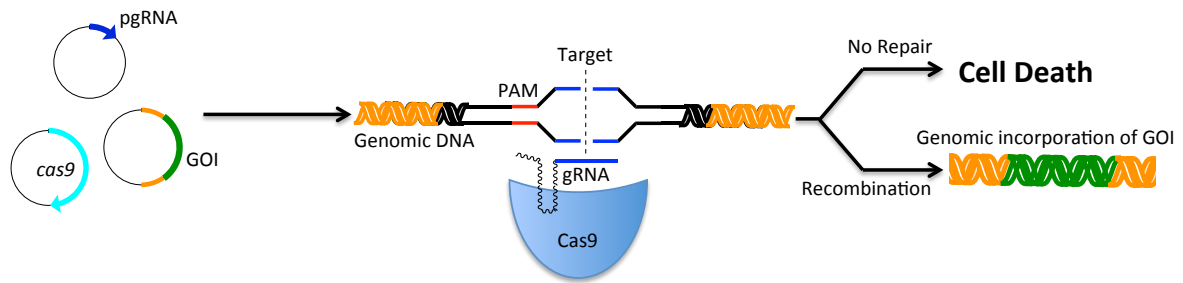


Figure 17 An overview of CRISPR-Cas9 for genome editing. *cas9* and plasmid pgRNA are transformed into a bacterial cell and expressed. Plasmid pgRNA transcribes the gRNA that serves as a genomic targeting strand and complexes with Cas9. Cas9 cuts the genome at the target site, if there is no repair the cell will die. A gene of interest (GOI) is provided along with homology arms matching areas flanking the target site. Homologous recombination can occur to repair the genome and incorporates the gene of interest.

Homology directed recombination is one strategy used by many organisms to repair a DSB.⁷⁵ It relies on using DNA that is a homologous match for a damaged area and repairing it by replacing the damage with a copy of the matching strand.⁷⁹ By targeting the location of the DSB with CRISPR and providing a homologous piece of DNA that can be used for recombination, cellular repair can occur and the cell will survive. Further, if the repair strand provided includes a mutation or a gene of interest, then that will also be included in the cellular DNA. The phage derived lambda red proteins can facilitate recombination in *E. coli*.⁷⁹ These proteins are able to protect a linear double stranded piece of DNA from degradation in the cell and exchange the DNA for homologous regions of the cellular DNA. Recombination with lambda red occurs with low efficiency without a means of selection, when coupled with CRISPR efficiencies of greater than 90% have been achieved.⁷⁵

Cas9 has previously been used to efficiently integrate GFP and the isobutanol pathway into the *E. coli* genome, which has been used as a heterologous host for erythromycin production.⁷⁵ Cas9 can be targeted to the genes or plasmids responsible for erythromycin biosynthesis in the organisms and a repair strand provided. The repair strand can also contain the selectivity-shifting mutations that have previously been discovered.

Further, CRISPR and recombineering can be used to genomically integrate the MatB mutants for the *in vivo* production of non-natural malonyl-CoAs.

Harnessing CRISPR will allow for efficient integration of PKS mutations in native or heterologous hosts. As CRISPR activity has never been demonstrated in *Aeromicrobium erythreum*, targeted integration of GFP in the genome of all three organisms would be an easily detectable proof of concept and control for the efficacy of CRISPR across hosts. As a primary target for erythromycin biosynthesis, the KR of module 6 can be inactivated so the polyketide retains a ketone. This position is normally glycosylated in erythromycin A and the inactivation will render that reaction impossible. Further, the product of the PKS would be the ketolide derivative of erythromycin, which is a useful precursor for the semi-synthesis of novel antibiotics.⁸⁰ Additional targets include the AT domains, where a single DSB event could allow for the installation of a library of mutations to affect the selectivity. *In vivo* engineering of PKS hosts will open the door to the production of novel bioactive compounds from PKS's with altered substrate specificity.

REFERENCES

1. R. B. Woodward, E. Logusch, K. P. Nambiar, K. Sakan, S. *et al.* Asymmetric Total Synthesis of Erythromycin. 1. Synthesis of an Erythronolide A Seco Acid Derivative via Asymmetric Induction. *J. Am. Chem. Soc.* **103**, (1981).
2. R. B. Woodward, E. Logusch, K. P. Nambiar, K. Sakan, S. *et al.* Asymmetric Total Synthesis of Erythromycin. 2. Synthesis of an Erythronolide A Lactone System. *J. Am. Chem. Soc.* **103**, 3213–3215 (1981).
3. Woodward, R. B. *et al.* Asymmetric Total Synthesis of Erythromycin. 3. Total Synthesis of Erythromycin. *J. Am. Chem. Soc.* **103**, 3215–3217 (1981).
4. Seiple, I. B. *et al.* A platform for the discovery of new macrolide antibiotics. *Nature* **533**, 338–345 (2016).
5. Wu, J. *et al.* Toward improvement of erythromycin a production in an industrial *Saccharopolyspora erythraea* strain via facilitation of genetic manipulation with an artificial attB site for specific recombination. *Appl. Environ. Microbiol.* **77**, 7508–7516 (2011).
6. Weissman, K. J. Uncovering the structures of modular polyketide synthases. *Nat. Prod. Rep.* **32**, 436–53 (2015).
7. Robbins, T., Kapilivsky, J., Cane, D. E. & Khosla, C. Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* **55**, 4476–4484 (2016).
8. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–7 (2014).
9. Rawlings, B. J. Type I polyketide biosynthesis in bacteria (part A). *Nat. Prod. Rep.* **18**, 190–227 (2001).
10. Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. a & Cane, D. E. Structure and

mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* **76**, 195–221 (2007).

11. Bravo-Rodriguez, K. *et al.* Substrate Flexibility of a Mutated Acyltransferase Domain and Implications for Polyketide Biosynthesis. *Chem. Biol.* **22**, 1425–1430 (2015).
12. Barajas, J. F., Blake-Hedges, J. M., Bailey, C. B., Curran, S. & Keasling, J. D. Engineered polyketides: Synergy between protein and host level engineering. *Synth. Syst. Biotechnol.* 1–20 (2017). doi:10.1016/j.synbio.2017.08.005
13. Gupta, S. *et al.* Generation of Novel Pikromycin Antibiotic Products Through Mutasynthesis. *ChemBioChem* **2**, 1609–1616 (2008).
14. Cane, D. E., Kudo, F., Kinoshita, K. & Khosla, C. Precursor-Directed Biosynthesis: Biochemical Basis of the Remarkable Selectivity of the Erythromycin Polyketide Synthase toward Unsaturated Triketides. *Chem. Biol.* **9**, 131–142 (2002).
15. John R. Jacobsen, C. Richard Hutchinson, D. E. C. & Chaitan Khosla. Precursor-Directed Biosynthesis of Erythromycin Analogs by an Engineered Polyketide Synthase. *Science*. **277**, 367–369 (1997).
16. Long, P. F. *et al.* Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Mol. Microbiol.* **43**, 1215–1225 (2002).
17. Yuzawa, S., Eng, C. H., Katz, L. & Keasling, J. D. Broad Substrate Specificity of the Loading Didomain of the Lipomycin Polyketide Synthase. *Biochemistry* **52**, 3791–3793 (2013).
18. Marsden, A. F. A. *et al.* Engineering Broader Specificity into an Antibiotic-Producing Polyketide Synthase. *Science*. **279**, 199–202 (1998).
19. Sheehan, L. S. *et al.* Engineering of the Spinosyn PKS: Directing Starter Unit Incorporation. *J. Nat. Prod.* **69**, 1702–1710 (2006).

20. Bayly, C. L. & Yadav, V. G. Towards precision engineering of canonical polyketide synthase domains: Recent advances and future prospects. *Molecules* **22**, 1–18 (2017).
21. Akey, D. L. *et al.* Structural basis for macrolactonization by the pikromycin thioesterase. *Nat. Chem. Biol.* **2**, 537–542 (2006).
22. Tripathi, A., Choi, S., Sherman, D. H. & Kim, E. Thioesterase domain swapping of a linear polyketide tautomycin with a macrocyclic polyketide pikromycin in *Streptomyces* sp. CK4412. *J. Ind. Microbiol. Biotechnol.* **43**, 1189–1193 (2016).
23. Keatinge-Clay, A. T. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. *chem. rev.* **117**, 5334–5366 (2017).
24. Dunn, B. J., Cane, D. E. & Khosla, C. Mechanism and Specificity of an Acyltransferase Domain from a Modular Polyketide Synthase. *Biochemistry* **52**, 1839–1841 (2013).
25. Poust, S. *et al.* Understanding the role of histidine in the GHSxG acyltransferase active site motif: Evidence for histidine stabilization of the malonyl-enzyme intermediate. *PLoS One* **9**, 2–5 (2014).
26. Williams, G. J. Engineering polyketide synthases and nonribosomal peptide synthetases. *Curr. Opin. Struct. Biol.* **23**, 603–612 (2013).
27. Yuzawa, S. *et al.* Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth. Biol.* **6**, 139–147 (2017).
28. Dunn, B. J. & Khosla, C. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *J. R. Soc. Interface* **10**, 1–13 (2013).
29. Nigam, A. *et al.* Modification of rifamycin polyketide backbone leads to improved drug activity against rifampicin-resistant *Mycobacterium tuberculosis*. *J. Biol. Chem.* **289**, 21142–21152 (2014).

30. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176 (2005).
31. Menzella, H. G., Carney, J. R. & Santi, D. V. Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases. *Chem. Biol.* **14**, 143–151 (2007).
32. Musiol-Kroll, E. M. *et al.* Polyketide Bioderivatization Using the Promiscuous Acyltransferase KirCII. *ACS Synth. Biol.* **6**, 421–427 (2017).
33. Walker, M. C. *et al.* Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science* **341**, 1089–94 (2013).
34. Kim, B. S., Sherman, D. H. & Reynolds, K. A. An efficient method for creation and functional analysis of libraries of hybrid type I polyketide synthases. *Protein Eng. Des. Sel.* **17**, 277–284 (2004).
35. Chemler, J. A. *et al.* Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination. *J. Am. Chem. Soc.* **137**, 10603–10609 (2015).
36. Vecchio, F. Del *et al.* Active-site residue, domain and module swaps in modular polyketide synthases. *J Ind Microbiol Biotechnol* **30**, 489–494 (2003).
37. Sundermann, U. *et al.* Enzyme-directed mutasynthesis: A combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chem. Biol.* **8**, 443–450 (2013).
38. Reeves, C. D. *et al.* Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry* **40**, 15464–15470 (2001).
39. Koryakina, I., McArthur, J. B., Draelos, M. M. & Williams, G. J. Promiscuity of a modular polyketide synthase towards natural and non-natural extender units. *Org. Biomol. Chem.* **11**, 4449 (2013).

40. Klopries, S., Sundermann, U. & Schulz, F. Quantification of N-acetylcysteamine activated methylmalonate incorporation into polyketide biosynthesis. *Beilstein J. Org. Chem* **9**, 664–674 (2013).
41. Koryakina, I. *et al.* Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering. *ACS Chem. Biol.* **12**, 114–123 (2016).
42. Beck, B. J., Aldrich, C. C., Fecik, R. A., Reynolds, K. A. & Sherman, D. H. Iterative Chain Elongation by a Pikromycin Monomodular Polyketide Synthase. *J. Am. Chem. Soc.* **125**, 4682–4683 (2003).
43. Oh, H.-S. & Kang, H.-Y. Total Synthesis of Pikromycin. *J. Org. Chem.* **77**, 1125–1130 (2012).
44. Lu, T. *et al.* Effect of chloramphenicol, erythromycin, moxifloxacin, penicillin and tetracycline concentration on the recovery of resistant mutants of *Mycobacterium smegmatis* and *Staphylococcus aureus*. *J. Antimicrob. Chemother.* **52**, 61–64 (2003).
45. Almutairi, M. M. *et al.* Co-produced natural ketolides methymycin and pikromycin inhibit bacterial growth by preventing synthesis of a limited number of proteins. *Nucleic Acids Res.* **45**, 9573–9582 (2017).
46. Hansen, D. A. *et al.* Biocatalytic Synthesis of Pikromycin, Methymycin, Neomethymycin, Novamethymycin, and Ketomethymycin. *J. Am. Chem. Soc.* **135**, 11232–11238 (2013).
47. Kittendorf, J. D., Beck, B. J., Buchholz, T. J., Seufert, W. & Sherman, D. H. Interrogating the Molecular Basis for Multiple Macrolactone Ring Formation by the Pikromycin Polyketide Synthase. *Chem. Biol.* **14**, 944–954 (2007).
48. Schultheis, S. E. Probing the Scope and Utility of Polyketide Synthases. (North Carolina State University, 2015).

49. Hansen, D. A., Koch, A. A. & Sherman, D. H. Substrate Controlled Divergence in Polyketide Synthase Catalysis. *J. Am. Chem. Soc.* **137**, 3735–3738 (2015).
50. Mortison, J. D., Kittendorf, J. D. & Sherman, D. H. Synthesis and biochemical analysis of complex chain-elongation intermediates for interrogation of molecular specificity in the erythromycin and pikromycin polyketide synthases. *J. Am. Chem. Soc.* **131**, 15784–15793 (2009).
51. Koryakina, I. & Williams, G. J. Mutant malonyl-CoA synthetases with altered specificity for polyketide synthase extender unit generation. *ChemBioChem* **12**, 2289–2293 (2011).
52. Thuronyi, B. W. & Chang, M. C. Y. Synthetic biology approaches to fluorinated polyketides. *Acc. Chem. Res.* **48**, 584–592 (2015).
53. Ambrish Roy, Alper Kucukural & Yang Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
54. Borders, C. L. *et al.* A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Sci.* **3**, 3541–548 (1994).
55. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chem. Biol.* **14**, 931–943 (2007).
56. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124–9 (2006).
57. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct. Funct. Bioinforma.* **57**, 702–710 (2004).
58. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J.*

