Visualization of a DH:DH Dimerization Interface in Juvenimicin PKS Module 5

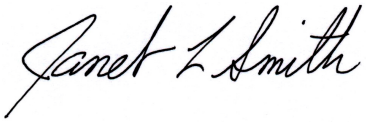
Anya Dhar

In partial fulfillment of the requirements for the degree of Bachelor of Science in Honors Biochemistry.

University of Michigan

3 April 2023

This thesis has been read and approved by Dr. Janet Smith, PhD.

Signed :  Date: 15-April-2023

Faculty advisor e-mail: janetsmi@umich.edu Phone: 734-615-9564

Table of Contents

Abstract

The field of combinatorial chemistry is highly interested in determining the structure of polyketide synthases (PKS) to manipulate existing PKS systems into producing macrolide antibiotics. To do so, we utilized x-ray crystallography, a method of structure determination that has been widely used across biological systems. Here, we determine the structure of the dehydratase (DH) domain within module 5 from the juvenimicin modular polyketide synthase pathway and confirm the presence of a DH:DH dimerization interface in the excised protein. Structure determination was achieved using molecular replacement with the PHENIX software suite. Our results provide insight into the molecular mechanism of PKS biosynthesis and may be useful in the development of novel drugs. Furthermore, our study demonstrates the utility of x-ray crystallography in the study of PKSs and highlights the importance of studying the structure and function of individual PKS domains.

1. Introduction

Protein Structure and Function

A major component of cells, proteins are responsible for carrying out many of the functions that are required to sustain life. Determination of the function of a protein is one of the core concepts in biochemistry and molecular biology as a field. Originating from its primary sequence, a protein's structure depends on a variety of factors, including pH environment, functional groups, covalent and noncovalent interactions, and more[1]. With protein function being so intimately linked to its structure, the aim shifted from purely direct observation of function to a visualization of its structure. Over the years, several methods of structural visualization have been utilized, both of two-dimensional and three-dimensional varieties. One of the most relied upon methods of protein structure visualization is x-ray crystallography.

X-ray crystallography aims to obtain a three-dimensional molecular structure from a crystalline form of the target. This technique can be applied to a variety of targets, including but not limited to, viruses, immune complexes, and protein-nucleic acid complexes[2]. For macromolecular crystallography, a protein product is expressed and purified at high concentration, before being exposed to a variety of conditions to force the protein out of solution and into crystalline form. From there, the crystals are harvested, mounted on a goniometer, and exposed to an x-ray beam. The resulting diffraction patterns, typically obtained from 180-360° of rotation, can be processed to determine information about the asymmetric unit of the input protein structure and its crystalline symmetry group, before using that information to determine "structure factors"[2]. From the diffraction data, a 3-D map of the electron density of the target can be calculated, into

which a protein model is iteratively built and refined. After rounds of refinement, a thermodynamically favorable conformation is selected as the structure of the protein[2].

There are several advantages to using x-ray crystallography as a method of structure determination, specifically the greatest atomic-level detail at the highest resolution of current methods. However, there are many drawbacks as well. The target must be crystallizable and only provides a static view of one potential conformation of the target. Despite its disadvantages, x-ray crystallography remains one of the most utilized techniques for protein structure visualization that is utilized in the fields of structural biology, molecular biology, and biochemistry. For the purposes of this research, x-ray crystallography was the best option for protein structure visualization and analysis.

Polyketide Synthases

Polyketides are a structurally and chemically diverse class of natural products with equally variable bioactivity. They are formed through the condensation of an acyl-thioester unit, such as malonyl-CoA and methylmalonyl-CoA, resulting in the formation of a diverse class of metabolites with varying structures[3]. Polyketides are derived from animals, plants, fungi, and bacteria and have been found to have extreme medicinal importance[4].

Reduced polyketides form the scaffold of a plethora of drugs, such as antibiotics (erythromycin[5]), immunosuppressants (rapamycin[6]), cholesterol-reducers (lovastatin[7]), insecticides (spinosyn[8]), and veterinary drugs (tylosin[9]). As a natural product, polyketides have

an advantage in scaffolding when interacting with biological targets, which make them a very attractive pharmacological target. The annual sales of pharmaceuticals that are derived from polyketides reaches twenty billion dollars[10].

Polyketides are produced by polyketide synthases (PKS), which are very attractive in the realm of combinatorial chemistry. Due to the complex nature of PKSs and the varying classes, as will be discussed, combinatorial chemistry seeks to engineer specific, non-natural PKSs with selected enzymes. The use of the natural products rather than attempting to generate these drugs synthetically is necessary; the large ring scaffolds contain a plethora of function groups, chiral centers, and more that make pure chemical synthesis difficult. In this way, generating highly specific drug intermediates, with chirality, acyl groups, and more selected for, can be accomplished through the newly engineered PKS. For this reason, PKS structures are of high importance to both the pharmaceutical and medicinal chemistry industries.

Despite the diversity of reduced polyketides, the biosynthetic machinery of PKSs is highly conserved across biosynthetic pathways, with each polyketide formed from carboxylic acid precursors[11]. There are, broadly speaking, three classes of PKSs to date.

Type 1 PKSs are multifunctional peptides, with linearly arranged domains that are covalently fused[4]. Type I PKSs are comprised of two groups within themselves; the iterative type, mostly found in fungi (iPKS), or the non-iterative modular type, found primarily in bacteria (modPKS)[4,12]. iPKSs repeatedly catalyze many rounds of elongation, while modPKSs function in an assembly line manner, similar to how nonribosomal peptide synthases function[4].

Type II PKSs are also multienzyme complexes, but they exclusively function iteratively, containing at least a ß-ketoacyl synthase (KS) and an acyl carrier protein (ACP). Found predominantly in bacteria, they are comprised of monofunctional proteins and tend to produce aromatic polyketides[4].

Type III PKSs or chalcone synthase-like PKSs are found predominantly in plants. At a basic level, they function as condensing enzymes that lack an ACP[11]. Rather, they use CoA as the anchor for chain extension[13]. Additionally, they exist as homodimers, rather than the complex multiprotein complexes of the Type I or II PKSs. However, as with the Type II PKSs, they function in an iterative manner[4].

Even with the variations in the PKS biomachinery, the diversity of the polyketide is largely attributed to the variegation of the ß-keto-processing domains, causing differing amount of ß-keto reduction[12]. The enzyme domains of PKS systems share functions with the enzymes of fatty acid synthases and suggest a common ancestor[14]. Fatty acid synthase function to catalyze the de novo synthesis of fatty acids through an iterative process of repeated condensations with a malonyl-CoA substrate[15]. It exists as a homodimer, with each monomer containing seven domains with a specific function[15].

In a very similar manner to the metazoan fatty acid synthase (FAS), many PKS systems form their products through several identical rounds of polyketide extension and modifications. However, the type I modular PKSs are well-characterized as well. These are more complex,

having a distinct multi-enzyme for each round of extension and modification, making them exciting targets to engineer toward the synthesis of natural products or analogs with an enhanced yield, activity, or bioavailability[16–18]. In most modular type I PKS, each active site is only used one time, causing a non-iterative process.

