Determining the Conformations of Several Polyketide Synthases in the Presence of Various Substrates

Dennis A. Hu

The University of Texas at Austin

Department of Chemistry and Biochemistry

Undergraduate Class of 2010

An Honors Thesis for the

Dean's Scholars Honors Program

Submitted May 2010

Adrian T. Keatinge-Clay, Ph.D.

Chemistry and Biochemistry

Supervising Professor

John F. Stanton, Ph.D.

Chemistry and Biochemistry

Honors Thesis Advisor

Abstract

Today, many microorganisms are becoming resistant to the drugs that doctors are administering to their patients at an exponential rate. If this rate continues, it won't be long until all microorganisms are resistant to the drugs we have developed. As a result, there has been a surge in the amount of research being conducted on antibiotics found in nature. In particular, polyketides have recently been discovered, and the Keatinge-Clay group, like many other groups around the world, has been focusing their efforts on this area. In this paper, the mechanisms by which different polyketides are made by machines known as polyketide synthases (PKSs) will be mentioned. This will then be followed by an examination of how these different PKSs incorporate and allow their substrates to enter into the active site groove. As of now, the Keatinge-Clay group has been able to obtain a crystal structure of Amphotericin ketoreductase (KR) domain 2 in the presence of 2-methyl-3-oxopentanyl-SNAC and $NADP^+$ at a 2.3Å resolution. However, the electron density of this protein will be used as a model for how 2-methyl-3-oxopentanyl-SNAC enters the active site groove. Meanwhile, further work will involve obtaining crystal structures for erythromycin KR1, tylosin KR1, and pikKR3 in the presence of the aforementioned substrates. In addition, efforts have been placed on crystallizing PikKS-AT-ACP_0 and producing ACP proteins which can be co-crystallzed later with the various KRs

I. Introduction:

Life on Earth began approximately 3.5 billion years ago and along with it came various infectious diseases caused by viruses, fungi, and protozoans. Then, approximately, 120,000 years ago, the first modern human beings were born and with them came the first treatments known to mankind. Specifically, the ancient Chinese used various plants to treat these infectious diseases while the ancient Egyptians, Greeks, and medieval Arabs used molds. However, it wasn't until 1928 when Sir Alexander Fleming made his discovery of penicillin from the mold, *Penicillium notatum*, that the development of modern antibiotics began. Currently, the majority of antibiotics are semi-synthetic, meaning that they are chemically modified from natural components; however, there are still many that are found in nature.

Of particular interest to the Keatinge-Clay group are polyketides, of which a few can be found in **Figure 1**. These polyketides are produced by a variety of bacteria, fungi, and plants, and comprise a large and diverse class of secondary metabolites.[12, 14] Also, a large portion of these polyketides possess important biological and medicinal properties such as antibacterial, antifungal, anticancer, and immunosuppressive activity. As a result, many researchers are interested in learning more about how these polyketides are made.

Specifically, polyketides are produced by multienzyme megasynthases known as polyketides synthases (PKSs), which use one of the simplest building blocks in biochemistry, carboxylic acids, in order to make some of the most elegant and complex compounds known to mankind. Furthermore, there are three types of polyketides however, only the first type will be described in this paper. Type 1 polyketide synthases are large, highly modular proteins which process polyketides via modules. This type of PKS is then subdivided into two categories: iterative and modular. For the iterative PKS, a single multienzyme is used repeatedly. On the other hand, the modular PKS uses separate modules that are linked to each other, and each module performs a particular step of the polyketide biosynthesis pathway. An example of a modular Type 1 PKS can be found in **Figure 2**.

Modularity is a highly sought after feature in engineering design and is defined by Khosla et. al as a multi-component system that can be divided into smaller subsystems, which interact with each other and can be predictably interchanged for functional flexibility and variety.

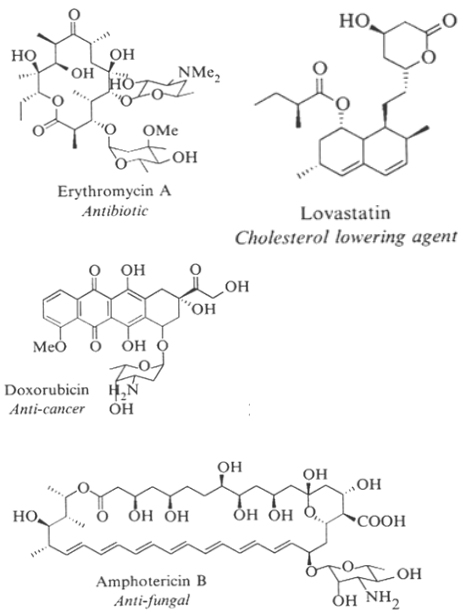

Figure 1: Several polyketides along with their medicinal properties are shown here.

As a result of this property, many researchers have tried to swap domains of certain modules with those of another module in order to learn more about each individual domain. Specifically, there are four ways by which a researcher can manipulate a PKS. They are (i) length of the polyketides chain, which is determined by the number of modules in the genetically engineered polyketides; (ii) choice of primer and extender units; (iii) extent of polyketides reduction; and (iv) stereochemistry of the carbons carrying alkyl and hydroxyl substituents. From these manipulations, more information can potentially be obtained on the fundamental mechanistic underpinnings of PKS catalytic processes such as the relative timing of alcohol group epimerization, stereochemical course of KS and KR-catalyzed reactions, and the unexpected configurational stability of an ACP-bound polyketide intermediate.[7] However, there are several limitations to these manipulations. In particular, it is very possible for downstream modules to not accept or process efficiently the products produced by the upstream modules. Two models have been considered recently to explain the factors controlling the transfer of a growing polyketide. The first involves protein-protein recognition being the primary factor while the transfer of the growing polyketide chain is nonselective. Meanwhile, the other model proposes that acceptor modules recognize key structural features of their natural substrates and discriminate functionally modified analogs to a certain degree. As of now, aspects of both models are accepted by the majority of researchers.[6]

Another difficulty is that the organisms that naturally produce these polyketides are not compliant to the procedures utilized in genetic engineering. However, this problem has been solved thanks to the researcher's best friend, *E. coli*. Currently, *E. coli* has been shown to be capable of supporting large amounts of polyketides, and thanks to the highly advanced molecular biology tools used for this organism, *E. coli* is now the ideal host for polyketide researchers.

Meanwhile, another problem with genetically engineering polyketide synthases lies with the PKS genes. Here, the genes are extraordinarily long (35 to >200 kb), highly repetitive, rarely contain conserved restriction sites, and contain approximately 70% G+C bases. As of now, Menzella et al. has experimented with total gene synthesis so that codon composition could be controlled and unique restriction sites could be introduced into the genes.[23] Meanwhile, another solution to this problem involves designing plasmids that can carry large DNA fragments.

Yet, despite these possible hurdles, researchers are still determined to find out more about these promiscuous enzymes. In particular, many research groups, such as the Keatinge-Clay group, are interested in building an exhaustive library of polyketides by using building blocks of various lengths and branches, which can be tolerated by PKSs. By building this library of polyketides, many potential drugs can be discovered in the future. Furthermore, organic chemists can use these constructed libraries for their own research since chirality, a property that distinguishes polyketides of the same chain length and structure, is something that can not be easily introduced into newly synthesized compounds.

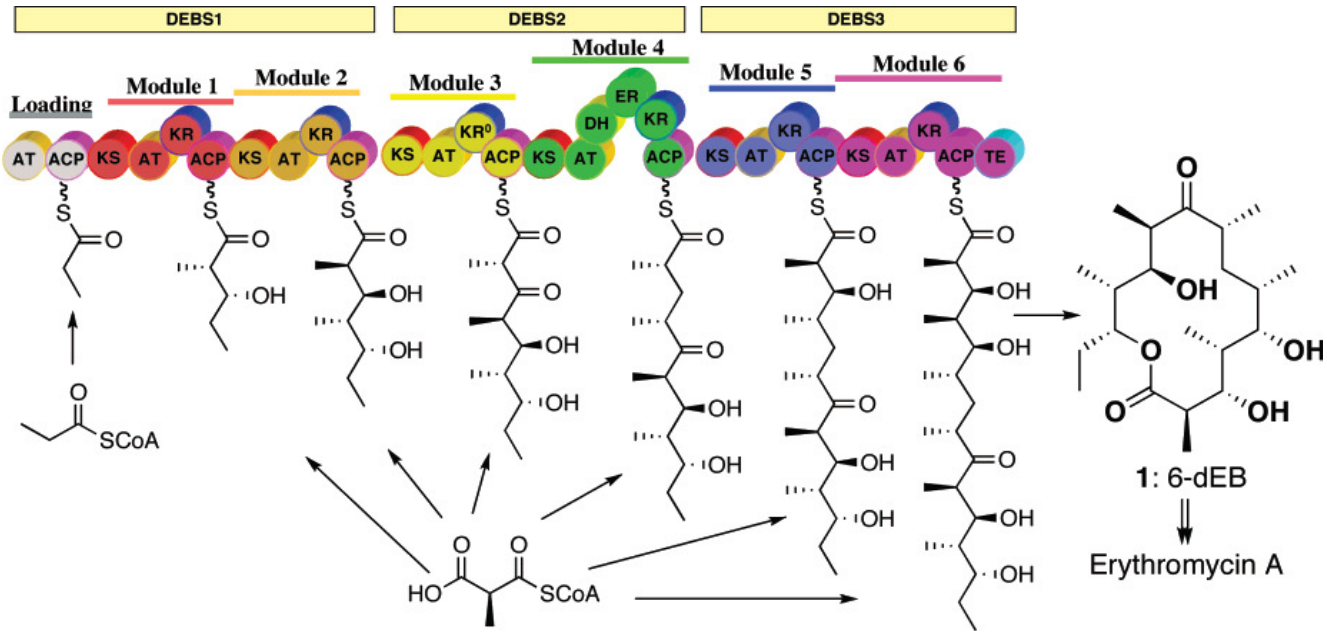

Figure 2: Modules of the prototypical Type I PKS: DEBS (Deoxyerythronolide B synthase). Note: KR3 is incompetent of reduction polyketides.

However, before any of these aforementioned goals can be achieved, much more information about the structure and catalytic activity of each domain of a PKS must be obtained. In the following pages, I will be describing the conformational changes of a type I ketoreductase domain after a polyketide and substrate have been incorporated into the active sites. This conformational change was obtained via crystallization and X-ray diffraction. Furthermore, efforts on other type I ketoreductase domains, ACP domains, and the loading module of Pikromycin will be described.

II. Background

A. DEBS

In polyketide research, the prototypical Type 1 modular PKS is the 6-deoxyerythronolide B synthase, also known as DEBS, from *Saccharopolyspora erythraea*. This enzyme is composed of three homodimeric polypeptides, each of which contains two covalently linked catalytic modules (each module weighs more than 300 kDa) and produces 6-deoxyerythronolide B, the macrolide aglycone core of the antibiotic erythromycin, from a propionyl-CoA starter unit and six methylmalonyl-CoA extender units along with a few cofactors.[6, 9, 28] (**Figure 1**) Furthermore, the research that has been performed on the structure and function of DEBS, whether that be through radioisotope and stable isotope labeling experiments, heterologous expression, directed mutagenesis, or *in vitro* studies on partly active proteins, has helped to elucidate the mechanism by which other PKSs construct their polyketides.[30]

I. Fatty Acid and Polyketide Biosynthesis

The mechanism by which polyketides are synthesized via PKSs in each module is similar to that used by fatty acid synthases (FAS) with a few differences. In fatty acid synthesis, FAS starts of by accepting an acetyl-CoA and then uses two-carbon units donated by the decarboxylation of an activated malonate in order to extend the pre-existing fatty acyl chains every cycle. The cycle then repeats seven more times until the 16-carbon saturated palmitate is produced. In addition, each step in this cycle results in a reduction of each acetate unit, which differs from that seen in PKSs where not every single carbonyl carbon of the

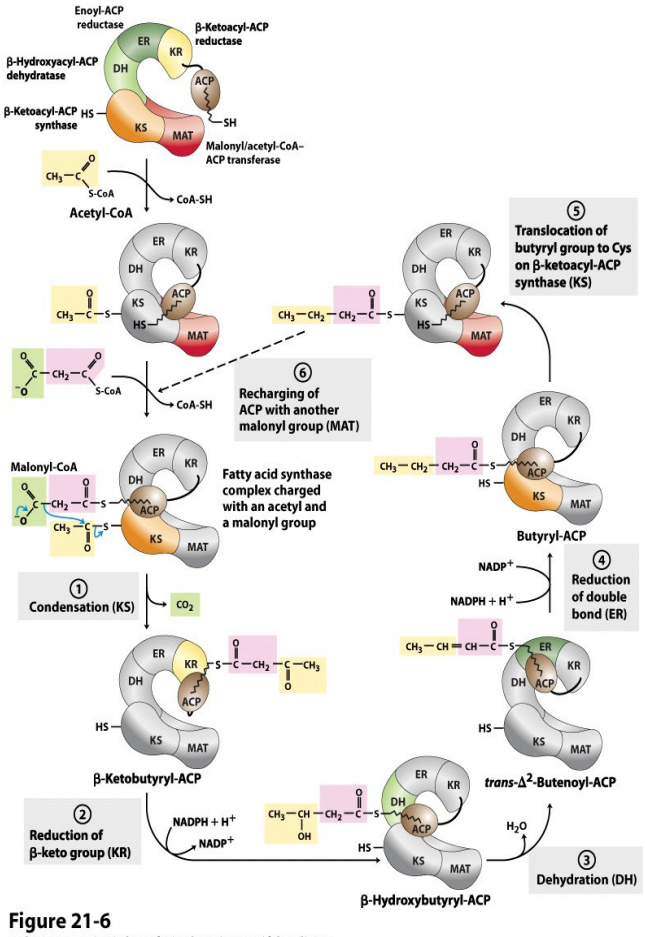

Figure 21-6
Lehninger Principles of Biochemistry, Fifth Edition

Figure 3: Fatty acid biosynthesis

extender unit is reduced. As a result, several polyketides can potentially be produced due to the many possible combinations of reduction that may occur.