- Comput. Chem.* **31**, 455–461 (2009).
59. Pyne, M. E. *et al.* *Synthetic biology and metabolic engineering in plants and microbes. Part A, Metabolism in microbes.* (2016).
 60. Huang, F. & Nau, W. M. A Conformational Flexibility Scale for Amino Acids in Peptides. *Angew. Chemie* **42**, 2269–2272 (2003).
 61. Zheng, J. *et al.* Structure and Function of the Macrolide Biosensor Protein, MphR(A), with and without Erythromycin. *J. Mol. Biol.* **387**, 1250–1260 (2009).
 62. Leslie Cuthbertson & a Justin R. Nodwella, B. The TetR Family of Regulators. *Microbiol. Mol. Biol. Rev.* **77**, 440–475 (2013).
 63. Möhrle, V., Stadler, M. & Eberz, G. Biosensor-guided screening for macrolides. *Anal. Bioanal. Chem.* **388**, 1117–1125 (2007).
 64. Koryakina, I., Neville, J., Nonaka, K., Van Lanen, S. G. & Williams, G. J. A high-throughput screen for directed evolution of the natural product sulfotransferase LipB. *J. Biomol. Screen.* **16**, 845–51 (2011).
 65. Tang, S.-Y. *et al.* Screening for Enhanced Triacetic Acid Lactone Production by Recombinant Escherichia coli Expressing a Designed Triacetic Acid Lactone Reporter. *J. Am. Chem. Soc.* 10099–10103 (2013).
 66. Jiang, M., Fang, L. & Pfeifer, B. A. Improved heterologous erythromycin A production through expression plasmid re-design. *Biotechnol. Prog.* **29**, 862–869 (2013).
 67. Brikun, I. A., Reeves, A. R., Cernota, W. H., Luu, M. B. & Weber, J. M. The erythromycin biosynthetic gene cluster of Aeromicrobium erythreum. *J. Ind. Microbiol. Biotechnol.* **31**, 335–344 (2004).
 68. El-Enshasy, H. A., Mohamed, N. A., Farid, M. A. & El-Diwany, A. I. Improvement of

- erythromycin production by *Saccharopolyspora erythraea* in molasses based medium through cultivation medium optimization. *Bioresour. Technol.* **99**, 4263–4268 (2008).
69. Fang, L., Guell, M., Church, G. M. & Pfeifer, B. A. Heterologous erythromycin production across strain and plasmid construction. *Biotechnol. Prog.* 1–6 (2017). doi:10.1002/btpr.2567
 70. Selle, K. & Barrangou, R. Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends Microbiol.* **23**, 225–232 (2015).
 71. Prashant Mali, Kevin M Esvelt, G. M. C. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).
 72. Makarova, K. S. *et al.* The CRISPR–Cas modules are adaptive immune systems that are present in most archaea and many bacteria. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
 73. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell* **60**, 385–397 (2015).
 74. Jennie R. Fagen, Daphne Collias, A. K. S. and & Chase L. Beisel. Advancing the design and delivery of CRISPR antimicrobials. *Curr. Opin. Biomed. Eng.* **4**, 57–64 (2017).
 75. Bassalo, M. C. *et al.* Rapid and Efficient One-Step Metabolic Pathway Integration in *E. coli*. *ACS Synth. Biol.* **5**, 561–568 (2016).
 76. Leenay, R. T. & Beisel, C. L. Deciphering, Communicating, and Engineering the CRISPR PAM. *J. Mol. Biol.* **429**, 177–191 (2017).
 77. Jiang, Y. *et al.* Multigene Editing in the *Escherichia coli* Genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
 78. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of

Escherichia coli. *Nucleic Acids Res.* **44**, 4243–4251 (2016).

79. Mosberg, J. A., Lajoie, M. J. & Church, G. M. Lambda Red Recombineering in *Escherichia coli* Occurs Through a Fully Single-Stranded Intermediate. *Genetics* **186**, 791–799 (2010).
80. Putnam, S. D., Sader, H. S., Farrell, D. J., Biedenbach, D. J. & Castanheira, M. Antimicrobial characterisation of solithromycin (CEM-101), a novel fluoroketolide: activity against staphylococci and enterococci. *Int. J. Antimicrob. Agents* **37**, 39–45 (2010).

APPENDICES

Appendix A

LC-HRMS Gradient and HESI scan range

HESI Source Parameters		LC Gradient	
		Time (Min)	% B
Spray voltage	3.5 kV	0.0	25.0
Capillary temperature	350°C	1.0	25.0
Heater temp	300°C	10.0	59.0
S Lens RF level	70 V	11.0	85.0
Sheath gas flow rate	60	12.0	85.0
Resolution	70,000	12.5	25.0
Scan Range	100-1000 m/z	16.5	25.0

Appendix B.

