UC Irvine
UC Irvine Electronic Theses and Dissertations

Title

Structural and Functional Studies Involved in Polyketide and Fatty Acid Biosynthesis

Permalink

https://escholarship.org/uc/item/0162m9hf

Author

Jackson, David

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Structural and Functional Studies of Enzymes Involved in Polyketide and Fatty Acid
Biosynthesis

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSPOHY

in Chemistry

By

David Richard Jackson

Dissertation Committee:
Professor Shiou-Chuan (Sheryl) Tsai, Chair
Professor Thomas Poulos
Professor Jennifer A. Prescher

2015

DEDICATION

This dissertation is dedicated to my family. To my grandparents, for always being supportive of me when embarking on the many adventures in my life. I have learned more from them than anyone else. They are awesome. To my brother Jeff, sister-in-law Amy, Uncle Jim, Aunt Carol, and three nieces, Ellie, Maddie, and Lily, for their love and support and always making me feel close to home no matter how far away from Columbiana I may be. To my father, who showed me the beauty of nature from an early age, and taught me that if a job is worth doing, it is worth doing well. To my mother, for the countless hours she spent helping me grow into the person I am today. I cannot express my gratitude enough for her constant encouragement and endless love. To my wife Ji Youn, for her love and support during the long years of graduate school, especially when we were miles apart. I cannot imagine life without you and I look forward to many more happy years together (with less late nights spent in the lab). 저는 지연이랑 있으면 항상 행복합니다.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would to express my sincere thanks to my graduate advisor Dr. Sheryl Tsai for her constant support over the past years. I am grateful for the freedom that she allowed me when it came to choosing research problems to tackle. During my time spent in the lab under her guidance, I have learned to approach scientific problems confidently and independently. I could not have asked for a better Ph.D. advisor or lab atmosphere.

I would like to thank my thesis committee members Dr. Jennifer Prescher and Dr. Tom Poulos for taking the time to help me over the years and reading through this dissertation. I have been lucky enough to see Dr. Prescher progress through the early stages of her career, and her hard work and creativity are inspiring. I had a lot of fun being the Head TA for her Chem Bio Lab and it was a great learning experience. Dr. Poulos may not realize this, but graduate students take notice of his presence at work on the weekends. I have heard on more than one occasion people talking about how seeing Dr. Poulos working on the weekends is awesome and that we should work harder.

I would like to thank the current and past members of the Tsai lab. Dr. Pouya Javidpour and Dr. Ming Lee taught me the ways of the mycelium. Dr. Grace Caldara guided me during my rotation and continued to work with me on our di-domain ARO/CYC project. Dr. Beth Shillito taught me how to process X-ray diffraction data. Dr. Joel Bruegger and Dr. Chi Nguyen taught me the practical aspects of protein crosslinking and many different types of experiments. I would like to thank Jesus Barajas and Tim Valentic for their endless conversations on life and research projects while enjoying many pitchers at the pub. Dr. Gaurav Shakya and Alex White helped me understand many concepts in organic chemistry. I would like to thank Gabe Moreno for the many philosophical discussions on biosynthesis. Andrew Schaub taught me many cool things about computers. Now he has the honor of representing the great state of Ohio in the Tsai lab after I graduate. Jake Milligan took on a big project with me during his rotation and we solved the AcpP=FabB crystal structure. I am grateful for all his assistance and useful suggestions along the way. I am looking forward to seeing the project progress under his direction. Due to the countless efforts of all the Tsai lab members, I have learned more than I could ever have imagined. Many life-long friendships have been forged within the walls of Natural Sciences 1.

I am grateful to my previous undergraduate researchers and rotation students. Joaquin Ponce-Zepeda always knew where to put the protein elutions. Avinash Patel and MyChi Nguyen aided in mutagenesis of ARO/CYCs in Chapter 2. Avinash Patel also aided in solving DpsC and the BexE crystal structures, which is described in Chapters 3 and 4, respectively. A lot of time was spent in the cold room for the BexE work, stay cool Avi! Stephanie Tu optimized AuaEII crystals, which is described in Chapter 4. I would like to thank all my friends and colleagues at UCI for their friendship and support over the years. I would not have been able to complete this dissertation without the help of all of you.

I would like to thank Dr. Michael Burkart, Dr. Hung-Wen Liu, Dr. Rolf Muller, Dr. Christian Hertweck, Dr. Chaitan Khosla, Dr. Bradley Moore, Dr. John Crosby, and Dr. Jurgen Rohr for supplying me with interesting enzymes to study and useful experimental suggestions. I would

also like to thank the students and post docs in these labs that have supplied me with compounds and DNA over the years.

The most important acknowledgement I have to make is to my family. If it was not for their years of love and support I don't know where I would be right now.

The interconnectedness of all life is humbling and inspiring. The lab is everywhere. When experiments aren't going well or life's challenges seem overwhelming, a simple walk outside will reveal beauty and inspiration around every corner.

"In every walk with nature one receives far more than he seeks."
-John Muir

David R. Jackson

CURRICULUM VITAE

EDUCATION

University of California, Irvine, Irvine, CA   2009 - present
Ph.D. candidate, *Chemistry*

Case Western Reserve University, Cleveland, OH   2005 - 2009
Bachelor of Science, *Biochemistry*

RESEARCH EXPERIENCE

Graduate Student, University of California, Irvine, *Chemistry*   2009 - present
  PI - Dr. Sheryl Tsai
      *Structural and functional characterization of enzymes and enzyme
      complexes from polyketide and fatty acid biosynthetic pathways*

Graduate Student, University of California, Irvine, *Pharmacology*   Jan. 2009 - Mar. 2009
  PI - Dr. Naoto Hoshi
      *Studied voltage-gated ion channel function using mammalian cell
      culture, patch-clamp electrophysiology, and western blotting*

Undergraduate Researcher, Case Western Reserve University, *Biochemistry*   2007 - 2009
  PI - Dr. Vivien Yee
      *Studied the molecular basis of RNaseL activation using
      macromolecular X-ray crystallography*

TEACHING EXPERIENCE
*All teaching assistant positions were at the University of California, Irvine

Teaching Assistant for Biopharmaceutics and Nanomedicine Lab (fall)	2014
Teaching Assistant for Organic Chemistry Lab (spring)	2014
Head Teaching Assistant for Chemical Biology Lab (winter)	2014
Head Teaching Assistant for Chemical Biology Lab (winter)	2013
Teaching Assistant for Chemical Biology Lab (spring)	2011
Head Teaching Assistant for General Chemistry Discussion (winter)	2011
Teaching Assistant for Organic Chemistry Lab (fall)	2010
Teaching Assistant for Organic Chemistry Lab (spring)	2010
Teaching Assistant for Chemical Biology Lab (winter)	2010

INVITED PRESENTATIONS AND POSTERS

"Using Mechanism-based Crosslinking to Determine Type II FAS Complex Crystal Structures"
Synthetic Biology: Engineering, Evolution, and Design Annual Meeting (poster)
2014

"Catching Daunorubicin Biosynthesis in Action via Atom Replacement Strategy"
University of California, Irvine, **Vertex Day** (talk)
2013

"Catching Daunorubicin Biosynthesis in Action via Atom Replacement Strategy"
Scripps Institute of Oceanography, **Natural Products Affinity Group** (talk)
2013

"Defining the Structural Basis for Reduction, Cyclization, and Aromatization During BE-7585A Biosynthesis"
Scripps Institution of Oceanography, **Natural Products Affinity Group** (talk)
2012

"Defining the Structural Basis for Reduction, Cyclization, and Aromatization During BE-7585A Biosynthesis"
American Society of Pharmacognosy (ASP) Annual Meeting (poster)
2012

"Defining the Structural Basis for Reduction, Cyclization, and Aromatization During BE-7585A Biosynthesis"
University of California, Irvine, **Vertex Day** (talk)
2011

"Type II Polyketide Biosynthesis"
University of California, Irvine, **Pfizer Chemistry Symposium** (poster)
2010

"The Crystallization of RNaseL Bound to a Small Molecule Activator"
Case Western Reserve University, **Department of Biochemistry Retreat** (poster)
2009

HONORS AND ACHIEVEMENTS

Teaching Award for Contributions to the Chemistry Department Teaching Program
University of California, Irvine
2013

Dean's Honors List
Case Western Reserve University
2008

Dean's High Honors
Case Western Reserve University
2007

FELLOWSHIPS

Medicinal Chemistry and Pharmacology Fellowship
University of California, Irvine
2009 - 2010

PROFESSIONAL AFFILIATIONS

American Society for Microbiology
2014 - present

PUBLICATIONS

Pre-publication

Jackson, D. R.; Festin, G.; Aguilar, S.; Patel, A.; Nguyen, M.; Sasaki, E.; Valentic, T. R.; Barajas, J. F.; Vo, M.; Khanna, A.; Liu, H. W.; Tsai, S. C. Structure and Function of Two Di-domain ARO/CYCs From Type II Polyketide Synthases. *In preparation*.

Milligan, J.; **Jackson, D. R.**; Beld, J.; Burkart, M. D.; Tsai, S. C. Crystal Structure of the *E. coli* AcpP/FabB Complex. *In preparation.*

Jackson, D. R.; Mohammed, L.; Vasilakis, K.; Wattana-amorn, P.; Long, P. F.; Crump, M. P.; Crosby, J.; Tsai, S. C. Determining the Molecular Basis for Starter Unit Selection During Daunorubicin Biosynthesis. *In preparation.*

Jackson, D. R.; Yu, X.; Patel, A.; Barajas, J. F.; Metsa-Ketala, M.; Liu, H. W.; Rohr, J.; Tsai, S. C. Insights Into Complex Oxidation During BE-7585A Biosynthesis: Structural Determination and Analysis of the Polyketide Monooxygenase BexE. *In preparation*.

Jackson, D. R.; Tu, S. S.; Nguyen, M.; Barajas, J. F.; Krug, D.; Pistorious, D.; Müller, R.; Tsai, S. C. Structural Insights Into Anthranilate Priming During Type II Polyketide Biosynthesis. *In preparation.*

Barajas, J. F.; Phelan, R. M.; Schaub, A. J.; Kliewer, J. T.; Kelly, P. J.; **Jackson, D. R.**; Luo, R.; Keasling, J. D.; Tsai, S. C. Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for Production of Highly Reduced Hydrocarbons. *In preparation*.

Barajas, J. F.; Shakya, G.; Topper, C. L.; Moreno, G.; Vagstad, A.L.; **Jackson, D. R.**; Townsend, C. A.; Burkart, M.; Tsai, S. C. Structural Insights Into Chain Length Predisposition and First-ring Cyclization of Fungal Polyketide Intermediates in the Product Template Domain. *In preparation*.

Published

Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R. L.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; **Jackson, D. R.**; Ishikawa, F.; O'Dowd, B.; McCammon, J. A.; Opella, S. J.; Tsai, S. C.; Burkart, M. D. Trapping the Dynamic Acyl Carrier Protein in Fatty Acid Biosynthesis. *Nature* **2013**, *505*, 427-431

ABSTRACT OF THE DISSERTATION

Structural and Functional Studies of Enzymes Involved in Polyketide and Fatty Acid

Biosynthesis

By

David Richard Jackson

Doctor of Philosophy in Chemistry

University of California, Irvine, 2015

Professor Shiou-Chuan (Sheryl) Tsai, Chair

Polyketides and fatty acids (FAs) are naturally produced molecules, which are both biosynthesized from acyl-CoA building blocks and have similar enzymatic machinery. However, the structures and functions of polyketides and fatty acids are very different. Polyketides are secondary metabolites that have been used as pharmaceuticals and leads in drug discovery, whereas fatty acids are the components of cell membranes and have roles as signaling molecules. The goal of this dissertation is to gain a better understanding of how the enzymes in type II polyketide synthases (PKSs) and fatty acid synthases (FASs) generate their diverse products.

Type II PKSs generate reactive poly-β-ketone intermediates, which are regiospecifically cyclized and aromatized by aromatase/cyclase (ARO/CYCs) enzymes. In Chapter 2, the structural and functional studies of two functionally divergent di-domain ARO/CYCs are discussed. This work reveals the role of each domain in these di-domain enzymes, and identifies residues in the active site pockets that are important for catalysis.

Type II PKSs typically have an acetyl-starter unit; however, some polyketide natural products exist that contain non-acetate starter units. The incorporation of non-acetate starter unit

adds molecular diversity to the poly-β-ketone scaffold. DpsC is the enzyme responsible for the enzymatic incorporation of a propionyl starter unit during daunorubicin biosynthesis. Structural studies of DpsC that reveal insight into its mechanism and substrate specificity are discussed in Chapter 3. AuaEII is the CoA-ligase responsible for the generation of anthranoyl-CoA, which is used as a starter unit during the biosynthesis of a class of hybrid type II polyketides named the aurachins. Chapter 4 discusses the crystal structure of AuaEII and its implications for starter unit specificity of the aurachin biosynthesis pathway. An additional enzymatic strategy to generate structural diversity in type II polyketide natural products is polyketide oxidation. BE-7585A is a type II polyketide natural product that is proposed to catalyze an oxidative rearrangement during biosynthesis. The FAD/NADPH-dependent oxygenase BexE is proposed to catalyze this oxidation. The crystal structure and initial *in vitro* characterization of BexE are discussed in Chapter 5.

The acyl carrier protein AcpP plays a central role in type II FA biosynthesis in *E. coli*. AcpP carries the nascent FA chain and delivers intermediates into the active sites of over 12 other enzymes during FA biosynthesis. The protein-protein interactions between AcpP and its target enzymes are transient, and this has hampered efforts to gain information about how AcpP interacts with target enzymes. Chapter 6 describes the application of mechanism-based crosslinkers to generate stable complexes between AcpP and two different partner enzymes: the dehydratase FabA and the ketosynthase FabB. Comparative analysis of the AcpP=FabA and AcpP=FabB crystal structures reveals differential AcpP interactions with partner enzymes during FA biosynthesis. This work serves as a basis for engineering protein-protein interactions to modulate the affinity of AcpP for different partner enzymes and alter fatty acid profiles *in vivo* for biofuel production.

Chapter 1

Introduction

1.1 Natural Products: History, Medicine, Ecology, and Diversity

Natural products are a diverse class of molecules produced by living organisms and have been used for medicinal purposes for thousands of years. The Egyptians collected and catalogued plants with medicinal uses as early as 1550 BC. Undoubtedly, throughout human history, plants, bacteria, and fungi have played major roles in treating disease(1). In China, during the Eastern Han dynasty (25-220 AD), a collection of information regarding medicinal plants and animals was recorded and included as instructions for the use of plants such as *Ephedra sinica,* which contains the alkaloid ephedrine(1). Ephedrine use has spanned thousands of years into modern times, and it is currently used as a stimulant, decongestant, appetite suppressant, and recreational drug(2). Another example from traditional Chinese medicine is the natural product artemisinin, which is isolated from *Artemesia* plants. It is an FDA approved and widely used anti-malarial drug(3). Historically, humans did not know why drinking liquid concoctions such as *Ephedra sinica* tea would bring on a sense of euphoria, or why *Artemasia* extracts would ward off malaria. The ancient people just accepted the fact that certain plants had certain medicinal effects. The development of modern chemical and biological sciences has led to the discovery of specific, naturally produced small molecules as the true reasons behind the therapeutic effects of plants, bacteria, fungi, and animals.

In more recent years, natural products have revolutionized medicine and saved the lives of millions of people worldwide. In 1928, Alexander Flemming discovered that the fungus *Penicillium rubens* produces the antibiotic commonly known as penicillin, and this discovery
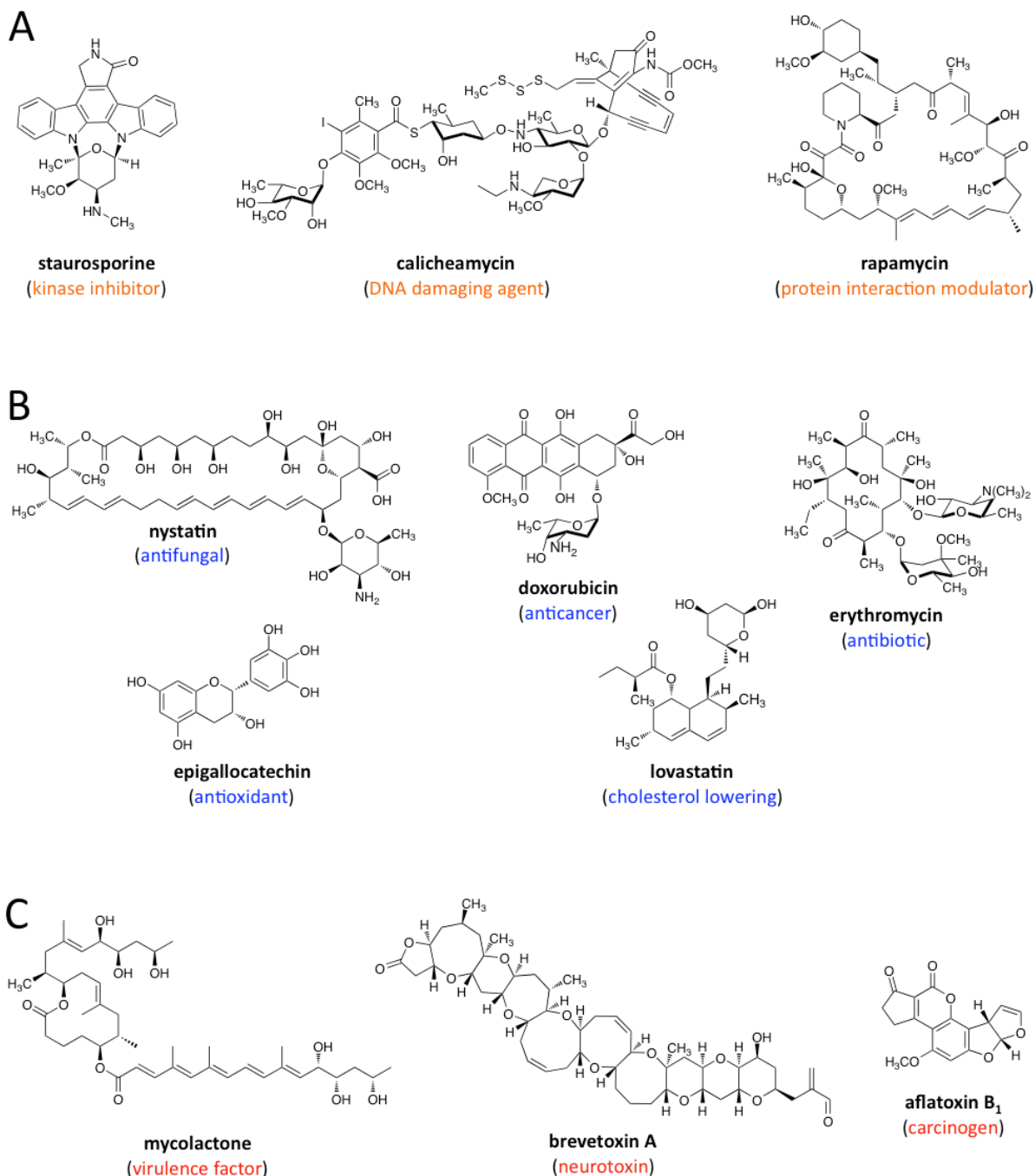
Figure 1-1. Structurally diverse and bioactive natural products. (A) Staurosporine, calicheamycin, and rapamycin are structurally complex natural products with unique bioactivity that have inspired both chemists and medical researchers for decades. (B) Polyketides with desirable medicinal properties and diverse chemical structures. (C) Polyketides that are detrimental to human health.

ushered in the era of antibiotics (from the 1940s – 1970s) that provided many of the natural

product-based antibiotics we use today such as chloramphenicol, tetracycline, erythromycin, vancomycin, and rifamycin(4, 5). Although many potent antibiotics have been isolated and identified from natural sources, antibiotic resistance is a growing problem in the world, and there is a need for the development of new molecules with augmented therapeutic potential. Natural product discovery has been accelerated recently by advances in DNA sequencing, DNA synthesis, and synthetic biology. Hopefully, these advances will lead to further discovery of novel therapeutics(6).

Antibiotics were not the only class of molecules that were isolated from natural sources in the later half of the 20th century. Natural products with diverse bioactivities such as anticancer (doxorubicin), immuno-modulating (rapamycin), and cholesterol-lowering (lovastatin) were also discovered from bacteria and fungi, each with unique modes of action (Figure 1-1)(7-9). Because these compounds are primarily used for therapeutic purposes, they represent an essential tool for the understanding of drug mechanism of action, as well as further development of new drugs. The alkaloid staurosporine from *Streptomyces staurosporeus,* for example, was found to be a kinase inhibitor that competes for the ATP binding site. Serendipitously, staurosporine was shown to be effective towards specific classes of kinases because of subtle structural variations of the ATP binding site, which inspired the development of a new class of selective kinase inhibitors(10). The novel chemistry and biology of staurosporine has paved the way for the development of selective kinase inhibitors, which was previously thought to be impossible(11). Another important class of medicinal natural products is the calicheamicins, a group of extremely cytotoxic enediyne polyketide natural products isolated from *Micromonospora echinospora* that cause DNA cleavage. The mechanism of DNA cleavage by calicheamicins is highly specific and mechanistically unprecedented. Calicheamicin DNA cleavage involves the

generation of a 1,4-didehydrobenze di-radical which reacts with and damages a DNA backbone(12). Knowledge of this elegant DNA cleavage chemistry would not exist if the calicheamycins had not been isolated from nature. The calicheamicins are currently used in antibody-drug conjugates for cancer treatment(13). Another extraordinary polyketide natural product is the immunosuppressant rapamycin, which is produced by *Streptomyces hygroscopicus* and has been implicated in lifespan elongation in mice(8). Rapamycin's mode of action is striking in that it simultaneously binds the surface of a cytosolic protein FKBP12 and also the mTOR complex, which ultimately results in the inhibition of the mTOR pathway(14). Rapamycin is the drug of choice for patients who undergo organ transplants and are at risk for rejection because of its immunosuppressant properties(15). Molecules such as staurosporine, the calicheamicins, and rapamycin are great examples of how natural products serve not only as useful therapeutics but also reveal novel therapeutic mechanisms of action that may be difficult to imagine otherwise. Currently, approximately 75% of FDA approved drugs are natural products or derived from natural product-based molecular scaffolds, which highlights the importance of these compounds in disease control and prevention(16).

A single gram of healthy soil can contain as many as 10 million microbes, which interact constantly by means of chemical communication(17). Ecologically, natural products play indispensible roles as signaling molecules and serve as a means for chemical defense, quorum sensing, and pathogenesis(18). Acyl-homoserine lactones, for example, allow bioluminescent bacteria to synchronize with each other and produce large flashes of light. Additionally, complex siderophores control the swarming behavior of the microbe *Vibrio alginolyticus*(19, 20). For millions of years, a symbiosis between Attine ants and their fungal food source has evolved, and natural products play a key role in their obligate mutualism(21). Attine ants have a special
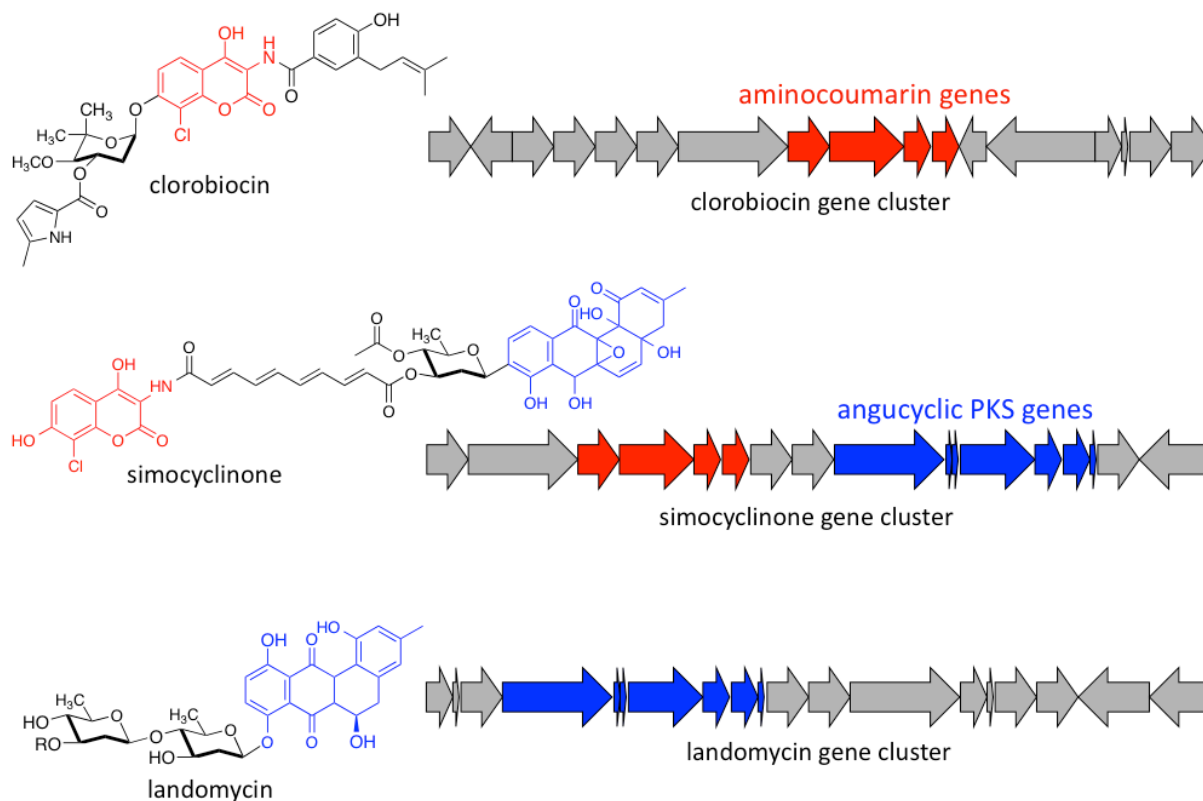
Figure 1-2. Shuffling of biosynthetic gene clusters leads to rearrangements and natural product diversification. Clorobiocin and simocyclinone share similar genes to produce an aminocoumarin moiety but are otherwise structurally unrelated. Simocyclinone and landomycin share a similar angucyclic polyketide moiety that is produced by similar genes in their respective gene clusters, yet their overall structures are quite different from simocyclinone, which contains a polyene-linked amincoumarin, while landomycin is linked to a polysaccharide (Figure adapted from reference 26).

compartment in their bodies that houses a type of *Actinomycete* bacteria belonging to the *Pseudonocardia* genus, which produces a class of antifungal compounds that are biosynthesized by a hybrid non-ribosomal peptide/polyketide synthase. These antifungal compounds selectively inhibit growth of a parasitic fungus that attacks the ant's fungal food source(22). Bacterially produced small molecules can also cause cellular differentiation and changes in morphology of eukaryote cells(23). Although researchers tend to focus on the medicinal properties of natural products, much less is known about the ecological function and evolution of natural products.