Another difference between FASs and PKSs involves how the polyketide is processed in each cycle. Specifically, FAS works by having the same five enzymes catalyze each cycle of fatty acid elongation in a process known as iterative. On the other hand, the domains of Type 1 PKSs, as stated before, are organized into modules, where each module catalyzes the extension and modification of each extender unit and then passes on the polyketide chain to the next module in a process similar to that seen in an assembly line.

II. Polyketide Biosynthesis Mechanism

Before PKSs can begin their repetitive Claisen-like condensations in order to create these magnificent polyketdies, methylmalonyl-CoA must first be present in order to initiate this whole process. However, methylmalonyl-CoA is not commonly found in the environment. As a result, bacteria use various routes in order to produce this precursor.

In one route, propionate initially becomes propionyl-CoA via a two step reaction that involves adenylation with ATP along with an attack by the thiol group of coenzyme A.[24] This propionyl-CoA is then converted to L-methylmalonyl-CoA by a carboxylation reaction catalyzed by propionyl-CoA carboxylase. This enzyme contains the cofactor biotin, which is commonly seen in carboxylation reactions, and obtains its energy via the cleavage of ATP to ADP and P_i. DEBS is now ready to select only the (2S)-methylmalonyl-CoA in order to begin building a polyketide chain.

As stated before, DEBS is composed of three separate polypeptides, named DEBS 1, 2, and 3, which contain two modules each and are about 330-370 kDa in size. As a result, there are a total of six modules in DEBS. Each module also includes all the catalytic domains required for one round of chain extension and modification. Specifically, each of these domains are composed of a set of three core domains: the β-ketoacyl-ACP synthase (KS), which transfers a polyketide chain from the appropriate upstream module to a cysteine residue and catalyzes a decarboxylative condensation between this polyketide intermediate and an ACP-bound (2*S*)-methylmalonyl-CoA substrate with an inversion of stereochemistry at the carbon atom derived from C-2 of the extender unit; the acyl-carrier protein (ACP) which carries the growing polyketide from domain to domain via its phosphopantetheine arm; and the acyltransferase(AT) domain, which loads the proper extender unit substrate, methylmalonyl-CoA, onto the arm of the ACP.[26] Once the β-ketoacylthioester is on the ACP domain, several

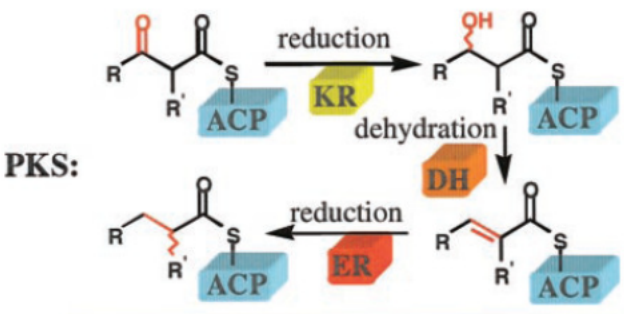

Figure 4: Simplistic overview of functions of the KR, DH, and ER domains

modifications can be performed on the polyketide via three more domains. These three domains are: a ketoreductase (KR) domain, which catalyzes the reduction of the β-ketoester to a β-hydroxyester; a dehydratase (DH) domain, which dehydrates the β-hydroxyester to form a cis or trans double bond; and an enoylreductase (ER) domain, which reduces the double bond. In addition, these domains do not have to modify the polyketide chain for every cycle; however, if these domains do act, then the modifications are stereospecific, meaning that different stereocenters can be produced in the final polyketide product.[8, 26, 30] As a result of the various stereocenters that can be produced along with the various number and type of extender units that can be used in each elongation step, various polyketides with many properties can be produced. This, in turn, is what fuels every polyketide researcher's ambition to build a polyketide library.

Eventually, the polyketide will reach module 6, the last module of DEBS, and when it is ready to leave the PKS, it will be attached to a serine on the thioesterase (TE) domain on the C terminus. This domain then catalyzes the cyclization and release of the polyketide in order to form the aglycone precursor. Afterwards, additional complexity may be introduced by "post-polyketide" enzymes which carry out oxidations, alkylations, glycosylations, and other transformations.[12]

One more important thing to note about this whole process is that the polyketide intermediate is passed from one module to the next while remaining covalently linked to the enzyme complex throughout the whole process. By doing so, the reactive polyketide intermediates are protected against potential degradation, and thus can become stable products.[26]

III. Assigning Stereochemistry to Polyketides

Before we look into the structures of various ketoreductase domains and acyl-carrier proteins when a substrate is incorporated into the active site, we must first understand how to assign stereochemistry. Generally, stereochemistry is assigned by giving priority to substituents of a stereocenter with larger substituents getting higher priority. Then, the chirality is determined by going from the lowest priority substituents to that of the highest priority, with the lowest priority being behind the plane. If one goes clockwise, then the stereocenter is an R. On the other hand, if one goes counter-clockwise, then the stereocenter is an S. Meanwhile, priority for polyketides is assigned differently. When looking at the chirality at the α-position, the hydrogen is given the lowest priority followed by the α-substituent. Meanwhile, for the chirality at the β-position, the hydrogen is once again given the lowest priority followed by the substituents at the γ-position. Once the order of priority is determined, the chirality, R or S, is determined just as stated before.[14]

IV. Ketoreductase domain of DEBS

The ketoreductase (KR) domain (expected M_r of eryKR1, 54,900; observed apparent M_r of eryKR1, 58,800) is a member of the SDR family of NAD(P)H-linked ketoreductases, in which proton transfer is thought to occur via a tyrosine to the carbonyl oxygen of the substrate while hydrogen bonds from a serine and lysine stabilize the polyketide intermediate. Also, the enzymes in this family have a well-defined Rossmann fold.[2] In addition, KRs are responsible for reducing the β-ketoester polyketide intermediate to a β-hydroxyester.[30] However, in the process of doing so, a chiral center is produced, which in turn may lead to the production of a wide variety of

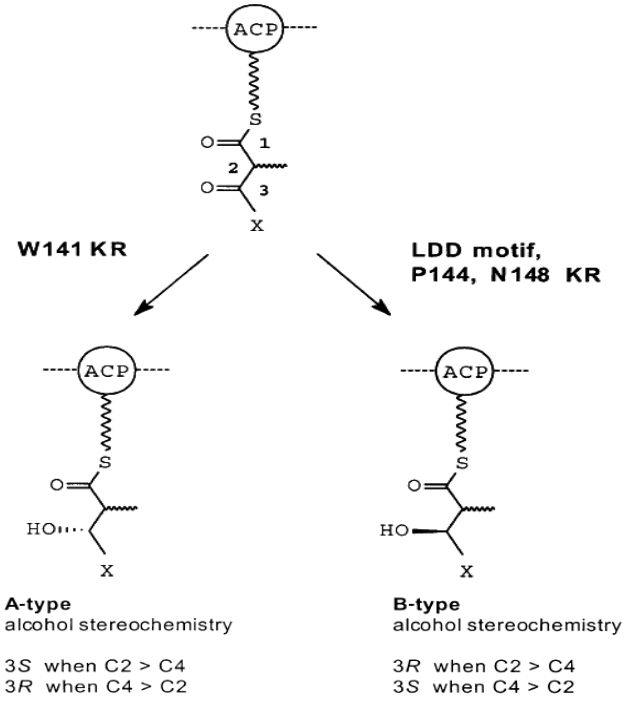

Figure 5: Important residues that are characteristic of A- and B- type KRs are shown here. Furthermore, the process by which stereochemistry is assigned to the C3 of both types of KRs are shown here. Note: The previous nomenclature has been expanded into A1-, A2-, B1-, and B2- type KRs. Also, a new class, C1- and C2-, has been created. Examples of these classes will be given later on in this paper.

polyketides. It is this diversity that drives researchers to study the KR domain.

Previous domain exchange experiments have shown that some DEBS KR domains are able to discriminate between the (2*R*)- or (2*S*)-methyl-3-ketoacyl substrates, and can then reduce these intermediates to one of two KR-intrinsic alcohol stereochemistry. Currently, Caffrey, et al. has named these classes "Type A and B." Type A KRs produce a (3S) isomer. In other words, they generate alcohols with the same alcohol stereochemistry as the (3S)-3-hydroxy-acyl-CoA intermediates in fatty acid breakdown. Meanwhile, B-type KRs are exactly the opposite and produce products with the same alcohol stereochemistry as the (3R)-3-hydroxyacyl-ACP chains in fatty acid biosynthesis. (**Figure 5**) With regards to residues found in A- and B-type KRs, there are very few consistent differences found between these two types. B-type KRs contain a LDD motif from residues 93 to 95. In this motif, D95 is strictly conserved. B-type KRs also differ from that of A-type in residues 141 to 148.

These aforementioned residues are all located close to active site residues and may contribute to catalytic activity. Recently, Reid et al. has shown that S136 and Y149 are important for DEBS KR6 activity by replacing Y149 with F and obtaining no activity. These same residues are found in DEBS KR1. It was then suggested that Y149 donates a proton to the β-carbonyl oxygen atom, which in turn stabilizes the substrate for attack by NADPH at the carbonyl carbon.

Even with all these previously mentioned residues, one can not always accurately predict what type of alcohol will be produced; however, the major indicators are the W141 and LDD motif for A-type and B-type, respectively. Recently, an expanded nomenclature for KRs has been implemented, and these KRs can now be classified as A1, A2, B1, B2, C1, or C2-type KRs. These new types of KRs will be described more in detail later on in this paper.[13] With regards to how A- and B- type KRs use NADPH to reduce the polyketide, experiments were performed on DEBS KR1, KR2, KR5, and KR6. It was then unexpectedly found that both types of KRs used the 4-*pro*-S hydride ion of nicotinamide adenine dinucleotide phosphate (NADPH). As a result, researchers have proposed three different models to account for how these two types of KR can produce polyketides of different chiralities. In the first model, Reid et al. proposed that the relative positions of the nicotinamide ring and Y149 are fixed. Therefore, the difference in stereochemistry between the two types of KRs comes from the polyketide substrate entering the KR active site from opposite directions. This model then gave rise to another model which states that the differences in stereochemistry for A- and B- type KRs comes from dimeric KR module being structured so that the both KR domains are oriented 180° with respects to each other. Specifically, it is believed that certain residues

