AUGUST 14, 2015

ORPHAN POKLYKETIDE SYNTHASE IDENTIFICATION

STREPTOMYCES SP. MG1

NEIL KLENK
THE UNIVERSITY OF TEXAS AT AUSTIN

Summary:

The goal of this project is to determine the structure of a polyketide synthase (PKS) product that has yet to be identified. Polyketides represent a diverse class of secondary metabolites that have vast applications. Polyketide products have already been employed to combat bacterial and fungal infections, immune disorders, and to name a few.

Characterizing PKS through usual laboratory techniques is not always practical, often due to chemical properties of the products**.** One work around is to characterize them computationally. This is made possible thanks to the collinear relationship between the synthase and resulting polyketide. This colinearity is referring to the ability to determine the structure of the polyketide result from the PKS amino acid sequence and vice versa. Computational identification is possible thanks to the assembly style manufacturing process through which the PKS makes the polyketides. There are a limited number of specific enzymatic domains that the PKS employs. These individual domains perform specific reactions at defined regions of the polyketide as is determined by the location of the domain in the module, and in turn its location in the PKS.

The first step will be finding a previously characterized PKS with a large amount of diversity in its domains. Once a PKS is decided, a fragment of its amino acid sequence will be used to find an uncharacterized PKS that has a high degree of homology. The program clustermine360 and DoBISCUIT will be used to help find an initial PKS sequence in its entirety. The NCBI database will then be searched for similar proteins by performing a Blastp search on the initial sequence while excluding the PKS that it is from. This provides results for PKS that have a high degree of homology to the initial, and an uncharacterized one will be selected. The search sequence will then be submitted to AntiSMASH which is domain recognition software for natural product biosynthesis. The AntiSMASH results will then be combined with manual curation of the sequence to build the structure of the polyketide that is produced from the orphan PKS.

Introduction:

Polyketide synthases (PKSs) are massive (>MDa) enzyme assembly lines encoded within the bacterial chromosome. They are composed a limited set of tethered domains that act in a consecutive, modular, fashion to construct polyketide substrates as they progress through the synthase from the N to C terminal end. Each of the domains performs a specific action on the growing PKS at a very well defined region of the polyketide. Where and how each of these domains acts is totally dependent on its location within the synthase. This collinear relationship between PKS sequence and polyketide structure, and the ability to determine one from the other is at the foundation of orphan PKS Identification.

PKSs are responsible for the production of a wide range of biological compounds classified as secondary metabolites. These organic compounds are not directly involved in the cells immediate viability, but instead affect the organism by improving such things as its survivability in its environment or its actual reproductive success. They have been found to be an incredibly useful group of naturally occurring molecules. Their functions are very diverse, and have been used for antifungal, antibacterial, insecticides, and anticancer purposes. Not only do polyketides have a wide range of functions, they are often really good at what they do. Out of the 7000 known PKS structures, more than 0.3% have been commercialized. Compared to the typical rate of <0.001%, this is excellent[1].

PKSs are believed to be the result of a gene duplication event of the bacterial fatty acid synthase (FAS)[2]. Once the bacteria had two sets of fatty acid synthases, the second became free to mutate without harming the organism's viability. The mutations continued until it reached a point at which it became beneficial to the organism, resulting in a functional secondary metabolite and a conserved PKS structure. The structural similarity between the two figures can be seen in Figure 1.

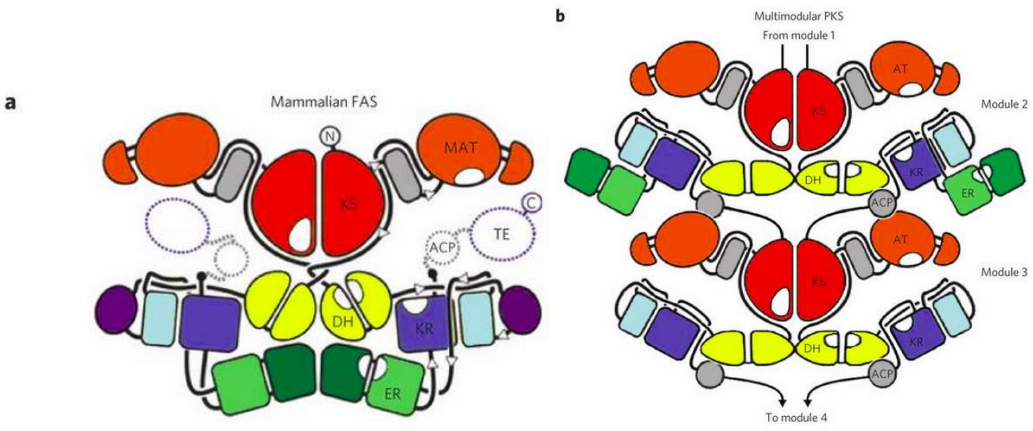


Figure 1. Mammalian FAS (Left) Modular PKS (Right)[3]