5

One advantage of using natural products for drug development is that we already know they contain bioactive scaffolds as a direct result of natural selection over millions of years. Unfortunately, in the vast majority of cases, we are unsure about what was selected for originally. Therefore, studying the enzymes and gene clusters involved in natural product biosynthesis provides a better understanding of how natural products evolve and are biosynthesized, which has a direct impact on therapeutic development.

All organisms contain metabolic pathways associated with primary metabolism, which are necessary for survival. Additionally, secondary metabolic pathways exist and produce bioactive natural products that confer an advantage to an organism but are not essential for survival (as discussed above)(24). Secondary metabolism uses chemical building blocks from primary metabolic pathways to generate structurally complex natural products. The bioactivity of natural products is a direct result of their complex molecular architecture, and the combinations of various enzymes inside living organisms generate that architecture. Typically, in bacteria and fungi, natural products are synthesized by sets of enzymes whose genes are located near each other and these collective gene groups are defined as biosynthetic gene clusters(25). Biosynthetic gene clusters can easily be transferred between organisms and can undergo recombinatorial events to yield new gene clusters and generate new products (Figure 1-2)(26). The evolutionary elasticity of biosynthetic gene clusters is readily apparent in natural products such as daunorubicin (anticancer), erythromycin (antibiotic), and nystatin (antifungal) that are all polyketides but have remarkably different structures and bioactivities. (Figure 1-1)(27, 28). Although these polyketides display diverse bioactivity, they contain similar gene clusters, which are most likely derived from a common ancestral gene cluster. The chemical complexity and resulting bioactivity present in secondary metabolism is amazing and inspiring.

Secondary metabolic pathways gain access to a wide variety of building blocks from primary metabolic pathways by using and modifying primary metabolites to evolve complex chemical structures. (Figure 1-3) For example, both acetyl-CoA and malonyl-CoA are both used as extender units in polyketide and fatty acid biosynthesis(29). Tryptophan,
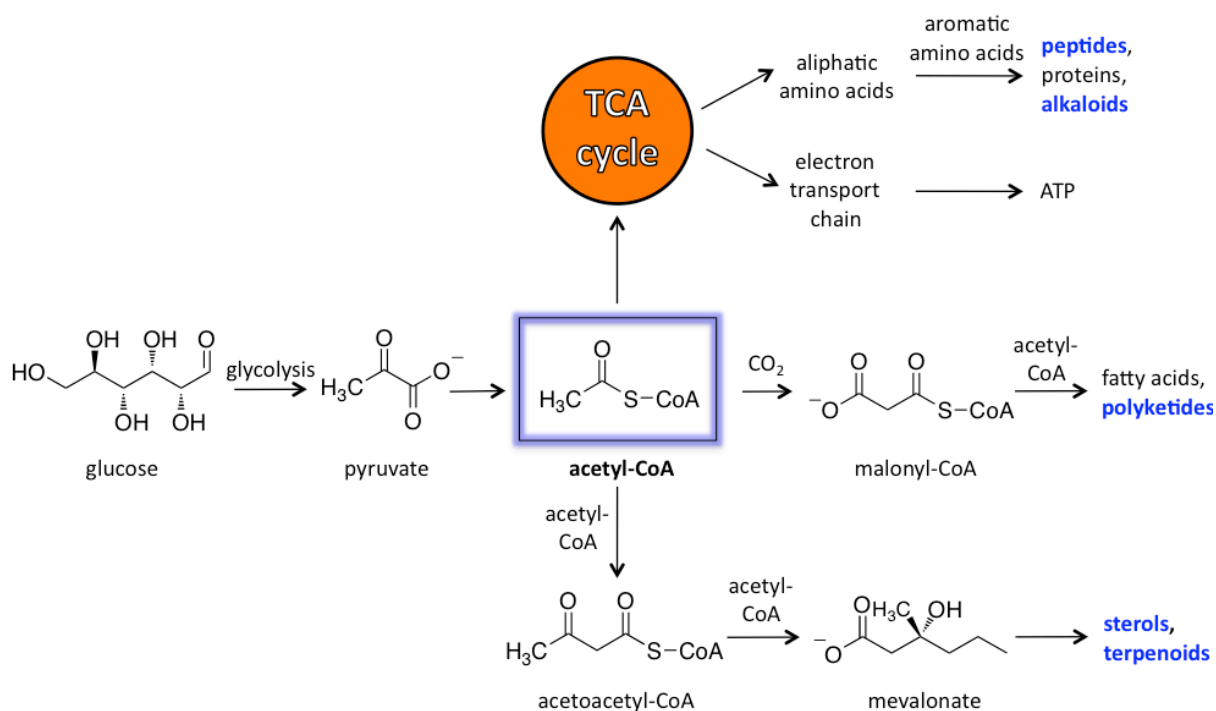
Figure 1-3. Acetyl-CoA is at the center of both primary and secondary metabolism. Glucose is converted to pyruvate during glycolysis then processed to yield acetyl-CoA. Acetyl-CoA can enter the TCA cycle and used to generate amino acids and ATP, which can be used to produce peptides, proteins, and alkaloids. Acetyl-CoA can be converted to malonyl-CoA and then used to generate fatty acids and polyketides. Acetyl-CoA can also be routed through the mevalonate pathway to produce sterols and terpenoids.

and other amino acid derivatives are commonly recruited for use in alkaloid biosynthesis(30). Various sugars such as glucose are commonly modified and appended to all types of natural products, and glycosylation has been found to be very important for the bioactivity of many natural product based drugs(31). There are also sets of biosynthetic enzymes that accept simple amino acids such as tryptophan and turn them into indole alkaloids with anticancer properties, or
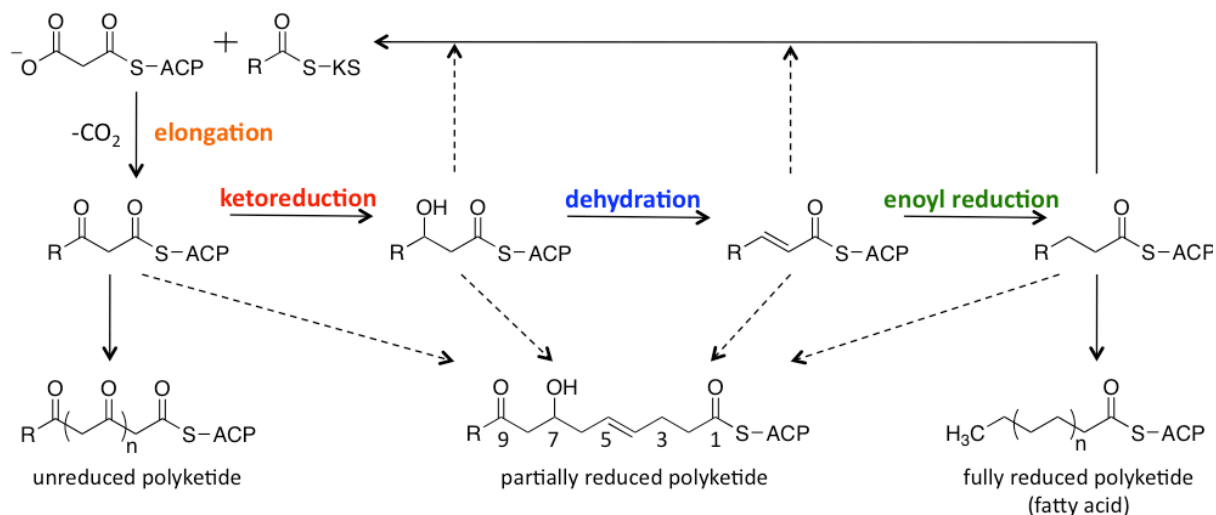
Figure 1-4. The process of elongation, reduction, and dehydration, which occurs (to differing extents) in polyketide and fatty acid biosyntheses. Malonyl-ACP is condensed with a primed ketosynthase and iterative condensations lead to an unreduced polyketide. When ketoreduction, dehydration, and enoyl reduction occur after each step of elongation, a fully reduced fatty acid is generated. In many polyketide biosynthetic pathways, varying degrees of reduction and dehydration occur that leads to partially reduced products. The example of a partially reduced polyketide is numbered (for reference) in the standard manner for polyketides and fatty acids.

giant polyketide megasynthases that accept acyl-CoA building blocks and turn them into antibiotics such as erythromycin. These enzymes are the links between primary and secondary metabolism. Understanding the structures and functions of enzymes involved in polyketide biosynthesis is a major aim of this dissertation.

Polyketides are typified by a poly-β-ketone substructure with varying degrees of reduction and are biosynthesized by a type I, type II, or type III polyketide synthase (PKS)(32). Type I PKSs are generally large multi-domain enzymes that produce polyketides in an assembly line fashion to generate compounds such as the antibiotic erythromycin. They are commonly found in bacteria and fungi(33). Type I PKSs produce polyketides with varying degrees of reduction, and often produce macrocyclic compounds. Type II PKSs are discrete enzymes that produce polycyclic aromatic compounds such as daunorubicin, and they are commonly found
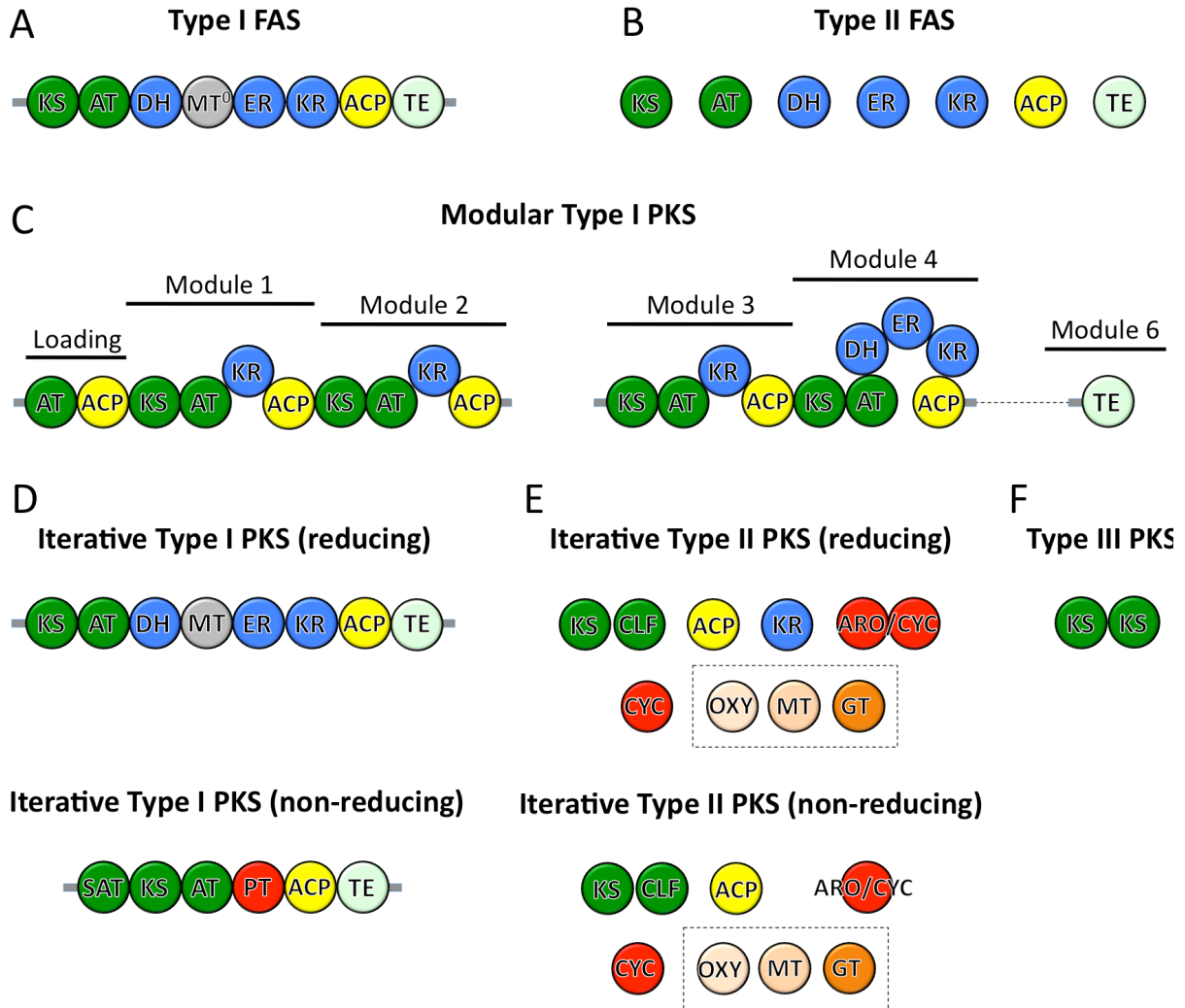
Figure 1-5. Domain architecture of PKSs and FASs. (A) Type I FASs are large multi-domain enzymes encoded on single polypeptide chains in eukaryotes. (B) Type II FASs are present in bacteria, chloroplasts, and mitochondria, and are composed of individual enzymes. Both type I and type II FASs act iteratively to produce fatty acids. (C) Modular type I PKSs consist of multiple large multi-domain enzymes and each domain typically modifies a growing substrate once in a linear assembly line-like manner. (D) Iterative type I PKSs can be reducing (contain reductive domains) or non-reducing (lack reductive domains) and iteratively elongate a growing polyketide. (E) Iterative type II PKSs can be reducing (contain a KR) or non-reducing (absence of KR) and iteratively elongate a polyketide. Iterative type II PKSs can also contain a variety of tailoring enzymes such as oxygenases (OXY) methyltransferases (MT), and glycosyltransferases (GT). (F) Type III PKSs are found in plants and do not contain an acyl carrier protein; ketosynthases (KSs) elongate a growing polyketide iteratively using malonyl-CoA.

9

soil dwelling in bacteria(34). Type III PKSs are typically found in plants and produce small aromatic polyketides derived from malonyl-CoA (35). Polyketides and fatty acids are synthesized in a very similar manner and share many of the same types of enzymes and domain architectures (Figures 1-4 and 1-5).

Although the motivation for this dissertation is primarily to understand natural product biosynthesis, a considerable portion of this dissertation will focus on fatty acid biosynthesis. I would like to make it clear to the reader that the term "natural product" is typically not used to describe a fatty acid, but I view fatty acids as specialized natural products for membrane formation and signaling. Therefore, it would be advantageous for the reader to consider this view while reading this dissertation in order to better understand the relationship between polyketides and fatty acids. In contemplating the biosynthesis of polyketides and fatty acids, it is useful to think of fatty acids as fully reduced polyketides, or conversely, polyketides as fully unreduced fatty acids (Figure 1-4). There is an evolutionary link between the two systems; therefore, using this type of mental framework is helpful in guiding one's thought process when trying to understand the structure and function of enzymes from both pathways. The following sections of the introduction will give a brief description of key enzymes involved in polyketide and fatty acid biosynthesis. Fatty acid synthases (FASs) were characterized before PKSs, and PKS enzymes have been named and classified based on related their FAS counterparts. Therefore, the organization, biochemistry, structure, and function of FASs will be discussed first, followed by PKSs. Grab a cold beer (or a drink of your choice) and enjoy!
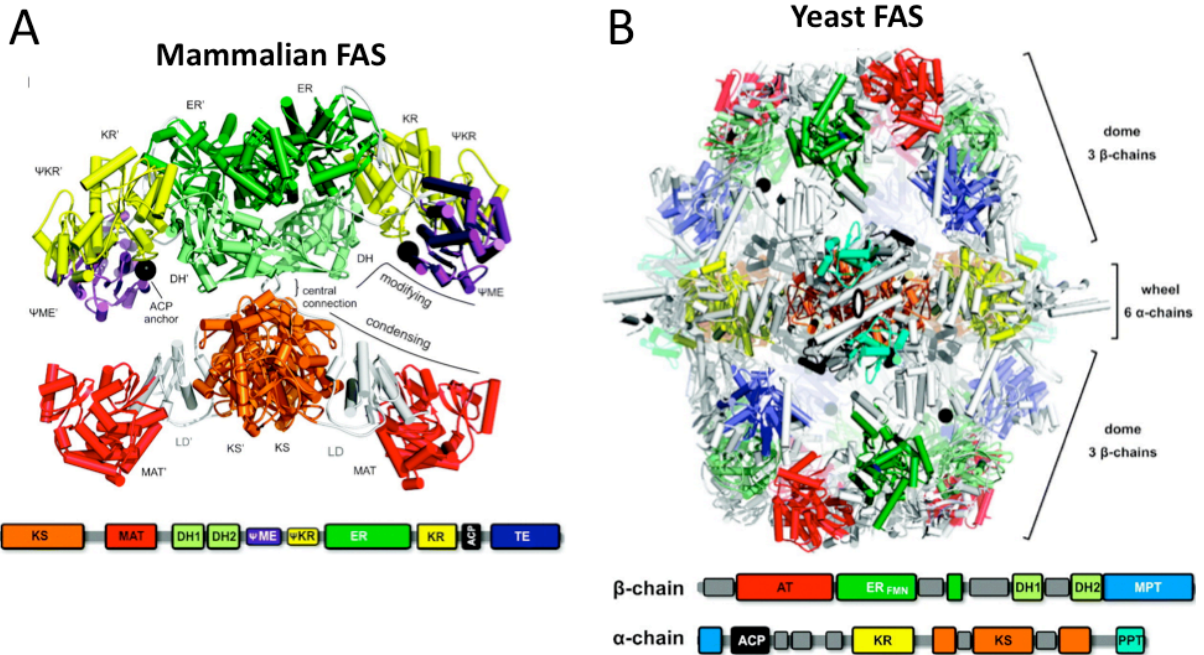
Figure 1-6. The structure and domain arrangement of the mammalian and yeast type I FAS. (A) The dimeric mammalian FAS crystal structure reveals the structure and orientation of domains (except the ACP and TE due to lack of electron density). (B) The hetero-dodecameric yeast FAS structure is composed of three α- and three β-chains that form a large central reaction chamber. (Figure adapted from reference 37)

1.2 Fatty Acids and Fatty Acid Biosynthesis

Fatty acids constitute the cell membranes of all living organisms, and fatty acid derivatives can act as potent signaling molecules (36). Most eukaryotes produce fatty acids (FAs) from acetyl- and malonyl-CoA using a large multi-domain enzyme called the type I fatty acid synthase (FAS)(37). The mammalian type I FAS consists of the 10 domains listed above, which are covalently linked on a single polypetide chain (Figure 1-6)(38). In contrast, bacteria produce fatty acids using a type II FAS, which consists of KS, AT, ACP, KR, DH, ER, and TE as discrete enzymes that need to form transient complexes for FA biosynthesis(39). The enzymes listed above for both type I and type II FAS perform very similar functions when compared to their PKS counterparts. In the following sections, the eukaryotic type I FASs will only be discussed
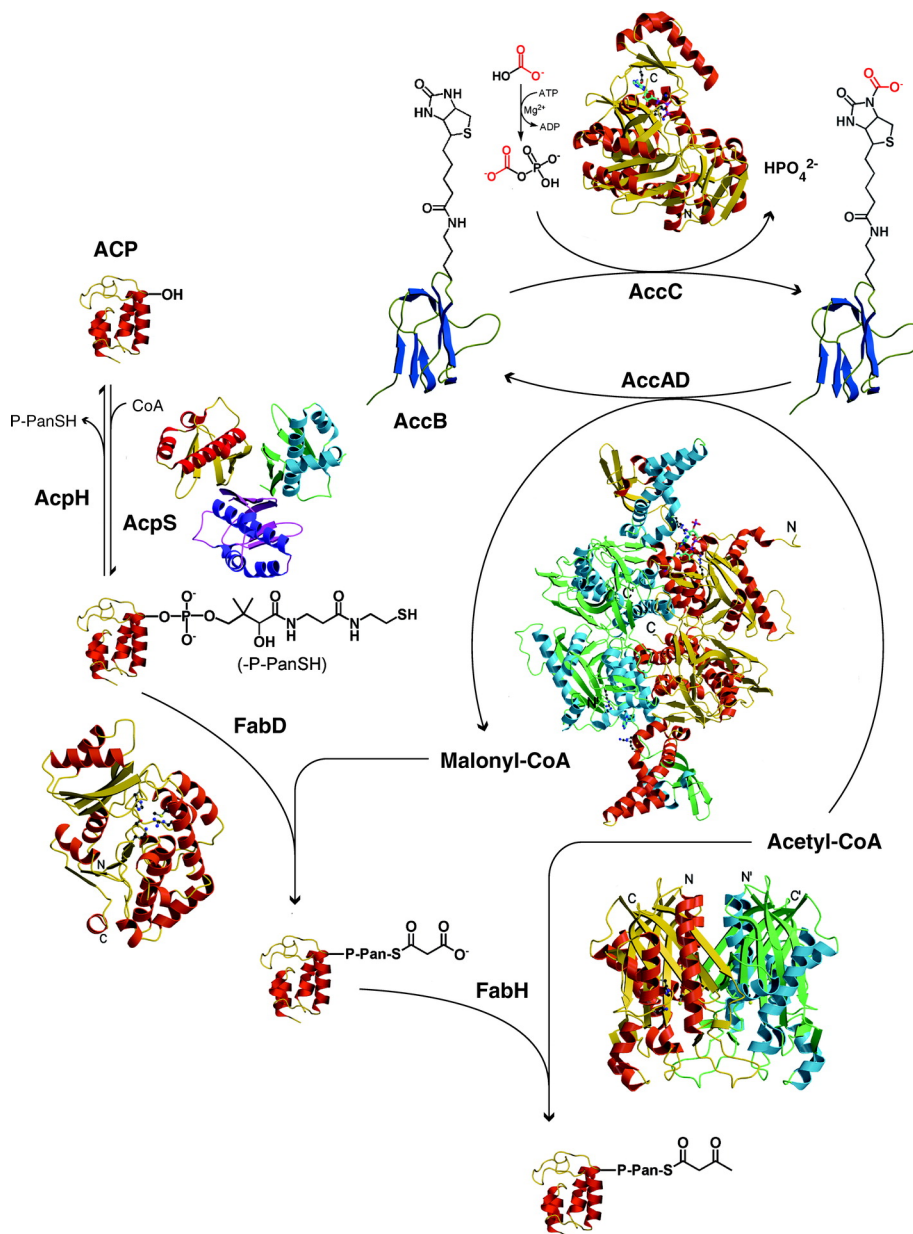
Figure 1-7. Type II FAS ACP activation and initial condensation reactions. The ACP is activated by posttranslational modification of Ser36 with phosphopantetheine (PPT) by the enzyme AcpS using CoA as a substrate to generate *holo*-ACP. The acyltransferase (AT) FabD, also known as malonyl-CoA:ACP transacylase, transfers a malonyl unit from malonyl-CoA onto the terminal thiol of the PPT on ACP to generate malonyl-ACP. The malonyl-CoA used in this process is generated by carboxylation of acetyl-CoA by acetyl-CoA carboxylase (ACC). Malonyl-ACP is then condensed with acetyl-CoA primed ketosynthase (KS) FabH, which is also known as 3-oxoacyl-ACP synthase, to form acetoacetyl-ACP that is elongated by other KSs. (Figure from reference 39)

12

briefly; bacterial type II FASs will be the main focus for an introduction to the structures and mechanisms of FAS enzymes.

1.2.1 Mammalian and Yeast Type I FAS

The crystal structures of the mammalian and yeast type I FASs have been solved (Figure 1-6)(38). The mammalian FAS consists of the domains listed above and is a homodimer in solution and in the crystal structure. In the mammalian FAS crystal structure, the ACP and TE could not be visualized due to high flexibility associated with the dynamic character of ACP. The yeast FAS consists of six α-chains (AT, ER, DH, and malonyl/palmitoyl transferase (MPT)) and a six β-chains (ACP, KR, KS, and phosphopantetheinyl transferase (PPT))(40). The heterododecameric yeast FAS forms a large cylindrical structure with an inner cavity, which acts as a reaction chamber. Collectively, these structures show that type I FASs are large multimeric enzymes which act in an iterative fashion to produce fatty acids and certain domains (such as the ACP and TE in the mammalian FAS) are highly mobile.