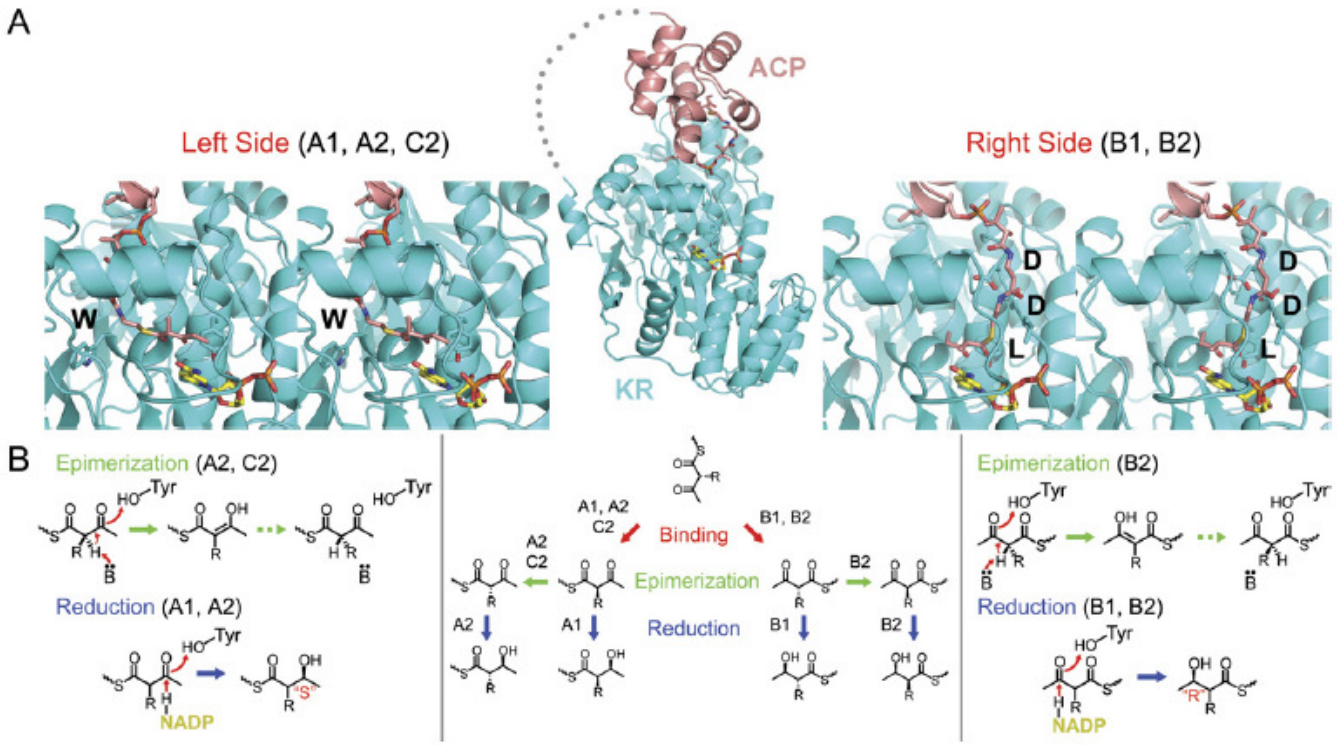

Figure 6: KR catalysis
(A) For A- and C2- type KRs, the phosphopantetheinyl arm slips behind the lid helix into the active site groove so that it can contact the conserved tryptophan. By doing so, the polyketide enters from the left. Meanwhile, for B- type KRs, the arm tries to slip into the active site groove from the left side; however, this time there is a LDD motif, which prevents the arm from going any further. Consequently, the polyketide can only enter the active site from the right.
(B) The mechanism by which polyketides epimerize and are reduced for A-, B-, and C- type KRs are shown here. It is believed that for epimerization, the polyketide can tautomerize to either the epimerized or nonepimerized forms. If the polyketide is epimerized, it is then reduced since the α-hydrogen can not be plucked by the base to undergo epimerization again. However, if the polyketide tautomerizes back to the nonepimerized form, then the epimerization reaction can reoccur.

such as the LDD motif might favor the polyketide intermediate going to the KR of one direction while other residues like P144 and N148 cause the intermediate to go to the KR of the opposite direction. In the last model, McPherson et al. proposed that the conserved residues of the two types of KR cause the nicotinamide ring and the Y149 to orient in different directions. As a result, reduction of the substrate occurs at two different orientations, which in turn gives rise to two stereochemically different products.[4] In either case, all three models support the idea that alternative modes of binding to the active site are responsible for different stereochemistries, and researchers have shown that the energetic differences between these two binding modes are very small.[26] Currently, polyketide research has reached the point where domain swapping experiments to produce products of different stereochemistry and domains can now be expressed as soluble and functionally active proteins in *E. coli* as long as domain boundaries have been correctly selected.[16] As a result of these advancements, researchers have been able to measure the kinetics and stereochemistries of eryKR1, KR2, KR5, KR6, and tylosin KR1. Also, models as to which diketide substrates enter into the active site are being proposed. So far, results have shown that at least some isolated KR domains are intrinsically capable of showing high specificity and stereospecificity towards a mixture of substrates while other KR domains are less

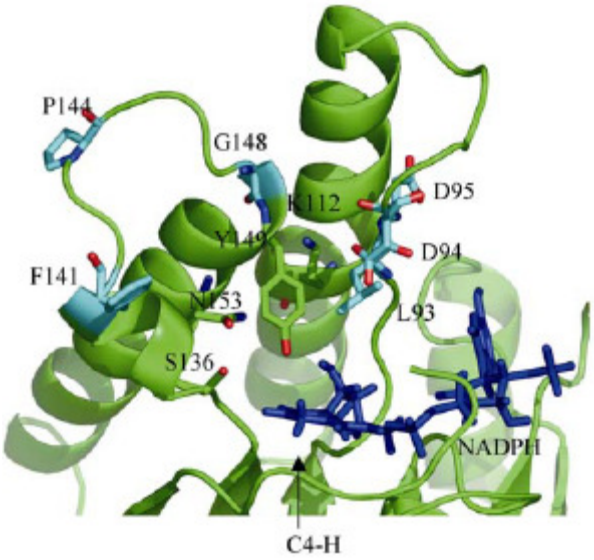

Figure 7: Homology model of eryKR1. All the conserved amino acid residues from the LDD motif to the residues on the loop containing Y149 are shown here.

discriminate.[31] Furthermore, it has been shown that one can not assume a KR in a hybrid PKS will maintain its intrinsic stereocontrol, and thus, the chiral centers of the product may not always be predicted as seen when the KR is confronted with an unnatural substrate.[26]

V. EryKR1

Once the (2R)-2-methyl-3-ketoacyl chain is produced, an epimerization reaction occurs before the diketide enters the eryKR1 domain to be reduced. Exactly how this epimerization reaction works is unknown as of now; however, it has been suggested that the KR1 and KR3 domains may play a role in the methyl group epimerization.

More details about the possible mechanisms will be described later on in this paper.[7] It also turns out that even if the (2R)-2-methyl-3-ketoacyl chain is not epimerized before entering the KR1 domain, the KR1 domain is still able to select the 2R enantiomer and then reduce it to the (2S, 3R) product. Furthermore, by selecting the (2S) enantiomer, the KR is able to control the stereochemistry of the methyl

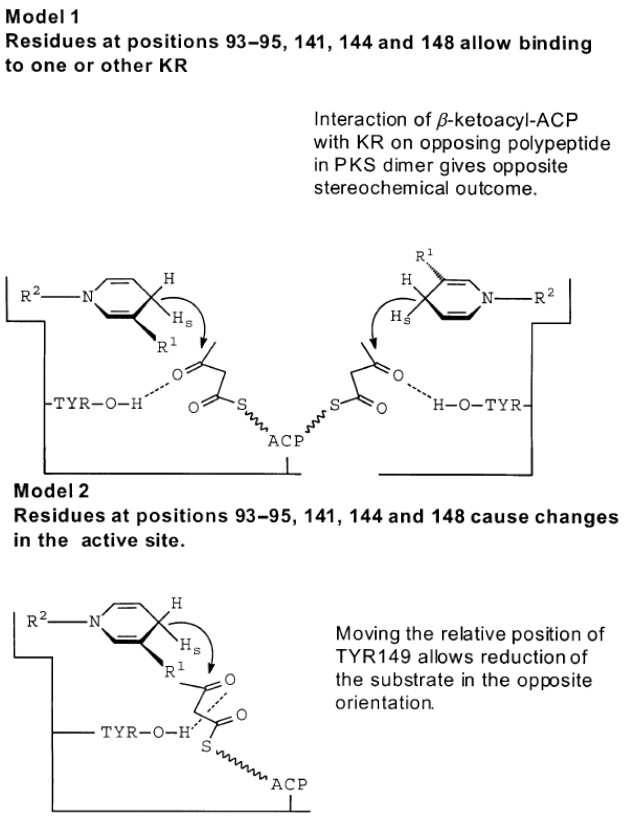

Figure 8: Models of how different types of KRs use NADPH to reduce polyketide intermediates are shown here. Model 1 is derived from the model proposed by Reid et al. Meanwhile, model 2 is proposed by McPherson

group at C-2. With regards to how a polyketide is reduced by eryKR1, the mechanism is the same as that found in B2-type KRs. In these types of KRs, the unepimerized polyketide is guided by the leucine on the LDD motif to enter the active site groove from the right. Meanwhile, a tyrosine from the αF helix abstracts an acidic α-hydrogen in order to enolize the polyketide. This is all made possible due to the proline on the αF helix, which causes the tyrosine to be free from the helix and thus have more flexibility. The enolized polyketide then tautomerizes back to the keto form in order to generate the epimerized polyketide. This polyketide then stays in this conformation because the α-hydrogen is now in position so that it is inaccessible to the aforementioned tyrosine. Once the epimerization is completed, a leucine or glutamine along with the lid helix work together to select the

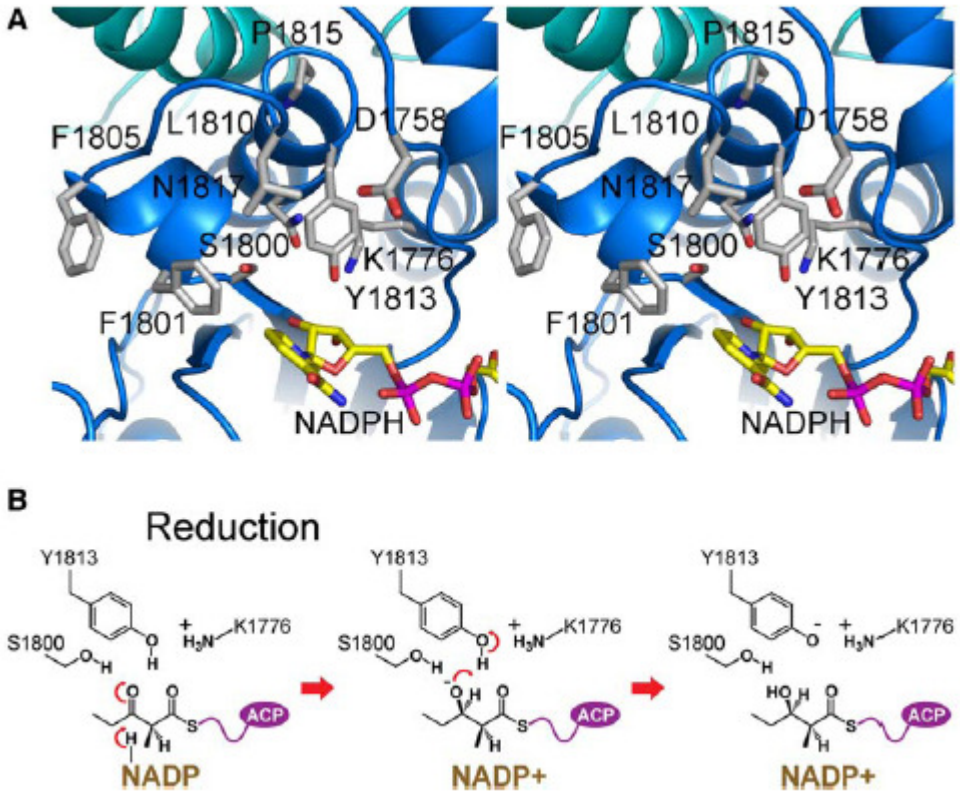

Figure 9: Crystal structures of eryKR1.
(A) All the important catalytic amino acid residues along with the NADPH in the active site groove are shown here.
(B) The mechanism by which eryKR1 reduces a polyketide to produce the "R" product is shown here.

epimerized polyketide for reduction to the "R" hydroxyl product. In addition, the leucine and the valine from the lid helix create a hydrophobic pocket in order to stabilize the epimerized α-methyl group.