Primers Used for Sequencing or Mutagenesis

Primers for **PikAIII**

D624 Reverse	GACGCGGTCGAGCGTGG
D624E Forward	GAGGTCGTCCAGCCCGTGACC
E658 Reverse	GCCCTGCGAGTGGCCG
E658D Forward	GATATCGCCGCCGCGTACGTCGC
E658V Forward	GTTATCGCCGCCGCGTACGTCGC
I684X Forward	NDKGCCGCCAGCTCGCCGGCAA
I684X Reverse	GGACTTGCTGCGCAGGGTGACGA
L689V Forward	GTTGCCGGCAAGGGCGGC
L689V Reverse	GTGGGCGGCGATGGAC
L808 Reverse	GTTGCGGTACCAGTAGGTGCC
L808A Forward	GCACGCCATCGCGTGGGC
L808G/C Forward	NGTCGCCATCGCGTGGGC
L808V Reverse	GTCCGCCATCGCGTGGGC
P752V Forward	GTCGTCGACTACGCCTCCC
P752V Reverse	GATGATCCGCGCACG
V753X Forward	ATCATCCCGDNTGACTACGCCTCCCACAGCCGGC AGGTCGAGATCA
V753X Reverse	GGCGTAGTCANHCGGGATGATCCGCGCACGGACG CCGTCGGCCTCG
S757X Forward	NNKACAGCCGGCAGGT
S757X Reverse	GGCGTAGTCGACCGGGATGAT
Y755V Reverse	GTCGACCGGGATGATCCGC

Y755V Forward	GTCGCCTCCCACAGCCGG
Y755X Forward	GCCTCCCACAGCCGGCAGGTC
Y755X Forward	GCCTCCCACAGCCGGCAGGTC
PikSequencing1	CAGCGGCCTCGGCGTCTT
PikSequencing2	CTCGCGCCGGGTGACGTGGAC
PikSequencing3	CTGCCTCGATGGCCGAATGCGAG
PikSequencing4	CACCCACTTCATCGAGGTCAG
PikSequencing5	CCTGTGGTCCGTCACCCAGGG
PikSequencing6	CATGCGCACCCCTCCTCGACG
PikSequencing7	GACGCCGTCGAGCAGAAGGCTG

**Primers for
PikAIV**

Pik4 seq1	cgtgaccgtggacacgg
Pik4 seq2	gcgttcggcgtgggc
Pik4 seq3	tcgacgacgccgctcg
Pik4 seq4	gctcgtactggatcagccc
Pik4 seq5	ttcctgcggctcagcacc
Pik4 Y753X Forward	NNKgcctcccacagcgccc
Pik4 Y753X Reverse	gtcgacgggatgatccgtgc

Primers for Gibson assembly of PikAIV AT into PikAIII

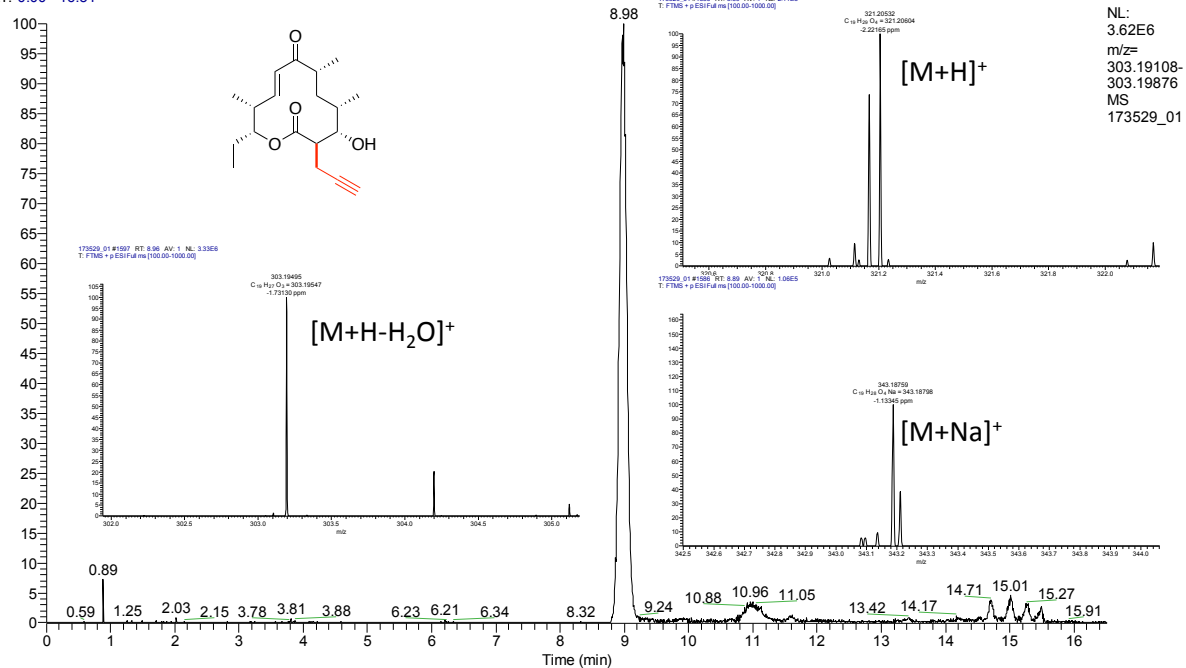
Fragment.Forward	CATCAGCgggacgaacgcg
Fragment.Reverse	CGCTCGGTctggaaggcgtacgtcgggag
Vector.Forward	cttcagACCGAGCGCTTC
Vector.Reverse	ttcgtcccGCTGATGCCGAA