The bacterial modPKS systems function in a very similar manner to an assembly line, organized into catalytic modules. The modPKS contain several domains with distinct and defined functions, each separated by interdomain linkers[4]. Each module sequentially extends and reduces the emerging polyketide intermediate(s) in a noniterative manner, while regenerating its various functional groups[4]. The KS in each module catalyzes the extension of the nascent polyketide via an acyl building block (e.g., malonyl), selected by an acyltransferase (AT) from the acyl-CoA pool. Then, the polyketide product may be variably reduced by a variety of enzymes, including ketoreductase (KR), dehydratase (DH), or enoylreductase (ER) domains. An ACP domain within the module then utilizes a covalent phosphopantetheine cofactor to bind and transfer intermediates across or within modules to the appropriate domain and processing to be carried out. Finally, a thioesterase (TE) domain in the terminal module severs the bond between the polyketide and PKS to offload the pathway product, usually via hydrolysis or macrolactone formation. An example of modular type I PKSs include those found in the tylactone-based macrolide antibiotic pathway[19] (Figure 1). Among characterized type I PKSs to date, each module contains at least one ß-KS, one AT, and one ACP to generate a ß-ketoacyl-*S*-ACP intermediate[11].

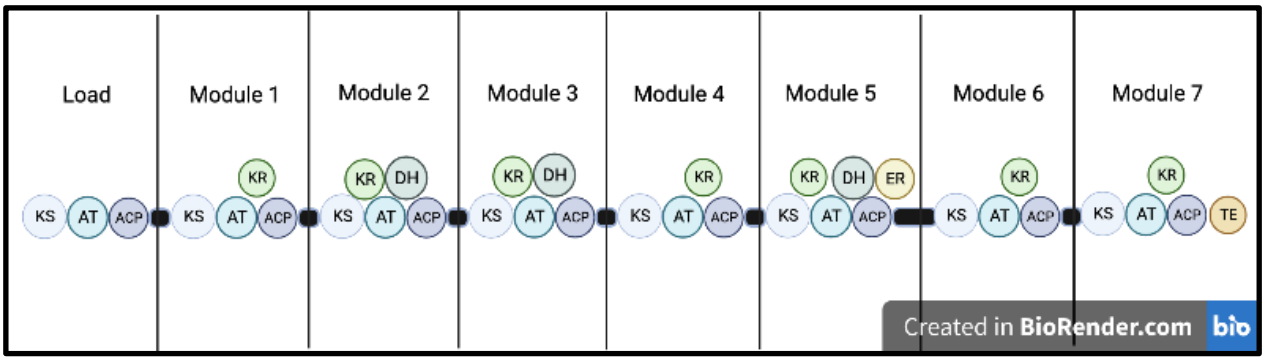


Figure 1: Example of modular PKS from tylactone-based macrolide antibiotic pathway[19]. Adapted from Lowell et. al's work and generated in BioRender. Visualization of load and varying modules in a modPKS. Each module minimally contains a KS, AT, and ACP domain, with varying reducing regions to provide a noniterative manner of polyketide synthesis.

In modPKS, there are two distinct methods of extension. Embedded or *cis*-AT versus dissociated or *trans*-AT provide the building blocks for extension with the module. Interestingly, the *cis*-AT modPKS shared a common ancestor with the metazoan FAS, evidenced by the split KR domain, which consists of both structural (KR_S) and catalytic (KR_c) subdomains. Additionally, it contains the common domain order within its megasynthase of KS-AT-DH-KR_S-ER-KR_c-ACP. PKS modules are dimeric, once again, similar to FAS, and were presumed to have a 3D architecture similar to the dimeric FAS[20]. Efforts to visualize PKS modules have resulted in crystal structures for all canonical PKS enzymes as excised, single domains[21,22], or as didomain constructs[23].

The size (reaching up to ~500 kDa) and flexibility of the modules of PKS are some of the most significant challenges posed to visualizing the multidomain architecture of the module using x-ray crystallography. The first moderate resolution structure of a full PKS module came from the actinobacteria pikromycin's PKS. Specifically, PikAIII (KS-AT-KR-ACP) was visualized, being

the penultimate module of the entirety of the pikromycin PKS pathway[24,25]. More advanced cryo-EM technologies were applied to engineered structures, designed for greater conformational stability, which resulted in higher-resolution structures for two additional PKS modules[26,27]. Structures for fungal iPKS[28], bacterial iPKS[29], and a *trans*-AT[30] have been reported as well. These structures provide evidence for key differences in function between various PKS systems, as well as the challenges that arise from imaging large, multidomain, flexible systems. To date, the visualization of the structure of a modular PKS reducing region with all enzyme domains (DH-ER-KR) has not been reported.

Cis-AT PKS modules, the metazoan FAS, and iPKS have two distinct structural regions. The first is an N-terminal extension region, comprised of the KS-AT domains, with the second being a reducing region, comprised of the DH-KR_s-ER-KR_c domains and an ACP, which is tethered flexibly to the C-terminus of the protein. Crystal structures of FAS[20] and a bacterial iPKS known as MAS[29] provided initial views of an entire reducing region. They exhibited a dimer architecture formation that are mediated by DH:DH and ER:ER contacts. However, crystal structures of excised ER domains[31] and a KR_S-ER-KR_C didomain[23] from modPKS suggested that the ER domains do not have dimeric architecture. A more complete understanding of the structure of the reducing region of the modPKS has clinical importance, as it could provide for a more effective design of engineered modules or chimeras from varying biosynthetic pathways.

While PKS systems exist as dimers naturally, the individual domain interfaces themselves are very weak, especially in type I PKS modules[30]. Therefore, excised domains don't necessarily behave as dimers once separated from the dimeric modules. Because of this, we sought to

confirm the presence of a DH:DH dimer interface within the JuvEIII DH-ER-KR tridomain, leading us to excise and attempt to solve the structure of the DH domain by itself.

Here, the structure of the modPKS DH domain from module 5 of the juvenimicin biosynthetic pathway of the actinobacterium *Micromonospora chalcea* (spp. *Izumensis*[19]) was solved and the DH:DH dimerization interface was confirmed. This tridomain (JuvEIII DH-ER-KR) (Figure 2) elucidates the architecture of the modPKS reducing region with a canonical DH:DH dimer interface, featuring DH:KR and ER:KR domain interfaces, mediated by surface-associated inter-domain linkers. Additionally, the overall structure of the JuvEIII DH-ER-KR lacks the ER:ER dimerization within the full reducing region, which is a departure from the existing reducing region domain orientation of FAS and iPKS.

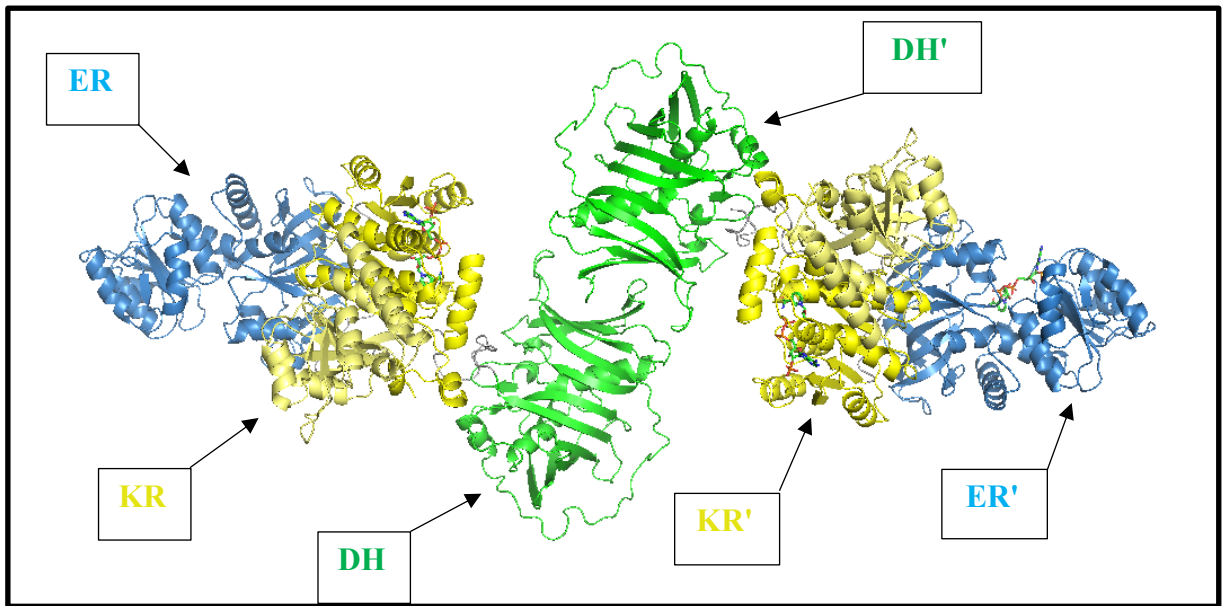


Figure 2: Full structure of JuvEIII DH-ER-KR reducing region. Generated in PyMol. Chain A is labeled with DH, ER, and KR domains, respectively. Chain B is denoted via prime notation.