Several experiments have also been performed on eryKR1 to determine which residues are critical to catalytic activity. In particular, all three residues in the LDD motif have been mutated, and it was found that L93 was the least tolerant to mutation, which makes sense since it is only 4Å from the catalytic Y149. All of these mutations were also found to reduce and increase the yield of the (2S, 3R) and (2R, 3S) isomers, respectively.[26] Furthermore, mutation of two of these residues was found to cause a change in alcohol stereochemistry in eryKR1.[2] Due to the high cost of CoA thioesters, methods have been developed to synthesize thioesters of N-acetylcysteamine as acyl donors for PKSs.[6, 12]

These newly synthesized thioesters have then been commonly used in the research lab. For example, Siskos et al. has examined what reduced products are produced by eryKR1 from a racemic mixture of (2R, S)-2-methyl-3-ketodiketide N-acetylcysteamine thioester (SNAC), and it was found that eryKR1 reduced the mixture with complete stereoselectivity and stereospecificity, even though the SNAC was not attached to an ACP or an intact PKS. Specifically, it was found that eryKR1 selected the 2S enantiomer and reduced it exclusively to *syn*-(2S, 3R)-diketide-SNAC.[7, 26, 31] Nevertheless, there are drawbacks to using these substrates when dealing with multiple rounds of mutagenesis of PKS domains. Specifically, enzymes may deal with these surrogates differently than they would with their natural substrates[26]

With regards to the future uses of these newly created KRs, it still remains to be seen if

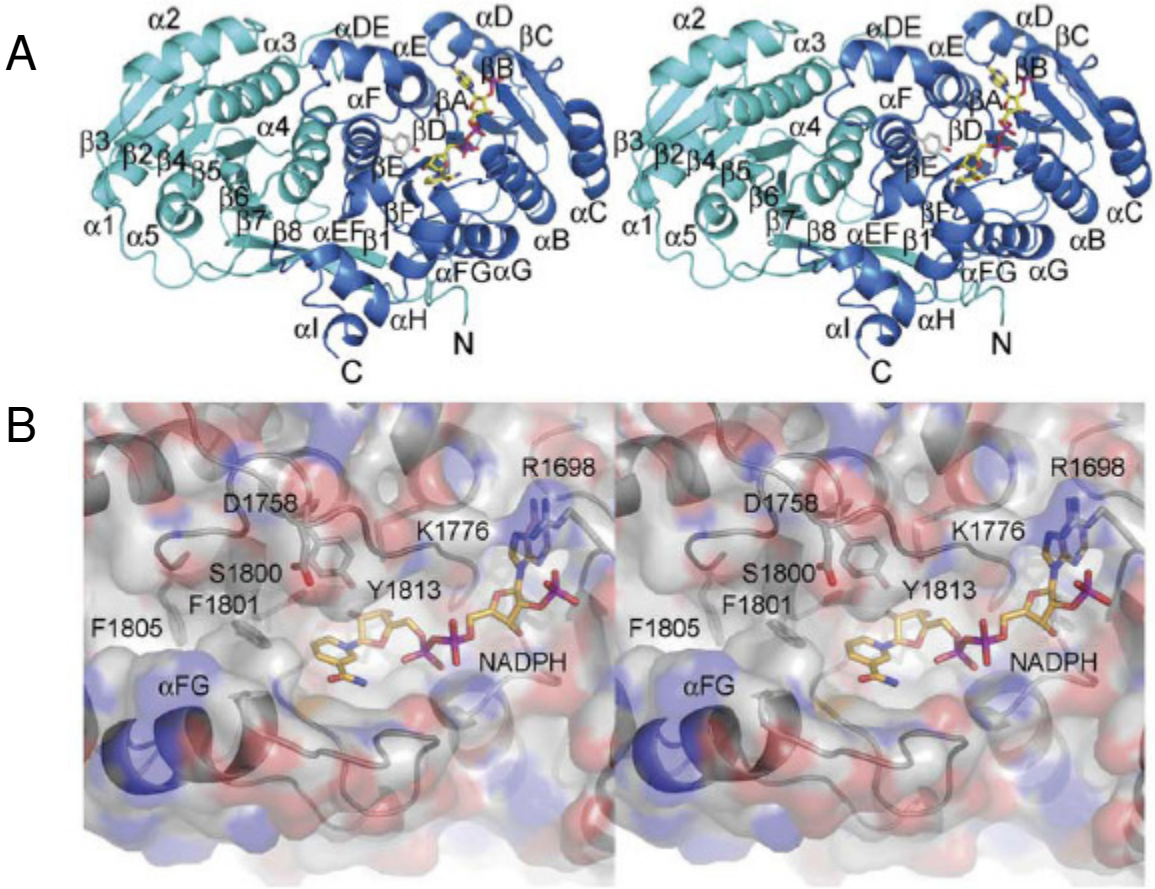

Figure 10: Crystal Structure of EryKR1.
(A) Stereodiagram showing the secondary structural elements of the KR domain. NADPH and the catalytic tyrosine are represented as sticks.

(B) Stereodiagram of the active site groove. All the important catalytic residues along with NADPH are shown here. Also, the α FG helix is believed to be involved with epimerization of the polyketide. Also, F1805 prevents the polyketide from entering the groove from the left.

these altered properties will stay when inserted into modular PKS multienzymes. If they do, then these new KRs can be used to expand our current library of polyketides.

VI. Linkers and Crystal Structure of EryKR1

In a PKS module, intrapolypeptide and interpolypeptide linkers make up a significant fraction of the total protein. Specifically, intrapoplypeptide linkers are composed of about 20 amino acids and separate an ACP of one module from a KS of the next. Meanwhile, interpolypeptide linkers are about 80-130 amino acids at the C-terminus of one module and interacts with a 30-50 amino acid sequence at the N-terminus of downstream module. From these linkers, module-module interactions can then occur via coiled coil interactions between the helical regions of these linker domains.[1, 22] Also, domains within the same module can interact with each other as seen when KS, AT, and KR domains interact with each other via inter-domain linkers. However, there are some linkers like the N-terminal coiled coil linker of module 5 that don't engage in protein-protein interactions with the remainder of the module and can thus be interchanged with other linkers without loss of function.[16]

As a result of the possible impact of linkers, many researchers have constructed hybrid synthases with or without the large interdomain linker. It was determined that the PKS required

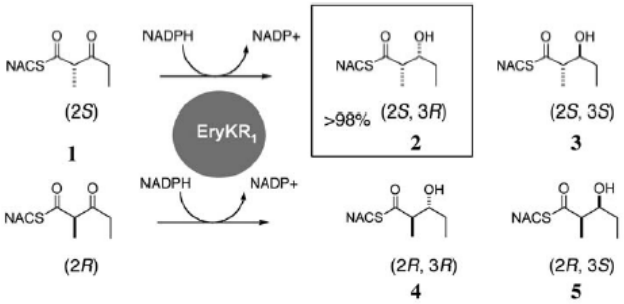

Figure 11: Stereochemistry of products produced by EryKR1. As can be seen here, the 2S enantiomer of a racemic mixture is stereospecifically selected by EryKR1 and the product is almost exclusively the (2S, 3R) enantiomer. Meanwhile, all other products are minor.

the linker in order to be present. More research has also been performed by Keatinge-Clay et al. who have determined the structure of DEBS KR1 with the large interdomain linker to 1.79Å resolution. The KR was found to be monomeric but composed of two subdomains, where each subdomain resembled an SDR monomer and were found to function as either a structural or functional subdomain. Specifically, the structural subdomain starts off with the bridging β-strand β1, which contains the sequence (H/L/M/F/Y) XXXW. This sequence is important since there is hardly any sequence conservation in this subdomain.[15] The only conserved sequences in this subdomain are T1572, G1594, E1602, R1649. Also, interdomain linkers were found to be a part of this domain. This structural half then ends with a second bridging strand, β8, which goes towards a loop that begins the Rossmann fold of the catalytic subdomain. Here, the KR domain and the 70 residues C-terminal to it comprise the catalytic subdomain and contain the catalytic residues found in SDR enzymes. This subdomain is then followed by a series of short helices, which ends 12 residues from the ACP.[15]

In the crystal structure of EryKR1 with NADPH, the adenine ring of NADPH stacks against R1698 and hydrogen bonds with D1726 and V1727. Meanwhile, the phosphates on the NADPH hydrogen bond to S1699 and form a salt bridge with R1698. With regards to the catalytic tyrosine, Y1813, and serine, S1800, they were positioned in the same orientations as that found in SDR enzymes. However, Y1813 is not in helix αF as seen in other SDR enzymes due to a P1815. Also, compared to other SDR structures, K1776, and N1817 have swapped positions. However, the amine in K1776 still hydrogen bonds with the backbone carbonyl that is 2 residues N-terminal of the catalytic serine. Meanwhile, the N1817 side chain continues to hydrogen bond to the Y1813 backbone carbonyl.[15]

In the Keatinge-Clay lab, the ternary complex of eryKR1 in the presence of NADPH and 2-methyl-3-oxopentanyl-CoA and 2-methyl-3-oxopentanyl-diketide is of interest and current progress on this project will be reported later on.

VII. Important Residues for "*R*" and "*S*" Reduction

Several residues have been found to help researchers predict which stereochemistry, *R* or *S*, will be found on the alcohol of the final product. For the "*S*" reduction, a tryptophan was found in the active site groove. Keatinge-Clay et al. hypothesize that the tryptophan may help orient the backbone of the protein so that the glutamine NH_2 can hydrogen bond to the thioester carbonyl, and thus guide the polyketide into the groove from the left side.[15]

Meanwhile, "*R*" reduction, as seen in eryKR1, was found to have a conserved D1758 and no tryptophan. Currently, the mode of action by which D1758 guides the polyketide is unclear. Furthermore, the glutamine found in "*S*" reduction is replaced with a leucine here, which cannot help guide the polyketide to come in from one or the other side. However, there is a phenylalanine (F1801), which is thought to help with moving the polyketide from the right side. Specifically, Keatinge-Clay et al. believe that due to phenylalanine's size, the steric effects block the

"S" reaction and instead cause the "R" reaction to occur via hydrophobic interactions.[15]

VIII. Epimerization:

As mentioned before, the actual mechanism behind the epimerization of the α methyl group is not known. However, it is believed that the KR domains of the first and third modules are responsible for this epimerization since 1) no module without a KR was able to catalyze an epimerization 2) A module with an inactive KR and no other β-processing enzymes was still able to catalyze an epimerization 3) Tyrosine is conserved in all the KRs that catalyze epimerizations. It is also believed that the epimerase activity is independent of NADPH since module 3 which contains an inactive KR that does not bind NADPH can still epimerize the polyketide intermediate.[15]

In the first module, the catalytic tyrosine, Y1813, lies outside the αF helix due to the presence of P1815. As a result, the tyrosine is hypothesized to have more freedom to move and thus, it can pluck the acidic hydrogen on the α-carbon in the polyketide intermediate. The intermediate can then tautomerize to one of two different forms. In one form, the intermediate is now epimerized. In the other form, the intermediate is the same as the original substrate and consequently goes through the same reaction until it tautomerizes to the epimerized form.[15] For the third module, N1817 is replaced by a serine. As a result, the αF helix may have more freedom for the tyrosine to move and epimerize the polyketide intermediate.[15]

With regards to the absence of a KR, it is believed that the DH domain catalyzes the epimerization reaction since its mechanism involves the abstraction of an α-proton.[15]

IX. ACP domain of DEBS

In each of the six modules along with the loading domain, there is a75- to 90- amino acid acyl carrier protein (ACP) domain (approximately 10 kDa), which interacts with all other domains during polyketide biosynthesis. The overall structure of the ACP domain is comprised of a right-hand twisted bundle made up of three major α-helices connected by two loops. Helix I (residues 14-32) runs anti-parallel to helices II (residues 55-68) and III (residues 83-94) and lies at an angle relative to helices II and III. In addition, helices II and III are also at an angle relative to each other. Furthermore, residues 14-42 are structured but less well defined than residues 43-94. Also, this helical bundle (residues 76-79) is stabilized by hydrophobic interactions between the core leucine, valine, and alanine residues.[1]

Meanwhile, as stated before, the purpose of this domain is to participate in the elongation steps of polyketide biosynthesis and to carry the polyketide intermediate from one domain to the next. Specifically, it accomplishes this goal by using a 20 Å-long phosphopantetheinyl group derived from CoA covalently attached to a conserved serine residue. This flexible arm, which interacts only transiently, if at all, with the polypeptide portion of the ACP, is then covalently bound to the monomer units and the growing acyl chain intermediates via a cysteine thiol group on the end of the arm.[1, 6, 20]

While attached to the phosphopantetheinyl-arm of the ACP, the configuration of the polyketide intermediate is stable and spontaneous lactonization or other intramolecular reactions are prevented. One explanation for this stability is that the ACP-bound polyketide is bound by the ACP itself and consequently does not swing freely in solution. Another explanation is that a kink is present in the polyketide chain and therefore prevents the β-ketone and thioester carbonyl groups from becoming coplanar. This in turn raises the pK_a by 10-15 orders of magnitude, which is enough to prevent the loss of stereochemistry. A "switch-blade" model has also been proposed to explain this stability. In this model, it is believed that the phosphopantetheinyl residue folds back on the

protein so that the attached polyketide is bound in a groove in the ACP and then flickers out of this groove when bound by another domain.[7]

In addition, all seven DEBS ACP domains appear to have some sequence similarity (45%-55%) and are predicted to have a 4-helix topology. However, there is growing evidence that these domains are not interchangeable and that each ACP has a preference for the KS of the same module.[7] Also, comparison of the steric and electrostatic surfaces at the interaction interface of the ACP domain suggests that protein-protein recognition is present.[1] With regards to ACP1, a D53 and a E59 are both found in the N-terminal half of helix II. As a result, these two negatively charged residues give ACP1 a negative electrostatic potential surface, which appears to be involved in conserved domain-domain binding interactions; however, it is believe that this surface is not likely to contribute to determining specificity of substrates. In addition, ACP1 contains a leucine at position 58.[1] Meanwhile, when dealing with research on how various ACP-bound substrates are recognized by KS domains, these various substrates must be posttranslationally phosphopantetheinylated in order to be functional.[5] In order to do so, a Sfp phosphopantetheinyl transferase is used to load these malonyl-CoA analogues directly onto the ACP. As a result, one is able to bypass the strict substrate specificity of the AT domain, which naturally loads the substrate onto the ACP.[6] Last but not least, much research has been conducted on whether KR domains have any intrinsic specificity towards the ACP domain of the same module. It was found that for the most part no intrinsic specificity existed (Results have shown that DEBS KR2 and KR6 had a lack of discrimination for their corresponding ACP domains) and that any sensitivity it did have had no observable effect on the overall rate of formation of the reduced diketide. This then suggests that the phosphopantetheinyl arm is enough for substrate recognition by the KR.[9]

One project currently being worked on in the Keatinge-Clay lab is purifying ACP domains of various polyketide synthases and co-crystallizing these domains with other isolated domains such as KR and DH domains. As of now, TylACP1, EryACP1, PikACP3, AmpACP2, and SpiACP3 are being isolated and purified.