1.2.2 Bacterial Type II FAS

FASs are present in all bacteria, and the *E. coli* FAS has served as a model type II FAS for many years. The *E. coli* type II FAS consists of a priming ketosynthase (KS) FabH, two elongating KSs (FabB and FabF), malonyl-CoA:ACP transacylase (AT) (FabD), acyl carrier protein (ACP) (AcpP), ketoreductase (KR) (FabG), dehydratase (DH) (FabA), and enoyl reductase (ER) (FabI). All of the above domains interact during FA biosynthesis to elongate and modify growing fatty acids(39). *E. coli* produces a variety of saturated and unsaturated FAs, with the most common FA being palmitate (a 16-carbon fatty acid)(41). Due to the requirement of
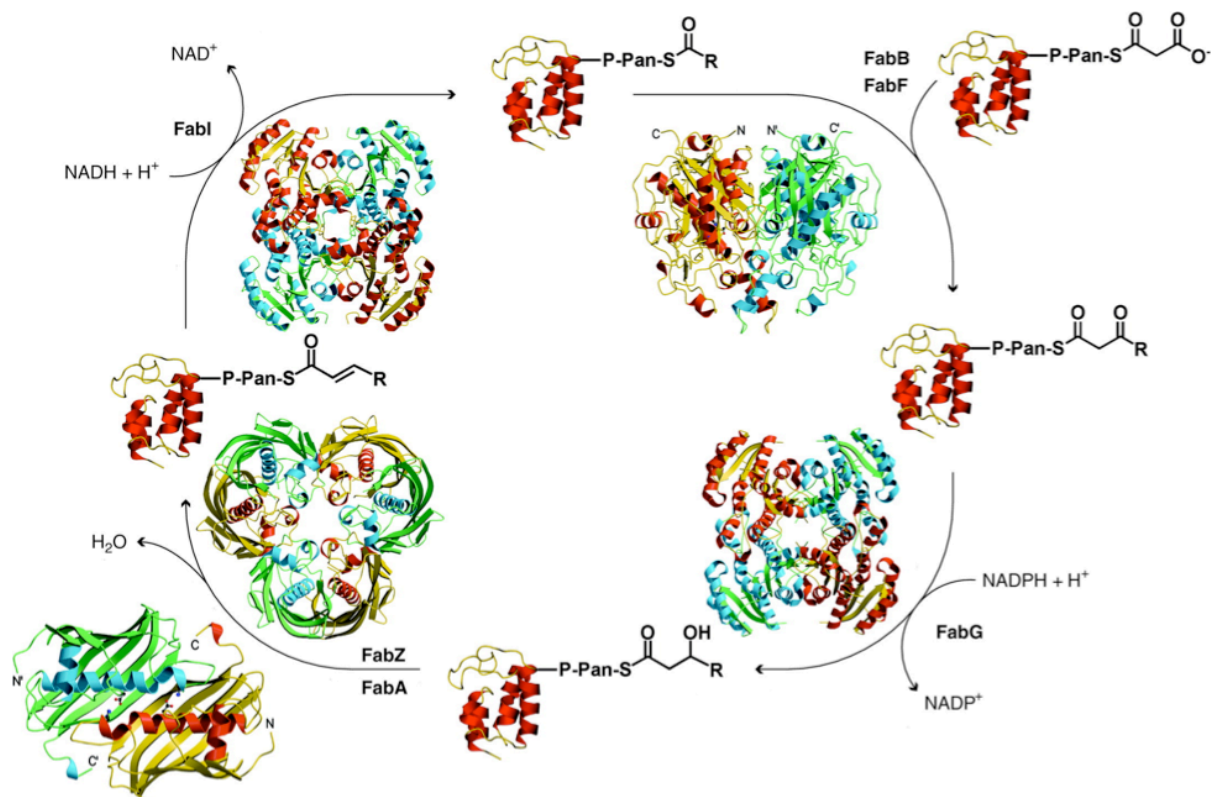
Figure 1-8. The elongation, reduction, and dehydration in bacterial type II FAS. Acyl-ACP is transferred to the active site of the ketosynthase (KS) FabB, also known as 3-oxoacyl-ACP synthase, and condensed with malonyl-ACP. The newly elongated substrate is transferred back onto PPT of ACP. The NADPH-dependent ketoreductase (KR) FabG, also known as 3-oxoacyl-ACP reductase, reduces the C3 carbonyl to an alcohol. The resulting alcohol is undergoes dehydration by the dehydratase (DH) FabA, also known as 3-hydroxyacyl-ACP dehydratase, to form and alkene between C2 and C3. The C2-C3 alkene is reduced by the NADH-dependent enoyl reductase (ER) FabI. This entire process can be repeated until a mature fatty acid is produced and then released from ACP. (Figure from reference 39)

FAs for cell viability, type II FAS enzymes are the targets of multiple antibiotics such as isoniazid (ER) and platensimycin (KS)(42, 43). Additionally, *E. coli* is the choice organism for the production of FA derived biofuels such as volatile medium-chain alkanes, alcohols, and fatty acid methyl esters (FAMEs)(44). Extensive research efforts have been aimed at understanding the biochemistry and structural biology of the *E. coli* type II FAS and these efforts have yielded the crystal structures of all enzymes in the pathway, as well as complete in vitro reconstitution and kinetic analysis(45). Despite this large body of work, very little is known about how type II

14
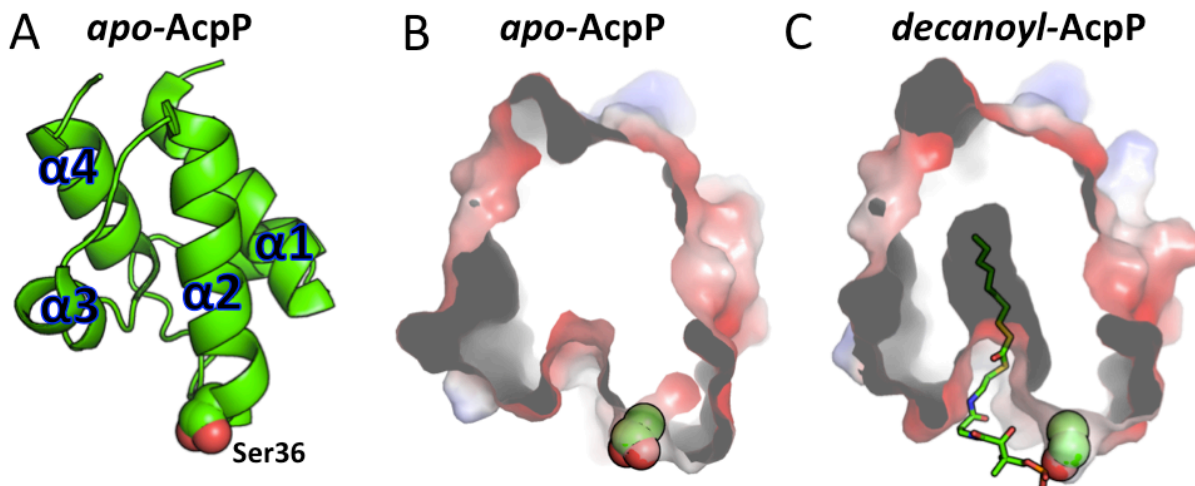
Figure 1-9. The structure and function of the *E. coli* type II FAS ACP (AcpP). (A) *Apo*-AcpP (PDB ID: 1T8K) is a four-helix bundle with a conserved serine residue (Ser36) at the end of helix 2. (B) Surface representation of *apo*-AcpP showing lack of a fully formed internal cavity. (C) Surface representation of decanoyl-AcpP (PDB ID: 2FAE) showing the formation of an internal cavity which sequesters a decanoyl fatty acid intermediate linked to Ser36 .

FAS enzymes interact *in vivo* and form complexes during FA biosynthesis. Below is a brief description of the structure and functions of *E. coli* type II FAS enzymes.

The ACP of the *E. coli* type II FAS is called AcpP, which carriers the growing fatty acid during elongation and delivers the fatty acid intermediates to the active site of each enzyme in the FAS pathway (Figures 1-6 and 1-7). AcpP is composed of four-helicies (helix 1, 2, 3, and 4) and has a conserved serine residue (Ser36) that is postranslationally modified with phosphopantetheine (PPT)(46). Before AcpP is phosphopantetheinylated, it is termed "*apo*-AcpP"; after phosphopantetheinylation it is termed "*holo*-AcpP". The terminal thiol of *holo*-AcpP is responsible for carrying the growing intermediates during FA biosynthesis (Figure 1-8). AcpP is a small (~10 kDa) and highly acidic (pI ~ 4) protein. AcpP is extremely important in *E. coli* and serves not only to shuttle intermediates in the FAS but also is involved in lipid A and biotin biosynthesis(47, 48). Multiple crystal structures of the *E. coli* AcpP have been solved,
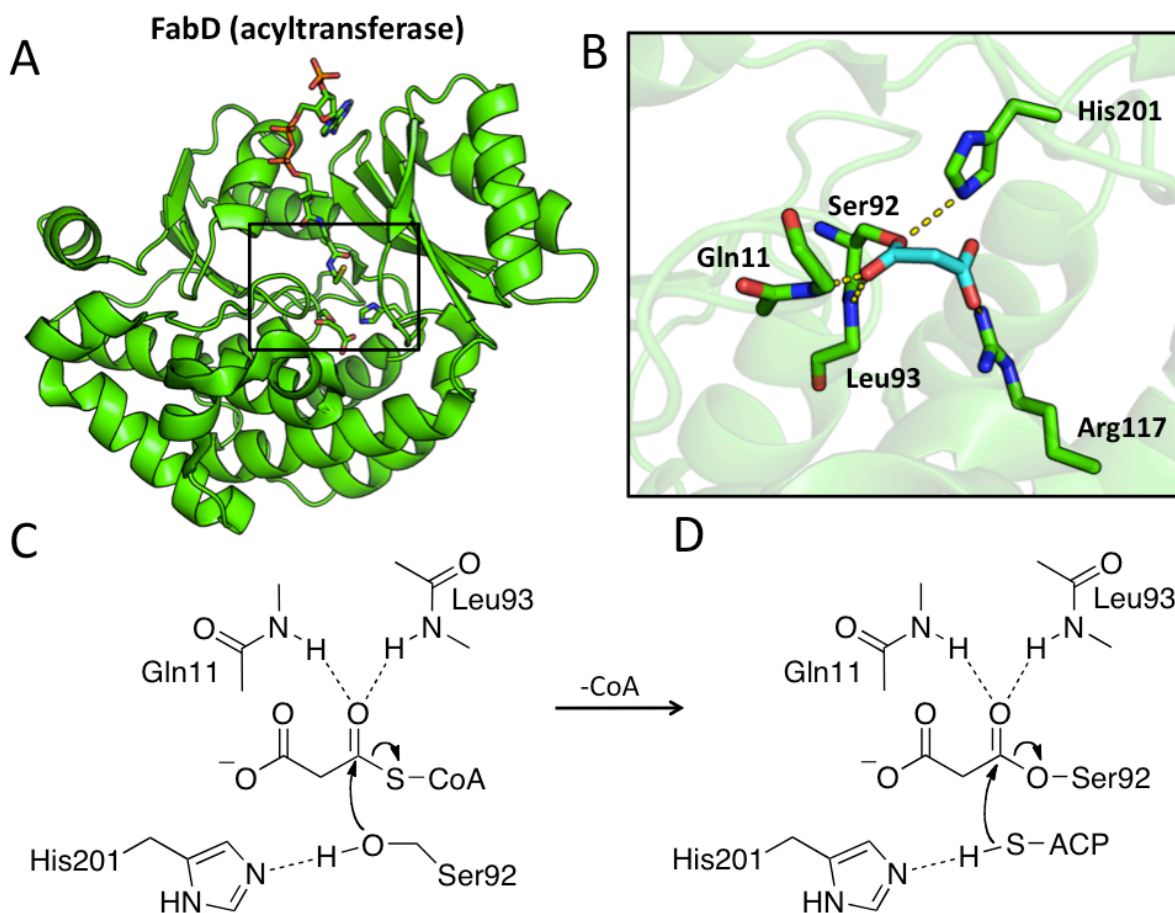
Figure 1-10. Structure and mechanism of the *E. coli* type II FAS acyltransferase FabD. (A) The crystal structure of FabD (PDB ID: 2G2Z) was solved after incubation with malonyl-CoA, showing the active site Ser92/His201 catalytic dyad with Ser92 acylated. (B) The malonyl-linked Ser92 ester carbonyl forms hydrogen bonds with the Gln11 and Ser92 amide hydrogens. The terminal carboxylate of the Ser92-linked malonyl moiety forms a salt bridge with Arg117. (C) Ser92 is activated by the neighboring His201 for nucleophilic attack onto the thioester bond of malonyl-CoA. The resulting negatively charged tetrahedral intermediate is stabilized by amide hydrogens of Gln11 and Ser92, which collapses to form the Ser92 acyl intermediate and releases CoA. (D) The PPT thiol of ACP enters the active site and reacts with the Ser92 linked acyl intermediate to transfer the malonyl group to the PPT thiol of ACP. This transfer involves the formation of a negatively charged tetrahedral intermediate, which is stabilized by amide hydrogens of Gln11 and Ser92.

including the apo- (PDB ID: 1T8K), decanoyl- (PDB ID: 2FAE), and multiple acyl-ACP forms

(Figure 1-9)(49, 50). Crystallographic and NMR structure studies have revealed that AcpP-

linked intermediates can be sequestered into an internal hydrophobic cavity, which may help

stabilize reactive intermediates. The sequestration of specific intermediates may also influence
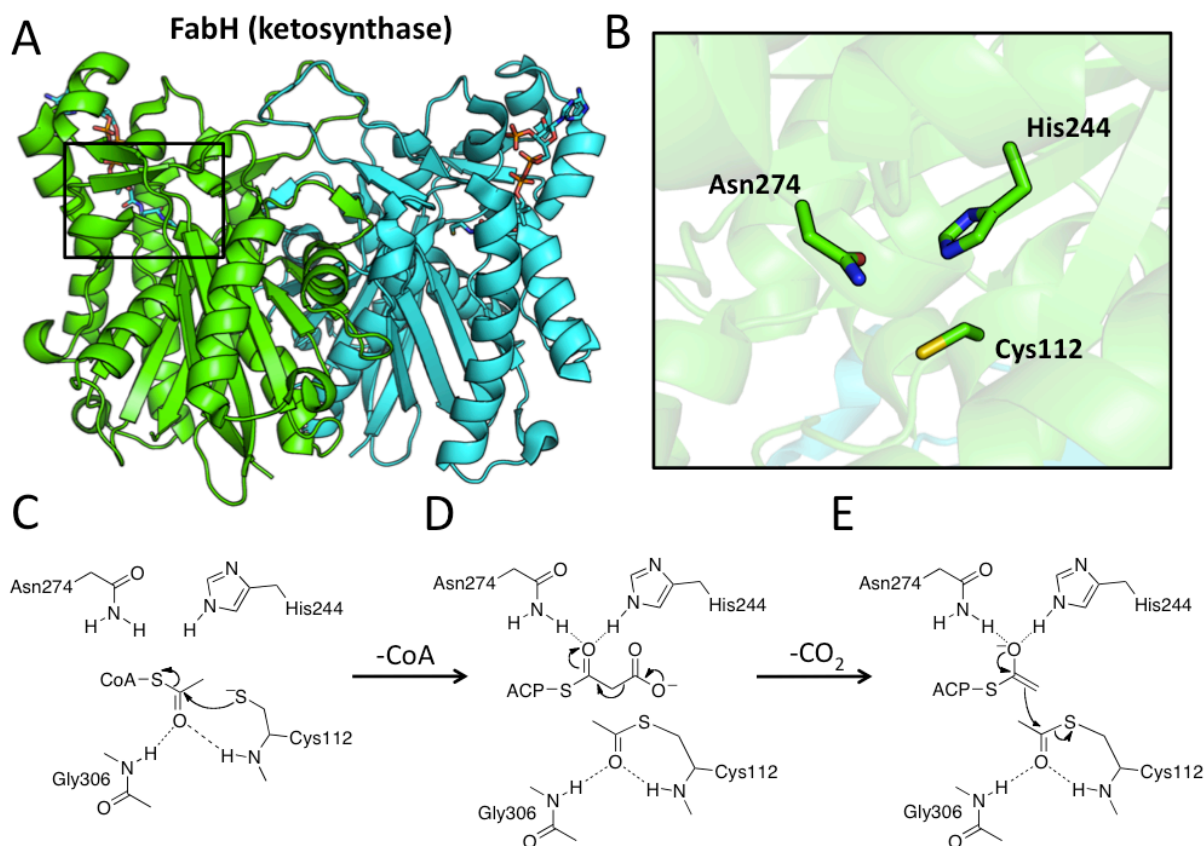
Figure 1-11. Structure and mechanism of the *E. coli* type II FAS priming ketosynthase FabH. (A) The crystal structure of FabH (PDB ID: 1EBL) was solved and shows that FabH forms a tight homodimer, with each monomer containing a functional active site. (B) The active site of FabH contains Cys112, Asn274, and His244 that are located at the end of the active site tunnel. Cys112 is located at the end of a long α-helix and the helix dipole is proposed to lower the pKa of the Cys112 thiol and increase its nucleophilicity. (C) Cys112 attacks the thioester bond of acetyl-CoA, forming a negatively charged tetrahedral intermediate, which is stabilized by the amide hydrogens of Cys112 and Gly306. The tetrahedral intermediate collapses to transfer the acetyl group onto Cys112 and release CoA. (D) The elongation occurs after Cys112 is primed by acetyl-CoA, and malonyl-ACP enters the active site. Asn274 and His244 facilitate the decarboxylation of malonyl-ACP by stabilizing the negative charge of the resulting enolate. (E) Once formed, the stabilized enolate undergoes condensation with acyl-Cys112. The resulting tetrahedral intermediate is stabilized by the amide hydrogens of Cys112 and Gly306, and collapses to release acetoacetyl-ACP.

AcpP structure and thus fine-tune the protein-protein interactions between AcpP and its FAS enzyme partners.

The elongation by FAS requires malonyl-CoA to be first transferred to the terminal thiol of PPT on AcpP to form malonyl-AcpP, which is catalyzed by FabD (Figure 1-10). FabD first binds malonyl-CoA and transfers the malonyl group onto the active site serine to form an acyl enzyme intermediate. AcpP then binds FabD and the malonyl group is transferred from the FabD active site serine to the terminal thiol on the PPT of AcpP. The primed malonyl-AcpP is then elongated by the appropriate KS (FabH, FabB, or FabF). The crystal structure of the *E. coli* FabD has been solved and revealed that the active site contains a Ser-His catalytic dyad (51). This type of active site is common in hydrolases and also other acyltransferases (ATs), but the hydrolases typically contain an active site triad where the histidine is further stabilized by a neighboring negatively charged residue (aspartic acid or glutamic acid)(52). The main difference between ATs like FabD and other members of the hydrolase family with similar active sites is the stability of the acyl enzyme intermediate. In the hydrolases, the acyl intermediate is rapidly hydrolyzed, which is facilitated by two amide hydrogens that form a strong oxyanion hole to stabilize the resulting tetrahedral intermediate, resulting in nuclephilic attack of water onto the acyl intermediate. In the case of FabD, the acyl intermediate is long-lived and is not readily hydrolyzed. Therefore, it was hypothesized that it contains a weaker oxyanion hole when the acyl intermediate is present(51). It is proposed that upon binding malonyl-CoA or AcpP, a small conformational change occurs to generate a stronger oxyanion hole and favor the transacylation. FabD is specific for the substrate malonyl-CoA, but other similar ATs exist, especially in the secondary metabolic pathways with relaxed substrate specificity that can load ACP with a variety of acyl-CoAs, leading to structurally diverse products(53).

The *E. coli* type II FAS has three enzymes that are responsible for elongating growing fatty acids: the priming KS FabH, and two elongating KSs, FabB and FabF. All three KSs are
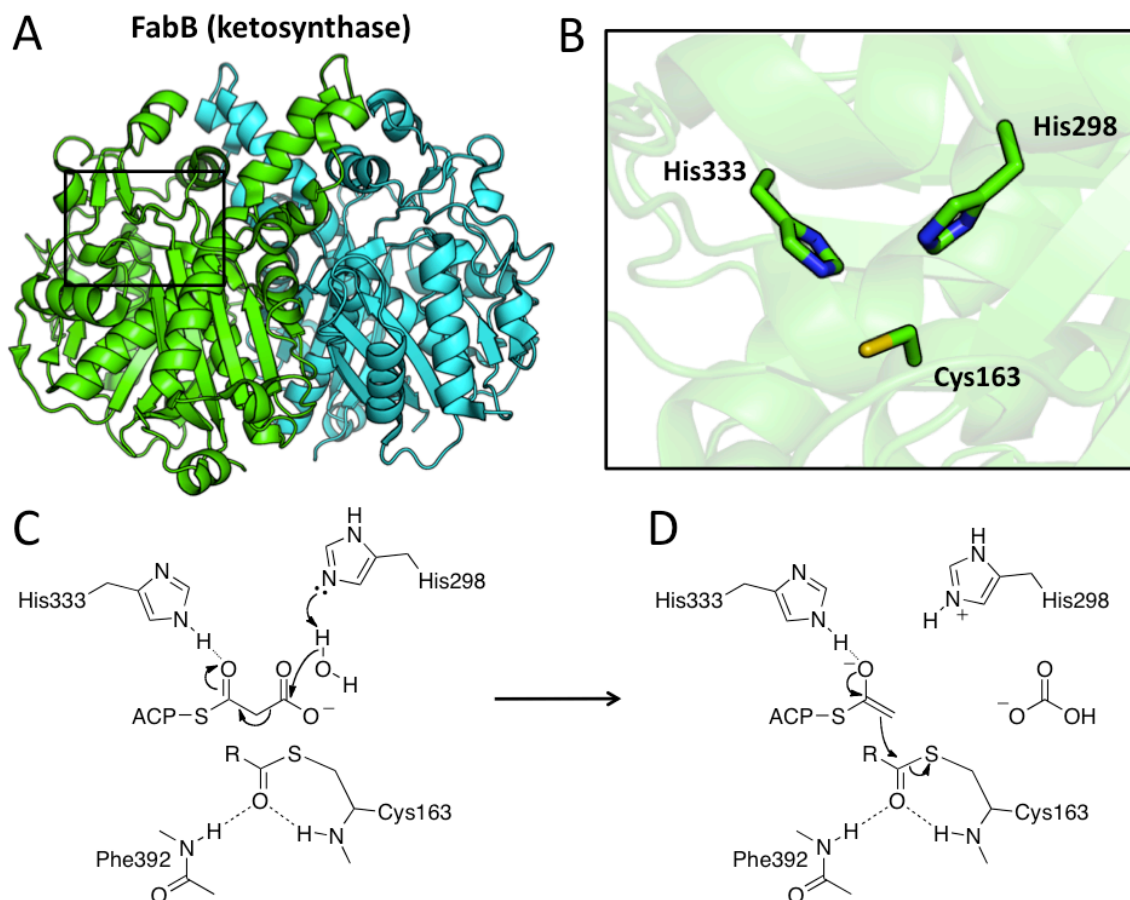
Figure 1-12. Structure and mechanism of the elongation ketosynthase FabB. (A) The crystal structure of FabB (PDB ID: 2VB9) was solved and showed that FabB forms a tight homodimer, with each monomer containing a functional active site. (B) The active site of FabB contains Cys163, His333, and His298, which are located at the end of the active site tunnel. Cys163 is located at the end of a long α-helix, and the helix dipole is proposed to lower the pKa of the Cys163 thiol and increase its nucleophilicity. (C) Cys163 is primed by an acyl-ACP, and the resulting acyl-Cys163 undergoes condensation with malonyl-ACP. His333 and His298 facilitate the decarboxylation of malonyl-ACP by stabilizing the negative charge on the resulting enolate. (D) Once formed, the stabilized enolate can undergo condensation with acyl-Cys163. The resulting tetrahedral intermediate is stabilized by the amide hydrogens of Cys163 and Phe392, and it collapses to release an acyl-ACP that is elongated by two carbons.

dimeric and have the same thiolase fold, but they have different active site residues. The crystal structures of the *E. coli* FabH, FabB, and FabF have all been determined(54-56). FabH is responsible for catalyzing the initiating step of fatty acid biosynthesis and is primed by acetyl-CoA, which is then condensed with malonyl-ACP to form acetoacetyl-ACP (Figure 1-11). FabB

and FabF catalyze the elongation of longer acyl-ACP intermediates and exhibit differences in substrate specificity. FabB is capable of elongating intermediates between the lengths of C4 to C14, whereas FabF can elongate C16 intermediates(41). FabB is also implicated in unsaturated FA biosynthesis and is responsible for elongating *cis*-3-decanoyl-ACP, a reaction that is not observed with FabF(57).

All three KSs contain a buried active site with a conserved cysteine at the end of a long pocket, which is accessed by the PPT of ACP. The conserved cysteine is located at the end of a long α-helix and the helix dipole is responsible for increasing the nucleophilicity of the active site cysteine(39). This highly conserved cysteine is responsible for accepting acyl-ACP intermediates and is the site of acyl-enzyme formation with an overall mechanism similar to that of FabD. The acyl intermediate undergoes condensation with an incoming malonyl-ACP. The active site of FabH facilitates the decarboxylation of an incoming malonyl-ACP by stabilizing the resulting enolate with an asparagine and histidine(55). The resulting enolate condenses with the cysteine-linked acyl intermediate and completes one round of elongation. In the case of both FabB and FabF, the residues that facilitate malonyl-ACP decarboxylation are two histidines (Figure 1-12). Similar to the FabD enzyme, FabH, FabB, and FabF contain an oxyanion hole formed by two amide hydrogens in order to stabilize the tetrahedral intermediate, which is formed when the enolate attacks the acyl-enzyme intermediate. Both FabB and FabF contain a long internal tunnel located near the dimer interface, which is where elongated fatty acid intermediates bind. This tunnel has been visualized in the crystal structures of both FabB and FabF, where the inhibitor cerulenin is covalently bound to the active site cysteine(58, 59). In summary, the *E. coli* type II FAS contains three condensation enzymes (FabH, FabB, and FabF),
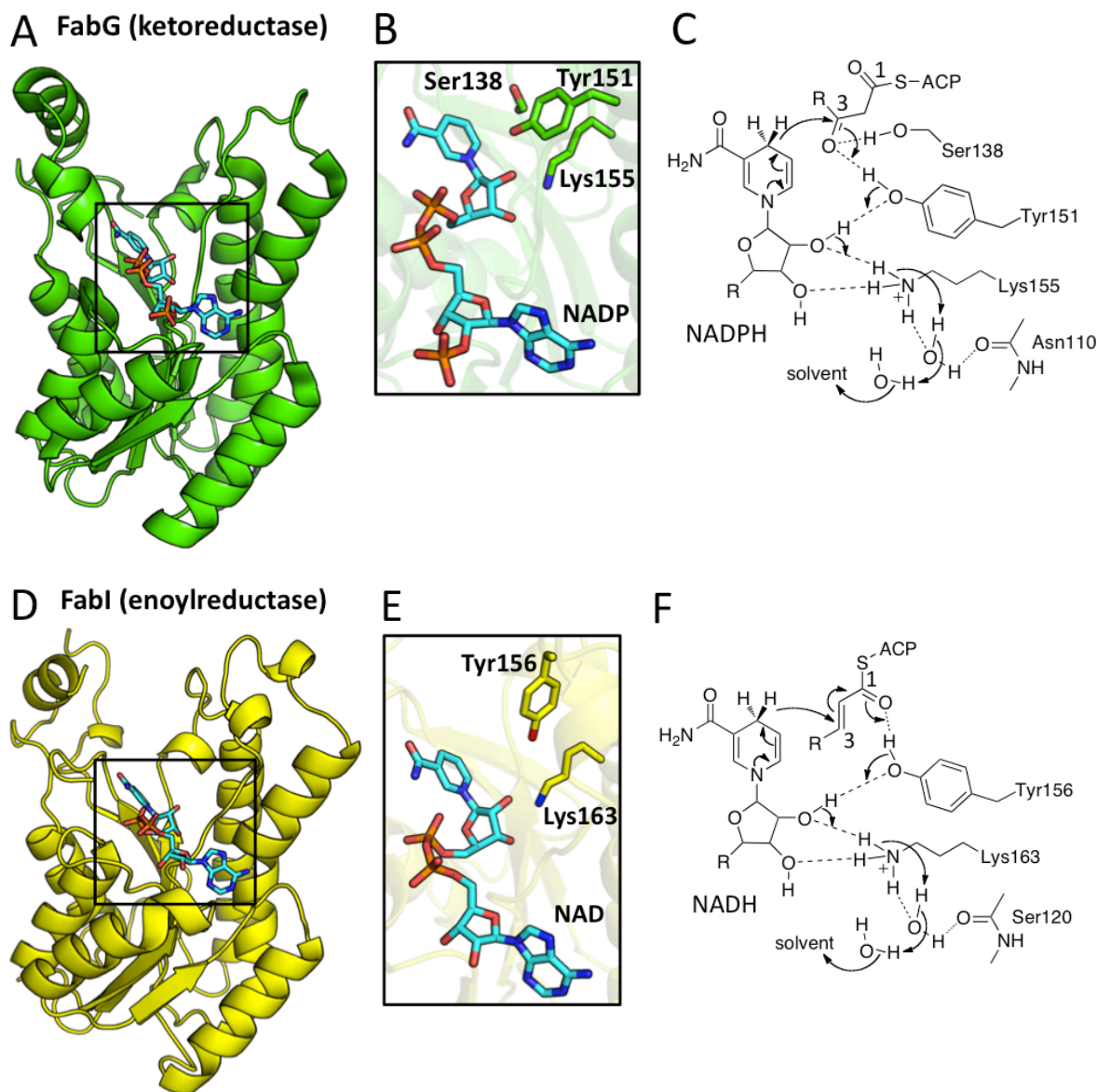
Figure 1-13. The crystal structures of the *E. coli* type II FAS dehydratase FabG and enoyl reductase FabI. (A) The crystal structure of FabG (PDB ID: 1Q7B) shows a classic Rossman fold and NADPH binding site. (B) The active site is comprised of Ser138, Tyr151, and Lys155 that aid in intermediate stabilization and proton transfer. (C) Acyl-ACP enters the FabG active site and positions the C3 carbonyl next to the NADPH nicotinamide ring. A hydrogen bond network is formed between Tyr151, the 2′ and 3′ hydroxyls of the NADPH ribose ring, Lys155, Asn110, and multiple waters connected to the solvent. Ser138 hydrogen bonds to the C3 carbonyl and stabilizes the negative charge formed upon hydride addition. (D) The crystal structure of FabI (PDB ID: 1DFI) shows a similar overall structure and fold to FabG but binds NADH instead of NADPH. (E) The FabI active site is composed of Tyr156 and Lys163. (F) FabI utilizes a similar hydrogen bond network to achieve intermediate stabilization and proton transfer after hydride addition.

which harbor the same overall fold but have different catalytic activities and substrate specificities during FA biosynthesis.