2. Materials and Methods

Construct Design

The JuvEIII DH-ER-KR coding region was generated from a full-length plasmid encoding *juvEIII* (ARW71485). It was inserted into pMCSG7 via ligation independent cloning (LIC) to produce an expression plasmid pTMM22, which encodes JuvEIII residues 2478-3541. Residues 2478-2753 (the DH domain) was amplified from pTMM22 and inserted into pMCSG7 using LIC to create pTMM23. Both pTMM22 and pTMM23 contain an N-terminal 6x His tag followed by a cleavage site for tobacco etch virus (TEV) protease and the target protein. It also encodes an N-terminal 6x His tag, followed by the Mocr solubility fusion partner. Primers are listed in Table 1. JuveIII DH-ER-KR and JuvEIII DH expression plasmids were verified using nanopore sequencing via Plasmidsaurus.

Primer Name	**5' – 3'**
JuvEIII_DHERKR_Forward	5'-CATCCGCTGCTGGGCGGG-3'
JuvEIII_DHERKR_Reverse	5'-CACCAGGTCGGCGAGGAGCG-3'
JuvEIII_DHERKR_Forward_LIC	5'-**TACTTCCAATCCAATGCA**CATCCGCTGCTGGGCGGG-3'
JuvEIII_DHERKR_Reverse_LIC	5'-**TTATCCACTTCCAATGTTA**CACCAGGTCGGCGAGGAGCG-3'
JuvIII_DH_Nhelix_F	5'-GGGGACGTGGCCTCCGCCGGG-3'

JuvEIII_DH_Reverse	5'-CACCGGCACCGCGCGCGCCAC-3'
JuvIII_DH_Nhelix_F_LIC	5'-**TACTTCCAATCCAATGCA**GGGGACGTGGCCTC CGCCGGG-3'
JuvEIII_DH_Reverse_LIC	5'-**TTATCCACTTCCAATGTTA**CACCGGCACCGCG CG CGCCAC-3'

Table 1: Summary of primers used.

Protein Expression and Purification

Plasmids encoding JuvEIII DH-ER-KR and JuvEIII DH were expressed in the same manner. Cells of *Escherichia coli* strain BL21 (DE3) were transformed with either pTMM21 or pTMM22. Transformed cells were grown in 0.5 L of TB media with 100 μg/mL ampicillin to an OD_{600} = 2.0 at 37°C. Cultures were cooled to 20°C over the span of 1 hour, induced with 200 μM IPTG and expressed for 16 hours. Cells were harvested via 30-minute centrifugation at 8,000 RPM and cell pellets were stored at -80°C.

Cell pellets containing JuvEIII DH-ER-KR or JuvEIII DH were resuspended in 35 mL lysis buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole, 0.1 mg/mL lysozyme, 0.05 mg/mL DNase, 2 mM $MgCl_2$), incubated on ice for 45 minutes, lysed by sonification, and centrifuged at 38,760 x g for 30 minutes at 4°C. The soluble fraction was filtered through a 0.45 μm syringe filter, loaded onto a 5 mL HisTrap column (Cytiva Life Sciences), washed with 10 column volumes of buffer A (50 mM HEPES pH 7.4, 300 mM NaCl, 10% (v/v) glycerol), and eluted with a linear gradient of 20-400 mM imidazole with buffer A and

buffer B (50 mM HEPES pH 7.4, 300 mM NaCl, 10% (v/v), 400 mM imidazole). Fractions containing JuvEIII DH-ER-KR or JuvEIII DH were pooled. The 6x His tag was removed via overnight incubation at 4°C with TEV protease (1:30 ratio of protease:JuvEIII protein), with the resulting protein flowed through a 5-mL HisTrap column to remove any uncleaved JuvEIII protein and the protease. DH-ER-KR or DH was concentrated by filter centrifugation (Amicon) and further purified via size exclusion chromatography in buffer C (50 mM HEPES pH 7.4, 150 mM NaCl, 10% (v/v) glycerol). JuvEIII DH-ER-KR (HiLoad 16/60 Superdex S300) eluted at an apparent molecular weight of 132 kDa, between the molecular weights of the monomer (109.5 kDa) and dimer forms (219 kDa). Similarly, JuvEIII DH (HiLoad 16/60 Superdex S75) eluted at an apparent molecular weight of 40.2 kDa, between the monomer (32.4 kDa) and dimer (64.8 kDa) molecular weights.

Interestingly, JuvEIII DH-ER-KR (HiLoad 16/60 Superdex S300) eluted at an apparent molecular weight of 132 kDa, which is between the weight of the monomer (109.5 kDa) and the dimer (219 kDa). Similarly, JuvEIII DH (HiLoad 16/60 Superdex S75) eluted at an apparent molecular weight of 40.2 kDa, once again, between the monomeric (32.4 kDa) and dimeric (64.8 kDa) molecular weights. After the removal of the His tag on the N-terminal, however, the DH eluted as an apparent dimer at a molecular weight of 65 kDa.

Protein Crystallization and Structure Determination

JuvEIII DH-ER-KR was crystallized at 4°C by hanging drop vapor diffusion in a 1:1 mixture of 2 μL protein stock (10 mg/mL JuvEIII DH-ER-KR in buffer C with 2 mM NADP) and reservoir

solution (2.0 M $(NH_4)_2SO_4$, 100 mM Bis-Tris pH 6.5). The protein was incubated with 2 mM NADP for 1 hour on ice prior to attempting crystallization. Large, single, lens-shaped crystals grew within 7-10 days. Crystals were harvested without additional cryoprotection and were flash cooled directly into liquid nitrogen.

JuvEIII DH was crystallized at 4°C via sitting drop vapor diffusion in a 1:1 mixture of 0.5 μL of protein stock (10 mg/mL JuvEIII DH in buffer C) and varying reservoir solutions, with the best crystals formed in the following condition: (buffer: 0.1 M Tris-HCl pH 8.5, precipitant: 2.4 M $(NH_4)_2HPO_4$). A variety of crystals, including those of a large and single nature, grew within 7-10 days. Crystals were harvested without additional cryoprotection and were flash cooled directly into liquid nitrogen.

The JuvEIII DH-ER-KR and JuvEIII DH diffraction data were collected at 100 K on GM/CA beamline 23ID-B at the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL). Data were processed and scaled using XDS[32]. The structure of JuvEIII DH-ER-KR was solved by molecular replacement with Phaser[33] in the PHENIX[34] software suite. AlphaFold2[35] was used to predict the structure of full-length JuvEIII DH-ER-KR from which the DH, ER, and KR individual domains were excised and used as molecular replacement search models.