B. TylKR1

The precursor to the antibiotic tylosin, tylactone, is synthesized by *Streptomyces fradiae* by a tylosin polyketide synthase, composed of seven homodimeric modules that are organized into five large proteins with subunits weighing between 170 to 460 kDa.[8] Specifically, this synthase condenses one ethylmalonyl, four methylmalonyl, and two malonyl extender units onto a priming propionyl unit to produce the 23-carbon tylactone. In addition, this PKS contains the same domains as found in DEBS; however, there are a few differences. One difference is that tylosin PKS can select the 2*R* enantiomer from a racemic mixture of polyketides or (2*R, S*)-2-methyl-3-ketodiketide N-acetylcysteamine thioester (SNAC) and then reduce the enantiomer to the *anti*-(2*R*, 3*R*) corresponding product.[26, 31] Another difference is that the KR2 and DH2 domains from tylactone PKS help produce a tri-substituted double bond on the final tylactone product while erythromycin PKS does not introduce double bonds onto the final 6-DEB product.[8] Just as seen in DEBS, many researchers are interested in the KR domain due to its ability to create polyketides of different alcohol and α-substituent chiralities. However, tylosin KR1 differs from DEBS KR1 in several aspects. First of all, tylosin KR1 is a B1-type ketoreductase. This type of ketoreductase uses a leucine on the LDD motif to guide the unepimerized polyketide into the active site groove.[13] Glutamine then works with the lid helix residues (residues on the α-FG helix and the loop preceding it) in order to prevent the polyketide from spontaneously epimerizing before reduction can occur. The tylosin KR1 then reduces the polyketide to give a "*R*" hydroxyl group. In addition, the N-terminal end of the lid helix is 7 Å closer to the active site than seen in EryKR1. Also,

the residues on the LDD motif and the conserved M417, found on the loop preceding the lid helix, help to make better crystal contact.

With regards to similarities between TylKR1 and EryKR1, there are a few. Specifically,

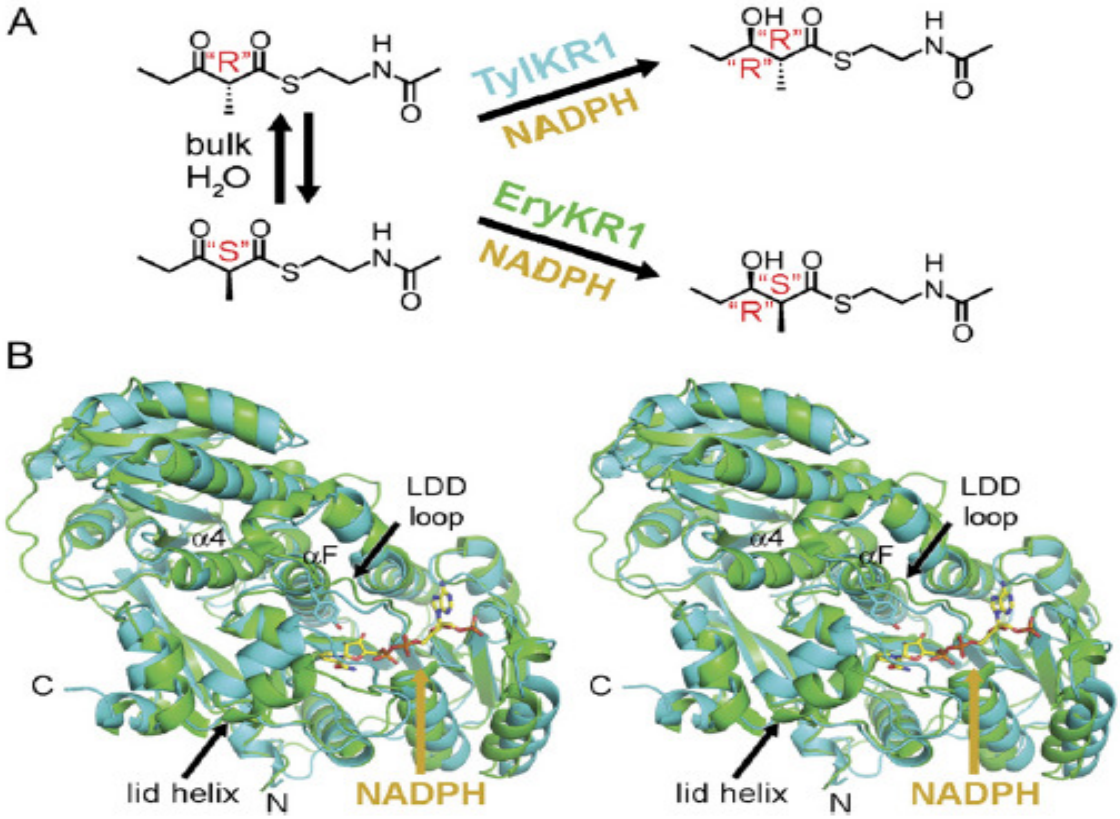

Figure 12: Tylosin and Erythromycin KR1.
(A) Stereochemistry of the products produced by TylKR1 and EryKR1 are shown here. Also, both ketoreductases can stereospecifically select the correct enantiomer to reduce from a racemic mixture.(B) Superimposition of TylKR1 (blue) over EryKR1 (green) is shown here. It can be seen here that there is a 7Å shift of the N-terminal end of the lid helix. Also, the NADPH is from the EryKR1 structure.

in the catalytic domain, the catalytic Y383 and conserved S370 both position the polyketide carbonyl to be reduced adjacent to the NADPH hydrogen just as seen in EryKR1. A K345 then causes Y383 to donate its proton to the carbonyl oxygen after hydride transfer. Meanwhile, another similarity involves the residues on the lid helix and their interactions with the active site. In particular, a E424 hydrogen bonds with Q380 in what is deemed the primary determinant of the stereochemistry of the α-substiuent; L426 contacts L411 and L431; and the carbonyl group of G422 hydrogen bonds with N377. Furthermore, the residues on the LDD motif are the same as that seen in EryKR1. The first aspartate caps the αF with the amide of the catalytic tyrosine while the second aspartate caps the lid helix.[13]

With regards to the model of TylKR1, Keatinge-Clay et al. have seen several favorable interactions between TylKR1 and the α-hydrogen and α-methyl group of the polyketide intermediate. In particular, E424 on the lid helix was found to hydrogen bond with the partially positive α-hydrogen. Meanwhile, the γ-methylene unit of E424 stabilizes the α-methyl group of the polyketide via hydrophobic interactions. Because of this ternary complex, an epimerized diketide can be prevented from forming since epimerization would lead to steric hindrance

between the α-methyl group and the E424 carboxylate. With regards to the interactions between TylKR1 and TylACP1, Keatinge-Clay et al. have used FAS ACP as a model to dock TylACP1 to TylKR1 while keeping the N terminus of TylKR1 close to the C terminus

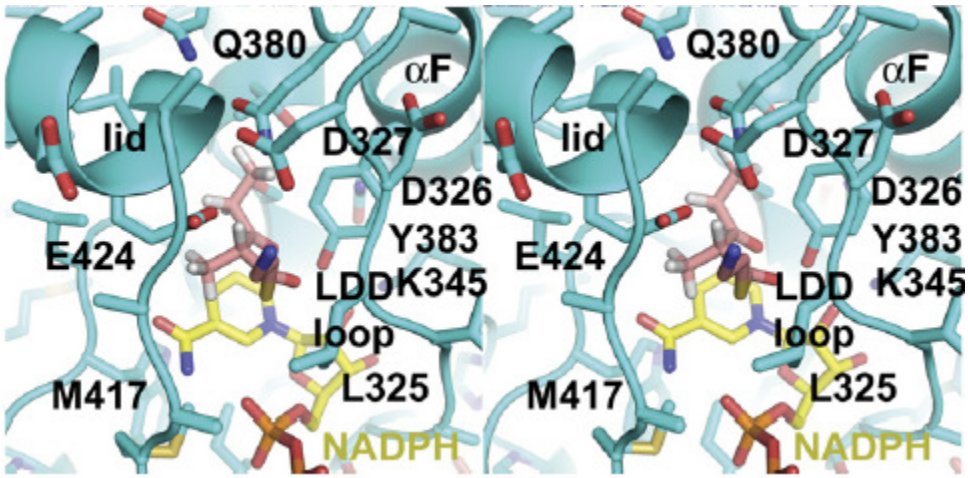

Figure 13: TylKR1 active site. NADPH and the diketide substrate are in the active site here. The L325 is believed to help guide the diketide into the active site groove from the right by making hydrophobic interactions with the phosphopantetheinyl arm. Also, an epimerization reaction may occur before reduction by NADPH. If the diketide has epimerized, then the α-metyl group may clash with the E424 carboxylate, which in may turn prevent the KR from binding the now epimerized diketide.

of the TylACP1. This was possible thanks to the approximately 20 residues that separate these two domains. It was then seen from this that the phosphopantetheinyl arm was able to enter the active site groove from the right side. From these results, the process by which the LDD motif guides polyketides has been elucidated. Specifically, it is believed that the conserved second aspartate on the LDD loop caps the lid helix. Thus, the phosphopantetheinyl arm can not slip between the LDD loop and the lid helix. Instead, the arm has to approach the active site groove via the other direction (right side) so that its hydrophobic end interacts with the leucine. Meanwhile, for A- and C2-type KRs, there is a conserved tryptophan instead of the LDD loop. As a result, the arm can slip between the lid helix and the LDD loop that were present in the B-type KRs and enter the active site groove. Meanwhile, the hydrophobic portion of tryptophan interacts with that of the arm in order to stabilize it.[13] Currently, the Keatinge-Clay group is interested in obtaining the ternary complex of TylKR1. Progress on this project will be discussed later on in the paper. Also, another project that will be discussed later in this paper involves TylKRACP1.

C. Pikromycin

Another macrolide antibiotic of interest to researchers is pikromycin, which is produced by the pikromycin PKS found in *Streptomyces venezuelae*. This PKS is composed of a loading domain and six extension modules found on four separate polypeptides (PikAI – PikAIV). Meanwhile, on each polypeptide is some combination of KS, AT, and an ACP along with optional KR, DH, and ER domains just like in DEBS. However, there are several key differences. One difference is that modules 5 and 6 are on separate polypeptides (PikAIII and PikAIV, respectively). In addition, a second TE domain is found on the PikAV polypeptide.[9] Furthermore, later modules were found to extend and process diketides nearly 3 orders of magnitude less efficiently than found in DEBS.[24]

The unique ability of pikromycin PKSs to produce two different polyketides under different environmental conditions has also been of

interest to fellow polyketide researchers. Specifically, these PKSs can produce a 14-membered ring macrolactone narbonolide and a 12-membered ring product 10-deoxymethynolide. Specifically, the 10-deoxymethynolide is produced in standard culture media while the narbonolide is produced in pure growth media.

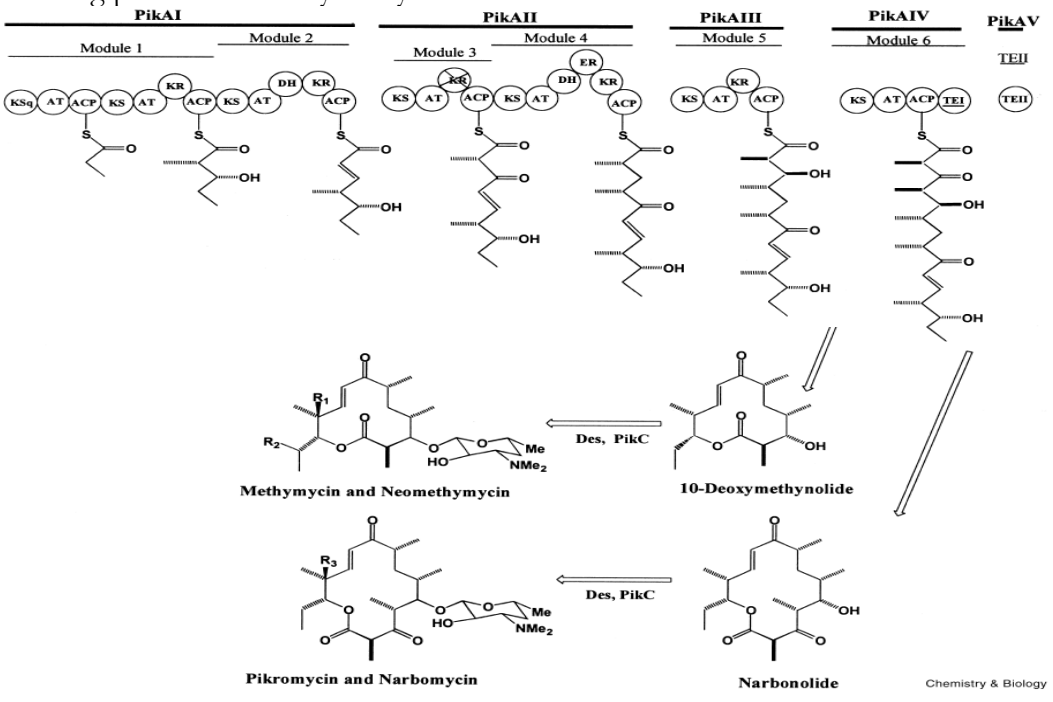

Figure 14: Domains of each module of Pikromycin PKS. The two possible products that can be produced by this PKS along with their derivatives are also represented here.