FabG is an NADPH dependent 3-ketoacyl-ACP reductase that reduces the C3 carbonyl of growing fatty acids intermediates to a C3 hydroxyl after each elongation cycle(60). FabG belongs to the short-chain dehydrogenase (SDR) superfamily and contains a Rossman fold that is responsible for binding the NADPH cofactor(61). Crystal structures of FabG with and without the NADP+ cofactor have been solved, and they reveal details of cofactor binding and residues involved in reduction and proton transfer (Figure 1-13)(60, 62).

Dehydration of the 3-hydroxyacyl-ACP intermediate formed by FabG is a key step in generating both saturated and unsaturated fatty acids. In the *E. coli* type II FAS, there are two dehydratases (FabA and FabZ), which differ in their substrate specificities and also catalytic capabilities. Both FabA and FabZ share the same hot dog fold, but FabA acts on shorter fatty acid intermediates (10 carbons or less), whereas FabZ acts on longer intermediates (more than 10 carbons). Further, FabZ can dehydrate long chain unsaturated intermediates(63). The FabA active site dyad is composed a histidine and aspartic acid. In contrast, the FabZ active site contains a histidine and glutamic acid(64, 65). FabA and FabZ both dehydrate 3-hydroxylacyl-ACP to form *trans*-2-enoyl-ACP, but only FabA can isomerize the double bond to form *cis*-3-decanoyl-ACP(63). The proposed dehydration mechanism of both FabA and FabZ employs aspartic acid (FabA) or glutamic acid (FabZ) for the protonatation of the C3 hydroxyl and histidine as the active site base to abstract a proton from C2, resulting in the formation of a C2-C3 double bond and elimination of water (Figure 1-14)(64).

FabI is an NADH dependent enoyl reductase that acts after FabA or FabZ in the FAS cycle to reduce the C2-C3 enoyl bond(66). Similar to FabG, FabI belongs to the SDR
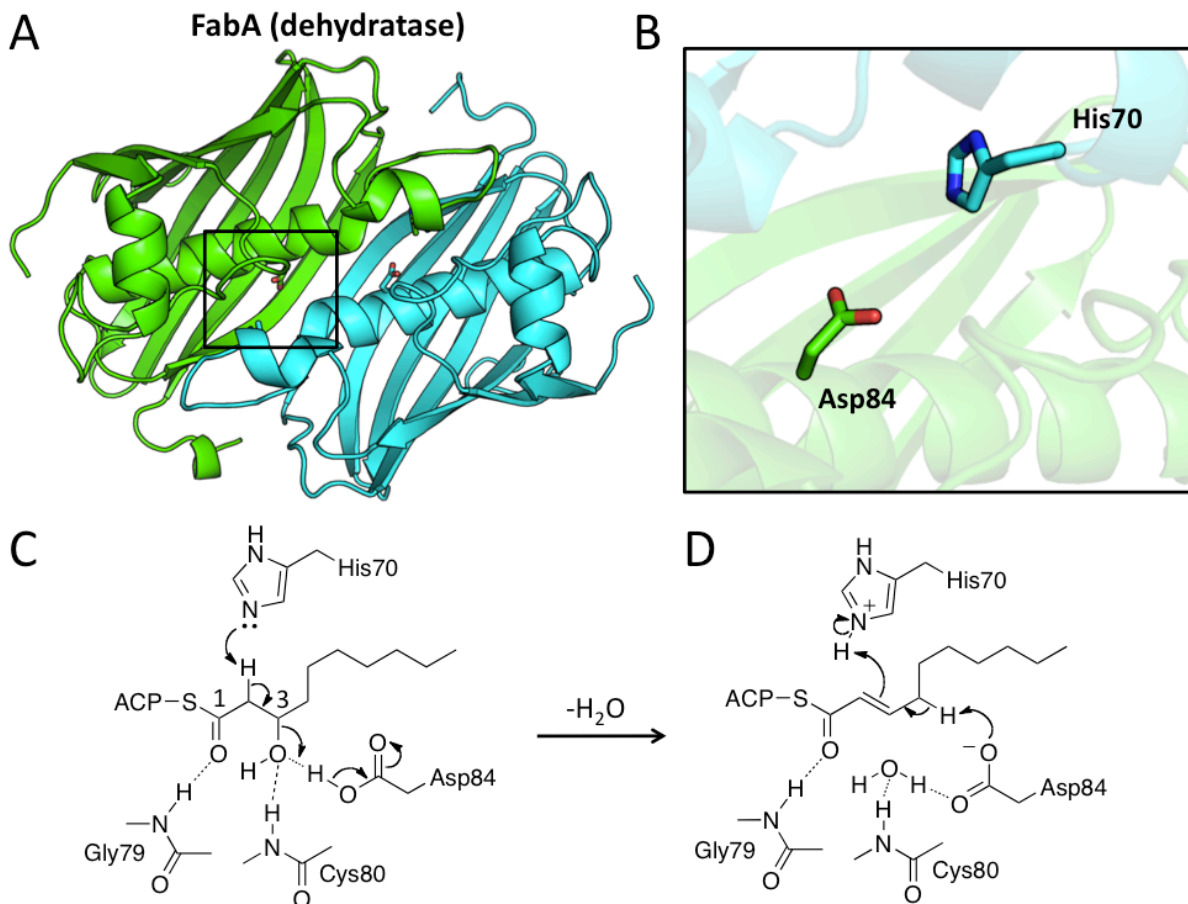
Figure 1-14 The structure and mechanism of the *E. coli* type II FAS dehydratase FabA. (A) The crystal of FabA (PDB ID: 1MKB) is a homodimer with two functional active sites at two monomer interfaces. (B) The FabA active site contains a His70/Asp84 catalytic dyad with each residue coming from a different monomer. (C) FabA catalyzes dehydration by removal of water across the C2-C3 bond. His70 acts as a base for deprotonation at the C2 position and Asp84 acts as an acid to protonate the C3 hydroxyl. The Gly79 amide hydrogen directs the substrate by forming a hydrogen bond with the C1 carbonyl and the Cys80 amide hydrogen helps coordinate the C3 hydroxyl and facilitates dehydration. (D) FabA can also catalyze isomerization of the C2-C3 double to form a *cis*-3-decanoyl-ACP intermediate. Asp84 is proposed to abstract a proton at the C4 position and His70 donates a proton at C2 to form *cis*-3-decanoyl-ACP. This reaction is a branch point between saturated and unsaturated fatty acid biosynthesis in *E. coli*.

superfamily and contains a Rossman fold for cofactor binding. Multiple crystal structures of FabI have been solved, including the ternary complexes with the FabI inhibitor triclosan(67, 68). FabI contains the same residues as FabG (except the active site serine) involved in cofactor binding,
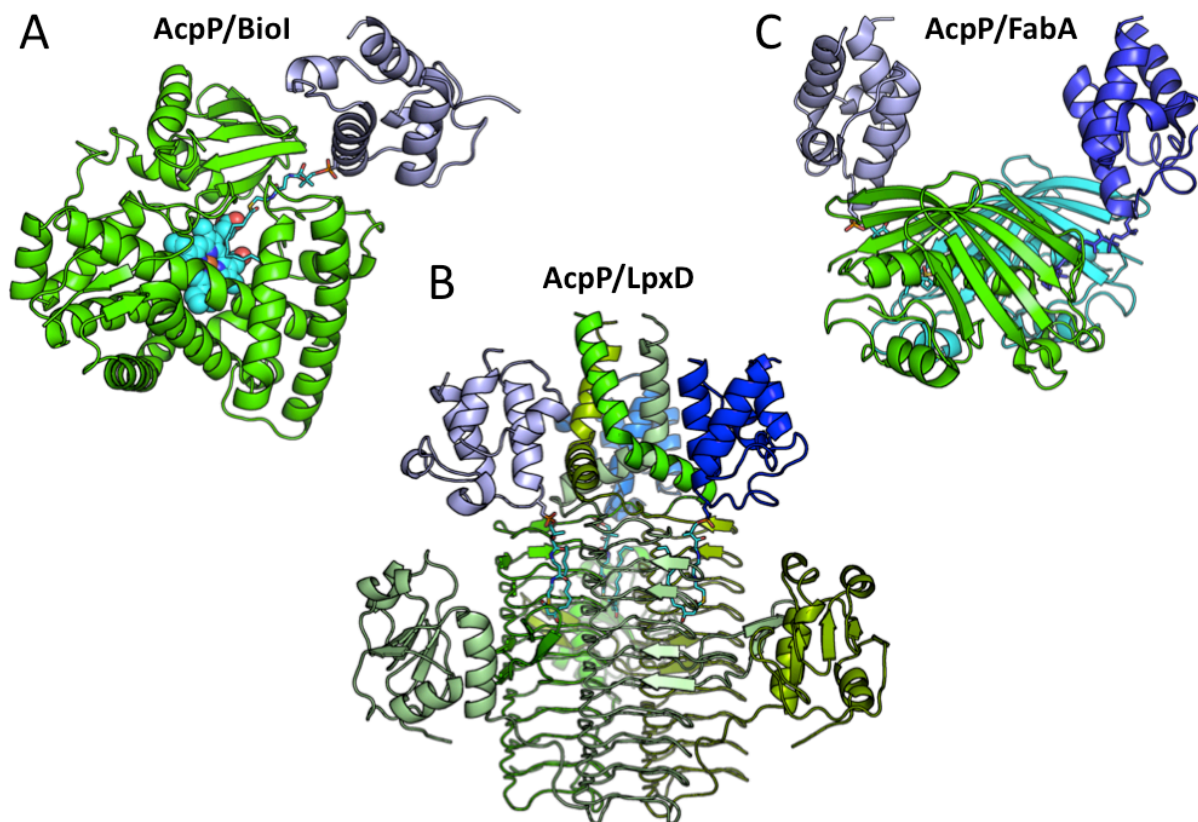
Figure 1-15. Crystal structures of AcpP-protein complexes from *E. coli* (AcpP is in shades of blue in all structures). (A) The crystal structure of AcpP with BioI, a cytochrome P450 enzyme (PDB ID: 3EJB). (B) The crystal structure of AcpP with LpxD, an enzyme involved in lipid A biosynthesis (PDB ID: 4IHG). (C) The crystal structure of AcpP with the dehydratase FabA (PDB ID: 4KEH).

reduction, and proton transfer is proposed to use the same mechanism as FabG, which is described above (Figure 1-13).

1.2.3 Protein-Protein Interactions in Type II FAS

In type II FAS (and type II PKS), the enzymes are stand-alone. They form transient protein complexes during the catalytic cycle. The ACP is at the center of catalysis and must carry the growing fatty intermediate to the active sites of more than 8 different enzymes(39). During this shuttling process, ACP sequesters the growing intermediate inside a hydrophobic internal

cavity located in the center of the four-helix bundle, which shields the intermediate from solvent but also may play a role in ACP-protein interactions. It is likely that depending on the chemical structure of the intermediate inside the ACP internal cavity, the ACP may change to a conformation that has a higher affinity for the next correct binding partner in the FAS pathway(69). Researchers studying the structural biology of type II FAS have tried for years to obtain high-resolution crystal structures of ACP in complex with different target enzymes, but only a small number of complexes from *E. coli* have been solved (Figure 1-15)(47, 70, 71).

Because of the differences between bacterial and human FAS structure and function, type II FAS enzymes have been targeted for antibacterial development. Current antibiotics include triclosan and isozanid (ER inhibitors), and multiple KS inhibitors have been identified for clinical investigation(72). By determining the structure of type II FAS/ACP complexes, we will provide a new opportunity for drug development.

Engineering microbes, especially *E. coli*, to produce energy dense biofuels has received much attention in the chemistry, engineering, biochemistry, and microbiology communities(73). *E. coli* has a fast growth rate, is genetically tractable, and can grow on simple substrates. These features have made this organism the go-to choice for metabolic engineering efforts aimed at biofuel production. The pathway of choice for biofuel production is the type II FAS pathway, and many different strategies have been applied to divert fatty acid production towards the production of fatty acid methyl esters (FAMEs), short and medium chain alkanes, short and medium chain alcohols, and other branched fatty acid derived molecules(74). One main strategy is changing the flux of the FAS to ensure longer residence time of shorter intermediates on ACP, which can be cleaved by TE and diverted to produce short chain biofuels(75). This process can tune the mRNA levels of individual FAS enzymes or introduce enzymes from other organisms

that interact with ACP. In both cases (and other strategies not discussed here), ACP-protein interactions are the key to the bioengineering effort. Without structural information of ACP/protein complexes, metabolic engineering efforts are lacking a key piece of information. This explains many failed results. Much time and resources have been spent on metabolic engineering of the *E. coli* type II FAS, yet we know very little about how the enzymes interact. To answer this question, we have used mechanism-based crosslinking to stabilize protein complexes coupled with X-ray crystallography to determine their structures.

The Burkurt group at UCSD generated mechanism-based crosslinkers that can be loaded onto ACP and form covalent complexes with target enzymes(76, 77). All biosynthetic intermediates are tethered to the ACP in both FASs and PKSs. As a consequence, the ACP directly interacts with all downstream target domains. The Burkart lab has developed a series of probes that can be chemoenzymatically loaded onto ACP that specifically targets the active site residues of KSs, DHs, and TEs. Our lab has made use of the chloroacryl-pantetheine amine probe for crosslinking ACP to KS, and the alkynyl sufonyl probes for crosslinking ACP to DH. These topics will be further expanded in Chapter 6.

1.3 Type I PKS

Type I PKSs can be divided into two main groups: modular, and iterative (Figure 1-5)(78). Iterative type I PKSs can further be divided into highly reducing and non-reducing PKSs. Modular type I PKSs are commonly found in both bacteria and fungi. Modular type I PKSs consist of multiple modules that each contain catalytic domains and produce polyketides with varying degrees of reduction. Each domain in each module of a type I PKS typically carries out one enzymatic function, and type I polyketides are often cyclized to yield macrolactone rings. A

fully reducing module would contain the additional DH and ER domains necessary to achieve a fully saturated carbon-carbon bond. Type I PKS priming is initiated once the AT loads CoA-derived starter units onto the ACP that can be derived from malonyl-CoA, methylmalonyl-CoA, or other acyl-CoAs. The elongation, dehydration, and enoyl reduction of type I PKSs will not be discussed in detail, but these reactions are carried out in a similar fashion as described above for FASs(79, 80). The modular type I PKS has become a prominent method for microbes to produce a wide variety of bioactive polyketides differing in both length and level of reduction. In the last module of a type I PKS, the thioesterase (TE) is responsible for releasing the product by hydrolysis or macrolactonization(81).

The iterative type I PKSs are predominantly found in fungi and produce a wide range of bioactive natural products. Iterative type I PKSs differ from modular systems, because they typically consist of only one large multi-domain module that acts in a repetitive fashion to elongate and modify a growing polyketide intermediate(82).

Non-reducing iterative type I PKSs typically contain SAT, KS, AT, ACP, the product template (PT), and TE/CYC domains. The SAT domain is responsible for loading the starter unit onto the ACP, which is often longer than the typical 2-carbon starter units derived from either acetyl- or malonyl-CoA(83). The PT domain is responsible for regioselectively cyclizing the growing non-reduced polyketide(84).

Highly reducing iterative type I PKSs, like non-reducing iterative type I PKSs, contain one large multi-domain module but include KR, DH, and ER domains that yield highly reduced polyketides and are often cyclized by the TE to form macrolactones(32). Although many type I PKSs fit well into these distinct categories, the rise in metagenomic DNA data has revealed
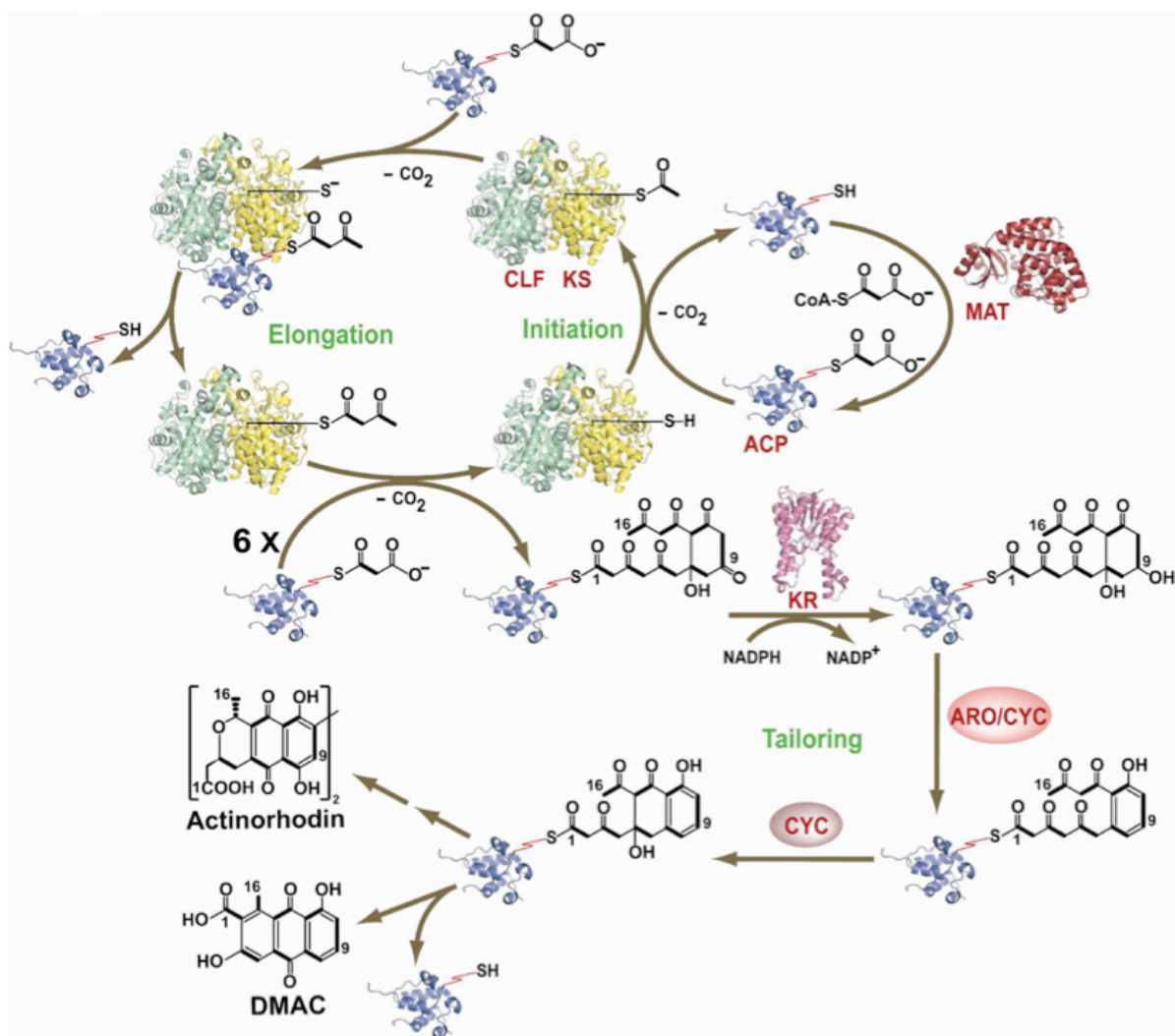
Figure 1-16. The priming, elongation, reduction, and cyclization reactions in a type II PKS using the actinorhodin biosynthetic pathway as a model. Malonyl-CoA:ACP transacylase (MAT, or AT) loads *holo*-ACP with a malonyl unit derived from malonyl-CoA. Many type II PKS pathways do not have a dedicated AT. Therefore, the AT is borrowed from the primary metabolic FAS pathway. The ketosynthase/chainlength factor (KS/CLF) is a functional heterodimer and catalyzes iterative elongation of acyl-ACP with malonyl-ACP. Once the growing poly-β-ketone is elongated to an octaketide intermediate, KR stereospecifically reduces the C9 carbonyl group to an alcohol and cyclizes the intermediate between C7 and C12. The C9-reduced, C7-C12 cyclized intermediate is then aromatized by an aromatase/cyclase (ARO/CYC) and further cyclized by a cyclase (CYC). Additional processing steps release the cyclized intermediate from ACP and yield actinorhodin. (Figure from reference 87)

many uncharacterized type I PKS gene clusters that blur the lines between the PKS classes listed above(85).

1.4 Type II PKS

Type II PKSs are commonly found in bacteria, especially members of the genus *Streptomyces*(34). Type II PKSs produce bioactive polycyclic aromatic natural products such as daunorubicin (anticancer) and tetracycline (antibiotic)(34). From a bioinformatic standpoint, type II PKS gene clusters are very easy to identify because of highly conserved genes such as the ketosynthase (KS) and chain length factor (CLF) genes. It is even possible to identify type II PKS gene clusters from environmental DNA using degenerate primers targeted at these highly conserved regions(86). Type II PKSs contain a minimal set of enzymes responsible for elongating a poly-β-ketone known as the "minimal PKS", which consists of the KS, CLF, ACP, and MAT(87). The minimal PKS catalyzes decarboxylative Claisen condensations of malonyl-CoA to synthesize a poly-β-ketone that then can undergo a series of enzymatic reactions to ultimately yield a bioactive natural product (Figure 1-16)(87). The KS and CLF form a functional heterodimer in solution, and the KS/CLF complex has an internal cavity that determines the length of the growing poly-β-ketone(88). Once the growing poly-β-ketone intermediate has reached the full length, regiospecific reduction by KR, and/or regiosepecific cyclization by an aromatase/cyclase (ARO/CYC) occurs to produce a cyclic intermediate(89, 90). This intermediate is then further cyclized by one or more cyclases (CYCs) and is spontaneously released from the ACP(91). The cyclized intermediate can undergo a series of other reactions not limited to, but including glycosylation, O-methylation, oxidation, reduction,
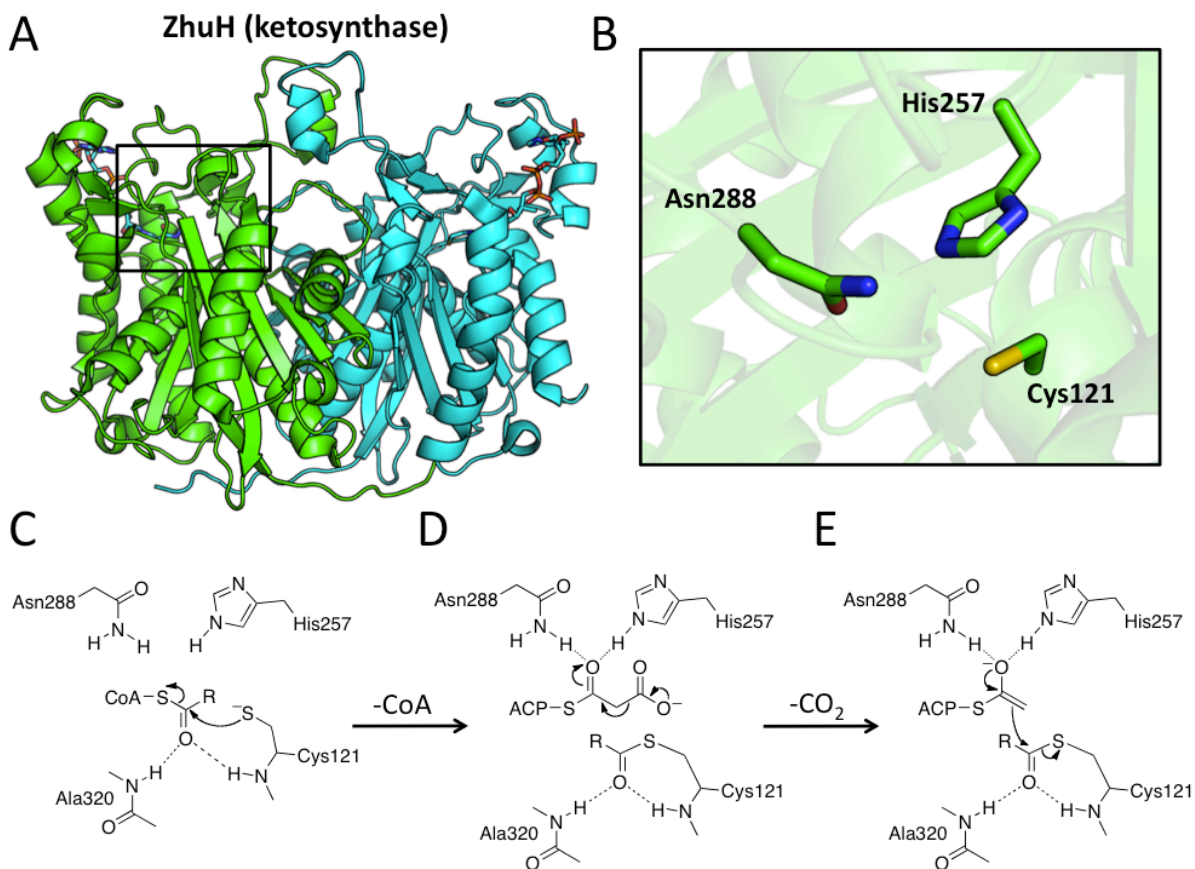
Figure 1-17. The structure and mechanism of the *Streptoymces R1128* type II PKS priming ketosynthase ZhuH. (A) The crystal structure of ZhuH (PDB ID: 1MZJ) shows that ZhuH forms a tight homodimer with a functional active site in each monomer. (B) The active site of ZhuH contains Cys121, Asn288, and His257 that are located at the end of the active site tunnel. Cys121 is located at the end of a long α-helix, and this is proposed to lower the pKa of the Cys121 thiol and increase its nucleophilicity. (C) Cys121 attacks the thioester bond of an acyl-CoA, forming a negatively charged tetrahedral intermediate, which is stabilized by the amide hydrogens of Cys121 and Ala320. The tetrahedral intermediate collapses to transfer the acyl group onto Cys121 and release CoA. (D) The elongation occurs after Cys121 is primed by an acyl-CoA and malonyl-ACP enters the active site. Asn288 and His257 facilitate the decarboxylation of malonyl-ACP by stabilizing the negative charge on the resulting enolate. (E) Once formed, the stabilized enolate undergoes condensation with acyl-Cys121. The resulting tetrahedral intermediate is stabilized by the amide hydrogens of Cys121 and Gly306 and collapses to release the elongated acyl-ACP.

or other ligation-type reactions(92). These final steps of biosynthesis are known as "tailoring" reactions and are carried out by a wide range of enzymes.