JuvEIII DH5 data was collected at APS and the data was processed using XDS[32] (Table 2). The structure of JuvEIII DH5 was solved at low-resolution using Phaser[33] in the PHENIX[34] software suite. Refinement of the structure was unsuccessful due to low resolution data and

streaky, multiple lattices visible on diffraction patterns. Several space group arrangements were used in an attempt to solve the structure, with data set 14 providing the best, albeit still low, quality solution. Visualization of the DH:DH dimer interface was done in Coot[36] and PyMOL.

4. Results

When attempting to generate the domain structure of a fully reducing type I PKS module, both crystallographic and cryo-EM approaches were utilized to elucidate a three-dimensional architecture. The tridomain termini were determined via structural and sequence alignments of PKS DH[37] and KR[38] domains. Testing across several modules, it was shown that the DH-ER-KR tridomain from module 5 of the juvenimicin PKS[19] (amino acid code: 2478-3541) yielded a highly pure, stable protein. JuvEIII DH-ER-KR eluted from a size-exclusion column as a monodisperse species at an apparent molecular weight of 132 kDa. The tridomain molecular weight as a monomer is 109.5 kDa, with the dimer molecular weight being 219 kDa. Similarly, JuvEIII DH eluted at an apparent molecular weight of 40.2 kDa. Again, this was between the monomeric (32.4 kDa) and dimeric (64.8 kDa) molecular weights. After the removal of the His tag on the N-terminal, however, the DH eluted as an apparent dimer at a molecular weight of 65 kDa. Both JuvEIII DH and DH-ER-KR yielded promising results in crystallization experiments, but not in electron microscopy screenings.

This elution as a dimer following the cleavage of the His tag demonstrates how weak the dimerization capacity is and reinforces our goal of visualizing the DH:DH interface for potential uses in combinatorial chemistry. Following solving a new type I PKS DH domain and after characterizing it's weak dimer interface, we attempted to evaluate all available PKS DH domains for characteristics that may result in or cause dimeric vs. monomeric behavior in solution (Table 3). A sequence alignment of all solved DH domain N-termini was generated (Figure 3A) which revealed that the fluvirucin DH[39] and others were crystallized with an extended N-terminus, including a dimer interface-participating helix (Figure 3B). We developed two constructs of the

JuvEIII DH domain, one including an extra helix on the N-terminal domain, and one without it, to determine if the helix itself participates in the dimer interface or if it stabilizes the dimer contact.

After expression and purification of the JuvEIII DH domain containing the N-terminal helix, we again see the maximum peak occurring at an elution volume of 15.7 mL, showing an apparent molecular weight of 40 kDa (Figure 4). This is again between the monomeric and dimer weights, illustrating the weak dimerization contacts of individual domains.

Name	Z-score	RMSD (to JuvIII_DH)	% Seq ID	Oligomer State	Modular State	Nres	% hydrophobicity	# Salt bridges
JuvIII_DH	N/A	N/A	N/A	Mono/Dim	Bimodule	51	45%	1
FluDH	35.1	1.7	43	?	Mono	55	45%	2
RifDH	33.8	1.7	43	?	Bimodule	56	48%	1
DEBSM4_DH	33.8	2.0	43	Mono/Dim	Bimodule	67	45%	0
CmiDH	33.7	1.7	45	Dimer	Split	55	45%	1
BorrDH	33.5	1.7	40	?	Mono	54	43%	0
AmbC_DH	32.3	2.0	44	Monomer	?	42	48%	0
GphF_DH	29.2	2.4	29	Monomer	?	50	32%	0
MAS DH	30.4	2.3	26	?	Mono	37	43%	0
CurF_DH	31.5	2.3	22	Dimer	Mono	54	35%	0
CurK_DH	31.4	2.3	22	Dimer	Mono	57	36%	3
CurJ_DH	30.0	2.3	20	Dimer	Mono	61	38%	0
CurH_DH	29.1	2.5	23	Dimer	Mono	50	36%	1

Table 3: Comparison of solved DH domain dimer interfaces with most similar three-dimensional structure (DALI[40]) with Tyler McCullough.

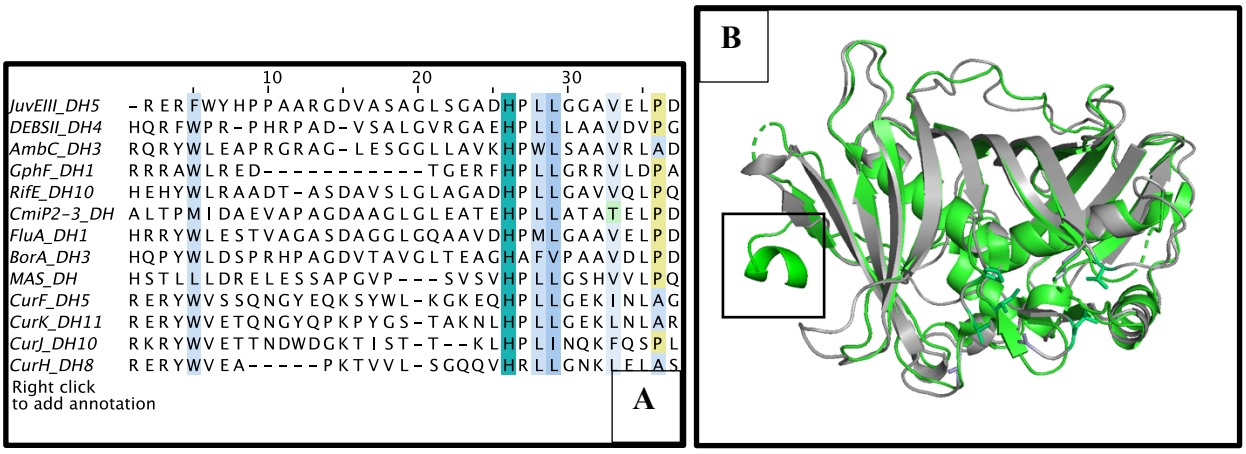


Figure 3: Comparison of JuvEIII DH and JuvEIII DH with N-terminus helix. A: Sequence alignment comparison, generated in Jalview. **B**: Structural image comparison of Flu and JuvIII DH excised domains, made in PyMol. There is high structural homology, barring the presence of the helix (within the box).