Primers for Gibson assembly of PikAIII AT into PikAIV

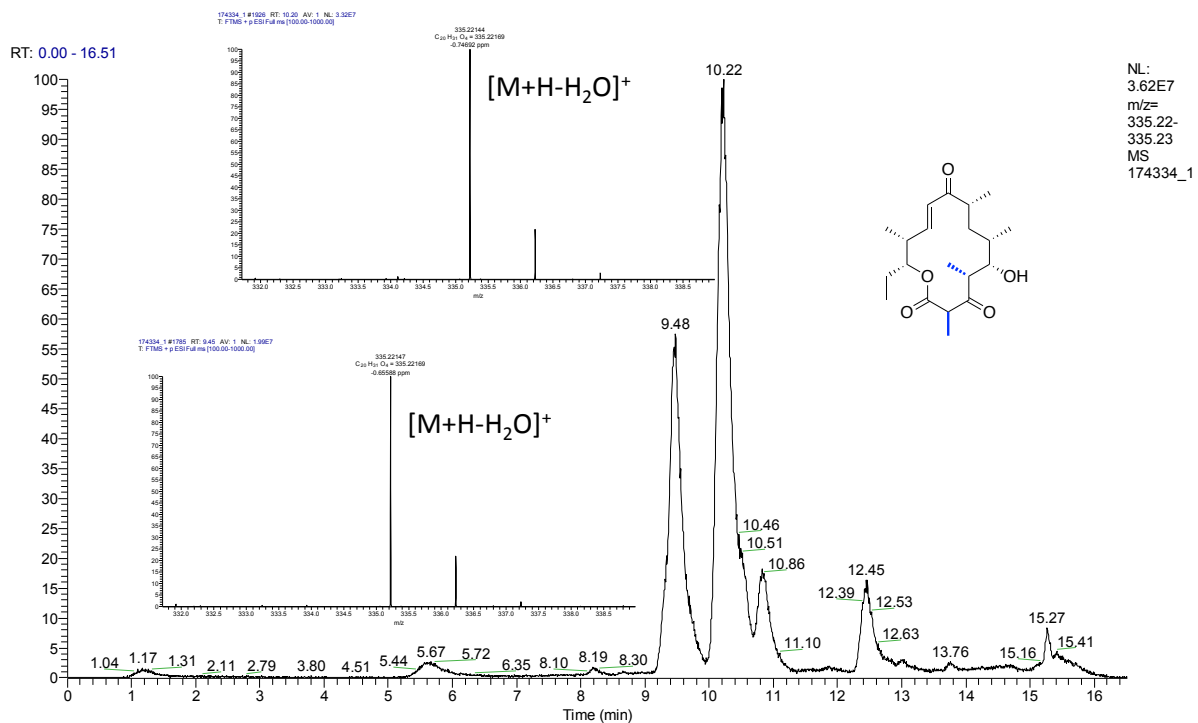
Fragment.Forward	cgtgggcGGGACGAACGCG
Fragment.Reverse	gagcgggtgCTGGAAGGCGTAGGTGGGG
Vector.Forward	CTTCCAGcaccgctcgtac
Vector.Reverse	TTCGTCCCgcccacgccgaa