PKSs have proven to produce extremely effective natural products. They are currently employed pharmacologically as Antibiotics, immunosuppressants, anticholesterol, and anticancer

drugs. With only a small percentage of PKSs characterized, it is probable that there are many more viable drugs being produced by unidentified PKSs. The goal of this project is to determine the structure of one of these unknown, or orphan, PKSs. This will be accomplished utilizing the collinearity rule[4]. This rule states that there is a direct relationship between the presence of specific domains, and the number of modules, and the product when it leaves the PKS. PKSs are multimodular synthases containing repeating domain patterns. Each of these domains have heavily conserved regions in their amino acid sequences that are crucial to its function, and can tell about the particular function of each domain. By determining the order of the domains and modules from the amino acid sequence, the structure of the product can be determined to a high degree of certainty. This "colinearity" between PKS and end product will be employed to take an orphan (unidentified PKS) and determine the structure, and probable function completely based off of its amino acid sequence.

PKSs are present in nature in a variety of different structures. The type I modular PKSs have their enzymes contained within multidomain polypeptides. This is a dramatic distinction from the type II and type III PKSs. Within type I PKSs there is further distinction between those that are modular, and those that are iterative. The iterative type I PKSs utilize a system where domain are reused in a cyclic fashion. This differs greatly from the modular type I PKSs that use a sequence of modules that do not repeat[3]. The human FAS is an example of an iterative type I PKS, but the Streptomyces sp. Mg1 PKS being investigated here functions as a modular type I PKS.

Methods and materials:

Before the alignment of the sequences could occur, an algorithm had to be selected. Alignments of multiple protein sequences has a lot of applications in the biological sciences. The most important aspects for this project included the identification of critical residues for the sake of structural prediction. The MUSCLE alignment algorithm has proven to be fast, efficient, and accurate for sets of sequences that are not extremely diverged, so it is what was chosen for this project.

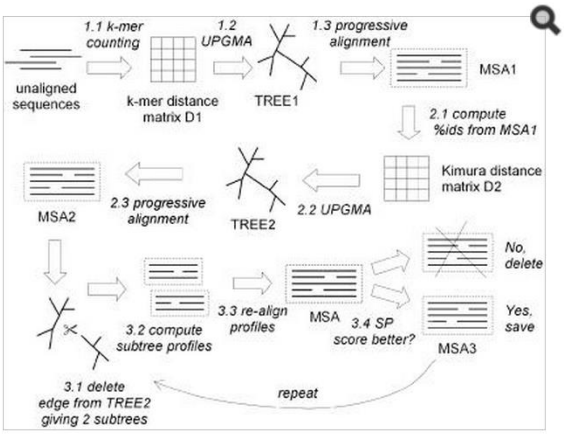


Figure 2. The Muscle Alignment Algorithm[5]

The MUSCLE alignment algorithm can be broken down into 3 stages, draft progressive, improved progressive, and refinement which can be seen in figure 2. In the *Draft progressive* stage speed is emphasized over accuracy to produce a multiple sequence alignment. The *k*mer distance is computed for each pair of input sequences which results in the creation of distance matrix D1. *k*mer is defined as a contiguous subsequence of length *k.* The matrix is then clustered with UPGMA (Unweighted Pair Group Method with Arithmetic Mean) to produce a binary tree with multiple sequence alignment 1 (MSA1) at the root via progressive alignments starting at the leaves. In the *Improved Progressive* stage the main goal is to produce an improved tree by reducing the main source of error; the approximate *k*mer distance measure. MUSCLE re-estimates the tree, but this time uses a Kimura distance. This is more accurate, but requires an alignment. The Kimura distance for each pair of input sequences is computed from MSA1 to give a new distance matrix, D2, which is then clustered with UPGMA again producing TREE2. The Kimura distance is a distance measure for an aligned pair. A second progressive alignment is produced from TREE2 to produce MSA2. This is optimized by only computing alignments for those subtrees whose branching orders changed relative to TREE1. Lastly the alignment is refined by dividing TREE2 into two subtrees and realigning the two profiles computed from the two subtrees. If the sum of pairs (SP) score is improved it's kept, otherwise discarded. This is repeated until divergence[5].

Many high value chemicals are products of microbial secondary metabolisms. For example, almost 70% of the drugs in medical use in the field of anti-infectives are secondary metabolites or their derivatives. The genes that encode the biosynthetic pathways to produce theses secondary metabolites are normally clustered together in biosynthetic gene clusters (BCGs) on the chromosome. Due to the spatial restrictions on the BCGs, genome mining of the gene clusters has become an effective way to discover novel compounds. Genomic mining is a subset of data mining in which existing annotated genes are used to predict the function of genes that have yet to be characterized[6]. This is akin to having a bunch of letters in a line, then going through a dictionary to determine the best locations to put the spaces in your sentences to result in the most logical sentence.