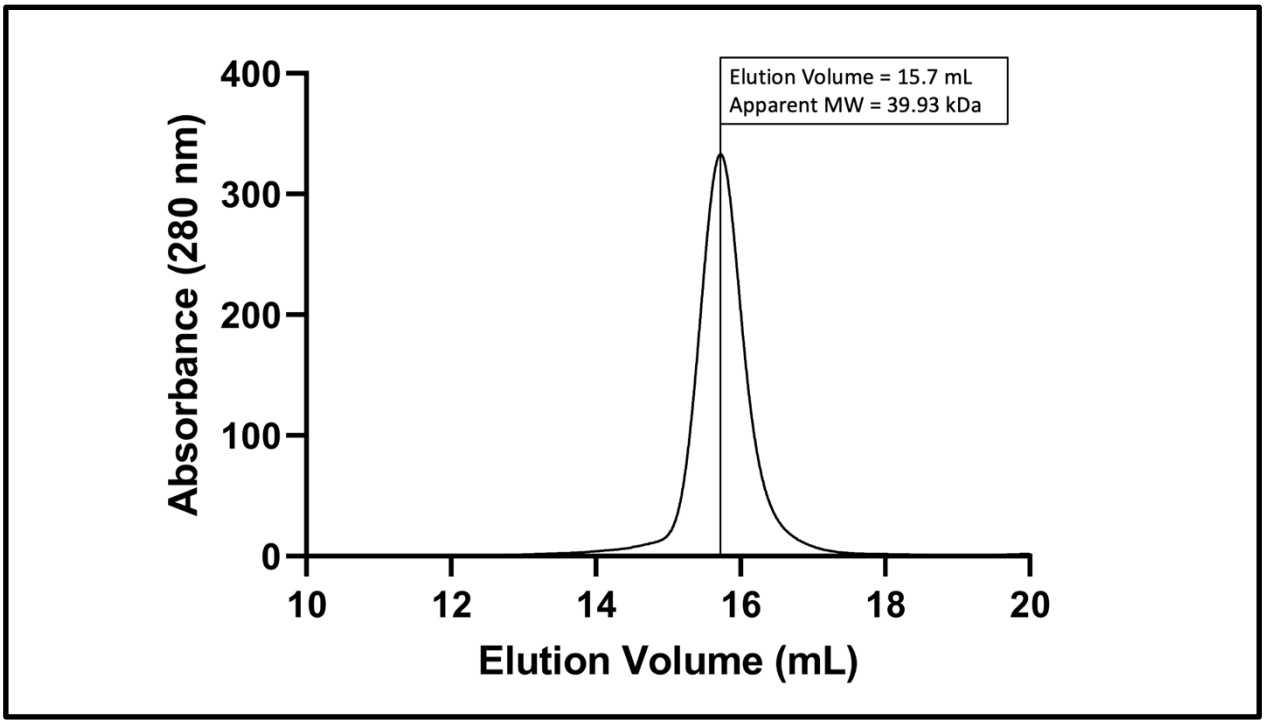


Figure 3: Analytical size exclusion graph of JuvEIII DH with N-terminal helix. Eluted at an apparent molecular weight of 40 kDa, between the monomeric and dimeric weights.

Crystallization Results of JuvEIII DH

JuvEIII DH5 with an extended N-terminal helix crystallized under various conditions at 4°C (Figure 4). A variety of crystal types were obtained, including large and single lens crystals, found during broad screen hits with conditions of buffer: 0.1 M Tris: HCl pH 8.5 and precipitant: 2.4 M $(NH_4)_2HPO_4$ (Figure 5A). All crystals were formed at a protein concentration of 10 mg/mL, in sitting drop solutions with a 0.5 µL: 0.5 µL ratio of protein solution to reservoir solution. In parallel, JuvEIII DH-ER-KR crystals were also formed, but these were not used when moving forward with structure determination, as the original conditions proved suitable for crystal formation and the broad screens were unnecessary (Figure 6).

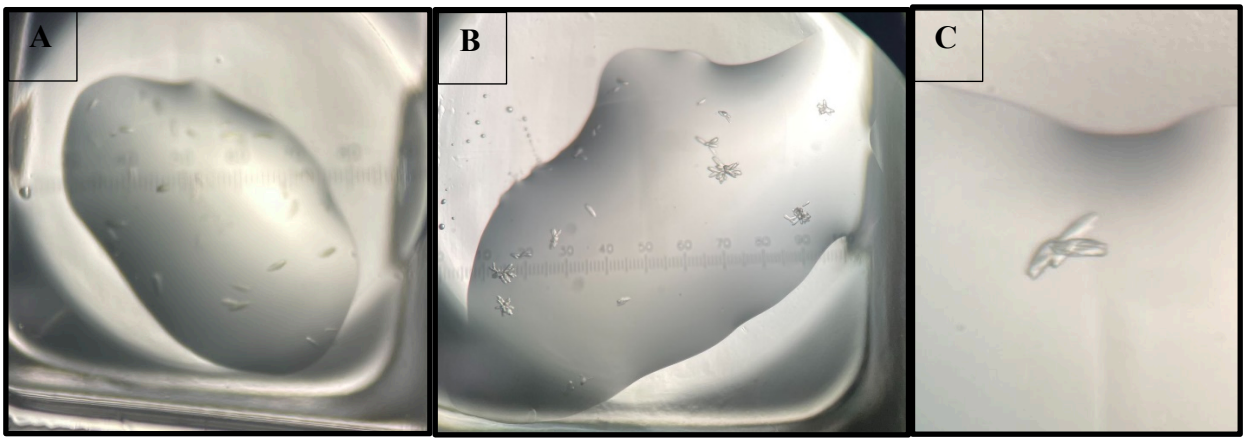


Figure 5: Cleaved JuvEIII DH N-terminus broad screen crystallization tray results. A: Large, single crystals formed in MCSG2 G2 condition (buffer: 0.1 M Tris: HCl pH 8.5, precipitant: 2.4 M $(NH_4)_2HPO_4$). **B:** Small, multi-lens crystals formed in the condition: buffer: 0.09 M HEPES: NaOH, pH 7.5, precipitant: 1.26 M sodium citrate, 10 (v/v) glycerol. **C:** Small, multilens crystals formed in: buffer: 0.1 M HEPES NaOH, pH 7.5, precipitant: 1.4 M sodium

citrate. All crystals obtained from broad screens diffracted between 3.11-3.00 Å and were used in designing follow-up conditions to optimize crystallization of the target.

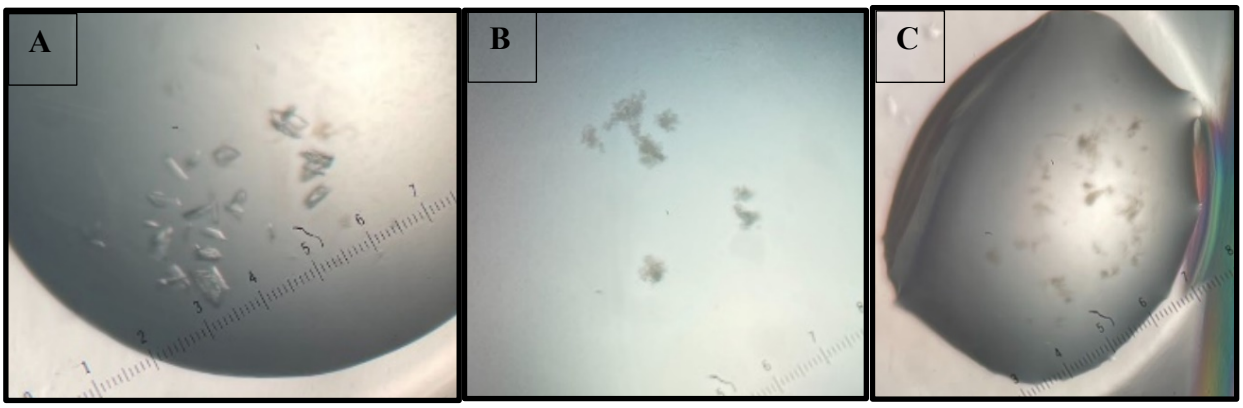


Figure 6: JuvEIII DH-ER-KR broad screen crystallization tray results. A: Large, fragmented crystals formed in condition: salt: 0.1 M magnesium formate, precipitant: 15% (w/v) PEG 3350. Provided the best crystal results and was used to generate the follow-up conditions. **B:** Small, multi-lens crystals formed in condition salt: 1.1 M sodium malonate, pH 7.0, buffer: 0.1 M HEPES: NaOH, pH 7, precipitant: 0.5% (w/v) jeffamine ED-2001, pH 7.0. **C:** Small, fragmented crystals formed in: salt: 0.1 M sodium acetate, buffer: 0.1 M HEPES NaOH, pH 7.5, precipitant: 22% (w/v) PEG 4000. These results were not used to move forward with JuvEIII DH-ER-KR structure determination, as the original condition was found to be sufficient for crystallization.