The biosynthetic logic of type II PKSs has evolved to facilitate ease of diversification through natural selection. The minimal PKS is capable of generating a highly reactive polyketide intermediate whose fate is decided by a variety of possible tailoring steps, which can then be slightly modified to yield new products over successive evolving generations of the producer microbe. Mutation of biosynthetic pathways is nature's way of conducting a large *in vivo* selection for bioactive compounds. It is quite common for one biosynthetic pathway to generate a suite of related compounds. If the particular set of compounds is toxic and acts as defense mechanism for the producer, generating multiple related compounds that all have the same biological target makes it harder for the victim of this molecular barrage to evolve resistance(93). This strategy, which has been employed by bacteria for millions of years, could be used in a clinical setting in order to treat bacterial infections with the goal of lowering the chances of resistance developing. We can learn many lessons from studying microbes, their genes, their enzymes, and the small molecules they biosynthesize. This dissertation will enlighten readers with some of these lessons.

1.5 Introduction to Type II PKS Enzymes

Type II PKS ACPs are similar to their type II FAS counterparts. The type II PKS ACP is a small four-helix bundle that is post-translationally modified with PPT. The terminal thiol of the *holo*-ACP linked PPT moiety is covalently linked to the polyketide intermediates, and it transports the growing chain into the active sites of partner enzymes. A major driving force for protein-protein interactions involving ACP is the electrostatic interactions with downstream
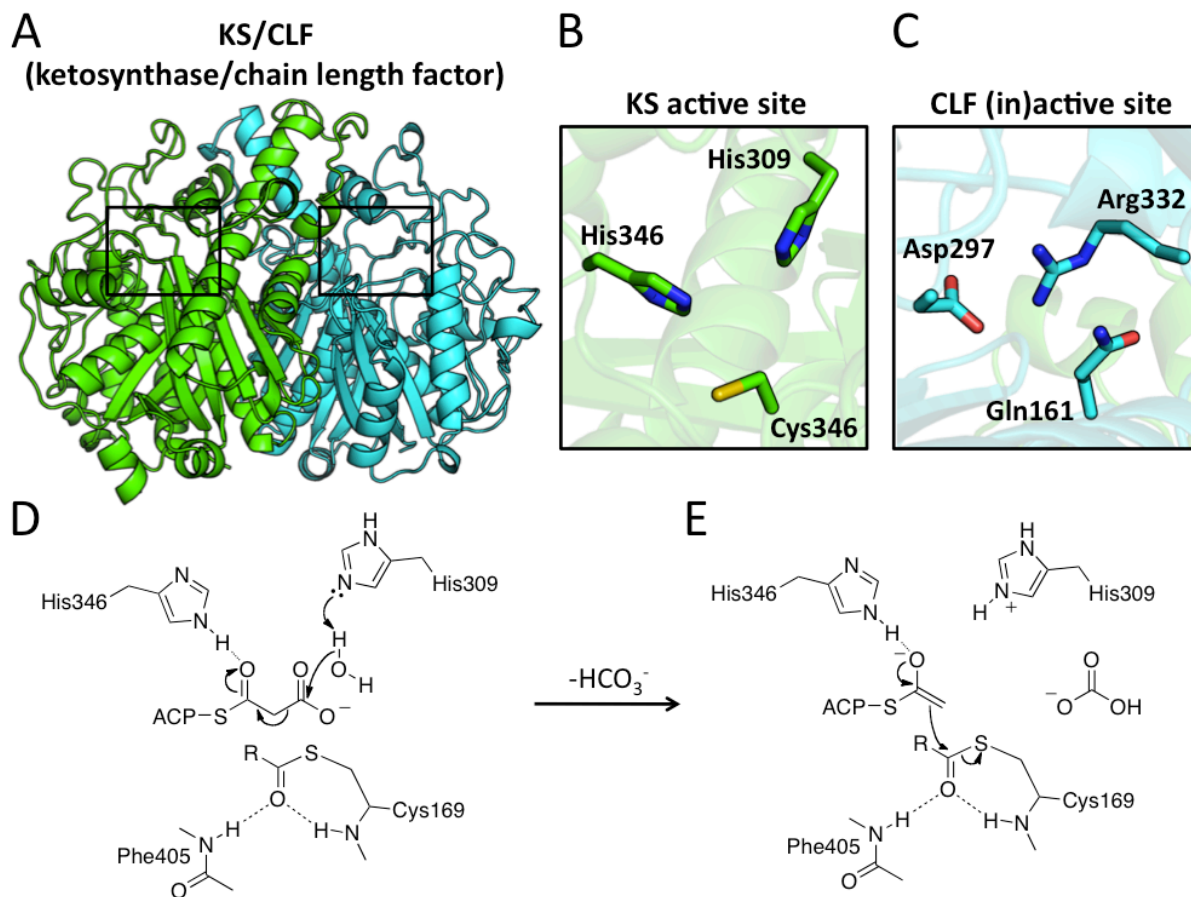
Figure 1-18. The structure and function the *Streptoymyces coelicolor* ketosynthase/chainlength factor (ActKS/CLF) from the actinorhodin biosynthesis pathway. (A) The crystal structure of ActKS/CLF (PDB ID: 1TQY) shows that ActKS/CLF forms a tight heterodimer containing only one functional active site in the KS. (B) The active site of the KS contains Cys169, His346, and His309, which are located at the end of the active site tunnel. Cys169 is located at the end of a long α-helix, and this is proposed to lower the pKa of the Cys169 thiol and increase its nucleophilicity. (C) The CLF has the same overall fold as the KS, and the corresponding active site residues of KS align with Gln161, Asp297, and Arg332 of CLF. These CLF active site residues are not catalytically active. (D) Cys169 is primed by an acyl-ACP, and the resulting acyl-Cys169 condenses with malonyl-ACP. His346 and His309 facilitate the activation of water for attack onto the carboxylate of malonyl-ACP to form the enolate and release biocarbonate. (E) Once formed, the stabilized enolate can undergo condensation with acyl-Cys169. The resulting tetrahedral intermediate is stabilized by the amide hydrogens of Cys169 and Phe405 and collapses to release an acyl-ACP that is elongated by two carbons.

partner enzymes(47, 70, 71). ACPs typically have many negatively charged surface residues (aspartic and glutamic acid), while the interacting partners contain patches of positively charged residues (lysine and arginine) on their surface near the active site entrance. In type I PKSs, the requirement for a highly negatively charged ACP and positively charged surface patch on a target enzyme is more relaxed, because ACP is covalently attached to most of its interacting partners, thus the type I ACP has a very high local concentration. In type II PKSs, however, ACPs always have negatively charged surfaces because of the requirement for interacting with a variety of partner enzymes containing positive surface patches.

The MAT is responsible for transferring malonyl-CoA onto the terminal thiol of PPT on ACP(94). In type II PKS, the MAT gene may exist in the PKS gene cluster, or the MAT may be borrowed from the organism's fatty acid pathway. In some PKSs, MAT homologues can load ACP with other acyl starter units such as a propionyl group(95).

The priming KSs use an acyl-CoA as a substrate to acylate an active site cysteine, in the same way as the type II FAS FabH is primed with acetyl-CoA(96). The starter unit undergoes condensation with malonyl-ACP. and the elongated intermediate is transferred to the KS/CLF for further chain elongation. The priming KSs play an important role in PKSs, because they allow for the incorporation of non-acetate starter units into polyketide biosynthesis, which leads to molecular diversity. The crystal structure of ZhuH, a priming KS from the R1128 pathway, reveals a thiolase fold similar to other KS enzymes(96). The overall structure and mechanism of ZhuH is similar to that of FabH and utilizes an active site cysteine for acyl transfer and an asparagine/histidine catalytic dyad for the decarboxylative elongation (Figure 1-17). Chapter 3 of this dissertation will dive deeper into the structural enzymology of type II PKS starter unit and reveal insights into propionyl priming during daunorubicin biosynthesis.

KS and CLF are two separate proteins that form a heterodimer, and it uses malonyl-ACP to catalyze the decarboxylative condensation in the KS active site to form the growing poly-β-ketone two carbons at a time. Interestingly, the KS and CLF both contain a thiolase fold, and their secondary structures overlay well. However, only the KS has a functional active site(88). The crystal structure of the KS/CLF complex from the actinorhodin biosynthetic pathway reveals an extensive dimer interface and a tunnel along the KS/CLF interface, which is proposed to determine the maximum length of the growing poly-β-ketone(88, 97). The KS active site contains a cysteine and two histidines (Figure 1-18). The cysteine is highly nucleophilic and attacks the incoming thioester of acyl-ACP intermediates, resulting in the transfer of the growing poly-β-ketone onto the cysteine. Malonyl-ACP then enters the active site, and decarboxylation is catalyzed by the two active site histidines to form a reactive enolate, which attacks the thioester bond of the cysteine linked intermediate to catalyze chain elongation. This cycle is repeated until the proper length of poly-β-ketone is generated. In the vast majority of known type II PKSs gene clusters, the genes of KS and CLF are located next to each other. They are often translationally coupled. To date, there have been no successful attempts to express a type II PKS KS/CLF heterologously in *E. coli*; the expression has only been successful in *Streptomyces* (Reference: personal communications with disgruntled type II PKS researchers).

Type PKS II gene clusters often contain one or more ketoreductases (KRs), which are enzymes that catalyze the regiospecific reduction of carbonyl groups of ACP-linked poly-β-ketone intermediates(98). PKS KRs use NADPH to stereospecifically reduce a ketone to the corresponding alcohol in the same way as FAS KRs(99). The most studied KR from a type II PKS is the actinorhodin KR (ActKR) that belongs the short chain dehydrogenase (SDR) superfamily(100, 101). Similar to other KRs, Act KR contains a Rossman fold for NADPH

34
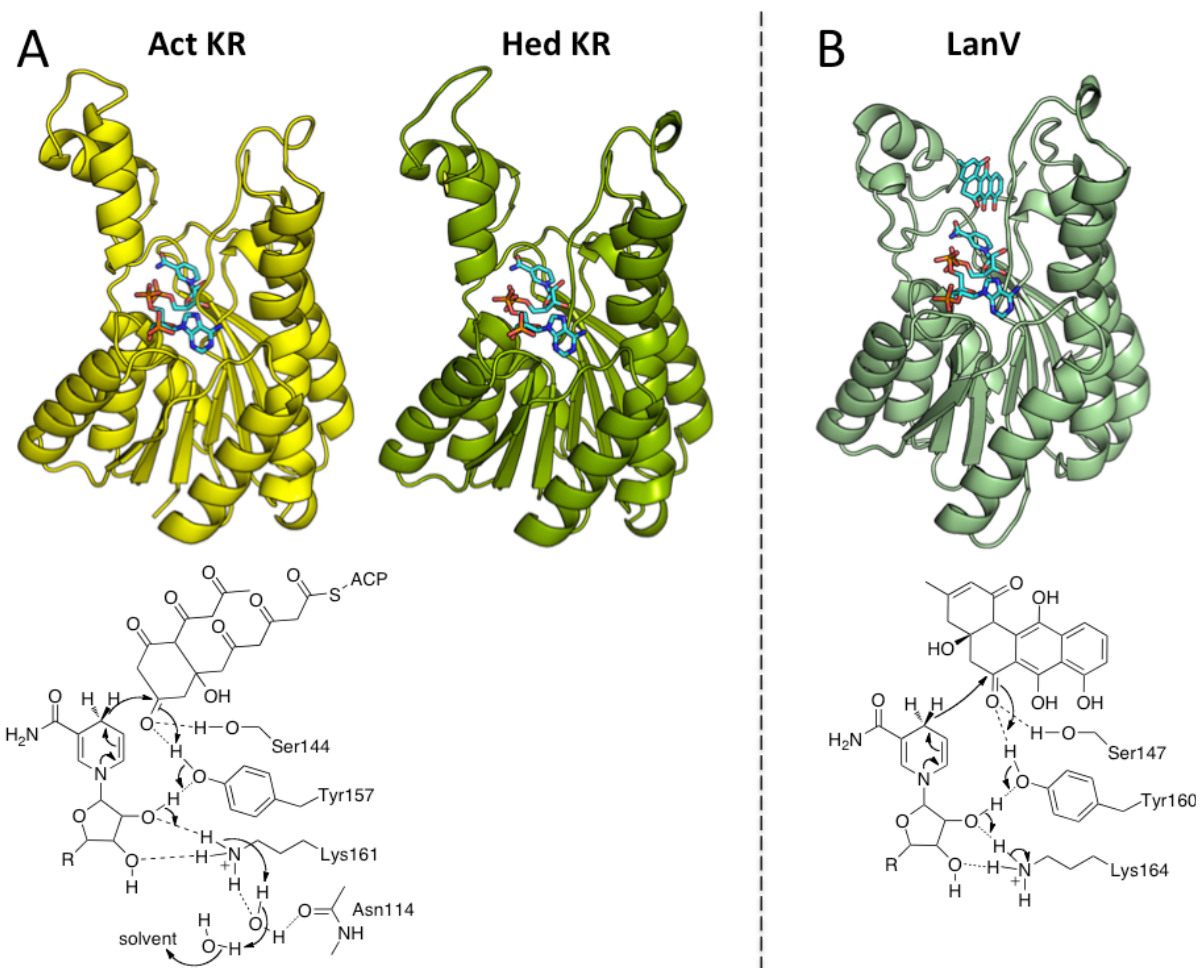
Figure 1-19. The structures and proposed mechanisms of type II PKS ketoreductases. (A) The crystal structures of ActKR (PDB ID: 1X7G) and HedKR (PDB ID: 3SJU) show a classic Rossman fold and NADPH binding site. The active site residues of ActKR consist of Ser144, Tyr157, Lys161, and Asn114, which aid in intermediate stabilization and proton transfer. Acyl-ACP enters the ActKR active site and positions the C9 carbonyl group next to the NADPH nicotinamide ring. A hydrogen bond network is formed between Tyr157, the 2' and 3' hydroxyls of the NADPH ribose ring, Lys161, Asn114, and multiple waters leading to the solvent. Ser144 hydrogen bonds to the C9 carbonyl group and stabilizes the negative charge formed upon hydride addition. The HedKR acts on polyketide intermediates with different chain lengths, but it also reduces at the C9 position and has the same active site residues as that of ActKR. (B) The crystal structure of LanV (PDB ID: 4KWI) shows a classic Rossman fold and NADPH binding site. The LanE active site consists of Ser147, Tyr160, and Lys164. In place of Asn114 of ActKR and HedKR, LanV has an ordered water molecule in the crystal structure. LanV is proposed to use a similar mechanism to Act and HedKR, but it acts on a non-ACP tethered substrate intermediate and reduces the C6 carbonyl instead of the C9 carbonyl group.
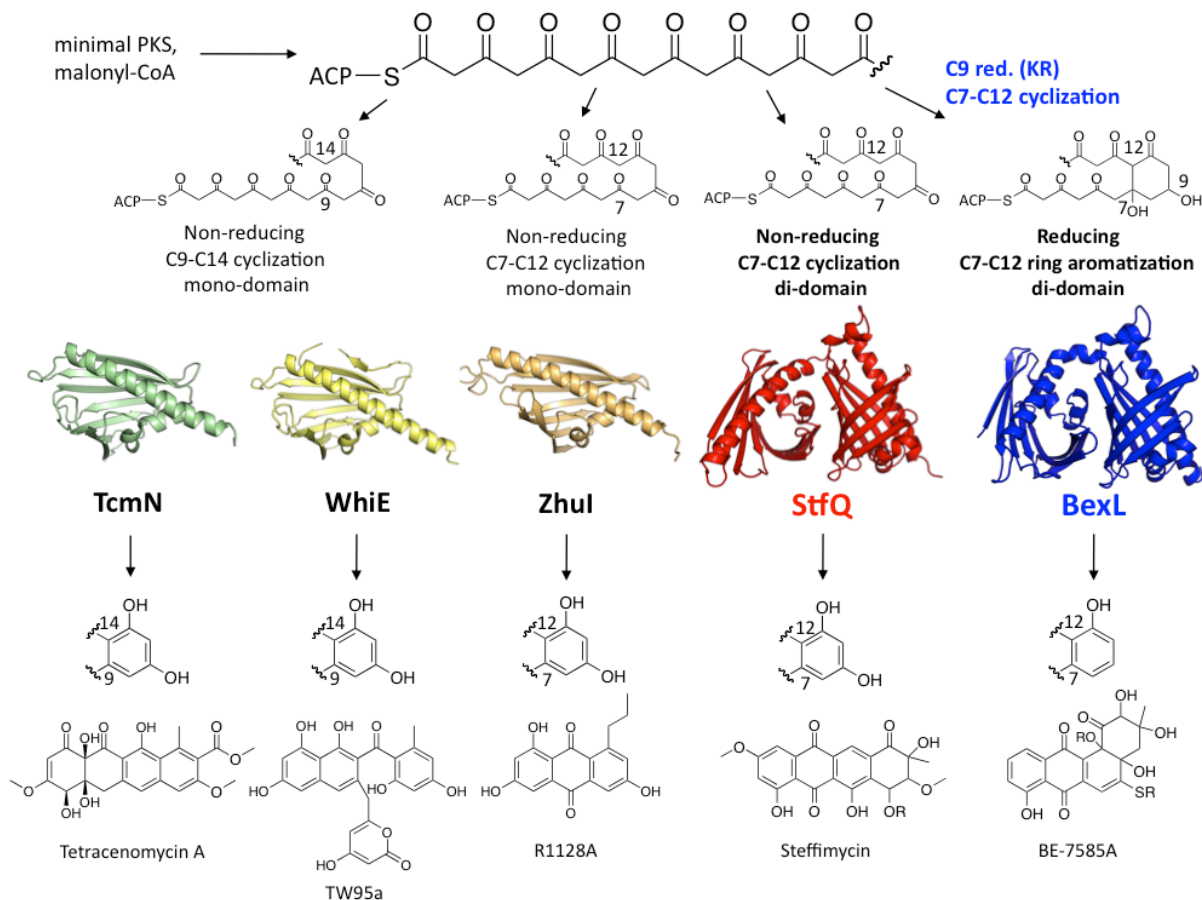
Figure 1-20. Schematic diagram of ARO/CYC activity and cyclization specificity in representative type II PKSs. The mono-domain ARO/CYCs TcmN (PDB ID: 2RER) and WhiE (PDB ID: 3TVR) act on unreduced polyketide intermediates to generate C9-C14 cyclized and aromatized products. The mono-domain ARO/CYC ZhuI (PDB ID: 3TFZ) and di-domain ARO/CYC StfQ act on unreduced polyketide intermediates to generate C7-C12 cyclized and aromatized products. The di-domain ARO/CYC BexL acts on a C9 reduced, C7-C12 cyclized intermediate and catalyzes aromatization of the C7-C12 cyclized ring by dehydrating the C9 hydroxyl.

binding, as well as conserved residues for proton transfer, and an extended proton relay network (Figure 1-19). HedKR is another type II PKS KR that also reduces C9 and is very similar to ActKR (Figure 1-19)(102). The LanV KR acts on a late-stage polyketide intermediate but reduces at the C6 position instead of C9 (Figure 1-19)(103). There is some evidence to suggest that Act KR is responsible for C7-C12 cyclization as well as C9 reduction, but it is unclear if KR itself promotes the cyclization or if the C9 reduced intermediate adopts a
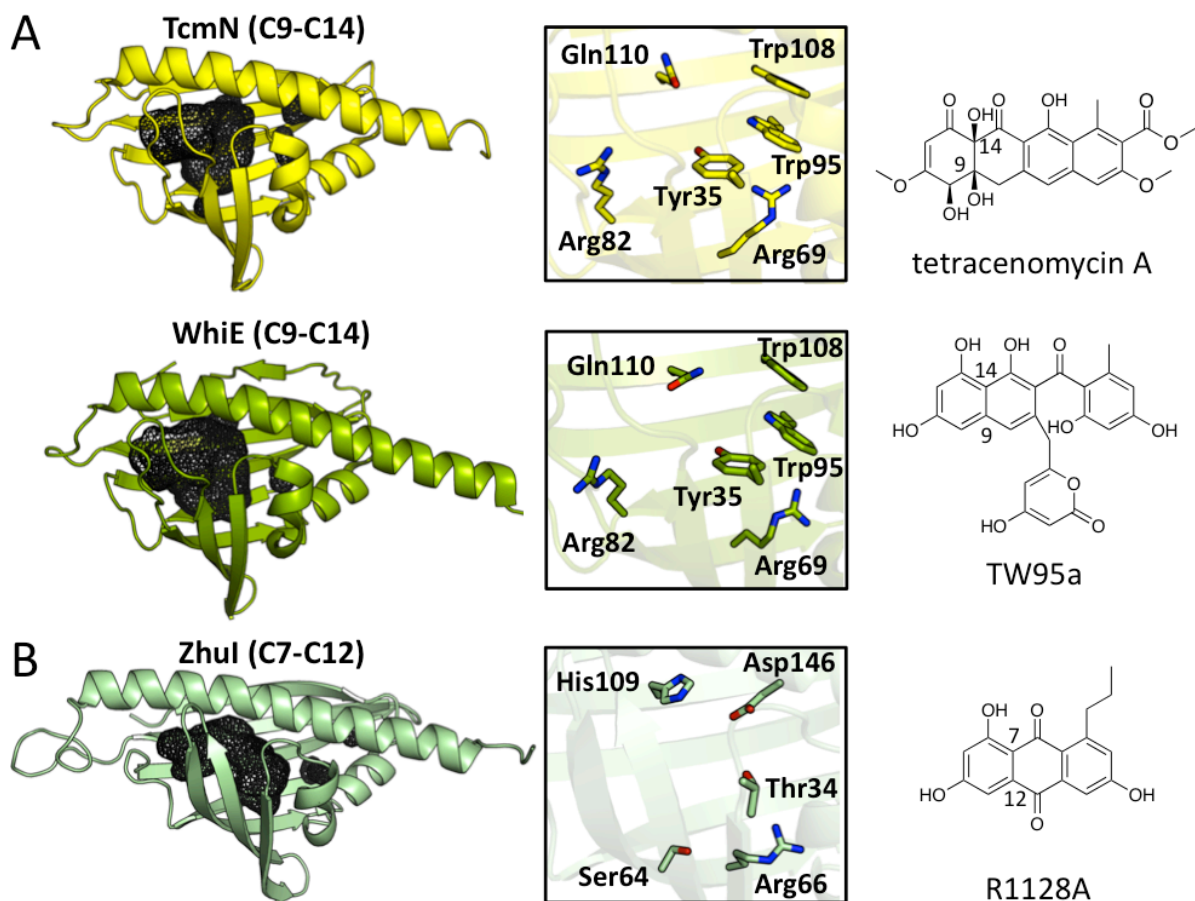
Figure 1-21. Internal cavity and active sites of the TcmN, WhiE, and ZhuI ARO/CYCs. (A) TcmN and WhiE ARO/CYC both have C9-C14 cyclization specificity and contain the same conserved active site residues. (B) ZhuI ARO/CYC has C7-C12 cyclization specificity and contains a different set of active site residues when compared to TcmN and WhiE. The only conserved residue between all three ARO/CYCs is the bottom pocket arginine (TcmN and WhiE: Arg69, ZhuI: Arg66).

conformation that promotes spontaneous cyclization to yield exclusively the C7-C12 cyclized product(90). By far the most common KR reduction pattern observed in type II PKSs is the reduction at the C9 position, which occurs in many PKS products, such as daunorubicin. Reduction at C9 is associated with dehydration of the resulting alcohol, leading to the formation of a terminal aromatic ring. The hydrophobic terminal aromatic ring allows C9 reduced/aromatized polyketides to intercalate DNA and interfere with DNA replication(104). Through sampling and diversification of regiospecific reduction in type II PKSs, nature has

evolved gene clusters to produce bioactive polyketides that can target critical cellular pathways such as DNA replication.

The aromatase/cyclases (ARO/CYCs) are responsible for the regiospecific cyclization and/or aromatization of ACP-linked polyketide intermediates. The Tsai lab has solved the structures of five different ARO/CYCs from different biosynthetic pathways (Figure 1-20)(89, 105, 106). ARO/CYCs can be classified based on the cyclization pattern they generate (C7-C12 or C9-C14), the number or ARO/CYC domains they contain (mono- or di-domain), if they are from a reducing (with KR) or non-reducing (without KR), and also the length of polyketide intermediate the ARO/CYC acts on. A brief overview of mono-domain ARO/CYCs with solved crystal structures will be presented here.