While there are a variety of programs that do this for specific classes of secondary metabolites, the antibiotic and Secondary Metabolite Analysis Shell (antiSMASH) does this for any BCG. antiSMASH facilitates the rapid genome mining of such a range of secondary metabolites allowing sfor the rapid and automatic genomic identification and analysis of BCGs of any type[7]. Due to its effectiveness and flexibility, it was the chosen genomic mining program for this project.

$$s(a,b) = \frac{1}{\lambda} \log \frac{p_{ab}}{f_a f_b}$$

Figure 3. The Formula used for Calculating Scores for Aligning Two Residues[8]

BLOSUM62 is the name of the matrix that is used in blast to find sequence similarities. The real goal of alignment s to determine if two different sequences are related evolutionarily or not. Therefore this is what the alignment score needs to reflect. The BLOSUM62 matrix takes

advantage of log-odds scoring, and the formula itself can be seen in figure 3. This provides for the comparison of the two hypotheses. s(a,b) is the score for aligning residues a and b. P_{ab} is the likelihood that the two residues are related because they are homologous. f_af_b is the likelihood that the two residues are occurring independently (are uncorrelated or unrelated). $1/\lambda$ is simply a scaling factor to get the result into integers that are easier to deal with[8].

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Figure 4. The BLOSUM62 Identity Matrix[16]

As can be seen in Figure 4, the values for aligning two residues varies greatly, from 11 to -4. This is due to the formula presented above. The rarity of the amino acid has a lot to do with it. For example tryptophan is a pretty rare amino acid, so when there are two paired it results in a large score of 11. The values for each amino acid pairing were determined from the overall probability of those two amino acids being aligned. This probability came from looking at a lot of known and trusted alignments that had previously been done, and counting the frequency at which each residue pair occurs[8].

The first step in the process of selecting an orphan PKS was to select a previously characterized polyketide. Mycolactone was chosen for a variety of reasons. It was well characterized, and it utilized all of the domains (KS, AT, DH, ER, KR). Both of these factors will improve the chances of a similar, but equally diverse orphan PKS. Once the base PKS was determined, a portion of its amino acid sequence had to be taken to try and find an orphan that had a similar structure.

1. Used Clustermine360, a database of microbial PKS/NRPS Biosynthesis to find the sequence for the well-known Mycolactone PKS[9]. This provided both the DNA sequence for the host organism (Mycobacterium Ulcerans Agy99), as well as the Mycolactone structure seen in figure 5. Having the structure is significant here because it will provide an approximation for what the orphan PKSs structure will resemble. However this was more information than was needed. Only the sequence of the associated PKS was needed, not the entire organism.

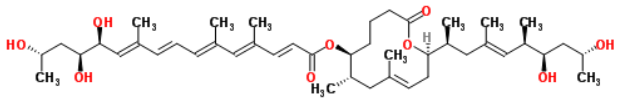


Figure 5. The Structure of Mycolactone[9]

2. DoBiscuit, a database of secondary metabolite biosynthetic gene clusters was then utilized to get the amino acid sequence of just the Mycolactone PKS[10].

3. A blastp of the above sequence was conducted, excluding the organism that the Mycolactone is located in (Mycobacterium Ulcerans Agy99 (taxid:362242)). This searched for amino acid sequences that were similar based off of both sequence similarity and query cover[11]. This search presented different proteins in a descending order based off of overall sequence similarity to the query sequence.

4. A quick search for similar sequences gave the M. Ulcerans PKS sequence (accession YP_906177). Blastp was used to determine similar sequences. Mycobacterium Streptomyces sp. Mg1 (accession WP_047960434) was found with an E value of 0 over a query cover of 99% with 52% sequence identity. This was selected as the PKS.

After the orphan PKS was selected, the identification could begin. The ascension number was taken and input to antiSMASH to determine the order of the modules and domains within those modules. After inputting the hypothetical protein sequence into antiSMASH there was a lack of a KS-AT (initiating) and no TE (terminating) modules. This indicated that the whole sequence wasn't present. Multiple sequences were then collected from NCBI and were aligned to attempt to get the whole sequence, but this was unsuccessful. The whole shotgun sequence for *Streptomyces sp. Mg1* was then sent through antiSMASH to try to find the sequence for the whole PKS. It located 28 different clusters, 3 of which were Type 1 PKS. Of those three, only one contained the module that was found in the previous Streptomyces sp. Mg1 PKS sequences that had been investigated. It contains the TE group indicating that it is in fact the whole sequence. It still lacked an obvious initiation module, but this will be touched on later. The result can be seen in figure 6 with figure 7 showing the information for the PKS of interest.