Diffraction Data Results from JuvEIII DH N-terminus

JuvEIII DH N-terminal crystals diffracted to a resolution of 3.06 Å (Table 2). Despite medium-resolution diffraction, the overall quality of the data was fairly low. However, the structure was solved via molecular replacement using an excised DH domain from the JuvEIII DH-ER-KR

structure (Figure 8). The asymmetric unit of space group $P3_221$ included two DH polypeptides (A and B). The JuvEIII DH-ER-KR structure was solved via molecular replacement as well, using AlphaFold2[35].

Protein	JuvEIII DH-ER-KR
Ligand	N/A
Diffraction data	
Space group	$P3_221$
Unit cell, a,b,c (Å)	62.34, 62.34, 122.62, 90, 90, 120
X-ray source	APS 23ID-B
Wavelength (Å)	1.033
d_{min} (Å)	3.06 (3.11 – 3.06)*
R_{meas}	0.258 (2.45)
Avg I/σ (I)	8.1 (0.9)
Completeness (%)	99.4 (99.8)
Redundancy	9.6 (9.6)
$CC_{1/2}$	0.995 (0.37)
Refinement	
Data range (Å)	50.0 – 3.06

* Values in parentheses pertain to the outermost resolution shell.

Table 2: Diffraction data from JuvEIII DH N-terminal region crystals.

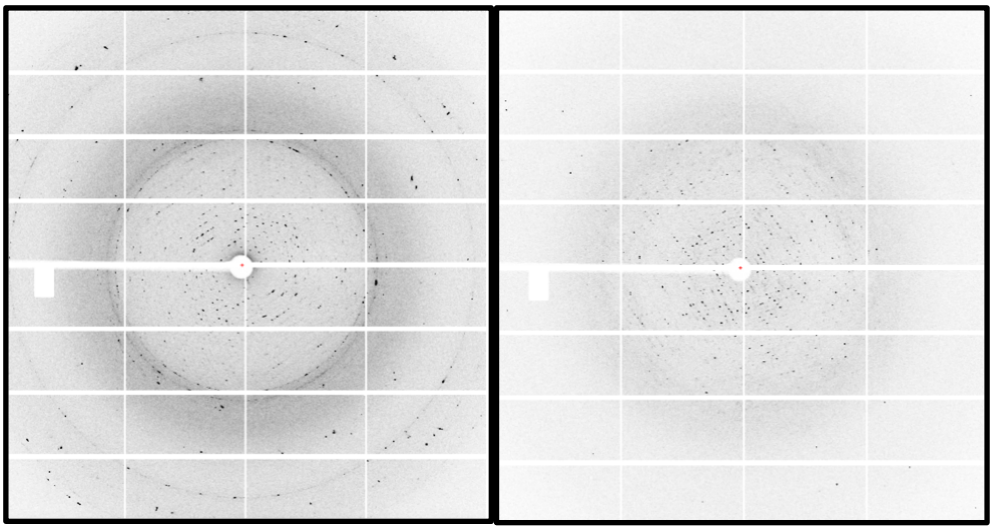


Figure 7: Diffraction data from JuvEIII DH N-terminus crystals. Taken from data set 14, showing spot-streaking from the diffraction pattern and potentially multiple crystals. However, we were able to visualize the DH:DH dimer interaction that we were looking for.

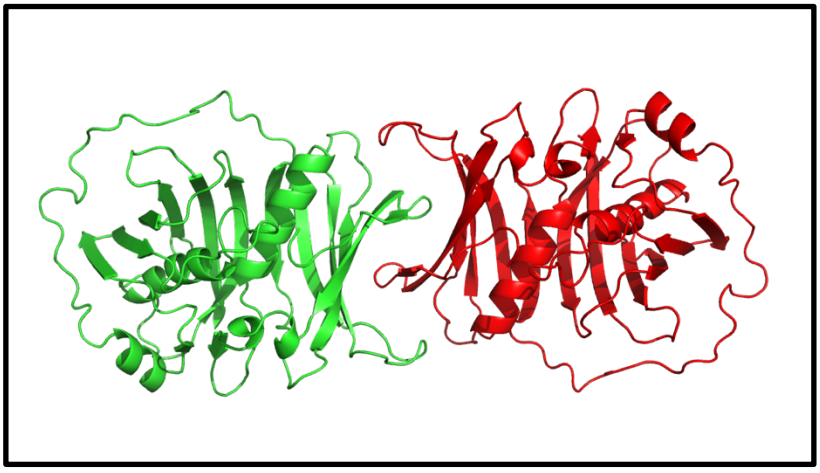


Figure 8: DH:DH interface from the JuvEIII DH-ER-KR reducing region. Generated in PyMol.

Dimerization of the DH:DH Interface

PKS modules are dimeric, so when identifying a JuvEIII dimer contact, the monomer-monomer DH:DH contact in the tridomain crystal was heavily considered. It consists of several hydrophobic side chains (Val2485, Leu2487, Pro2488, Val2494, Val2561, Trp2588).

Additionally, the DH:DH contact (Figure 9) is identical to dimer contacts in ten out of eleven structures of excised DH domains from PKS modules[38,40,41]. The flat, extended JuvEIII DH dimer has an interdomain angle of ~174°, which is very similar to the structures of excised DH domains from other type I PKS modules[37,41,42] (inter-domain angles of 172-207°) and distinct from the V-shaped DH-DH interface of porcine (113°)[20] and the LovB iPKS (124°)[28]. The DH domain, like the DH-ER-KR construct, eluted from a size-exclusion column with an apparent molecular weight of 40 kDa, between the monomeric and dimeric molecular weight. This is common amongst excised DH domains[41,43,44].

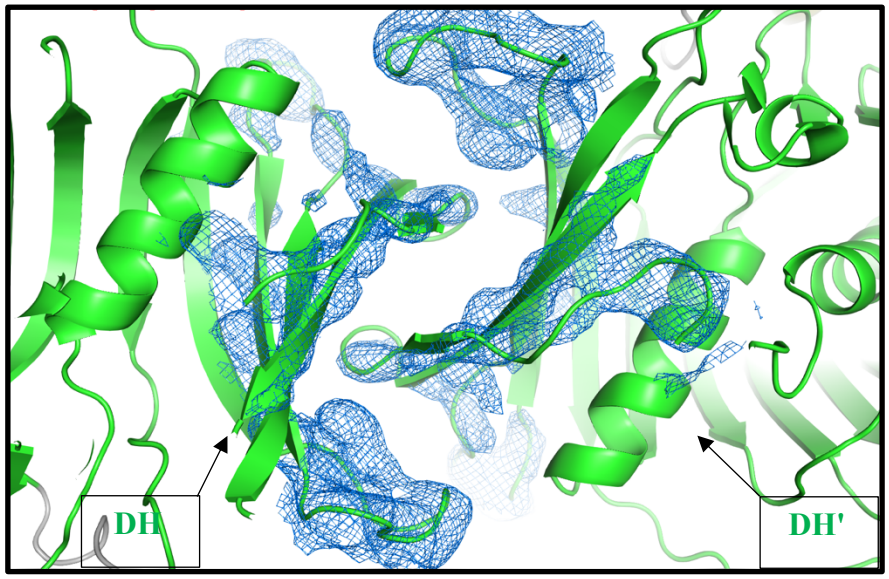


Figure 9: Visualization of DH:DH dimerization within DH-ER-KR reducing region via electron density mapping. Generated in PyMol.