It is currently believed that the narbonolide is produced by having the polyketide go through all six modules before being cut off by the TE while the 10-deoxymethynolide is produced due to a premature chain termination. It has been proposed that this early termination is due to an alternative start codon 600 nucleotides downstream of the normal *pik*AIV start codon. As a result, the N-terminal truncated PikAIV has only half of the KS_6 domain and therefore can not catalyze the final elongation step. As of now, there is evidence that does not fully support this model; however, the true mechanism is still unclear. It is believed though that the docking domain interactions between PikAIII and PikAIV play a vital role in the true mechanism.[10, 19, 33]

With regards to the KR3 domain found on PikAII, it is a C2-type ketoreductase. For these types of ketoreductases, an unepimerized polyketide is guided by a conserved tryptophan to enter the active site groove from the left side. An acidic α-hydrogen is then abstracted by either a lid helix residue or a water molecule in order to enolize the polyketide. Tautomerization would then occur to yield the epimerized polyketide. Enolization is then prevented from occurring since the α-hydrogen of the epimerized polyketide is inaccessible to the base. In addition, the dinucleotide binding motif is covered and the conserved asparagine is generally replaced by a smaller residue.[13]

Previous research on pikromycin PKS has for the most part involved examining the catalytic domains of PikAIII and PikAIV and the interactions between these two polypeptides. However, not much research has been performed on PikAII, specifically the KR3 domain. As a result, one of the projects worked on by the Keatinge-Clay lab involves examining the conformational changes of the domain when a

diketide is present in the active site groove, and the results of this project will be presented later in this paper. Meanwhile, another project of interest is that of PikKSATACP$_0$ in the loading module. If we are able to obtain crystals are obtained of this module, then we will be the first group to have ever crystallized an entire module of a PKS with all of its domains intact.

D. Amphotericin

Another polyketide of interest is the polyene macrolide, amphotericin, produced by *Streptomyces nodosus*. This polyketide is an antifungal antibiotic, and is able to kill off some viruses and protozoans by disrupting cell membranes and forming channels that lead to loss of small molecules and ions. However, treatment with this antibiotic is complicated by its low water-solubility and severe nephrotoxicity symptoms. As a result, much research has gone into developing amphotericin derivatives that have either eliminated or at least reduced these symptoms.[3, 29] Currently, Power et al. have been able to produce four derivatives and several amphotericin A and B analogs; however, there are still two more derivatives that this group is trying to produce.

In order to understand more about amphotericin, the machine in charge of synthezing these antibiotics, amphotericin PKS, should be extensively studied. This PKS resembles that of DEBS in that there are three core domains (KS, AT, ACP, and a TE) along with several accessory domains (KR, ER, DH). In addition, several modifications occur after the biosynthesis of amphotericin B. These include the addition of a mycosamine sugar, hydroxylation at C8 and oxidation of the methyl group (C41) to a carboxyl group. However, there are differences between amphotericin PKS and DEBS. One big difference is that module 1 of amphotericin PKS generates a (2S, 3S)-2-methyl-3-hydroxy butanoic acyl thioester while DEBS module 1 generates a(2S, 3R)-2-methyl-3-hydroxy pentanoic acyl thioester.[3] Another difference is that amphotericin PKS can be bimodular, trimodular, hexamodular, or contain only one module. Currently, the largest amphotericin PKS is hexamodular. One more difference is that the end products of amphotericin PKS are the heptaene amphotericin B and the tetraene amphotericin A. The only difference between these two is that the C28-C29 double bond is reduced in amphotericin A.[3, 29]

With regards to previous research on

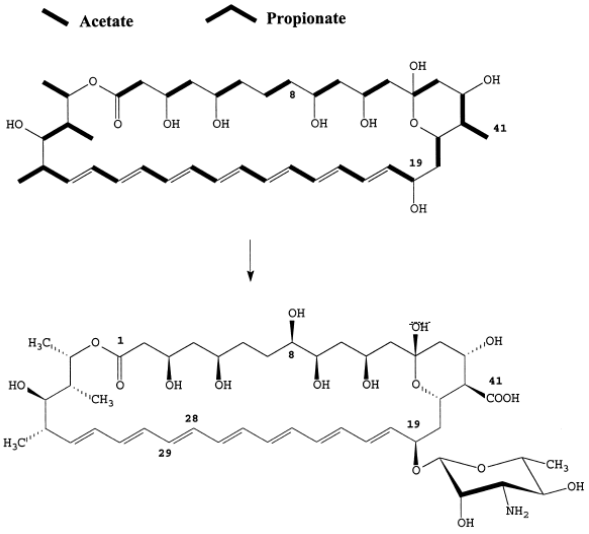

Figure 15: Biosynthesis of Amphotericin B. The positions where the acetate and propionate extender units are incorporated in amphotericin B are shown here.

amphotericin KR domains, KR1 was found to be vital for macrolactone formation, KR3-KR9 for biosynthesis of polyene units, KR10 for glycosylation, KR11 for formation of hemiketal ring, and KR2 for formation of stable transmembrane channel. In the Keatinge-Clay lab, the structure and function of amphotericin KR2 are of interest since little is known about the conformational changes of KR2 when various substrates enter into the active site. Amphotericin KR2 is an A1-type ketoreductase, meaning that a conserved tryptophan helps guide the unepimerized polyketide to enter the active site groove from the left. Afterwards, a glutamine interacts with the unepimerized α- substituent and a "S" hydroxyl group is produced. The conformational change that occurs in amphotericin KR2 when 2-methyl-3-oxopentanyl-

SNAC and NADPH enter the active site will be described later on in this paper.[13]

Currently, the Keatinge-Clay lab is very interested in obtaining the crystal structure of this PKS in the presence of a diketide. Specifically, one project being worked on involves obtaining the ternary complex of AmpKR2 along with NADPH and 2-methyl-3-oxopentanyl-CoA.

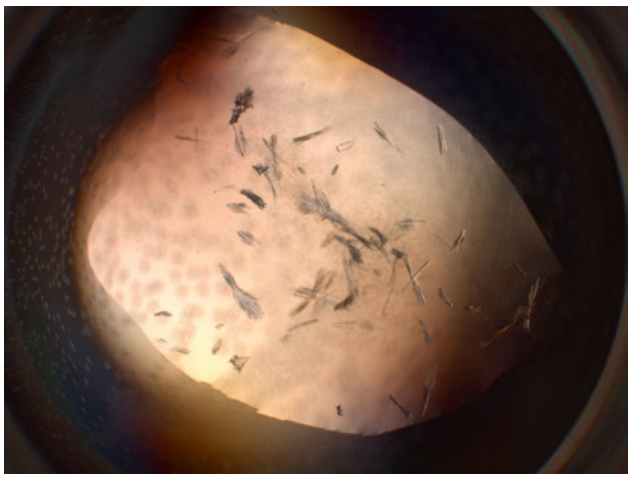

Figure 16: These are Tylosin KR1 crystals that are grown in 1.475 M ammonium sulfate, 0.1 M Tris-HCl, pH 8.0, and are found in a 1:1 drop ratio of protein to crystallization buffer.

III. Results and Discussion

A. Erythromycin Ketoreductase Domain 1

Previous work by Keatinge-Clay et al., has shown that erythromycin KR1 crystals grew around 35% PEG3350, 0.2 M guanidinium hydrochloride, and 0.1 M HEPES (pH 7.0).[15] As a result, the same conditions were set up here; however, the drop ratios were varied between 0.4 : 1 and 1 : 1 protein to crystallization buffer solution ratios and the concentration of PEG used ranged from 36.5% to 39%. No crystals were seen and we believed that the lack of substrate prevented the loop helix from being stabilized, which in turn led to no crystals being formed. As a result, erythromycin KR1 was grown with 5 mM NADPH and 5 mM 2-methyl-3-oxopentanyl-CoA under 96 different conditions and crystals were found to exist at 0.2 M ammonium acetate, 0.1 M Tris-HCl (pH 8.5), and 30% (v/v) isopropanol. EryKR1 was then grown under the same conditions with the exception of isopropanol being between 22.5 and 35% and the drop ratio being between 1:1 and 2.5:1. However, this time no crystals were found in any of the 24 wells. In future experiments, a higher concentration of substrates may be used and the Phoenix crystallization robot may be used again to provide us with possible conditions that EryKR1 may crystallize under.

B. Tylosin Ketoreductase Domain 1

Previous work by Keatinge-Clay et al. has shown that tylosin KR1 crystallizes at 1.5 M ammonium sulfate, 0.1 M Tris-HCl (pH 8.0), and with a drop ratio of 1:1 protein to crystallization buffer solution. From these crystals, the methionine residue on the loop was observed to have made contact with the 2-methyl-3-oxopentanyl-SNAC. However, whether this residue would also make crystal contacts with NADPH is unknown. Thus, the same crystallization conditions were set up except that the drop ratios of 1:1, 2:1, and 2.5:1 were used. Crystals were then seen between 1.45 M and 1.525 M ammonium sulfate at all drop ratios. These crystals were then soaked in 15 mM NADPH, 60 mM 2-methyl-3-oxopentanyl-SNAC, and 10% glycerol for 30 minutes and then sent off to the synchrotron. Unfortunately, no diffraction patterns were seen and efforts are currently being made to send more soaked crystals to the synchrotron.

C. Tylosin Ketoreductase-Acyl Carrier Protein Domain 1

Previous work in the Keatinge-Clay lab showed that the lid helix seemed to lift when the ACP came into the active site with the diketide. In order to confirm this, a tylosin KR1 directly attached to an ACP was expressed and purified. The Phoenix crystallization robot was then used; however, no crystals were produced. It was then

thought that possibly the KR domain may be interfering with the flexibility of the ACP. Consequently, this project was put on hold and efforts were placed on purifying and expressing ACP domains so that we could co-crystallize the ACP and KR domains together in the future.

D. Pikromycin Ketoreductase Domain 3

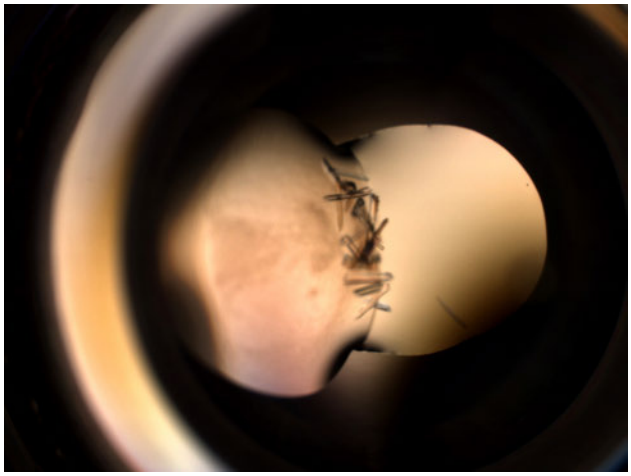

Figure 17: These are AmpKR2 crystals that were grown in 2.8 M ammonium sulfate, 20% glycerol, and were found in a 3:1 drop ratio of protein to well solution. In addition, these crystals were soaked with 10 mM $NADP^+$.

Crystallography trays were set up for pikromycin KR3 with a concentration of PEG4000 ranging from 31 to 35%, a drop ratio ranging from 1.4 : 1 to 2.4 : 1, 3% glycerol, 0.1 M Tris-HCl (pH 8), and 0.22 M sodium acetate. Crystals were seen a month later at in the well containing 34% PEG 4000, a drop ratio of 2.22 : 1, 3% glycerol, 0.1 M Tris-HCl (pH 8), and 0.22 M sodium acetate. A 24-well crystallography plate consisting of a drop ratio between 1.8 : 1 and 2.66 : 1 and a concentration of PEG4000 ranging from 33.5% to 34.5% along with 3% glycerol, 0.1 M Tris, pH 8, and 0.22 M sodium acetate was then set up to optimize these conditions. Currently, crystals have been seen in the well containing 33.5% PEG4000 and a drop ratio of 1.8 : 1; however, only two weeks have passed. Thus, more time will be given so that crystals in other wells may form. Another project being worked on involves soaking PikKR3 crystals with 2-methyl-3-oxopentanyl-phosphopantetheine. Currently, three crystals were sent to the synchrotron; however, there was mosaicity on two of the crystals. Meanwhile, the other crystal gave an approximately 3Å resolution. This resolution was not enough to see the electron density.