The TcmN mono-domain ARO/CYC is responsible for C9-C14 cyclization and aromatization of a non-reduced 20-carbon poly-β-ketone during tetracenomycin biosynthesis(107). The crystal structure of TcmN revealed a helix-grip fold that is structurally related to the DH hot dog fold(89). In ARO/CYCs, the large central α-helix is moved away from the β-sheet core of the protein, resulting in a much larger interior pocket when compared to the hot dog fold of DHs(108). The large interior pocket of TcmN ARO/CYC contains multiple active site residues (Gln110, Trp108, Arg82, Tyr35, Trp95, and Arg69) that are proposed to apply acid/base chemistry to perform a regiospecific intramolecular aldol cyclization (Figure 1-21). The reactive nature of poly-β-ketone intermediates has hampered efforts to determine the specific roles of each residue important for the activity of TcmN ARO/CYC.

The WhiE mono-domain ARO/CYC is responsible for C9-C14 cyclization and aromatization of a non-reduced 24-carbon poly-β-ketone during spore pigment synthesis in *Streptomyces coelicolor*(109). WhiE also has the helix-grip fold and contains a large interior
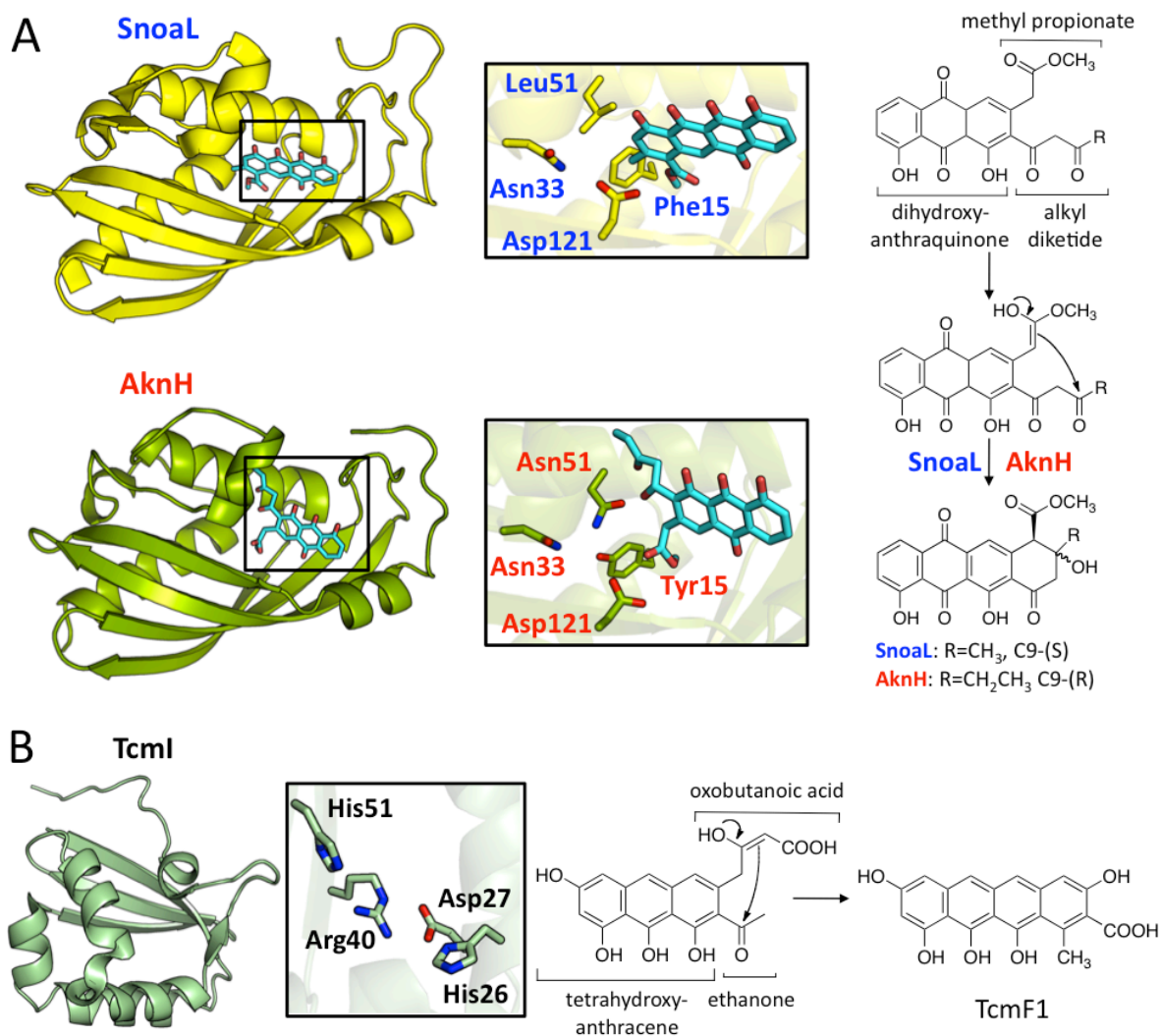
Figure 1-22. Structure and function of 4th ring cyclases (CYCs). (A) The crystal structures of SnoaL (PDB ID: 1SJW) and AknH (PDB ID: 2F99) act on similar substrates and share a similar fold. The SnoaL and AknH active sites both contain Asn33 and Asp121, which are proposed to be involved in catalysis. SnoaL contains a hydrophobic Leu51/Phe15 pair whereas AknH contains a hydrophilic Asn51/Tyr15 and these residues are proposed to be responsible for orienting the alkyl diketide and determining the stereochemical outcome of the reaction. (B) The structure of TcmI (PDB ID: 1TUW) reveals an internal cavity with proposed active site residues His51, Arg40, Asp27, and His26. The role of these residues in catalysis is not clear.

pocket with residues (Gln110, Trp108, Arg82, Tyr35, Trp95, and Arg69) proposed to use acid-base chemistry to catalyze regiospecific cyclization (Figure 1-21)(106). WhiE, unlike TcmN, has been implicated in chain length control of the growing poly-β-ketone. The WhiE minimal PKS

only produces 24-carbon poly-β-ketone intermediates in the presence of WhiE(110). In the absence of WhiE, aberrantly cyclized shorter shunt products are observed. This finding suggests that WhiE is not only responsible for regiospecific cyclization of the full length poly-β-ketone intermediate, but also interacts with one or more minimal PKS enzymes and stabilizes the growing poly-β-ketone. Further studies are needed to understand the molecular basis of this phenomenon.

The ZhuI mono-domain ARO/CYC is responsible for C7-C12 cyclization and aromatization of a non-reduced polyketide with varying chain lengths (due to the use of non-acetate starter units) during R1128 biosynthesis(111). ZhuI, like TcmN and WhiE, has a large interior pocket, although the pocket is slightly smaller than those of TcmN and WhiE because of a shift of the central helix (Figure 1-21)(105). In ZhuI, the catalytic residues involved in acid-base catalysis were determined (His109, Asp146, Thr34, Ser64, Arg66), which are different from those identified in WhiE and TcmN(105).

To summarize our previous ARO/CYC studies, the crystal structures of the mono-domain ARO/CYCs revealed that although these enzymes all act on poly-β-ketone intermediates, depending on their cyclization specificity and substrate size, the interior ARO/CYC pocket contains different residues that orient the substrate for proper cyclization. To quote a fellow lab member of mine, "I don't like to think of ARO/CYCs as enzymes, I like to think of them as polyketide chaperones that fold a linear polyketide intermediate in the right orientation to achieve a given cyclization motif". The focus of Chapter 2 of this dissertation is to present our newest results of di-domain ARO/CYCs from a reducing and non-reducing PKS for critical comparison with each other, and with the aforementioned mono-domain ARO/CYCs.

In type II PKSs, the cyclases (CYCs) catalyze 4th ring cyclization by intramolecular aldol condensations. The crystal structures of three CYCs have been solved, revealing insights about their cyclization mechanisms (Figure 1-22). AknH and SnoaL have an α + β fold, and both
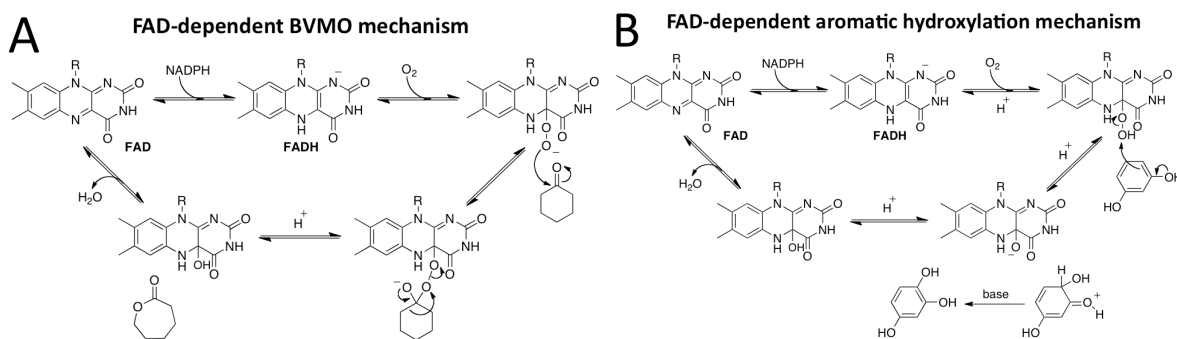
Figure 1-23. The mechanisms of FAD-dependent oxidation reactions related to type II PKS oxygenases. (A) Bayer-Villager monooxygenases catalyze the oxidation of ketones into esters. The FAD cofactor is first reduced by NAD(P)H and then reacts with O_2 to form a peroxy-flavin intermediate. The nucleophilic peroxy-flavin attacks the ketone to form a "Criegee intermediate" that collapses to generate an ester. (B) Aromatic hydroxylases oxidize activated aromatic rings but use a slightly different mechanism than that of BVMOs. The FAD is first reduced by NAD(P)H and then reacts with O_2 and H^+ to form an electrophilic flavin hydroperoxide, which can undergo electrophilic aromatic substitution. In both mechanisms, the loss of water regenerates the FAD cofactor.

contain internal pockets(112, 113). In the case of AknH and SnoaL, intramolecular aldol cyclization is proposed to be catalyzed by an active site aspartate acting as a base to generate an enolate, which then attacks a carbonyl group to form the terminal (4th) ring. The mechanism for TcmI, which has a different fold, is less clear, because mutagenesis of the putative active site acids and bases (histidine, aspartic acid, and arginine) did not lead to a loss of activity(114).

Type II PKS gene clusters often contain a variety of oxygenases capable of catalyzing hydroxylation, oxidation of alcohols to ketones, and Bayer-Villager monooxygenation(34). Oxidations typically take place after the polyketide is fully cyclized and released from the ACP. Type II PKS oxygenases are diverse and include P450s, FAD and FMN dependent oxygenases, and non-heme iron dependent enzymes. The addition of hydroxyl groups to polyaromatic type II
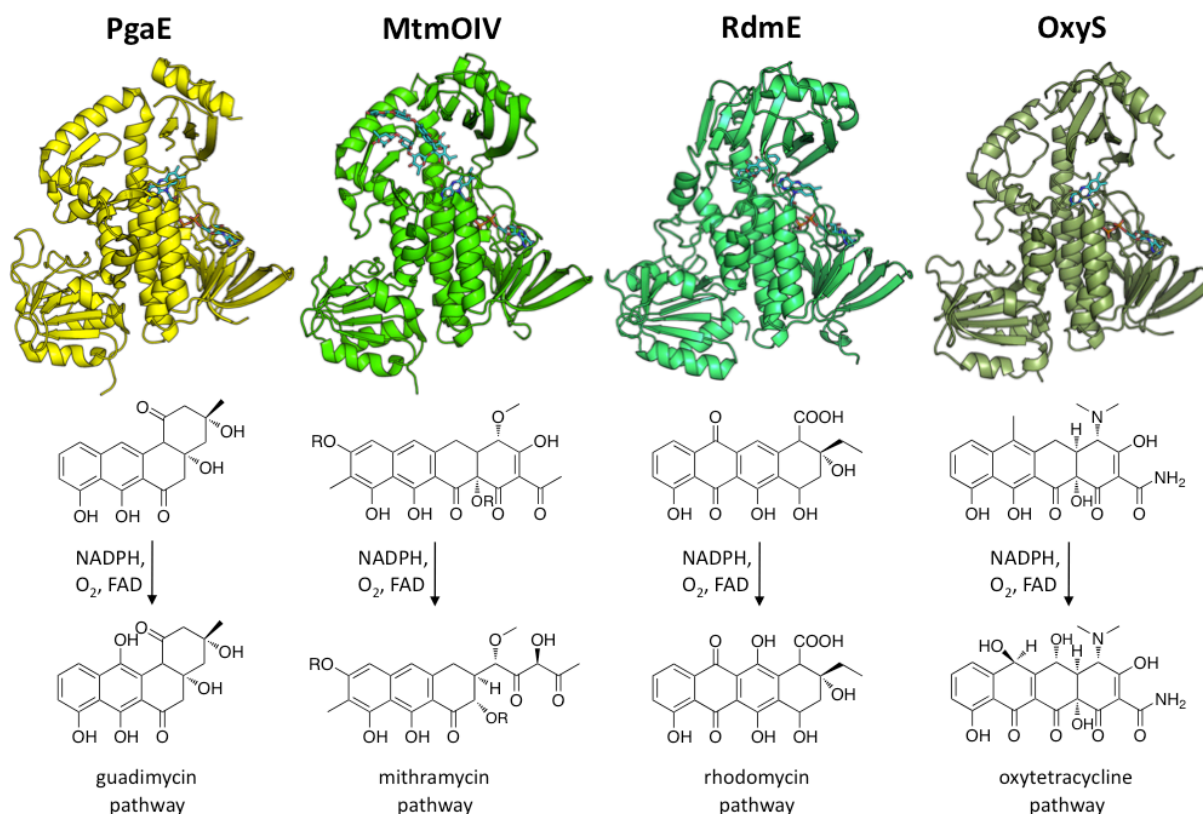
Figure 1-24. The crystal structure and reactions catalyzed by BVMOs and aromatic hydroxylases in type II PKS biosynthesis. PgaE (PDB ID: 2QA1) and RdmE (PDB ID: 3IHJ) are both aromatic hydroxylases that act on angucyclic and anthracyclic type II polyketide intermediates, respectively. OxyS (PDB ID: 3IHG) is an aromatic hydroxylase and can also oxidize the benzylic position of an intermediate in the oxytetracycline pathway. MtmOIV (4K5S) is a BVMO, which catalyzes oxidative ring expansion and cleavage during mithramycin biosynthesis.

PKS products increases their hydrophilicity and adds extra hydrogen bond donors/acceptors to the aromatic system that can influence bioactivity. The most studied type II PKS oxygenases are the FAD dependent oxygenases, which include the aromatic hydroxylases and Bayer-Villager monooxygenases (BVMOs) (Figure 1-23 and 1-24). These enzymes activate molecular oxygen and oxidize substrates by forming new carbon-oxygen bonds(115, 116). Chapter 5 describes the structure of BexE, a putative BVMO, which is proposed to catalyze the oxidative cleavage of a type II polyketide intermediate and then cyclize the cleaved product(117).

The last set of tailoring reactions includes methyl transfer and glycosylation reactions. The hydroxyl groups of polyketides are often methylated by S-adenosyl methionine dependent methyltransferases, and the presence or lack of a specific methylation can greatly impact bioactivity(34). The majority of type II PKS products are glycosylated, and it has been demonstrated that glycosylation is required for bioactivity(118). The enzymes involved in glycosylation are called glycosyl transferases (GTs) that transfer a variety of sugars to hydroxyl groups, amines, and nucleophilic carbons of type II polyketides. These enzymes are diverse and have been well studied. However, they are outside the scope of this dissertation and will not be discussed further.

1.6 Type II Starter Unit Selection and Diversity

Many type II PKS pathways harbor a dedicated ketosynthase that is responsible for loading the CoA-derived starter units onto the ACP, which are then elongated to yield diverse polyketides. The priming KS IIIs have been found to load a variety of CoA derived starter units onto ACP, not limited to, but including propionyl, butanoyl, and benozoyl groups (Figure 1-25) (119). Starter unit incorporation is initiated by the transfer of acyl-CoA derived starter unit onto an active site cysteine (or serine) of the KS followed by chain elongation with malonyl-ACP. The overall reaction results in one elongation by two carbons. The ACP-tethered starter unit is then elongated by the KS/CLF complex until a fully elongated intermediate is formed and then processed as described above to yield the final product. Using this method to incorporate non-acetate starter units adds an additional strategy to achieve molecular diversity in type II PKSs, which can lead to multiple structurally related bioactive products. Chapter 3 of this thesis is focused on understanding the structural basis for starter unit selection of an atypical 3-carbon
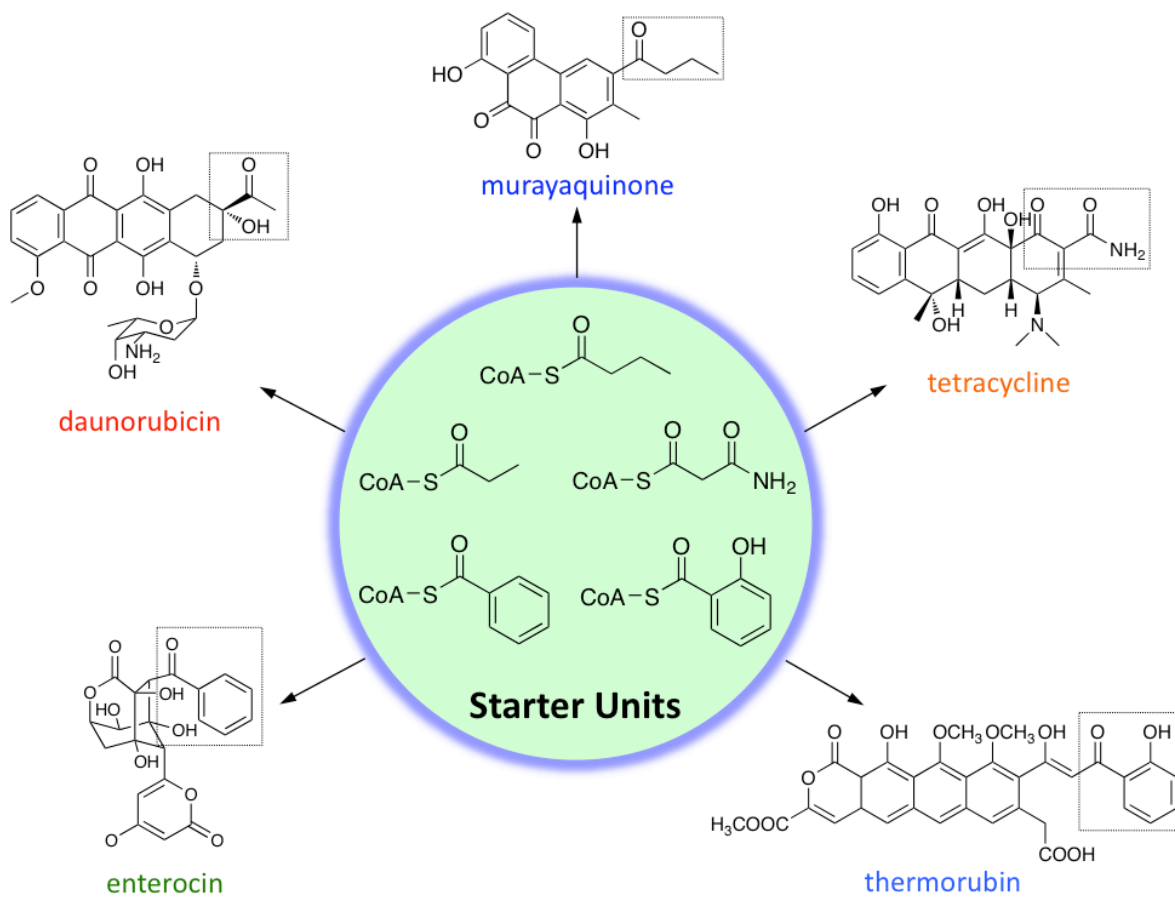
Figure 1-25. Starter unit diversity in type II PKSs. Daunorubicin, murayaquinone, tetracycline, enterocin, and thermorubin all use non-acetate starter units which leads to scaffold diversity of aromatic type II PKS products.

starter unit, which is incorporated into daunorubicin biosynthesis by the enzyme DpsC(120). In addition to KS III-mediated starter unit incorporation, recently, a novel CoA-ligase was discovered to have AT activity, and it transfers a CoA-derived starter unit onto ACP, which then undergoes elongation to yield a polyketide natural products with electron transport chain inhibitory activity called aurachins(121). Chapter 4 of this thesis is focused on understanding the structural basis of this novel starter unit system.

1.7 Introduction of Dissertation Chapters

Chapters 2-6 describe the results of my scientific efforts over the past five years. Chapters 2, 3, 4, and 5 are focused on the structural and mechanistic studies of PKS enzymes. Chapter 2 describes the crystal structures and *in vitro* functional studies of two di-domain ARO/CYC enzymes (BexL and StfQ) from type II PKSs. Chapter 3 discusses the structural determination and analysis of multiple crystal structures of the enzyme DpsC, an AT/KS III involved in daunorubicin biosynthesis. Chapter 4 focuses on the crystal structure and structural analysis of AuaEII, a novel AT from a hybrid type II PKSs pathway. Chapter 5 describes the structural determination of and initial *in vitro* assays of BexE, an oxygenase involved in polyketide biosynthesis. Chapter 6 focuses on type II fatty acid biosynthesis with an emphasis on the structural biology of protein-protein interactions. Chapter 6 begins with the published results of our lab's work on the crystal structure of the AcpP=FabA complex and is followed by my work on the recently solved AcpP=FabB complex crystal structure. Chapter 7 is a conclusion and provides future directions for the work presented in this dissertation.

References

1. Ji, H. F., Li, X. J., and Zhang, H. Y. (2009) Natural products and drug discovery Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia?, *Embo Rep 10*, 194-200.
2. Henry, K. L., Edwards, R. W., and Oetting, E. R. (2007) Use of ephedra among rural-dwelling U.S. adolescents, *Subst Use Misuse 42*, 949-959.
3. Klayman, D. L. (1985) Qinghaosu (Artemisinin) - an Antimalarial Drug from China, *Science 228*, 1049-1055.
4. Drews, J. (2000) Drug discovery: A historical perspective, *Science 287*, 1960-1964.
5. Lewis, K. (2013) Platforms for antibiotic discovery, *Nat Rev Drug Discov 12*, 371-387.
6. Wright, G. (2014) Perspective Synthetic Biology Revives Antibiotics, *Nature 509*, S13-S13.
7. Demain, A. L., and Sanchez, S. (2009) Microbial drug discovery: 80 years of progress, *J Antibiot (Tokyo) 62*, 5-16.
8. Seto, B. (2012) Rapamycin and mTOR: a serendipitous discovery and implications for breast cancer, *Clin Transl Med 1*, 29.
9. Alberts, A. W. (1988) Discovery, Biochemistry and Biology of Lovastatin, *Am J Cardiol 62*, J10-J15.
10. Nakano, H., and Omura, S. (2009) Chemical biology of natural indolocarbazole products: 30 years since the discovery of staurosporine, *J Antibiot (Tokyo) 62*, 17-26.
11. Tamaoki, T., Nomoto, H., Takahashi, I., Kato, Y., Morimoto, M., and Tomita, F. (1986) Staurosporine, a Potent Inhibitor of Phospholipid/Ca++Dependent Protein-Kinase, *Biochem Bioph Res Co 135*, 397-402.
12. Kren, V., and Martinkova, L. (2001) Glycosides in medicine: "The role of glycosidic residue in biological activity", *Curr Med Chem 8*, 1303-1328.
13. Ricart, A. D. (2011) Antibody-Drug Conjugates of Calicheamicin Derivative: Gemtuzumab Ozogamicin and Inotuzumab Ozogamicin, *Clin Cancer Res 17*, 6417-6427.
14. Sehgal, S. N. (1998) Rapamune (R) (RAPA, rapamycin, sirolimus): Mechanism of action immunosuppressive effect results from blockade of signal transduction and inhibition of cell cycle progression, *Clin Biochem 31*, 335-340.
15. Cobbold, S. P. (2013) The mTOR pathway and integrating immune regulation, *Immunology 140*, 391-398.
16. Newman, D. J., and Cragg, G. M. (2012) Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010, *J Nat Prod 75*, 311-335.
17. Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K. M., Kent, A. D., Daroub, S. H., Camargo, F. A. O., Farmerie, W. G., and Triplett, E. W. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity, *Isme J 1*, 283-290.
18. Davies, J. (2013) Specialized microbial metabolites: functions and origins, *J Antibiot 66*, 361-364.
19. Visick, K. L., Foster, J., Doino, J., McFall-Ngai, M., and Ruby, E. G. (2000) Vibrio fischeri lux genes play an important role in colonization and development of the host light organ, *Journal of Bacteriology 182*, 4578-4586.
20. Bottcher, T., and Clardy, J. (2014) A Chimeric Siderophore Halts Swarming Vibrio, *Angew Chem Int Edit 53*, 3510-3513.