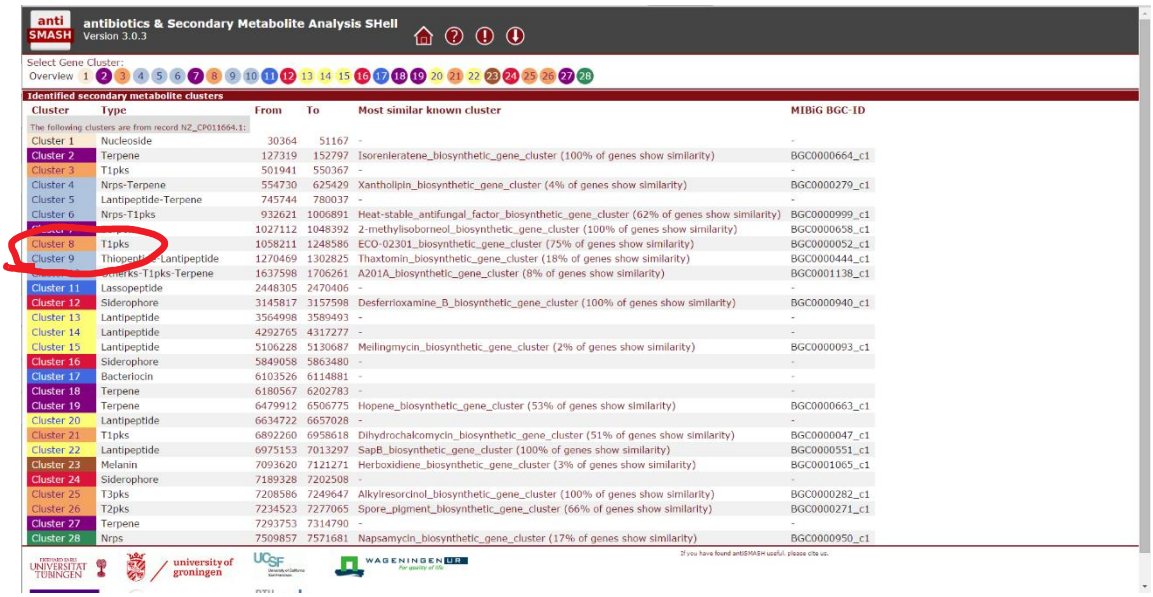


Figure 6: All the Clusters Detected in the Streptomyces sp. Mg1 whole genome shotgun by antiSMASH

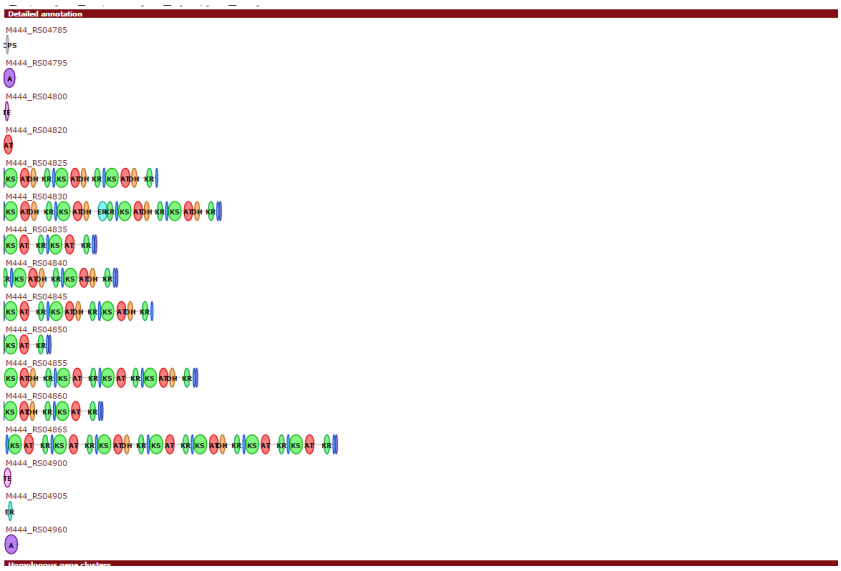


Figure 7: Sequence for the PKS of interest.

Now that the full sequence was present, sequences for each domain could be collected. antiSMASH makes this very easy by allowing users to simply click on This allowed me to get the amino acid sequence for each domain of the synthase. The KR and AT sequences were then aligned. Once the sequences are aligned, information from Keating-Clay's paper "*The Structures of Type I Polyketide Synthases*" was all used to determine what modifications each domain made to the product chain.

Results:

Streptomyces and M. Ulcerans are two different families, which doesn't make much sense for it to be the top hit, considering a well characterized member of one family was used to find the second. This could be explained by a horizontal gene transfer. Horizontal gene transfer, also referred to as lateral gene transfer, is the lateral movement of genetic movement of genetic information from donor to recipient bacteria. This is opposed to vertical transfer where genetic information is transferred from parent to offspring in sexual or asexual reproduction. This can confer a distinct evolutionary advantage to recipient organisms as this is the primary reason for antibiotic resistance[12]. The transfer could have happened from Mycobacterium marinum DL240490 (Accession ABF57667). When a Blastp was performed on the sequence with Mycobacterium marinum excluded, Streptomyces sp. Mg1 (accession WP_037794998) was the top result with an E value of 4e-78. Mycobacterium marinum and Mycobacterium tuberculosis are closely related[13], so the association between Mycobacterium tuberculosis and Streptomyces sp. Mg1 was checked as well since the Mycobacterium marinum sequence was only partial. Blastp was run on Mycobacterium tuberculosis(Accession WP_031709871), and showed a very close relationship to Streptomyces sp. Mg1 (WP_037794998.1)with it again being the top hit. With an E-value of 0.0 with 98% query cover, and 51% identity, it is possible that the horizontal gene transfer occurred from Mycobacterium tuberculosis. This will help the structure be determined. If the sequences are so closely related, then the final structure should be similar to the two.