2b

RT: 0.00 - 16.51

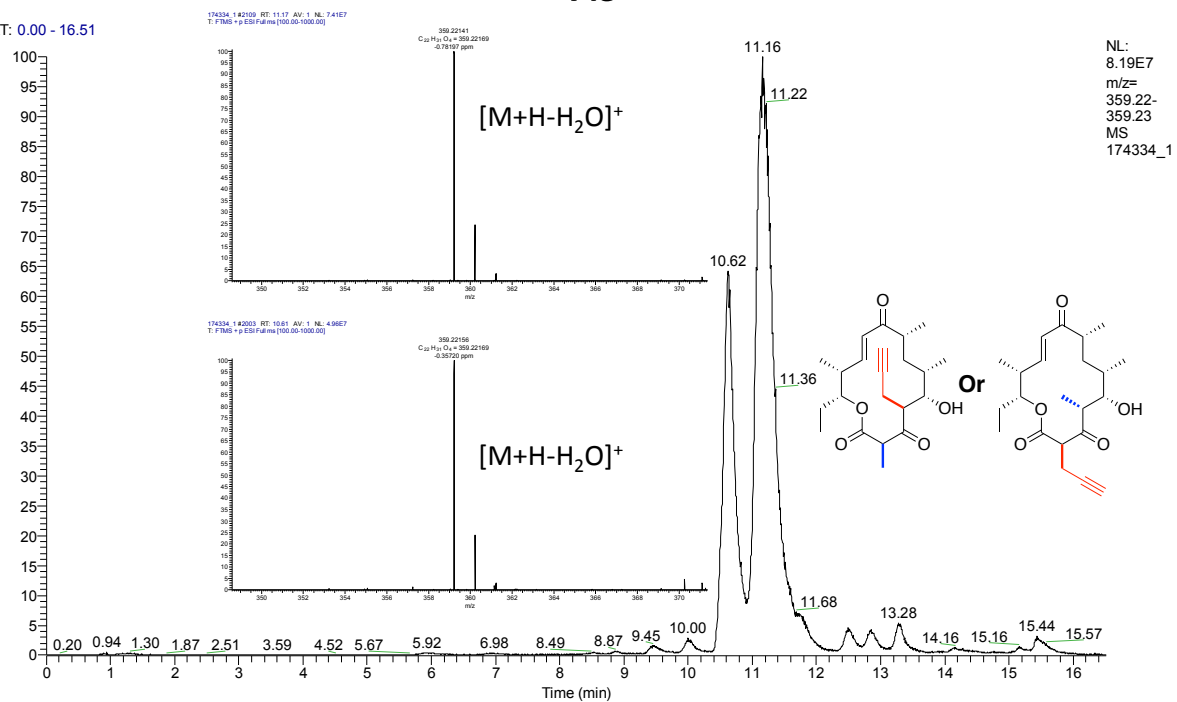


3



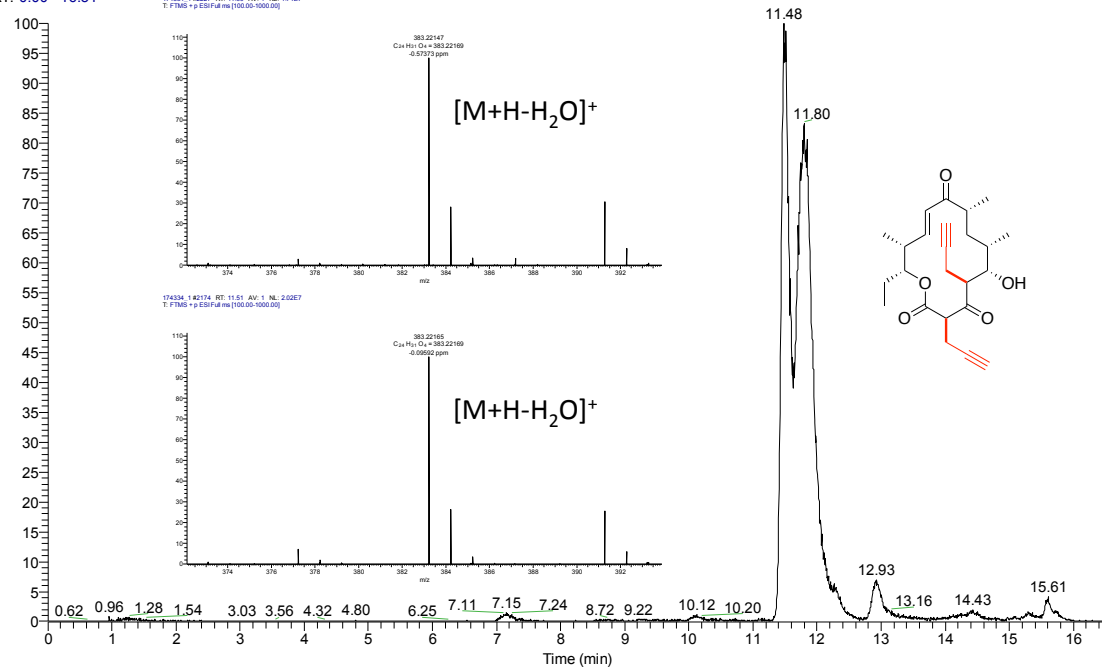
4b

RT: 0.00 - 16.51

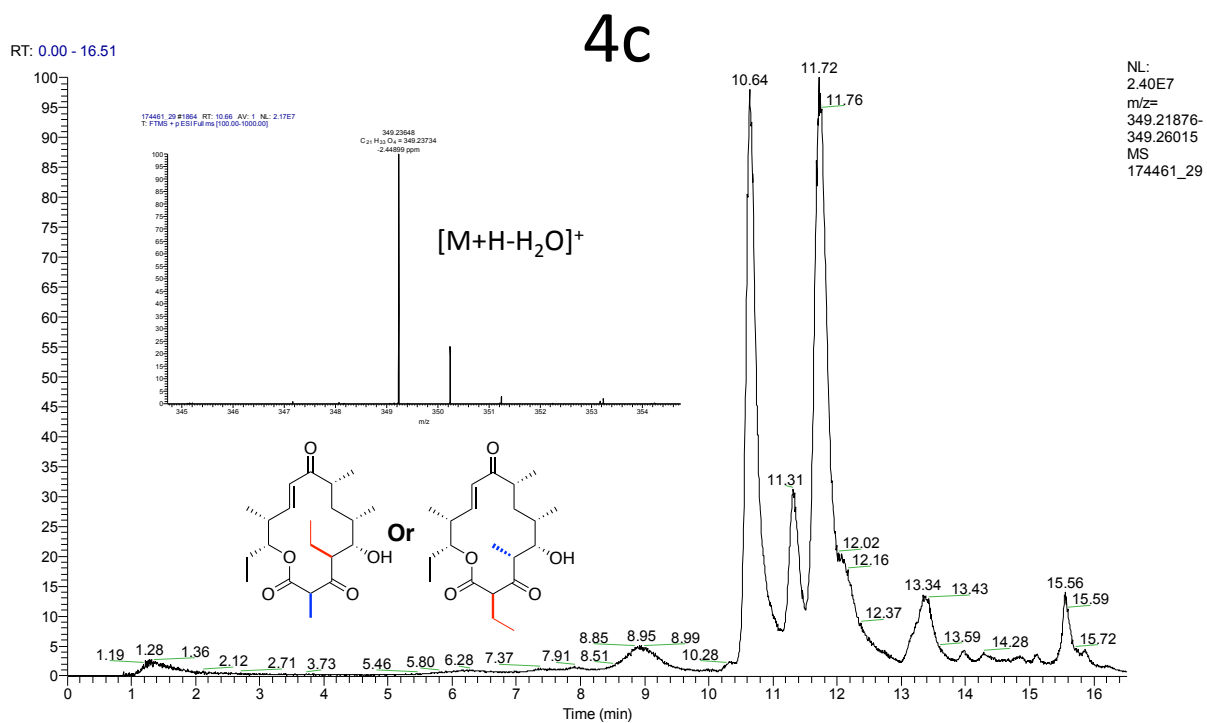


5b