21. Currie, C. R., Scott, J. A., Summerbell, R. C., and Malloch, D. (2003) Fungus-growing ants use antibiotic-producing bacteria to control garden parasites (vol 398, pg 701, 1999), *Nature 423*, 461-461.
22. Oh, D. C., Poulsen, M., Currie, C. R., and Clardy, J. (2009) Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis, *Nat Chem Biol 5*, 391-393.
23. Williams, P. (2007) Quorum sensing, communication and cross-kingdom signalling in the bacterial world, *Microbiol-Sgm 153*, 3923-3938.
24. O'brien, J., and Wright, G. D. (2011) An ecological perspective of microbial secondary metabolism, *Curr Opin Biotech 22*, 552-558.
25. Hopwood, D. A. (1997) Genetic contributions to understanding polyketide synthases, *Chem Rev 97*, 2465-2497.
26. Fischbach, M. A., Walsh, C. T., and Clardy, J. (2008) The evolution of gene collectives: How natural selection drives chemical innovation, *Proc Natl Acad Sci U S A 105*, 4601-4608.
27. Garrod, L. P. (1957) Erythromycin Group of Antibiotics, *Brit Med J 2*, 57-63.
28. Dobias, B., and Hazen, E. L. (1961) Nystatin, *Chemotherapy 3*, 108-&.
29. Hopwood, D. A. (1990) Molecular-Genetics of Polyketides and Its Comparison to Fatty-Acid Biosynthesis, *Annu Rev Genet 24*, 37-66.
30. Ziegler, J., and Facchini, P. J. (2008) Alkaloid biosynthesis: Metabolism and trafficking, *Annu Rev Plant Biol 59*, 735-769.
31. Thibodeaux, C. J., Melancon, C. E., and Liu, H. W. (2007) Unusual sugar biosynthesis and natural product glycodiversification, *Nature 446*, 1008-1016.
32. Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms, *Curr Opin Chem Biol 7*, 285-295.
33. Keatinge-Clay, A. T. (2012) The structures of type I polyketide synthases, *Nat Prod Rep 29*, 1050-1073.
34. Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork, *Nat Prod Rep 24*, 162-190.
35. Austin, M. B., and Noel, A. J. P. (2003) The chalcone synthase superfamily of type III polyketide synthases, *Nat Prod Rep 20*, 79-110.
36. Das, U. N. (2006) Essential fatty acids - A review, *Curr Pharm Biotechno 7*, 467-482.
37. Leibundgut, M., Maier, T., Jenni, S., and Ban, N. (2008) The multienzyme architecture of eukaryotic fatty acid synthases, *Curr Opin Struc Biol 18*, 714-725.
38. Maier, T., Jenni, S., and Ban, N. (2006) Architecture of mammalian fatty acid synthase at 4.5 angstrom resolution, *Science 311*, 1258-1262.
39. White, S. W., Zheng, J., Zhang, Y. M., and Rock, C. O. (2005) The structural biology of type II fatty acid biosynthesis, *Annu Rev Biochem 74*, 791-831.
40. 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.
41. Magnuson, K., Jackowski, S., Rock, C. O., and Cronan, J. E. (1993) Regulation of Fatty-Acid Biosynthesis in Escherichia-Coli, *Microbiol Rev 57*, 522-542.
42. Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., Collins, D., Delisle, G., and Jacobs, W. R. (1994) Inha, a Gene Encoding a Target for Isoniazid and Ethionamide in Mycobacterium-Tuberculosis, *Science 263*, 227-230.
43. Wang, J., Soisson, S. M., Young, K., Shoop, W., Kodali, S., Galgoci, A., Painter, R., Parthasarathy, G., Tang, Y. S., Cummings, R., Ha, S., Dorso, K., Motyl, M., Jayasuriya,

H., Ondeyka, J., Herath, K., Zhang, C. W., Hernandez, L., Allocco, J., Basilio, A., Tormo, J. R., Genilloud, O., Vicente, F., Pelaez, F., Colwell, L., Lee, S. H., Michael, B., Felcetto, T., Gill, C., Silver, L. L., Hermes, J. D., Bartizal, K., Barrett, J., Schmatz, D., Becker, J. W., Cully, D., and Singh, S. B. (2006) Platensimycin is a selective FabF inhibitor with potent antibiotic properties, *Nature 441*, 358-361.

44. Steen, E. J., Kang, Y. S., Bokinsky, G., Hu, Z. H., Schirmer, A., McClure, A., del Cardayre, S. B., and Keasling, J. D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass, *Nature 463*, 559-U182.

45. Yu, X. Y., Liu, T. G., Zhu, F. Y., and Khosla, C. (2011) In vitro reconstitution and steady-state analysis of the fatty acid synthase from Escherichia coli, *P Natl Acad Sci USA 108*, 18643-18648.

46. Chan, D. I., and Vogel, H. J. (2010) Current understanding of fatty acid biosynthesis and the acyl carrier protein, *Biochem J 430*, 1-19.

47. Masoudi, A., Raetz, C. R., Zhou, P., and Pemble, C. W. t. (2014) Chasing acyl carrier protein through a catalytic cycle of lipid A production, *Nature 505*, 422-426.

48. Agarwal, V., Lin, S., Lukk, T., Nair, S. K., and Cronan, J. E. (2012) Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis, *P Natl Acad Sci USA 109*, 17406-17411.

49. Qiu, X. Y., and Janson, C. A. (2004) Structure of apo acyl carrier protein and a proposal to engineer protein crystallization through metal ions, *Acta Crystallogr D 60*, 1545-1554.

50. Roujeinikova, A., Simon, W. J., Gilroy, J., Rice, D. W., Rafferty, J. B., and Slabas, A. R. (2007) Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates, *J Mol Biol 365*, 135-145.

51. Oefner, C., Schulz, H., D'Arcy, A., and Dale, G. E. (2006) Mapping the active site of Escherichia coli malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography, *Acta Crystallogr D 62*, 613-618.

52. Dodson, G., and Wlodawer, A. (1998) Catalytic triads and their relatives, *Trends Biochem Sci 23*, 347-352.

53. Piel, J. (2010) Biosynthesis of polyketides by trans-AT polyketide synthases, *Nat Prod Rep 27*, 996-1047.

54. Huang, W. J., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. (1998) Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from E-coli reveals the molecular architecture of condensing enzymes, *Embo J 17*, 1183-1191.

55. Davies, C., Heath, R. J., White, S. W., and Rock, C. O. (2000) The 1.8 angstrom crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from Escherichia coli, *Structure 8*, 185-195.

56. Pappenberger, G., Schulz-Gasch, T., Kusznir, E., Muller, F., and Hennig, M. (2007) Structure-assisted discovery of an aminothiazole derivative as a lead molecule for inhibition of bacterial fatty-acid synthesis, *Acta Crystallogr D 63*, 1208-1216.

57. Feng, Y. J., and Cronan, J. E. (2009) Escherichia coli Unsaturated Fatty Acid Synthesis COMPLEX TRANSCRIPTION OF THE fabA GENE AND IN VIVO IDENTIFICATION OF THE ESSENTIAL REACTION CATALYZED BY FabB, *Journal of Biological Chemistry 284*, 29526-29535.

58. Choi, K. H., Price, A., Heath, R., Li, Z. M., White, S., and Rock, C. (2001) Inhibition of beta-ketoacyl-[acyl carrier protein] synthase by thiolactomycin and cerulenin: Structure and mechanism, *Faseb J 15*, A191-A191.

59. Moche, M., Schneider, G., Edwards, P., Dehesh, K., and Lindqvist, Y. (1999) Structure of the complex between the antibiotic cerulenin and its target, beta-ketoacyl-acyl carrier protein synthase, *Journal of Biological Chemistry 274*, 6031-6034.

60. Price, A. C., Zhang, Y. M., Rock, C. O., and White, S. W. (2001) Structure of beta-ketoacyl-[acyl carrier protein] reductase from Escherichia coli: negative cooperativity and its structural basis, *Biochemistry 40*, 12772-12781.

61. Persson, B., Hedlund, J., and Jornvall, H. (2008) Medium- and short-chain dehydrogenase/reductase gene and protein families : the MDR superfamily, *Cell Mol Life Sci 65*, 3879-3894.

62. Price, A. C., Zhang, Y. M., Rock, C. O., and White, S. W. (2004) Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG, *Structure 12*, 417-428.

63. Heath, R. J., and Rock, C. O. (1996) Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in Escherichia coli fatty acid biosynthesis, *J Biol Chem 271*, 27795-27801.

64. Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., and Smith, J. L. (1996) Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site, *Structure 4*, 253-264.

65. Kimber, M. S., Martin, F., Lu, Y., Houston, S., Vedadi, M., Dharamsi, A., Fiebig, K. M., Schmid, M., and Rock, C. O. (2004) The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from Pseudomonas aeruginosa, *J Biol Chem 279*, 52593-52602.

66. Heath, R. J., and Rock, C. O. (1995) Enoyl-acyl carrier protein reductase (fabI) plays a determinant role in completing cycles of fatty acid elongation in Escherichia coli, *J Biol Chem 270*, 26538-26542.

67. Heath, R. J., Rubin, J. R., Holland, D. R., Zhang, E., Snow, M. E., and Rock, C. O. (1999) Mechanism of triclosan inhibition of bacterial fatty acid synthesis, *J Biol Chem 274*, 11110-11114.

68. Baldock, C., Rafferty, J. B., Sedelnikova, S. E., Baker, P. J., Stuitje, A. R., Slabas, A. R., Hawkes, T. R., and Rice, D. W. (1996) A mechanism of drug action revealed by structural studies of enoyl reductase, *Science 274*, 2107-2110.

69. Newmister, S. A., and Sherman, D. H. (2014) Crystal structures of acyl carrier protein in complex with two catalytic partners show a dynamic role in cellular metabolism, *Chembiochem 15*, 1079-1081.

70. Cryle, M. J., and Schlichting, I. (2008) Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(BioI) ACP complex, *Proc Natl Acad Sci U S A 105*, 15696-15701.

71. Nguyen, C., Haushalter, R. W., Lee, D. J., Markwick, P. R., Bruegger, J., Caldara-Festin, G., Finzel, K., Jackson, D. R., Ishikawa, F., O'Dowd, B., McCammon, J. A., Opella, S. J., Tsai, S. C., and Burkart, M. D. (2014) Trapping the dynamic acyl carrier protein in fatty acid biosynthesis, *Nature 505*, 427-431.

72. Heath, R. J., White, S. W., and Rock, C. O. (2002) Inhibitors of fatty acid synthesis as antimicrobial chemotherapeutics, *Appl Microbiol Biotechnol 58*, 695-703.

73. Lennen, R. M., and Pfleger, B. F. (2013) Microbial production of fatty acid-derived fuels and chemicals, *Curr Opin Biotechnol 24*, 1044-1053.
74. Janssen, H. J., and Steinbuchel, A. (2014) Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels, *Biotechnol Biofuels 7*, 7.
75. Lennen, R. M., Braden, D. J., West, R. A., Dumesic, J. A., and Pfleger, B. F. (2010) A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes, *Biotechnol Bioeng 106*, 193-202.
76. Meier, J. L., Haushalter, R. W., and Burkart, M. D. (2010) A mechanism based protein crosslinker for acyl carrier protein dehydratases, *Bioorg Med Chem Lett 20*, 4936-4939.
77. Worthington, A. S., Porter, D. F., and Burkart, M. D. (2010) Mechanism-based crosslinking as a gauge for functional interaction of modular synthases, *Org Biomol Chem 8*, 1769-1772.
78. Cox, R. J., and Simpson, T. J. (2009) Fungal type I polyketide synthases, *Methods Enzymol 459*, 49-78.
79. Keatinge-Clay, A. T. (2012) The structures of type I polyketide synthases, *Nat Prod Rep 29*, 1050-1073.
80. Smith, S., and Tsai, S. C. (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases, *Nat Prod Rep 24*, 1041-1072.
81. Boddy, C. N., Schneider, T. L., Hotta, K., Walsh, C. T., and Khosla, C. (2003) Epothilone C macrolactonization and hydrolysis are catalyzed by the isolated thioesterase domain of epothilone polyketide synthase, *J Am Chem Soc 125*, 3428-3429.
82. Fujii, I. (2010) Functional analysis of fungal polyketide biosynthesis genes, *J Antibiot (Tokyo) 63*, 207-218.
83. Crawford, J. M., Dancy, B. C., Hill, E. A., Udwary, D. W., and Townsend, C. A. (2006) Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase, *Proc Natl Acad Sci U S A 103*, 16728-16733.
84. Xu, Y., Zhou, T., Zhou, Z., Su, S., Roberts, S. A., Montfort, W. R., Zeng, J., Chen, M., Zhang, W., Lin, M., Zhan, J., and Molnar, I. (2013) Rational reprogramming of fungal polyketide first-ring cyclization, *Proc Natl Acad Sci U S A 110*, 5398-5403.
85. Foerstner, K. U., Doerks, T., Creevey, C. J., Doerks, A., and Bork, P. (2008) A computational screen for type I polyketide synthases in metagenomics shotgun data, *PLoS One 3*, e3515.
86. Wawrik, B., Kerkhof, L., Zylstra, G. J., and Kukor, J. J. (2005) Identification of unique type II polyketide synthase genes in soil, *Appl Environ Microbiol 71*, 2232-2238.
87. Das, A., and Khosla, C. (2009) Biosynthesis of aromatic polyketides in bacteria, *Acc Chem Res 42*, 631-639.
88. Keatinge-Clay, A. T., Maltby, D. A., Medzihradszky, K. F., Khosla, C., and Stroud, R. M. (2004) An antibiotic factory caught in action, *Nat Struct Mol Biol 11*, 888-893.
89. Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y., and Tsai, S. C. (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides, *Proc Natl Acad Sci U S A 105*, 5349-5354.
90. Javidpour, P., Bruegger, J., Srithahan, S., Korman, T. P., Crump, M. P., Crosby, J., Burkart, M. D., and Tsai, S. C. (2013) The determinants of activity and specificity in actinorhodin type II polyketide ketoreductase, *Chem Biol 20*, 1225-1234.

91.   Zhou, H., Li, Y., and Tang, Y. (2010) Cyclization of aromatic polyketides from bacteria and fungi, *Nat Prod Rep 27*, 839-868.
92.   Rix, U., Fischer, C., Remsing, L. L., and Rohr, J. (2002) Modification of post-PKS tailoring steps through combinatorial biosynthesis, *Nat Prod Rep 19*, 542-580.
93.   Hofle, G., and Kunze, B. (2008) Biosynthesis of aurachins A-L in Stigmatella aurantiaca: a feeding study, *J Nat Prod 71*, 1843-1849.
94.   Koppisch, A. T., and Khosla, C. (2003) Structure-based mutagenesis of the malonyl-CoA:acyl carrier protein transacylase from Streptomyces coelicolor, *Biochemistry 42*, 11057-11064.
95.   Waldman, A. J., and Balskus, E. P. (2014) Lomaiviticin biosynthesis employs a new strategy for starter unit generation, *Org Lett 16*, 640-643.
96.   Pan, H., Tsai, S., Meadows, E. S., Miercke, L. J., Keatinge-Clay, A. T., O'Connell, J., Khosla, C., and Stroud, R. M. (2002) Crystal structure of the priming beta-ketosynthase from the R1128 polyketide biosynthetic pathway, *Structure 10*, 1559-1568.
97.   Tang, Y., Tsai, S. C., and Khosla, C. (2003) Polyketide chain length control by chain length factor, *J Am Chem Soc 125*, 12708-12709.
98.   Bartel, P. L., Zhu, C. B., Lampel, J. S., Dosch, D. C., Connors, N. C., Strohl, W. R., Beale, J. M., Jr., and Floss, H. G. (1990) Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in streptomycetes: clarification of actinorhodin gene functions, *J Bacteriol 172*, 4816-4826.
99.   Javidpour, P., Korman, T. P., Shakya, G., and Tsai, S. C. (2011) Structural and biochemical analyses of regio- and stereospecificities observed in a type II polyketide ketoreductase, *Biochemistry 50*, 4638-4649.
100.   Korman, T. P., Tan, Y. H., Wong, J., Luo, R., and Tsai, S. C. (2008) Inhibition kinetics and emodin cocrystal structure of a type II polyketide ketoreductase, *Biochemistry 47*, 1837-1847.
101.   Korman, T. P., Hill, J. A., Vu, T. N., and Tsai, S. C. (2004) Structural analysis of actinorhodin polyketide ketoreductase: cofactor binding and substrate specificity, *Biochemistry 43*, 14529-14538.
102.   Javidpour, P., Das, A., Khosla, C., and Tsai, S. C. (2011) Structural and biochemical studies of the hedamycin type II polyketide ketoreductase (HedKR): molecular basis of stereo- and regiospecificities, *Biochemistry 50*, 7426-7439.
103.   Paananen, P., Patrikainen, P., Kallio, P., Mantsala, P., Niemi, J., Niiranen, L., and Metsa-Ketela, M. (2013) Structural and functional analysis of angucycline C-6 ketoreductase LanV involved in landomycin biosynthesis, *Biochemistry 52*, 5304-5314.
104.   Yang, F., Teves, S. S., Kemp, C. J., and Henikoff, S. (2014) Doxorubicin, DNA torsion, and chromatin dynamics, *Biochim Biophys Acta 1845*, 84-89.
105.   Ames, B. D., Lee, M. Y., Moody, C., Zhang, W., Tang, Y., and Tsai, S. C. (2011) Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway, *Biochemistry 50*, 8392-8406.
106.   Lee, M. Y., Ames, B. D., and Tsai, S. C. (2012) Insight into the molecular basis of aromatic polyketide cyclization: crystal structure and in vitro characterization of WhiE-ORFVI, *Biochemistry 51*, 3079-3091.
107.   Bao, W., Wendt-Pienkowski, E., and Hutchinson, C. R. (1998) Reconstitution of the iterative type II polyketide synthase for tetracenomycin F2 biosynthesis, *Biochemistry 37*, 8132-8138.

108. Iyer, L. M., Koonin, E. V., and Aravind, L. (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily, *Proteins 43*, 134-144.

109. Yu, T. W., and Hopwood, D. A. (1995) Ectopic expression of the Streptomyces coelicolor whiE genes for polyketide spore pigment synthesis and their interaction with the act genes for actinorhodin biosynthesis, *Microbiology 141 (Pt 11)*, 2779-2791.

110. Shen, Y., Yoon, P., Yu, T. W., Floss, H. G., Hopwood, D., and Moore, B. S. (1999) Ectopic expression of the minimal whiE polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes, *Proc Natl Acad Sci U S A 96*, 3622-3627.

111. Marti, T., Hu, Z., Pohl, N. L., Shah, A. N., and Khosla, C. (2000) Cloning, nucleotide sequence, and heterologous expression of the biosynthetic gene cluster for R1128, a non-steroidal estrogen receptor antagonist. Insights into an unusual priming mechanism, *J Biol Chem 275*, 33443-33448.

112. Kallio, P., Sultana, A., Niemi, J., Mantsala, P., and Schneider, G. (2006) Crystal structure of the polyketide cyclase AknH with bound substrate and product analogue: implications for catalytic mechanism and product stereoselectivity, *J Mol Biol 357*, 210-220.

113. Beinker, P., Lohkamp, B., Peltonen, T., Niemi, J., Mantsala, P., and Schneider, G. (2006) Crystal structures of SnoaL2 and AclR: two putative hydroxylases in the biosynthesis of aromatic polyketide antibiotics, *J Mol Biol 359*, 728-740.

114. Thompson, T. B., Katayama, K., Watanabe, K., Hutchinson, C. R., and Rayment, I. (2004) Structural and functional analysis of tetracenomycin F2 cyclase from Streptomyces glaucescens. A type II polyketide cyclase, *J Biol Chem 279*, 37956-37963.

115. Lindqvist, Y., Koskiniemi, H., Jansson, A., Sandalova, T., Schnell, R., Liu, Z., Mantsala, P., Niemi, J., and Schneider, G. (2009) Structural basis for substrate recognition and specificity in aklavinone-11-hydroxylase from rhodomycin biosynthesis, *J Mol Biol 393*, 966-977.

116. Gibson, M., Nur-e-alam, M., Lipata, F., Oliveira, M. A., and Rohr, J. (2005) Characterization of kinetics and products of the Baeyer-Villiger oxygenase MtmOIV, the key enzyme of the biosynthetic pathway toward the natural product anticancer drug mithramycin from Streptomyces argillaceus, *J Am Chem Soc 127*, 17594-17595.

117. Sasaki, E., Ogasawara, Y., and Liu, H. W. (2010) A biosynthetic pathway for BE-7585A, a 2-thiosugar-containing angucycline-type natural product, *J Am Chem Soc 132*, 7405-7417.

118. La Ferla, B., Airoldi, C., Zona, C., Orsato, A., Cardona, F., Merlo, S., Sironi, E., D'Orazio, G., and Nicotra, F. (2011) Natural glycoconjugates with antitumor activity, *Nat Prod Rep 28*, 630-648.

119. Moore, B. S., and Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units, *Nat Prod Rep 19*, 70-99.

120. Bao, W., Sheldon, P. J., Wendt-Pienkowski, E., and Hutchinson, C. R. (1999) The Streptomyces peucetius dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide, *J Bacteriol 181*, 4690-4695.

121. Pistorius, D., Li, Y., Mann, S., and Muller, R. (2011) Unprecedented anthranilate priming involving two enzymes of the acyl adenylating superfamily in aurachin biosynthesis, *J Am Chem Soc 133*, 12362-12365.

Chapter 2

Crystal Structure and Biochemical Studies of Two Di-Domain Aromatase/Cyclases (ARO/CYCs) During Polyketide Biosynthesis

2.1 Abstract

Aromatic polyketides comprise a large class of natural products with diverse bioactivity. During biosynthesis, linear poly-β-ketone intermediates are regiospecifically cyclized, yielding molecules with defined cyclization patterns that are crucial for the polyketide bioactivity. The aromatase/cyclases (ARO/CYCs) are responsible for regiospecific cyclization of bacterial polyketides. The two most common cyclization patterns are C7-C12 and C9-C14 cyclizations. We have previously characterized three mono-domain ARO/CYCs. The last remaining uncharacterized class of ARO/CYCs is the di-domain ARO/CYCs, which catalyze C7-C12 cyclization and/or aromatization. Di-domain ARO/CYCs can further be separated into two sub-classes, "non-reducing" that act on non-reduced poly-β-ketones, and "reducing" that act on C9 reduced poly-β-ketones. For years, the functional role of each domain in cyclization and aromatization for di-domain ARO/CYCs has remained a mystery. Here we present the first structural and functional analysis, along with an in-depth comparison, of the non-reducing (StfQ) and reducing (BexL) di-domain ARO/CYCs. This work completes the structural and functional characterization of mono- and di-domain ARO/CYCs in bacterial type II PKSs and lays the groundwork for engineered biosynthesis of new bioactive polyketides.
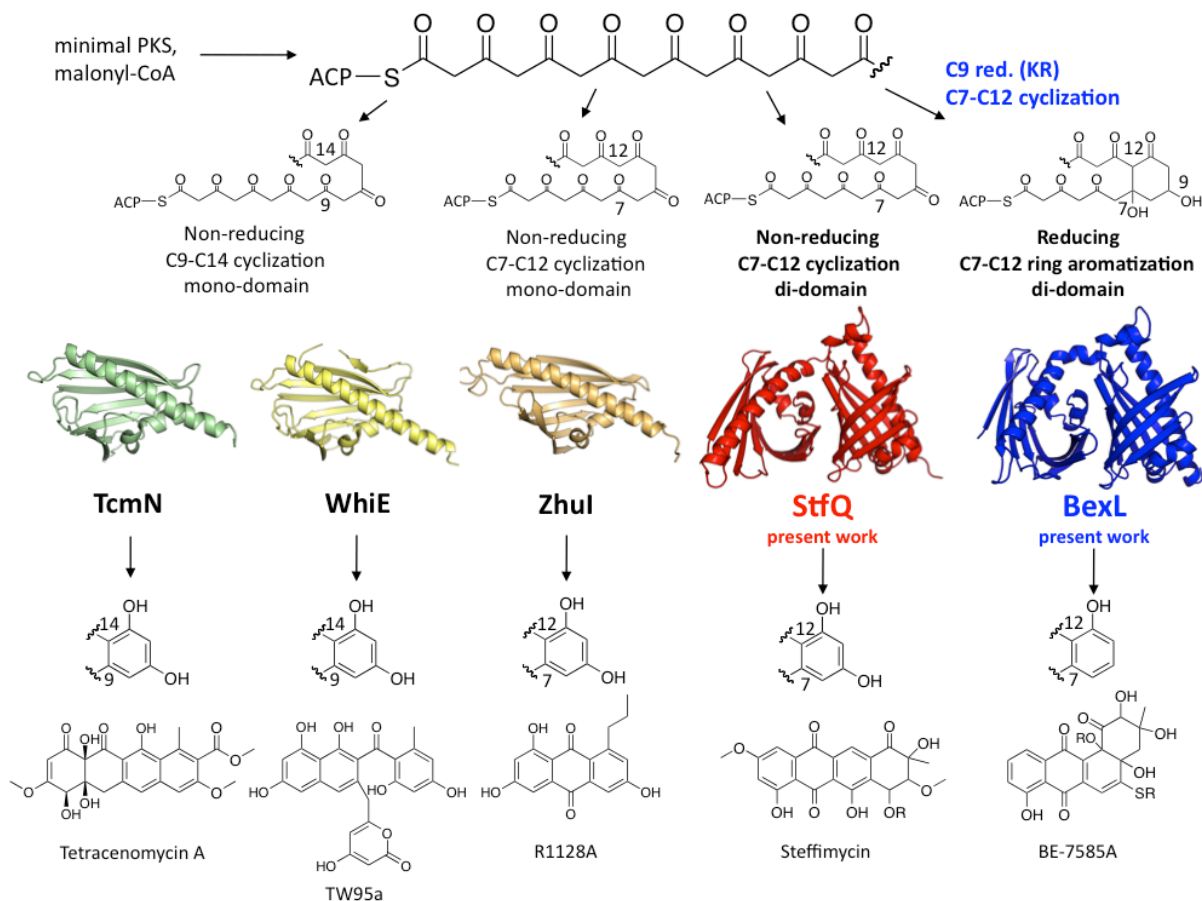
Figure 2-1. Schematic diagram of ARO/CYC activity and cyclization specificity in representative type II PKSs. The monodomain ARO/CYCs TcmN (PDB ID: 2RER) and WhiE (PDB ID: 3TVR) act on unreduced polyketide intermediates to generate C9-C14 cyclized and aromatized products. The monodomain ARO/CYC ZhuI (PDB ID: 3TFZ) and di-domain ARO/CYC StfQ act on unreduced polyketide intermediates to generate C7-C12 cyclized and aromatized products. The didomain ARO/CYC BexL acts on a C9 reduced, C7-C12 cyclized intermediates and catalyzes the aromatization of the C7-C12 cyclized ring by dehydration of the C9 hydroxyl group.

2.2 Introduction

The biosynthesis of type II aromatic polyketide natural products has been vigorously investigated due to the versatile pharmacological properties these compounds have exhibited(1-7). The type II polyketide synthase (PKS) is composed of dissociated enzymes that are utilized iteratively and are responsible for the elongation, cyclization, and modification of the polyketide chain (Figure 2-1)(3, 4, 8, 9). The mechanism that causes this cyclization is not understood in

54

detail. Regiospecific cyclization of linear poly-β-ketone intermediates is a critical step in the biosynthesis of bioactive polyketides. A complete understanding of this process at the molecular level is necessary for future biosynthetic engineering efforts. How and where the poly-β-ketone cyclization takes place had been speculated for decades(10-14). Without such knowledge, the polyketide cyclization pattern cannot be predicted and engineered.