The initiation module for this PKS seems to be missing. There is no typical KS-AT starter domain, so alternatives had to be investigated. One possibility that has a history with the *Streptomyces* bacteria family. Is the 4-guanidinylbutyrrl-CoA starter unit that is believed to originate from an arginine before preceding down a three-step precursor pathway leading to 4-guanidinylbutryl-CoA[14].

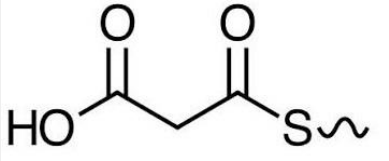
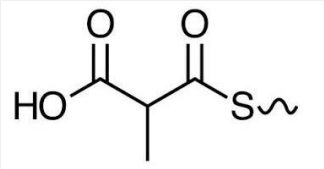
Each of the six primary enzymatic domains in these type I PKSs have a specific function, yet are responsible for nearly all of the diversity in polyketides. The six domains are the acyl carrier protein (ACP), ketosynthase (KS), acyl-transferase (AT), ketoreductase (KR), dehydratase (DH), and the enolreductase (ER). The first three domains are all necessary for each two carbon chain elongation. The KR, DH, and ER domains are all considered processing domains and function to give PKSs such diverse functional groups. These processing domains display highly conserved motifs in their sequence, to the extent that both their presence and the type of chemistry being performed can be determined.

The ACP functions as the shuttle for the growing polyketide. The ACP brings it to the active sites of the synthase. These domains contain a heavily conserved catalytic serine that the prosthetic group is post translationally added to, allowing the ACP to move the group through that module of the synthase before passing it off.

The KS domain is the largest and most conserved within PKSs. It is physically responsible for the elongation of the polyketide making it crucial for the existence of the PKS. Its major function is the formation of the carbon-carbon bonds. This region is so highly conserved that its presence is what tells search algorithms that they have encountered a PKS.

The AT domain is the last of the domains that is necessary for chain elongation. It functions to catalyze the transfer of the extender unit to the ACP of its module, which is what extends the chain. The extender units are usually either a malonyl-CoA or methylmalonyl-CoA. Which extender unit will be used can be determined by the looking for which of the two conserved motifs are present as can be seen in table 1[3].

Table 1. AT Domain Motifs, Resulting Extender Units and Structure.

Extender Unit	**Structure**	**Motif**
Malonyl-CoA		GHS(I/V)G and HAFH
Methylmalonyl-CoA		GHSQG and YASH

These motifs can be seen in the sequence alignments that were performed on SeaView alignment software using the MUSCLE alignment algorithm. The GHS-G and the -A-H motif can be seen in figure 8. From this information, it can be deduced that the C4 AT 1, C6 AT 3, C7 AT 1, and C7 AT 2 all have a methylmalonyl extender unit while all the others use a malonyl extender unit. In both figures the arrows are pointing to the methylmalonyl extender units

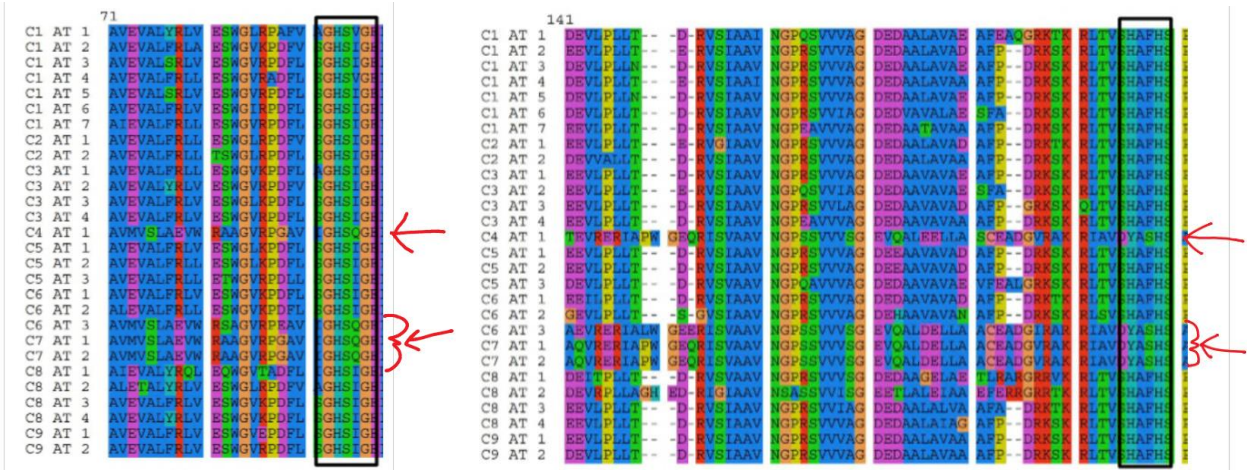
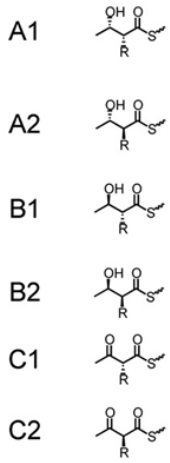