RT: 0.00 - 16.51

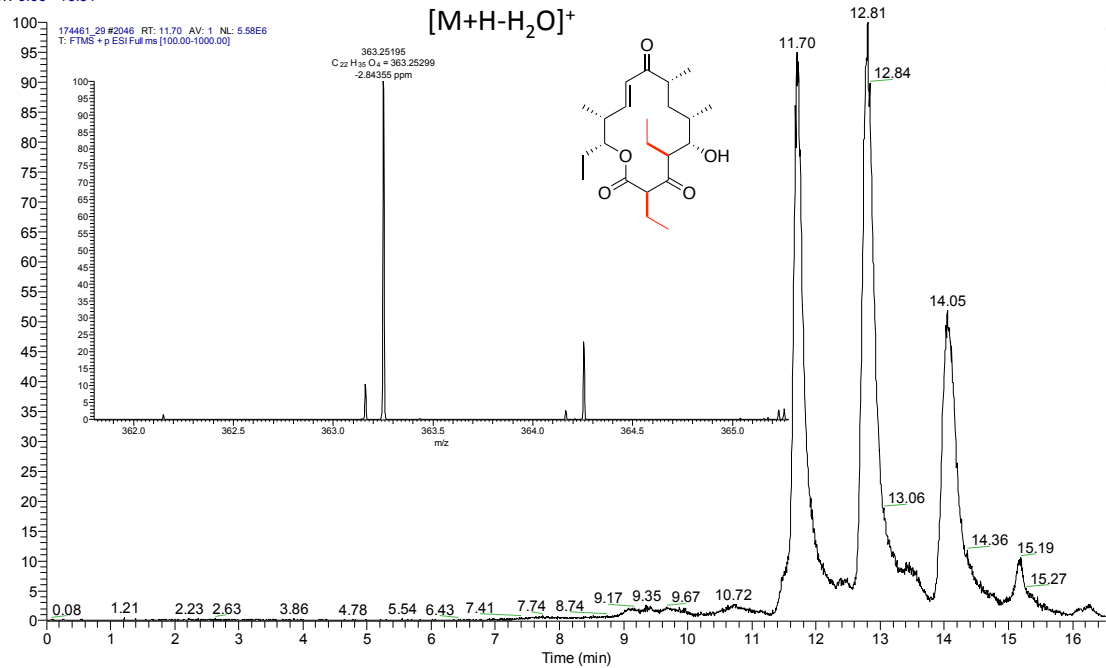


NL:
2.19E7
m/z=
383.22-
383.23
MS
174334_1



5c

RT: 0.00 - 16.51



NL:
6.09E6
m/z=
363.23733-
363.26030
MS
174461_29

Tp-Pentaketide