E. Pikromycin KS-AT-ACP_0

The loading module of pikromycin (Pikromycin KS-AT_0-ACP) has been purified several times and the FPLC chromatogram looks fairly clean. It can be found in **Figure 16**. The Phoenix crystallization robot was recently used and currently no crystals have formed.

F. Amphotericin Ketoreductase Domain 2

Previously in the Keatinge-Clay lab, amphotericin KR2 (AmpKR2) was grown in the presence of ammonium sulfate. However, due to the possibility of the thioester bond in 2-methyl-3-oxopentanyl-SNAC being cleaved by the amine group, amphotericin KR2 with 10 mM $NADP^+$ was initially grown in the presence of D,L-malic acid, pH 7.0. Specifically, trays were set up for AmpKR2 in the presence and absence of glycerol. In the absence, AmpKR2 was allowed to crystallize around 1.9 M D,L-malic acid, 0.1 M Sodium Cacodylate, pH 6.7, 0.2 M NaCl, and at approximately a 2 : 0.6 protein: well solution drop ratio. Meanwhile, the conditions in the presence of 5% glycerol were the same except that the well solution contained approximately 2.5 M D,L-malic acid. No crystals formed even when the conditions were slightly varied. As a result, we decided to try substituting ammonium sulfate for D,L-malic acid; however, before doing so, we had to determine how long it would take before the thioester bond of 2-methyl-3-oxopentanyl-SNAC would be cleaved. In order to accomplish this, a TLC and LC/MS experiment was performed. It turned out that after 30 minutes, the thioester bond still was not cleaved.

A detailed explanation of these results will be discussed later on. Crystallography plates were then set up with ammonium sulfate replacing

D,L-malic acid in the well solution. It turned out that AmpKR2 with 10 mM $NADP^+$ crystallized between 1.8 M and 2.4 ammonium sulfate, 0.1 M sodium cacodylate, pH 6.7. Meanwhile, in the presence of 20% glycerol, AmpKR2 and 10 mM NADP+ crystallized at 0.1 M sodium cacodylate, pH 6.7, 0.2 M NaCl, and between 2.6 and 3.2 M ammonium sulfate. These crystals were then soaked in 10% glycerol and 50 mM 2-methyl-3-

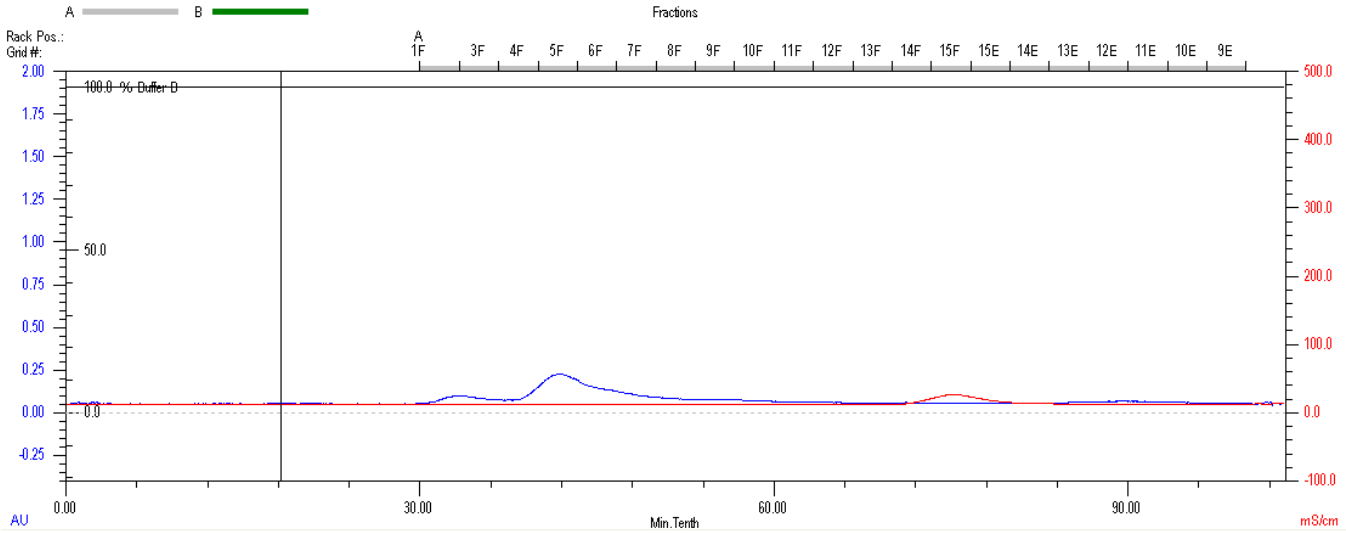

Figure 18: The FPLC chromatogram for PikKS-AT_0-ACP can be found here. As can be seen here, the chromatogram is fairly clean.

oxopentanyl-SNAC for 30 minutes and then sent to the synchrotron for data collection.

The diffraction patterns of these crystals were obtained at approximately 2.2 Å resolution and were then analyzed. The P1 space group was used with units of a = 61.489, b = 63.690, c = 71.710, α = 73.07, β = 67.49, and γ = 89.88. These units were then refined and were found to be: a = 60.45, b = 63.15, c = 71.08, α = 72.56, β = 66.99, and γ = 89.75. Integration then occurred, and the final units were determined to be: a = 61.047 ± 0.001, b = 63.492 ± 0.002, c = 71.505 ±0.002, α = 72.573 ± 0.001, β = 67.110 ± 0.001, and γ = 89.745 ± 0.001.

For the crystal structure of both monomers A and B, some electron density next to part of 2-methyl-3-oxopentanyl-SNAC was observed. However, the identity of this electron density was unknown. Also, three waters were seen next to this density. In addition, a methionine residue, which could not be seen in previous crystal structure, was observed and found to be fairly ordered. Furthermore, a glutamine residue was seen pointing away from the active site. This was also not seen before in previous crystal structures. We then decided to send more AmpKR2 crystals that were soaked in the same conditions as before to the synchrotron so that we could hopefully get better resolution.

In this newly solved crystal structure, the electron density around the 2-methyl-3-oxopentanyl-SNAC was still unclear; however, this structure was used as a model for polyketide binding by AmpKR2. In this crystal structure, the carbonyl group of glycine was found to interact with the substrates via hydrogen bonds. Furthermore, a nearby serine residue also participated in hydrogen bonding in order to stabilize the substrate.

F. ACP domains

Currently, TylACP1, EryACP1, PikACP3, AmpACP2, and SpiACP3 have been transformed into Top 10 BL21 cells and a plasmid purification MiniPrep has been performed. A restriction check digest has been run; however, no bands have been seen yet. As a result, another check digest will be performed. If the digest works and two bands representing the plasmid and the insert are present, then the plasmids will be transformed into BL21 cells, followed by purification of these proteins. Attempts will then be made to crystallize these proteins. However, if no bands are seen, then the plasmids may be retransformed into Top10 cells again.

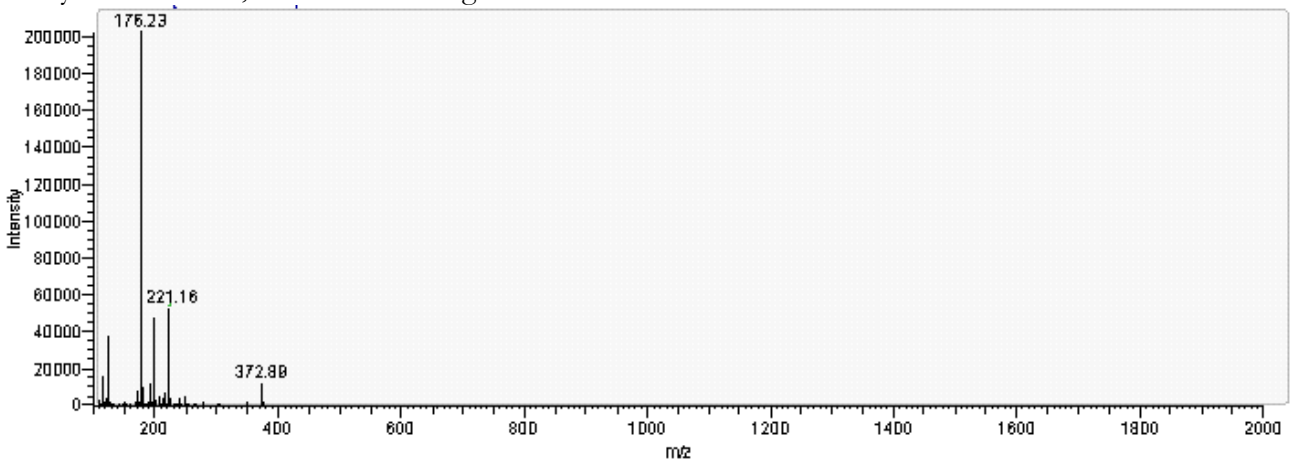

Figure 19: LC/MS of the reaction between ammonium sulfate and 2-methyl-3-oxopentanyl-SNAC.

G. LC-MS of Ammonium Sulfate With 2-methyl-3-oxopentanyl- SNAC

As previously mentioned, the attempt to crystallize AmpKR2 in the presence of 2-methyl-3-oxopentanyl-SNAC depended on whether the amine group of ammonium sulfate would hydrolyze the thioester bond of this diketide. As a result, we decided to figure out how long it would take for the thioester bond to be hydrolyzed so that we could determine how long to soak our crystals in 2-methyl-3-oxopentanyl-SNAC and $NADP^+$ for. Thin liquid chromatography was performed on two plates: a control plate which just had drops of 2-methyl-3-oxopentanyl-SNAC every five minutes and another plate which had drops containing 2-methyl-3-oxopentanyl-SNAC and ammonium sulfate every five minutes. After twenty minutes, the plates were examined; however, the products of the reaction did not separate well enough for identification.

Consequently, a LC-MS was performed, and the results can be seen in **Figure 19**. If ammonium sulfate did indeed cleave the 2-methyl-3-oxopentanyl-SNAC within 20 minutes, then two peaks representing SNAC and the propyl-propyl chain would be seen. However, in the spectrum, two peaks were seen. The huge peak at m/z = 176.23 represented 2-methyl-3-oxopentanyl-SNAC while the smaller peak at m / z = 221.16 and 372.89 represented impurities that were not either of the cleaved products. As a result, we were fairly sure that the thioester bond would not be cleaved when our AmpKR2 crystals were soaked in 2-methyl-3-oxopentanyl-SNAC and $NADP^+$.

IV. Conclusion

As of now, much progress has been made on determining the ternary complexes of the various ketoreductase and acyl carrier protein domains. Currently, all the ketoreductases except for erythromycin KR1 have been crystallized and much effort has been put into soaking these crystals with either $NADP^+$ or NADPH and 2-methyl-3-oxopentanyl-SNAC or 2-methyl-3-

oxopentanyl-CoA and then sending them off to the synchrotron to be diffracted. Also, the pET28b plasmid containing the various ACP domains have been transformed into Top 10 cells and will be transformed into BL21 cells should the check restriction digest look good. Furthermore, the Phoenix crystallization robot has been used on the PikKSATACP$_0$ protein and hopefully crystals will form soon. Last but not least, the AmpKR2 project has made the biggest progress towards our goal of obtaining a ternary complex. Although the electron density around the 2-methyl-3-oxopentanyl-SNAC was not very clear, we are pretty confident that the polyketide is in the active site. Thus, we are using this electron density as a model for how 2-methyl-3-oxopentanyl-SNAC binds into the active site groove.

Despite these hurdles, the potential of each of these projects is huge. For the ketoreductases, understanding the ternary complex will allow us to better understand how polyketide synthases work and thus be able to produce recombinant PKSs that can produce a wide array of polyketides, many of which could potentially have antibiotic properties. Meanwhile, if we were able to purify ACPs, then the processes by which ACPs interact with other domains such as the KR or KS could be examined and understood. Furthermore, the potential with PikKSATACP$_0$ is huge. Specifically, if the crystallization conditions for this loading module could be found, then the first ever entire module would be crystallized. Consequently, the interactions within this module would be understood and all of this acquired information could then be used to build other recombinant PKSs.