Figure 8. Aligned AT Domains

Figure 9. Different KR Domain Stereochemistry[3]



The KR domain is the first of the processing domains that will be discussed. It is responsible for setting up to two stereocenters up, one at the alpha carbon if a methyl is present, and the other at the beta carbon to form a beta-hydroxyl group. It comes in six different types (A1, A2, B1, B2, C1, C2). A type KRs result in a hydroxyl group with L-orientation while B type KRs result in a hydroxyl group with a D-orientation. C type KRs are reductase incompetent. The α-substituent allows the KR types to be characterized even further. If the α-substituent is D-orientation it is appended with a "1". If the α-substituent is L-oriented it is appended with a "2". Figure 9 demonstrates the different stereochemistry associated with each type of KR domain[3].

Table 2. The Motifs that Determine the Resultant Stereochemistry[3]

Conserved Domains:

KR Type	Motif Associated with KR Type
Type A1	Conserved W eight residues before the catalytic Y--N region
Type A2	Same as A1, but includes a conserved H three residues before the Y---N region
Type B (all)	Contains LDD instead of a W before catalytic Y---N region
Type B1	No P residue on the third residue of Y---N catalytic region
Type B2	Contains a P making it a conserved Y--P--N region

After collecting the sequences through antiSMASH, they were once again aligned using the SeaView program. The motifs described in Table 2 were looked for to determine the type of KR domain present in each module. This resulted in one type A1, no type A2, 25 type B1, two type B2.

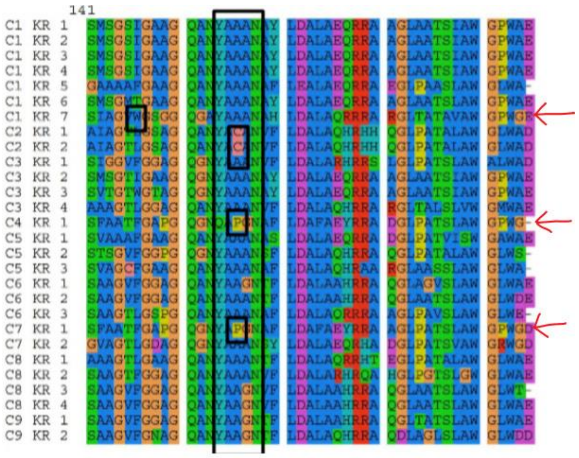


Figure 10. KR Domain alignment of Y---N Region

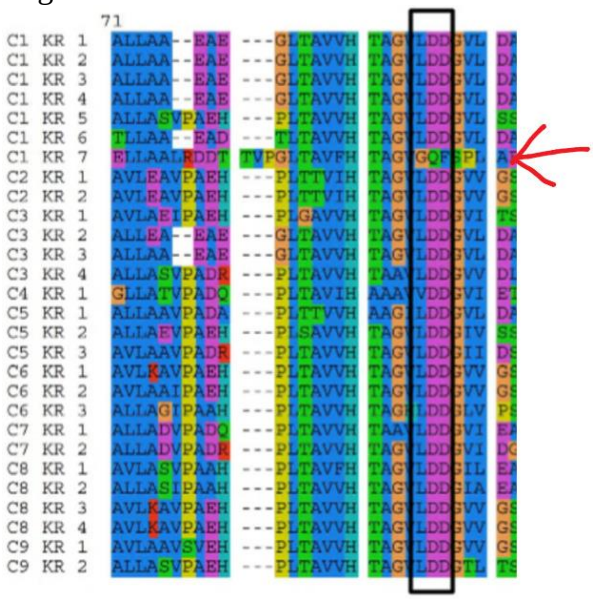


Figure 11. KR Domain Alignment of LDD Loop

Figure 10 shows the KR Domain alignment with figure 11 showing the presence or lack of the LDD loop.

All that do not have a red arrow follow the B1 motif.

The presence of tryptophan and LDD loop in C1 KR 7, but lack of histidine 2 residues N-terminal to the catalytic tyrosine indicate it's a type A1.

The cysteine in C2 KR 1 and 2 is abnormal, but the presence of a LDD loop combined with the lack of a proline two residues C-terminal to the catalytic tyrosine indicate that they are still type B1.

The presence of a proline two residues C-terminal to the catalytic tyrosine and presence of the LDD loop indicate they are type B2.