5. Discussion

Designing an engineered modPKS system for selective, efficient synthesis of novel polyketide depends on having a complete understanding of module structure, function, and connection beyond the individual catalytic domains[16,17]. PKS modules have inherent flexibility, providing challenges for visualization in x-ray crystallography. Structures of PKS modules, all composed of KS-AT-KR-ACP catalytic domains, obtained via cryo-EM, have highlighted the issues of visualizing such a dynamic system[24–27]. When comparing results to the metazoan FAS[20] and iPKS systems[28,29], the architecture of PKS modules reveal distinct features of modular and iterative systems. Providing the first view of a complete reducing region from a type I *cis*-AT modular PKS is an important addition to the overall understanding of the PKS module architecture. While the structure of the JuvEIII DH-ER-KR remains consistent with previously published literature, the multi-domain architecture is an extremely important discovery, providing a departure from the typical architecture of the iterative systems.

Visualizing the JuvEIII DH N-terminus region provided an avenue to observe the DH:DH dimerization interface. Determination of this interface, which was observed in both the DH-ER-KR full reducing region structure and in the isolated DH N-terminus structure, was successful overall. With the discovery of the multi-domain architecture, identifying interfaces that remain consistent with previously published literature provides insight into the ways that individual modules interact with each other. However, the sample collection was not pure, and the diffraction data collected was from multiple crystals, making assignment of a space group difficult.

PKS modules exhibit several sites of dimerization, including at the DH:DH interface. Cooperativity among several weak contacts enable the formation of the dimer in modPKS systems, which can include contributions from more than one module. The observation of the monomer: dimer mixture in solution after size-exclusion gives evidence for the weakness of the affinity of the individual dimer contacts. This remains consistent with previous literature for other PKS modules[45]. On the other hand, the lack of ER dimerization in JuvEIII is consistent with all previously reported observations of monomers in solution and in structures of ER domains excised from other PKS modules[31]. In contrast, the iterative systems also have dimer contacts of KS, DH, and ER domains, but not in the terminal TE. Here, the DH:DH contact is also weak, as we observed in the excised human FAS DH domain, which was a monomer in solution and in crystals.

We attempted to compare differences in dimerization strength in the DH interface across a variety of solved DH domains (Table 3). Several structural features were compiled, from the hydrophobic character of the domains to the presence of salt bridges that participated in forming the interface. The modular character itself was considered as well, trying to determine if pulling the DH from a mono-modular versus bi-modular system influenced dimerization strength. In the end, no clear pattern seemed to emerge. However, we did see that the fluvirucin DH domain contains an extended DH domain, allowing us to extend the juvenimicin DH domain and crystallize its N-terminal helix to characterize its effect in the dimer interface.

The JuvEIII DH-ER-KR structure can serve as a prototype for the reducing region of a type I *cis*-AT modular PKS system. Additionally, it illustrates how the structure of PKS modules differ

from their iterative homologs, such as the metazoan FAS[20] and two iPKS[28,29]. Even though the ordering of the catalytic domains is identical, the arrangement of the domains relative to one another is notably different in the modular and iterative synthases. Interestingly, the KR domain is positioned between the DH and ER domains, which is strikingly different from the reducing domains in the metazoan FAS or iPKS systems[20,28,29], wherein the DH and ER are adjacent. The modPKS enzyme domains of the reducing region are closely associated, with DH and ER at opposite surfaces of the KR_C subdomain in contacts that are strongly mediated by the respective linker peptides. In contrast, the iterative systems have both the DH and ER domains on the same side of the KR, as well as more extensive linkers that are not generally associated with surfaces of enzyme domains.

The interdomain linkers are also shorter in modPKS than in the iterative systems. For example, the ER-KR_C linker in the modPKS is roughly 8 residues but is roughly 30 residues in length in the FAS[20] and iPKS[29] systems. These linkers function as a "glue" between domains as hydrophobic, surface-associated residues. In iterative systems, however, the longer linkers tend to possess some secondary structure and are less associated with the enzyme surface than in the modPKS. Both the longer length and non-surface association of the iPKS systems permit a greater range of motion than in the modular systems. These observations suggest that the modPKS fully reducing region may be less dynamic than its iterative counterparts, largely due to the structure and function of their linker proteins. The JuvEIII DH-ER-KR represents the canonical, fully reducing region of *cis*-AT modular PKS systems, illustrating how the modPKS systems have diverged from the FAS and iterative PKS systems.

In conclusion, the JuvEIII DH domain behaved similarly to the full DH-ER-KR in solution. Despite an exhaustive literature search and structural comparison, there seems to be no correlated features of the DH dimer interface that predicts monomeric vs. dimeric solution state behavior. Therefore, it appears more likely that DH domains likely possess weak dimer interfaces that are labile in absence of additional upstream and downstream dimer contacts found in the context of a full module. This structure is important for the advancement of the design approach to engineered modPKS systems. We believe the low structure quality of the DH N-terminus region is due in part to poor selection and the presence of multiple crystals within the sample. Going forward, steps to optimize this structure include refining the crystallization conditions to minimize multi-crystal contamination.

6. Acknowledgements

Firstly, I would like to thank Dr. Janet Smith and everyone in her lab for investing so much time and energy toward my continuing growth as a scientist. I have learned an innumerable number of things from your laboratory and found a space where I could explore my passion for biochemistry without fear of failure. Your support through almost my entire undergraduate collegiate career will not be forgotten.

Additionally, I would like to name Tyler McCullough and Michael Rankin for all their support over the years. You have been some of the best mentors I have ever had, and I will forever be grateful. Even though I was just an undergraduate and came to you with almost no experience, you never treated me as anything less than an equal and were always there to lend a hand when I needed it.

I also would like to thank my family and friends for all their love and support over the years. Sam Edgcombe and Steven Dunne, you took a chance on a young one such as myself and helped me find my space and confidence as an aspiring scientist. You always were people I could turn to for help during my undergraduate career, and I will never forget it. To my family, thank you for always pushing me to be the best I can be, even when I thought I couldn't do it. Your support has gotten me to places I never thought I could go.

Lastly, thank you to anyone who has encouraged me thus far in my career. You got me through late night study sessions and tough exams, and I couldn't have done it without you.

7. References

(1) White, B., Kahriman, A., Luberice, L., and Idleh, F. (2010) Evaluation of software for introducing protein structure: Visualization and simulation. *Biochem. Mol. Biol. Educ. 38*, 284–289.

(2) Smyth, M. S. (2000) x Ray crystallography. *Mol. Pathol. 53*, 8–14.

(3) Ridley, C. P., Lee, H. Y., and Khosla, C. (2008) Evolution of polyketide synthases in bacteria. *Proc. Natl. Acad. Sci. 105*, 4595–4600.

(4) Wang, J., Zhang, R., Chen, X., Sun, X., Yan, Y., Shen, X., and Yuan, Q. (2020) Biosynthesis of aromatic polyketides in microorganisms using type II polyketide synthases. *Microb. Cell Factories 19*, 110.

(5) Staunton, J., and Wilkinson, B. (1997) Biosynthesis of Erythromycin and Rapamycin. *Chem. Rev. 97*, 2611–2630.

(6) Aparicio, J. F., Molnár, I., Schwecke, T., König, A., Haydock, S. F., Ee Khaw, L., Staunton, J., and Leadlay, P. F. (1996) Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: Analysis of the enzymatic domains in the modular polyketide synthase. *Gene 169*, 9–16.

(7) Tobert, J. A. (2003) Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors | Nature Reviews Drug Discovery. *Nat. Rev. Drug Discov. 2*, 517–526.

(8) Kim, H. J., Choi, S., Jeon, B., Kim, N., Pongdee, R., Wu, Q., and Liu, H. (2014) Chemoenzymatic Synthesis of Spinosyn A. *Angew. Chem. Int. Ed. 53*, 13553–13557.