V. Acknowledgements

This project would not be possible without the help of many members of the Keatinge-Clay lab and the Welch Foundation. Specifically, the Welch Foundation sponsored this project. Meanwhile, Jianting Zheng, Shawn Piasecki, and Clint Taylor provided help with laboratory methods. Also, I would like to thank Chris Fage who worked intimately with me to help get this project started and running. Last but not least, I want to thank Dr. Adrian Keatinge-Clay for his constant advice and supervision.

VI. Materials and Methods

Cloning

The pET28b plasmids that contained erythromycin KR1, tylosin KR1, amphotericin KR2, pikromycin KR3, and tylosin KRACP1 were isolated and provided by Dr. Adrian Keatinge-Clay and Jianting Zheng. Meanwhile, the pET21b plasmid containing the PikKSAT$_0$ACP domain was provided by Chris Fage. For the ACP domains, the ACP fragments were amplified with the following primers: 5'TAGACGTAGCATATGGCGTCGCTGCCCGCGCCC GAGCGCGA 3'and 5'AGACTGGAATTCTCAGGTCG CACCGCCGAGTTCGGCGGCCA 3'for EryACP1, 5'AG TAGTACTCATATGTCCATGCTGAACGAGACCGAA CGCCT 3'and 5' AGTTCAGAATTCTCAGGCTCCGGC GCCGGTCAGCCGGT 3'for TylACP1, 5' CTAGTGTCC ATATGGACGGGCTGCCCACCGCCGA 3'and 5'GTCC TAGAATTCTCAGGGTGCGGTGCCGTAGAGCAGC GAA3'for AmpACP2, 5'CTAAGTGCACATATGGGTG CGCTCACCGGCGCCGAACA 3'and 5'ATGTACGAAT TCTCAGGCCTGCTCGCCCAGGATCTCCGCGA 3'for PikACP3, and 5'CGAAGTGCACATATGGCGGAGCT GCCGGAAGCCCAACGA 3'and 5'ACTCGAGAATTCT CACGCGACATCACCCACCAGCTCCT 3'for SpiACP3. The amplified DNA was then cut with EcoRI and NdeI restriction enzymes and then inserted into pET28b plasmids. The ligation was then confirmed by doing a check digest and submitting the purified plasmid DNA for sequencing.

Protein Expression, Purification, and Crystallization

For all the ketoreductase domains purified, the plasmids were transformed into *E. coli* BL21 cells and these cells were then grown in a Luria broth medium at 37°C until an OD of approximately 0.35 was reached. The temperature was then turned down to 15°C and 1 mL of 1 mM IPTG was added in order to induce protein expression. For the ACP domains, the same procedures were performed; however, *E. coli* Top10 cells were initially used in order to amplify DNA expression. The purified DNA was then transformed into *E. coli* BL21 cells.

After approximately 14 hours of growing cells, the cells were harvested, resuspended in Lysis Buffer (0.5 M NaCl, 30 mM HEPES, pH 7.5), and lysed via a sonicator. The total cell lysate was then poured over a Ni-NTA column

that was already equilibrated with the previously made Lysis buffer. The bound proteins on the column were then washed with a 15 mM imidazole buffer and eluted with a 150 mM buffer. All proteins then underwent fast protein liquid chromatography and were separated on a Superdex 200 gel-filtration column that was previously equilibrated with 150 mM NaCl, 10 mM HEPES, pH 7.5, and 10% glycerol. All proteins were then concentrated to approximately 15 mg/mL.

For TylKRACP1, 0.12 μmol of KRACP was added to 0.525 μmol of 2-methyl-3-oxopentanyl-CoA, 0.0057 μmol of SFP, 10 mM of $MgCl_2$, 150 mM NaCl, 100 mM HEPES, pH 7.5, and 10% glycerol. The mixture was then placed in a 4°C fridge for 6 hours and ran again on the Superdex 200 gel-filtration column.

The Phoenix crystallization robot was then used to grow proteins under 384 different conditions in an attempt to determine which conditions produced crystals. Furthermore, for each condition, a 1:1 and 2:1 drop ratio of protein to well solution was constructed. For grown crystals, a 24-well plate was used to screen around these crystallization conditions. Once proteins crystals were seen in these 24-well plates, these crystals were co-crystallized with 50 mM 2-methyl-3-oxopentanyl-SNAC and 10 mM $NADP^+$ and then were sent off to the synchrotron to be diffracted.

LC-MS of Ammonium Sulfate With 2-methyl-3-oxopentanyl-*SNAC*

Proteins were initially not crystallized in the presence of ammonium sulfate due to the possibility that the amine group of the ammonium sulfate would attack and cleave the thioester bond in 2-methyl-3-oxopentanyl-SNAC. In order to test this hypothesis, a TLC experiment was performed. In the first experiment, a control was set up where a drop of 2-methyl-3-oxopentanyl-SNAC was placed on the silica plates every five minutes and then the plates were placed in ethyl acetate. Meanwhile, on another silica plate, a drop if 50 mM 2-methyl-3-oxopentanyl-SNAC and ammonium sulfate was placed on the silica plates every five minutes. The plate was then placed in ethyl acetate. Once the ethyl acetate reached the top of the silica plates, the TLC plates were examined and there was no clear separation. As a result, a LC/MS was performed. Specifically, a solution of 2-methyl-3-oxopentanyl-SNAC and ammonium sulfate was allowed to sit at room temperature for 20 minutes. Afterwards, the solution was extracted with 0.5 mL of ethyl acetate and then dried down. The sample was then resuspended in 50% methanol and 0.1% formic acid and sent to the LC/MS to be analyzed.

VII. References

1. Alekseyev, V.Y, Liu, C.W., Cane, D.E., Puglisi, J.D., and Khosla, C. (2007). Solution Structure and Proposed Domain-Domain Recognition Interface of an Acyl Carrier Protein Domain From a Modular Polyketide Synthase. Protein Science *16*, 2093-2107.

2. Baerga-Ortiz, A., Popovic, B., Siskos, A.P., O'Hare, H.M., Spiteller, D., Williams, M.G., Campillo, N., Spencer, J.B., and Leadlay, P.F. (2006). Directed Mutagenesis Alters the Stereochemistry of Catalysis by Isolated Ketoreductase Domains from the Erythromycin Polyketide Synthase. Chem. Biol. *13*, 277-285.

3. Caffrey, P., Lynch, S., Flood, E., Finnan, S., and Oliynyk, M. (2001). Amphotericin Biosynthesis in *Streptomyces nodosus*: Deductions From Analysis of Polyketide Synthase and Late Genes. Chem. Biol. *8*, 712-

4. Caffrey, P. (2003). Conserved Amino Acid Residues Correlating With Ketoreductase Stereospecificity in Modular Polyketide Synthases. Chem. Biochem. *4*, 649-662.

5. Caffrey, P. (2005). The Stereochemistry of Ketoreduction. Chem. Biol. *12*, 1060-1062.

6. Cane, D.E., Walsh, C.T., and Chaitan, K. (1998). Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. Science *282*, 63-68.

7. Castonguay, R., He, W., Cehn, A.Y., Khosla, C., and Cane, D.E. (2007). Stereospecificity of Ketoreductase Domains of the 6-Deoxyerythronolide B. Synthase. J. Am. Chem. Soc. *129*, 13758-13769.

8. Castonguay, R., Valenzano, C.R., Chen, A.Y., Keatinge-Clay, A., Khosla, C., and Cane, D.E. (2008). Stereospecificity of Ketoreductase Domains 1 and 2 of the Tylactone Modular Polyketide Synthase. J. Am. Chem. Soc. *130*, 11598-11599.

9. Chen, A.Y., Cane, D.E., and Khosla, C. (2007). Structure-Based Dissociation of a Type 1 Polyketide Synthase Module. Chem. Bio. *14*, 784-792.

10. Chen, S., Xue, Y, Sherman, D.H., and Reynolds, K.A. (2000). Mechanisms of Molecular Recognition in the Pikromycin Polyketide Synthase. Chem. Biol. *7*, 907-918.

11. Hertweck, C. (2009). Polyketide Biosynthesis. Angew. Chem. Int. Ed. *48*, 4688-4716.

12. Jacobsen, J.R., Keatinge-Clay, A.T., Cane, D.E., and Khosla, C. (1998). Precursor-Directed Biosynthesis of 12-Ethyl-Erythromycin. Bioorganic and Medicinal Chemistry *6*, 1171-1177.

13. Keatinge-Clay, A.T. (2007). A Tylosin Ketoreductase Reveals How Chirality is Determined in Polyketides. Chem. Biol. *14*, 898-908.

14. Keatinge-Clay, A.T. (2008). Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. J.Mol. Biol. *384*, 941-953.

15. Keatinge-Clay, A.T., Stroud, R.M. (2006). The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. Structure *14*, 743-748.

16. Khosla, C., Kapur, S., and Cane, D.E. (2009). Revisiting the Modularity of Modular Polyketide Synthases. Current Opinion in Chemical Biology *13*, 135-143.

17. Khosla, C., Tang, Y., Chen, A.Y., Schnarr, N., and Cane, D.E. (2007). Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. Ann. Rev. Biochem. *76*, 11.1-11.27.

18. Kim, CY.., Alekseyev, V.Y., Chen, A.Y., Tang, Y., Cane, D.E., and Khosla, C. (2004). Reconstituting Modular Activity from Separated Domains of 6-Deoxyerythronolide B Synthase. Biochemistry *43*, 13892-13898.

19. Kittendorf, J.D., Beck, B.J., Buchholz, T.J., Seufert, W., and Sherman, D.H. (2007). Interrogating the Molecular Basis for Multiple Macrolactone Ring Formation by the Pikromycin Polyketide Synthase. Chem Biol. *14*, 944-954.

20. Leibundgut, M., Jenni, S., Frick, C., and Ban, N. (2007). Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase. Science *316*, 288-290.

21. Luo, G., Pieper, R., Rosa, A., Khosla, C., and Cane, D.E. (1996). Erythromycin Biosynthesis: Exploiting the Catalytic Versatility of the Modular Polyketide Synthase. Bioorganic and Medicinal Chemistry *4*, 995-999.

22. Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A., and Santi, D.V. (2005). Combinatorial Polyketide Biosynthesis by *de novo* Design and Rearrangement of Modular Polyketide Synthase Genes. Nature Biotechnology *23*, 1171-1176.

23. Menzella, H.G., Reisinger, S.J., Welch, M., Kealey, J.T., Kennedy, J., Redit, R., Tran, C.Q., and Santi, D.V. (2006). Redesign, Synthesis, and Functional Expression of the 6-Deoxyerythronolide B Polyketide Synthase Gene Cluster. J. Ind. Microbiol. Biotechnol. *33*, 22-28.

24. Mortison, J.D., Kittendorf, J.D., and Sherman, D.H. (2009). 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.

25. Nelson, D.L., and Cox, M.M. Lehninger Principles of Biochemistry, 5th ed. New York: W.H. Freeman Company. 2008.

26. O'Hare, H.M., Baerga-Ortiz, A., Popovic, B., Spencer, J.B., and Leadlay, P.F. (2006). High-Throughput Mutagenesis to Evaluate Models of Stereochemical Control in Ketoreductase

Domains from the Erythromycin Polyketide Synthase. Chem. Biol. *13*, 287-296.

27. Pieper, R., Gokhale, R.S., Luo, G., Cane, D.E., and Khosla, C. (1997). Purification and Characterization of Bimodular and Trimodular Derivatives of the Erythromycin Polyketide Synthase. Biochemistry *36*, 1846-1851.

28. Pieper, R., Luo, G., Cane, D.E., and Khosla, C. (1995). Cell-Free Synthesis of Polyketides by Recombinant Erythromycin Polyketide Synthases. Nature *378*, 263-266.

29. Power, R., Dunne, T., Murphy, B., Lochlainn, L.N., Rai, D., Borissow, C., Rawlings, B., and Caffrey, P. (2008). Engineered Synthesis of 7-Oxo- and 15-Deoxy-15-Oxo-Amphotericins: Insights into Structure-Activity Relationships in Polyene Antibiotics. Chem. Biol. *15*, 78-86.

30. Schnarr, N.A., Chen, A.Y., Cane, D.E., and Khosla, C. (2005). Analysis of Covalently Bound Polyketide Intermediates on 6-Deoxyerythronolide B Synthase by Tandem Proteolysis-Mass Spectrometry. Biochemistry *44*, 11836-11842.

31. Siskos, A.P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, D., Popovic, B., Spencer, J.B., Staunton, J., Weissman, K.J., and Leadlay, P.F. (2005). Molecular Basis of Celmer's Rules: Stereochemistry of Catalysis by Isolated Ketoreductase Domains from Modular Polyketide Synthases. Chem. Biol. *12*, 1145-1153.

32. Weissman, K.J. Methods in Enzymology. Oxford: Elsevier Inc. 2009.

33. Wu, J., He, W., Khosla, C., and Cane, D.E. (2005). Chain Elongation, Macrolactonization, and Hydrolysis of Natural and Reduced Hexaketide Substrates by the Picromycin/Methymycin Polyketide Synthase. Angew. Chem. Int. Ed. *44*, 7557-7560.