The DH domain is the second of the processing domains in PKSs. When present, they catalyze the dehydration of the polyketide inter mediate, resulting in a double bond between the alpha and beta carbon via syn elimation of water. This domain is a versatile enzymatic scaffold that can produce both cis and trans bonds. The active site motifs of DHs that produce cis bonds are believed to be the same as those that produce cis bonds. The KR domain comes into play here again and is the deciding factor if it is a cis or trans bond. Those that are preceded by A type KRs result in cis double bonds due to the 120° difference in dihedral angle of the C-C_α-C_β-C_λ L-β- hydroxyl group bound for syn elimination compared to D-β-hydroxyacyl group bound for syn elimation of the same bond. This results in one cis bond as can be seen in figure 12.

The ER domain is last of the processing domains. It functions to reduce double bonds created by DH. The ER is stereoselective, and will only reduce trans-α,β-double bonds. There is only one ER present in this PKS, and it is contained within the same module as a trans-α,β-

double bond. This results in one totally reduced bonds in the structure as can be seen in figure 12.

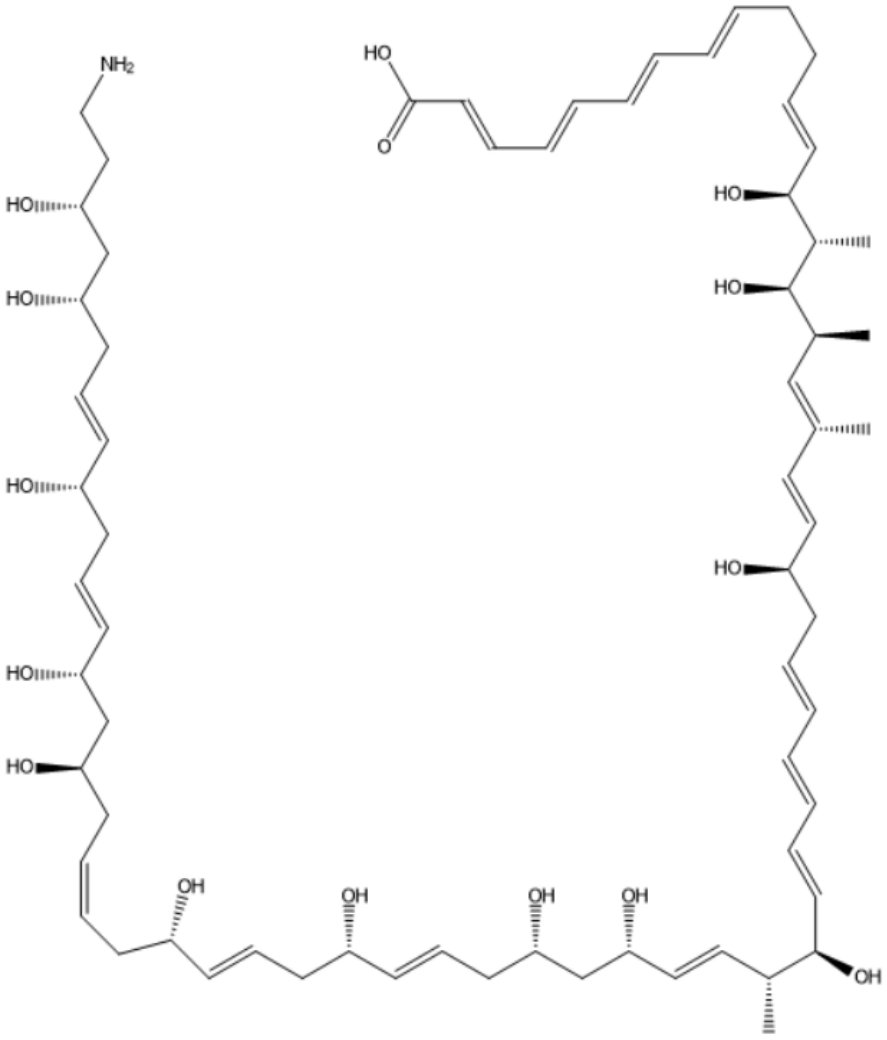


Figure 12. Resultant Polyketide Prior to Modifications by Tailoring Enzymes

After putting all of the information together, the full polyketide can be assembled with the exception of any modification made by tailoring enzymes. The result is a linear polyene with many hydroxyl groups. This is extremely similar to the structure of mediomycin, so the most likely case is that it also has strong antifungal properties[15].

Discussion:

After putting all of the information together, the full polyketide can be assembled with the exception of any modification made by tailoring enzymes. The result is a linear polyene with many hydroxyl groups. This is extremely similar to the structure of mediomycin a strong antifungal. Since function comes from structure, the most likely case is that it also has strong antifungal properties[15].

To gain more confidence about the function of this polyketide, a phylogenetic tree should be created. This tree would provide a visual representation of the evolutionary distance between the orphan PKS and different known PKSs whose function has already been determined. By knowing the evolutionary distance between a known PKS and the orphan, the function of the orphan can be deduced to a much higher degree of certainty.

Acknowledgements:

Thanks to Professor Keatinge-Clay served as my mentor throughout this project, and Drew Wagner who allowed me to bounce questions off of him whenever I hit a wall.