(9) Cundliffe, E., Bate, N., Butler, A., Fish, S., Gandecha, A., and Merson-Davies, L. (2001) The tylosin-biosynthetic genes of Streptomyces fradiae. *Antonie Van Leeuwenhoek 79*, 229–234.

(10) Gomes, E. S., Schuch, V., and Lemos, E. G. de M. (2013) Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Braz. J. Microbiol. 44*, 1007–1034.

(11) Cheng, Y.-Q., Tang, G.-L., and Shen, B. (2003) Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. 100*, 3149–3154.

(12) Chen, H., and Du, L. (2016) Iterative Polyketide Biosynthesis by Modular Polyketide Synthases in Bacteria. *Appl. Microbiol. Biotechnol. 100*, 541–557.

(13) Shimizu, Y., Ogata, H., and Goto, S. (2017) Type III Polyketide Synthases: Functional Classification and Phylogenomics. *ChemBioChem 18*, 50–65.

(14) Jenke-Kodama, H., Sandmann, A., Müller, R., and Dittmann, E. (2005) Evolutionary Implications of Bacterial Polyketide Synthases. *Mol. Biol. Evol. 22*, 2027–2039.

(15) Jensen-Urstad, A. P. L., and Semenkovich, C. F. (2012) Fatty acid synthase and liver triglyceride metabolism: Housekeeper or messenger? *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids 1821*, 747–753.

(16) Cai, W., and Zhang, W. (2018) Engineering modular polyketide synthases for production of biofuels and industrial chemicals. *Curr. Opin. Biotechnol. 50*, 32–38.

(17) Klaus, M., and Grininger, M. (2018) Engineering strategies for rational polyketide synthase design. *Nat. Prod. Rep. 35*, 1070–1081.

(18) Zargar, A., Lal, R., Valencia, L., Wang, J., Backman, T. W. H., Cruz-Morales, P., Kothari, A., Werts, M., Wong, A. R., Bailey, C. B., Loubat, A., Liu, Y., Chen, Y., Chang, S., Benites, V. T., Hernández, A. C., Barajas, J. F., Thompson, M. G., Barcelos, C., Anayah, R., Martin, H. G., Mukhopadhyay, A., Petzold, C. J., Baidoo, E. E. K., Katz, L., and Keasling, J. D. (2020)

Chemoinformatic-Guided Engineering of Polyketide Synthases. *J. Am. Chem. Soc. 142*, 9896–9901.

(19) Lowell, A. N., DeMars, M. D., Slocum, S. T., Yu, F., Anand, K., Chemler, J. A., Korakavi, N., Priessnitz, J. K., Park, S. R., Koch, A. A., Schultz, P. J., and Sherman, D. H. (2017) Chemoenzymatic Total Synthesis and Structural Diversification of Tylactone-Based Macrolide Antibiotics through Late-Stage Polyketide Assembly, Tailoring, and C—H Functionalization. *J. Am. Chem. Soc. 139*, 7913–7920.

(20) Maier, T., Leibundgut, M., and Ban, N. (2008) The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science 321*, 1315–1322.

(21) Weissman, K. J. (2015) Uncovering the structures of modular polyketide synthases. *Nat. Prod. Rep. 32*, 436–453.

(22) Keatinge-Clay, A. T. (2012) The structures of type I polyketide synthases. *Nat. Prod. Rep. 29*, 1050.

(23) Zheng, J., Gay, D. C., Demeler, B., White, M. A., and Keatinge-Clay, A. T. (2012) Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol. 8*, 615–621.

(24) Whicher, J. R., Dutta, S., Hansen, D. A., Hale, W. A., Chemler, J. A., Dosey, A. M., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., and Skiniotis, G. (2014) Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature 510*, 560–564.

(25) Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., and Skiniotis, G. (2014) Structure of a modular polyketide synthase. *Nature 510*, 512–517.

(26) Cogan, D. P., Zhang, K., Li, X., Li, S., Pintilie, G. D., Roh, S.-H., Craik, C. S., Chiu, W., and Khosla, C. (2021) Mapping the catalytic conformations of an assembly-line polyketide synthase module. *Science 374*, 729–734.

(27) Bagde, S. R., Mathews, I. I., Fromme, J. C., and Kim, C.-Y. (2021) Modular polyketide synthase contains two reaction chambers that operate asynchronously. *Science 374*, 723–729.

(28) Wang, J., Liang, J., Chen, L., Zhang, W., Kong, L., Peng, C., Su, C., Tang, Y., Deng, Z., and Wang, Z. (2021) Structural basis for the biosynthesis of lovastatin. *Nat. Commun. 12*, 867.

(29) Herbst, D. A., Jakob, R. P., Zähringer, F., and Maier, T. (2016) Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature 531*, 533–537.

(30) Tittes, Y. U., Herbst, D. A., Martin, S. F. X., Munoz-Hernandez, H., Jakob, R. P., and Maier, T. (2022) The structure of a polyketide synthase bimodule core. *Sci. Adv. 8*, eabo6918.

(31) Khare, D., Hale, W. A., Tripathi, A., Gu, L., Sherman, D. H., Gerwick, W. H., Håkansson, K., and Smith, J. L. (2015) Structural Basis for Cyclopropanation by a Unique Enoyl-Acyl Carrier Protein Reductase. *Structure 23*, 2213–2223.

(32) Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr. 66*, 125–132.

(33) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *Phaser* crystallographic software. *J. Appl. Crystallogr. 40*, 658–674.

(34) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr. 66*, 213–221.

(35) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature 596*, 583–589.

(36) Emsley, P., and Cowtan, K. (2004) *Coot* : model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr. 60*, 2126–2132.

(37) Akey, D. L., Razelun, J. R., Tehranisa, J., Sherman, D. H., Gerwick, W. H., and Smith, J. L. (2010) Crystal Structures of Dehydratase Domains from the Curacin Polyketide Biosynthetic Pathway. *Structure 18*, 94–105.

(38) Keatinge-Clay, A. T., and Stroud, R. M. (2006) The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure 14*, 737–748.

(39) Barajas, J. F., McAndrew, R. P., Thompson, M. G., Backman, T. W. H., Pang, B., de Rond, T., Pereira, J. H., Benites, V. T., Martín, H. G., Baidoo, E. E. K., Hillson, N. J., Adams, P. D., and Keasling, J. D. (2019) Structural insights into dehydratase substrate selection for the borrelidin and fluvirucin polyketide synthases. *J. Ind. Microbiol. Biotechnol. 46*, 1225–1235.

(40) Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Res. 50*, W210–W215.

(41) Sung, K. H., Berkhan, G., Hollmann, T., Wagner, L., Blankenfeldt, W., and Hahn, F. (2018) Insights into the Dual Activity of a Bifunctional Dehydratase-Cyclase Domain. *Angew. Chem. Int. Ed. 57*, 343–347.

(42) Fiers, W. D., Dodge, G. J., Sherman, D. H., Smith, J. L., and Aldrich, C. C. (2016) Vinylogous Dehydration by a Polyketide Dehydratase Domain in Curacin Biosynthesis. *J. Am. Chem. Soc. 138*, 16024–16036.

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

(44) Dodge, G. J., Ronnow, D., Taylor, R. E., and Smith, J. L. (2018) Molecular Basis for Olefin Rearrangement in the Gephyronic Acid Polyketide Synthase. *ACS Chem. Biol. 13*, 2699–2707.

(45) Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2013) Cyanobacterial Polyketide Synthase Docking Domains: A Tool for Engineering Natural Product Biosynthesis. *Chem. Biol. 20*, 1340–1351.