Literature Cited:

[1]Weissman, Kira J., and Peter F. Leadlay. "Combinatorial Biosynthesis of Reduced Polyketides." *Nature Reviews Microbiology Nat Rev Micro* 3.12 (2005): 925-36. Web. 10 Aug. 2015. <http://www.nature.com/nrmicro/journal/v3/n12/full/nrmicro1287.html>.

[2]O'brien, Robert V., Ronald W. Davis, Chaitan Khosla, and Maureen E. Hillenmeyer. "Computational Identification and Analysis of Orphan Assembly-line Polyketide Synthases." *J Antibiot The Journal of Antibiotics* 67.1 (2013): 89-97. Web. 10 Aug. 2015.

[3]Keatinge-Clay, Adrian T. "The Structures of Type I Polyketide Synthases." *Nat. Prod. Rep. Natural Product Reports* 29.10 (2012): 1050. Web. 1 June 2015.

[4]Callahan, B., M. Thattai, and B. I. Shraiman. "Emergent Gene Order in a Model of Modular Polyketide Synthases." *Proceedings of the National Academy of Sciences* 106.46 (2009): 19410-9415. *PNAS*. Web. 7 Aug. 2015. <http://www.pnas.org/content/106/46/19410.full#ref-19>.

[5]Edgar, Robert C. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research*32(5) (2004): 1792-797. *PMC*. Web. 7 Aug. 2015. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/>.

[6]Lee, Gir-Won, and Sang-Soo Kim. "Genome Data Mining for Everyone." *BMB Reports* 41.11 (2008): 757-64. *ResearchGate*. Web. 10 Aug. 2015. <http://www.researchgate.net/publication/23481711_Genome_data_mining_for_everyone._BMB_Rep>.

[7]Weber, Tilmann, Kai Blin, Srikanth Duddela, Daniel Krug, Hyun Uk Kim, Robert Bruccoleri, Sang Yup Lee, Michael A. Fischbach, Rolf Müller, Wolfgang Wohlleben, Rainer Breitling, Eriko Takano, and Marnix H. Medema. "AntiSMASH 3.0—a Comprehensive Resource for the Genome Mining of Biosynthetic Gene Clusters." *Nucleic Acids Res Nucleic Acids Research* 43.W1 (2015): n. pag. Web. 7 Aug. 2015. <http://nar.oxfordjournals.org/content/43/W1/W237>.

[8]Eddy, Sean R. "Where Did the BLOSUM62 Alignment Score Matrix Come From?" *Nature Biotechnology* 22 (2004): 1035-036. Web. 6 Aug. 2015. <http://www.nature.com/nbt/journal/v22/n8/full/nbt0804-1035.html>.

[9]http://www.clustermine360.ca/CompoundFamilyDetails.aspx?cfid=95

[10]Ichinkawa, N., M. Sasagawa, M. Yamamoto, H. Komaki, Y. Yoshida, S. Yamazaki, and N. Fujita. *DoBISCUIT*. Nucleic Acids Research Database Issue, 26 Nov. 2012. Web. 13 Aug. 2015. <http://www.bio.nite.go.jp/pks/>.

[11]*NCBI*. U.S. National Library of Medicine, n.d. Web. 13 Aug. 2015. <http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/BLAST/how_blastp_works.html>.

[12]Gyles, C., and P. Boerlin. "Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease." *Veterinary Pathology* 51.2 (2013): 328-40. *PubMed*. Web. 11 Aug. 2015. <http://www.ncbi.nlm.nih.gov/pubmed/24318976>.

[13]Rombouts, Y., L. Alibaud, S. Carrere-Kremer, E. Maes, C. Tokarski, E. Elass, L. Kremer, and Y. Guerardel. "Fatty Acyl Chains of Mycobacterium Marinum Lipooligosaccharides: STRUCTURE, LOCALIZATION AND ACYLATION BY PapA4 (MMAR_2343) PROTEIN." *Journal of Biological Chemistry* 286.38 (2011): 33678-3688. Web. 13 Aug. 2015.

[14]Hong, H., T. Fill, and PF. Leadlay. "A Common Origin for Guanidinobutanoate Starter Units in Antifungal Natural Products†." *Angew Chem Int Ed Engl* (2011): n. pag. *A Common Origin for Guanidinobutanoate Starter Units in Antifungal Natural Products*. Nov. 2011. Web. 13 Aug. 2015. <http://onlinelibrary.wiley.com/doi/10.1002/anie.201308136/abstract>.

[15]Cai, Ping, Fangming Kong, Pamela Fink, Mark E. Ruppen, R. Thomas Williamson, and Tabei Keiko. "Polyene Antibiotics from Streptomyces Mediocidicus." *ResearchGate*. N.p., 30 Oct. 2006. Web. 13 Aug. 2015. <http://cat.inist.fr/?aModele=afficheN&cpsidt=18606598>.

[16]Rost, Hannes. Licensed under CC BY 3.0 https://en.wikipedia.org/wiki/File:BLOSUM62.gif.