In 2008, we reported the crystal structure of the first ARO/CYC (TcmN ARO/CYC), which is a single-domain protein(15). Based on the structural analysis and mutagenesis results, we proposed that mono-domain ARO/CYCs have an active site and are capable of catalyzing polyketide cyclization and aromatization. Since then, we have further supported this hypothesis with structural and biochemical studies of two other mono-domain ARO/CYCs, WhiE and ZhuI (Figure 2-1)(16, 17). These studies strongly support the role of ARO/CYC as the site of polyketide cyclization. However, a large number of type II PKSs contain di-domain ARO/CYCs that have two seemingly identical domains(18-23). How and why these enzymes require two domains (as opposed to just one) to conduct the cyclization/aromatization is not understood.

The di-domain ARO/CYCs are involved in both non-reducing and reducing PKSs(18, 24). In non-reducing systems, the di-domain ARO/CYCs regiospecifically cyclize a polyketide from C7 to C12 followed by aromatization(18). In reducing systems, a KR first regiospecifically cyclizes the linear poly-β-ketone from C7 to C12, and then reduces the C9 carbonyl to a hydroxyl group(7, 25). A di-domain ARO/CYC then catalyzes the dehydration of the C9 hydroxyl followed by first ring aromatization (Figure 2-1)(14). Therefore, the growing poly-β-ketone intermediate is transported directly from the ketosynthase (KS) to the ARO/CYC in a non-reducing system, while the intermediate needs to be transported from KR to ARO/CYC in a reducing system. Prior to this study, there was no knowledge about whether there are any
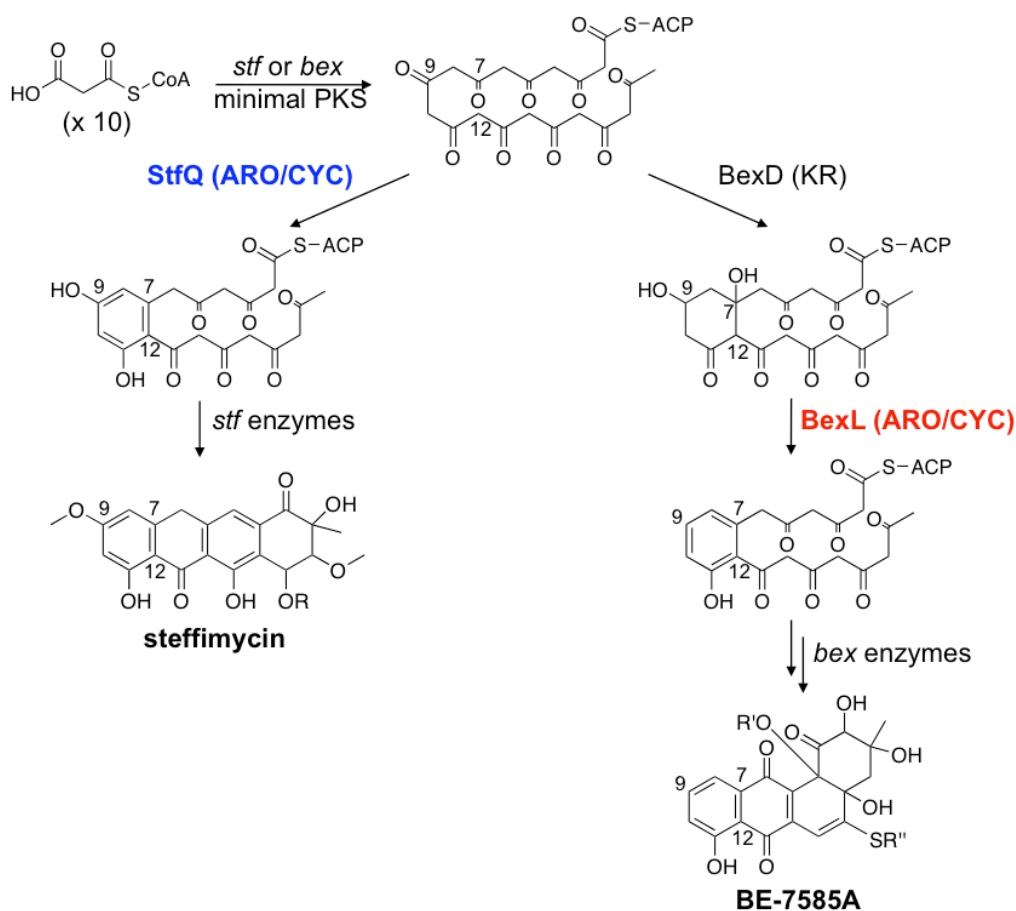
Figure 2-2. The role of ARO/CYCs in reducing and non-reducing type II PKSs. A reactive 20-carbon poly-β-ketone is a common intermediate during both steffimycin and BE-7585A biosynthesis. In the steffimycin pathway, the ARO/CYC StfQ cyclizes and aromatizes the terminal C7-C12 ring. The aromatized intermediate is then processed to yield steffimycin. In contrast, the ARO/CYC BexL aromatizes the terminal C7-C12 ring, which is formed by the KR. The aromatized intermediate is then processed to yield BE-7585A.

differences between the di-domain ARO/CYCs in reducing versus non-reducing PKSs, nor any information on the role that ARO/CYCs may play in the product specificity of reducing and non-reducing PKSs. The key issue is the difficulty in crystallization; multiple groups have vigorously pursued the crystallization of these di-domain ARO/CYCs with at best non-diffracting crystals.

To critically compare the reducing and non-reducing di-domain ARO/CYCs, we have chosen StfQ as a model non-reducing di-domain ARO/CYC and BexL as a model reducing di-

domain ARO/CYC (Figure 2-2). StfQ is a di-domain ARO/CYC from *Streptomyces steffisburgensis* and part of the non-reducing PKS responsible for the biosynthesis of the natural product steffimycin, which has exhibited promising antitumor properties(26-28). StfQ is the enzyme responsible for the C7-C12 first ring cyclization and aromatization of the elongated poly-β-ketone substrate. In contrast, BexL is a di-domain ARO/CYC in the reducing PKS responsible for the biosynthesis of the anticancer agent BE-7585A, which is first reduced at the C9 position by KR and then aromatized by BexL (Figure 2-2)(24). Both StfQ and BexL cyclize 20-carbon poly-β-ketone substrates; therefore, the structural enzymology of these two ARO/CYCs offers a great opportunity for side-by-side comparison and insight into the mechanisms of di-domain ARO/CYCs. Here we present the first structural and biochemical characterization of two di-domain ARO/CYCs, StfQ and BexL using x-ray crystallography, structure-based mutagenesis, *in silico* docking, and *in vitro* functional assays.

2.3 Results and Discussion

2.3.1 Overall Structures

Size exclusion chromatography revealed that BexL is a monomer in solution and StfQ is a dimer in solution (data not shown). This result matches with the oligomeric states found in the crystal structures and may have implications in ARO/CYC-ACP interactions (detailed below). The monomers of BexL and StfQ both contain 309 amino acids that consist of two separate domains, the N-terminal (N-term) and C-terminal (C-term) domains (Figure 2-3 and Figure 2-4A-B). In both enzymes, the N- and C-term domains adopt a helix-grip fold, which is found in
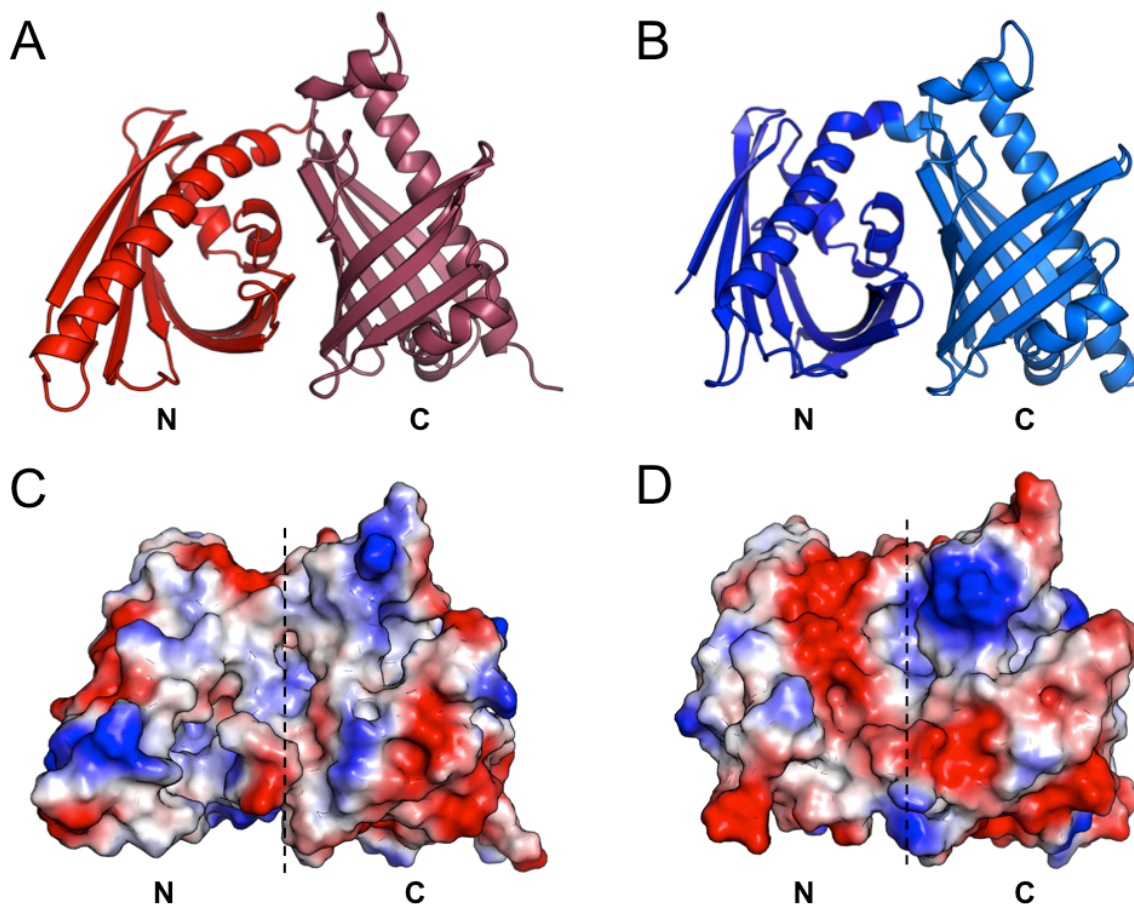
Figure 2-3. A comparison of the tertiary structures and electrostatic surfaces of BexL versus StfQ. (A) Overall structure of BexL shows two distinct domains: N-term (left) and C-term (right). Both domains have a helix-grip fold with a long central α-helix surrounded by a series of β-sheets. (B) Overall structure of StfQ similarly shows two distinct domains: N-term (left) and C-term (right). (C) Electrostatic surface potential map of BexL shows multiple positive patches (blue) that may bind ACP. (D) Electrostatic surface potential map of an StfQ monomer showing complementary positive (blue) and negative (red) patches, which form stabilizing interactions between the two StfQ monomers.

other type II PKS ARO/CYCs(15-17). The BexL N-term consists 146 amino acids, followed by a 4 amino acid linker, and the C-term has 160 amino acids. The StfQ N-term contains 144 amino acids, followed by a 6 amino acid linker, and the C-term has 157 amino acids. The overall sequence identity between BexL and StfQ is 32%. The identity of the BexL N-term (BexL
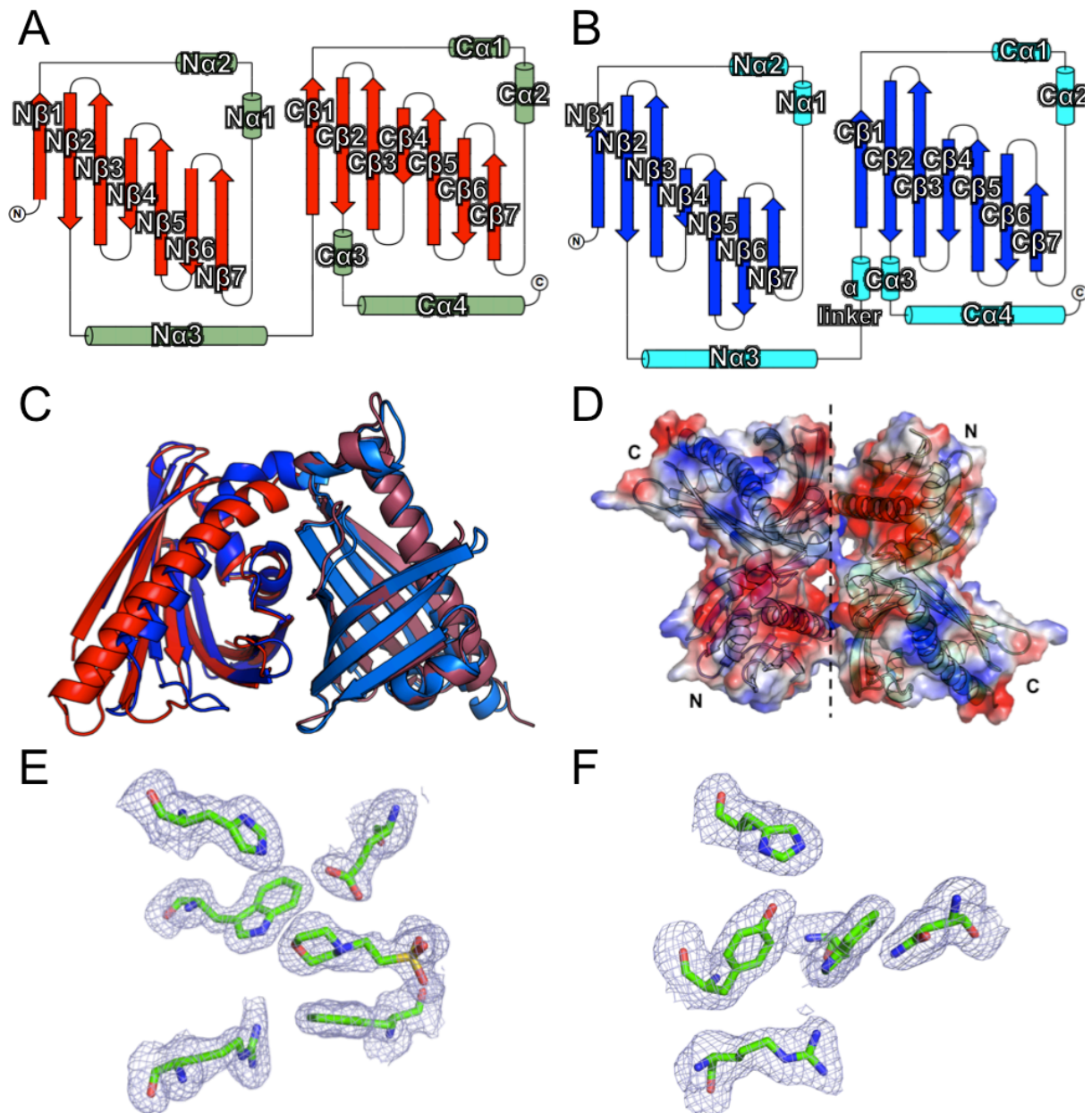
Figure 2-4. The secondary structure maps, structural overlays, and active site electron density maps of BexL and StfQ. (A) Secondary structure cartoon of StfQ. (B) Secondary structure cartoon of BexL. (C) Ribbon diagram overlay of StfQ and BexL. (D) StfQ dimer structure and surface. (E) SA-Omit map of representative BexL active site. (F) SA-Omit map of representative StfQ active site showing clear electron density for all side chains and MES.

residues 1-146) and StfQ N-term (residues 1-146) is 21%, whereas the C-term domains (BexL

residues 147-309, StfQ residues 147-309) are 39% identical (Figure 2-5).

The N-term domains of StfQ and BexL are comprised of seven anti-parallel β-sheets (Nβ1-Nβ7) and three α-helices (Nα1-Nα3) (Figure 2-4A-B). Nα1 and Nα2 are short helices that connect Nβ1 to Nβ7. Nα3 is the longest helix that spans a length of 31 Å in StfQ and 45.3 Å in BexL. Nα3 is cradled by Nβ1-Nβ7 in the helix grip fold, similiar to the secondary structural elements exhibited by the monodomain ARO/CYCs(15-17). In StfQ, Nα3 turns into a six amino acid α-helix linker to the C-terminus. BexL contains a short unstructured linker sequence between the N-term α-helix and the start of the C-term domain.

In both BexL and StfQ, the secondary structure arrangements are very similar for both C-term and N-term domains (Figure 2-4C). In each domain, there are seven anti-parallel β-strands (Cβ1-Cβ7) and four α-helices (Cα1-Cα4). Cα1 and Cα2 connect and span Cβ1-Cβ7. Cα3 is a short α-helix comprised of seven amino acids in StfQ and eight amino acids in BexL that extends from Cβ2 to Cα4. This extra α-helix Cα3 is directly above the C-terminal pocket entrance and positioned at the di-domain interface. Therefore, Cα3 may play a role in ACP binding interactions or regulate substrate access to the C-terminal pocket. Cα4 is a long central α-helix that spans the entire length of the domain; it is 32 amino acids long with a length of 43.5 Å in StfQ and 47 Å in BexL (Figure 2-3A-B).

The two domains of both proteins lay back to back such that the β-sheets interact through hydrophobic interactions. This involves Nβ5 –Nβ7 of the N-terminus and Cβ4-Cβ6 of the C-term domain. In BexL, both the N- and C-term pocket entrances are on the same face of the protein and border the axis that sits along the di-domain interface. Both the N- and C-term pocket entrances of BexL have relatively wide entrances and considerable positive charge, which may facilitate ACP docking (Figure 2-6). In contrast, the StfQ N- and C- term pocket entrances are pointed away from each other and open towards opposite faces of the StfQ dimer. The StfQ N-

A

```
                    1        10         20         30         40         50         60         70
        BexL  .....MREAEHDILVDAPADEVYRLVAEVANWPRIFPPTVFVDHVERGEGTERIRIWATANG.EPKNWTSRRELDPAARR
        StfQ  MPTLGKDTLQYTATVAAPAPLVFDLVAGLENWPQFHGPSVHAEPLGRRDGAEEFQHWVIDDRTVRTWRARWRFDSQALR

                   80        90        100        110        120        130        140        150
        BexL  ISFHQEVSTPPVAEMSGTWIVEPVSAATAKVRLLHAYRAVGDDPGGLAWIDRAVDTNSRSELAAALKHNVELVTNPELTFS
        StfQ  IGYTYDPAEPAAPAQHGEWTFRRLSDASTEVR..VEYELLGDE.GDAALADRELRELVDCVTDAAERHEERRD...LVVD

                  160       170        180        190        200        210        220        230
        BexL  FTDTVRIDAPAKDVYDFVDQAALWAERLPHVSSVDLREPSPGLQVLRMDTRAKDGSVHTTESVRVCFPHHRIVYKQTTLP
        StfQ  FEDPLFVAGAVDDAYTYLYEADKWPERIPHVARLVMEERVPNIQFFDMDTTTPDGSAHTTRSVRVCLPGDKIVYKQIQLP

                  240       250        260        270        280        290        300
        BexL  ALMTLHTGRWDFAEEPGGGTTASSEHTVVLNTANIAKVLGAGAGVAEAREFVRTALSTNSRATLGFAKDHAEARP..
        StfQ  KLLTGHTGHWKFTPTREG.FVLGARHTATIKPSALP.ILGEGTTVLDARKYLRRVLSANSMGNLRLAKAFAEERAGV
```

B

```
                         1        10         20         30         40         50         60         70
        BexL_N_term  .....MREAEHDILVDAPADEVYRLVAEVANWPRIFPPTVFVDHVERGEGTERIRIWATANGEPK.NWTSRRELDPAARR
        BexL_C_term  TNPELTFSFTDTVRIDAPAKDVYDFVDQAALWAERLPHVSSVDLREPSPGLQVLRMDTRAKDGSV.HTTESVRVCFPHHR
        StfQ_N_term  MPTLGKDTLQYTATVAAPAPLVFDLVAGLENWPQFHGPSVHAEPLGRRDGAEEFQHWVIDDRTVRTWRARWRFDSQALR
        StfQ_C_term  ERRDLVVDFEDPLFVAGAVDDAYTYLYEADKWPERIPHVARLVMEERVPNIQFFDMDTTTPDGSA.HTTRSVRVCLPGDK

                        80        90        100        110        120        130        140
        BexL_N_term  ISFHQEVSTPPVAEMSGTWIVEPVSAA........TAKVRLLHAYRAVGDDP...GGLAWIDRAVDTNSRSELAALKHNV
        BexL_C_term  IVYKQTTLPALMTLHTGRWDFAEEPGGGTTASSEHTVVLNTANIAKVLGAGAGVAEAREFVRTALSTNSRATLGFAKDHA
        StfQ_N_term  IGYTYDPAEPAAPAQHGEWTFRRLSDA..........STEVRVEYELLGDEG....DAALADRELRELVDCVTDAAERHE
        StfQ_C_term  IVYKQIQLPKLLTGHTGHWKFTPTREG.FVLGARHTATIKPSALP.ILGEGTTVLDARKYLRRVLSANSMGNLRLAKAFA

        BexL_N_term  ELV...
        BexL_C_term  EARP..
        StfQ_N_term  ......
        StfQ_C_term  EERAGV
```

C

	BexL N-term	BexL C-term	StfQ N-term	StfQ C-term
BexL N-term	-	22.60%	20.55%	11.64%
BexL C-term	22.60%	-	17.12%	38.65%
StfQ N-term	20.55%	17.12%	-	2.74%
StfQ C-term	11.64%	38.65%	2.74%	-

Figure 2-5. Sequence analysis of the BexL and StfQ that analyze N- and C-term domains separately. (A) BexL and StfQ sequence alignments. Key active site residues are highlighted with a black dot. (B) Alignment of the N- and C-term domains of BexL and StfQ. (C) Sequence identity matrix for BexL and StfQ N- and C-term domains. The overall sequence identity between BexL and StfQ is 32%.

term pocket entrance is smaller and has less positive character when compared to the StfQ C-term pocket entrance. The size difference of N- and C-term pocket entrance may have a direct effect on StfQ-ACP interactions, which would favor the ACP to deliver the growing polyketide into the C-term (as opposed to the N-term) for cyclization and aromatization.
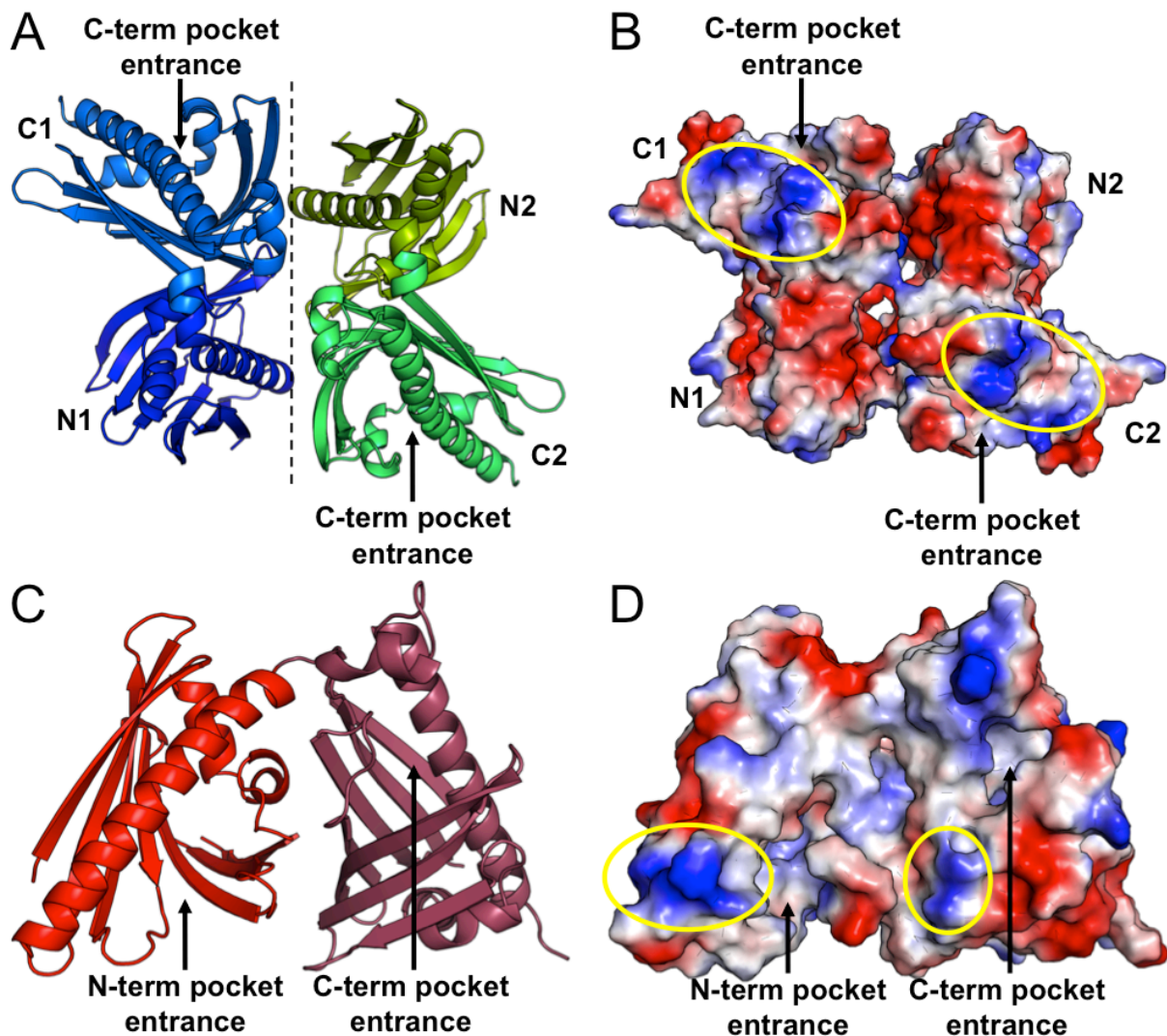
Figure 2-6. The locations of the active site pocket entrances of the StfQ dimer and BexL monomer. (A) The C-term pocket entrances both face the solvent for both StfQ monomers. The N-term pocket of StfQ is located at the bottom of the dimer interface and not easily accessible for substrate binding (not shown). (B) The positive patch (highlighted in yellow), which may facilitate ACP binding, is located near the C-term pocket entrance of StfQ. (C) Both the N- and C-term pocket entrances are located on the same face of the BexL monomer. (D) Multiple small positive patches (highlighted in yellow) are located on the concave surface between the N- and C-term pocket entrances of the BexL monomer.

The oligmoeric states of BexL and StfQ are also different. BexL exists as a monomer with the N- and C-term pockets relatively close to each other, while StfQ exists as a dimer with N- and C-term pockets on opposite side of each monomer (Figure 2-6).
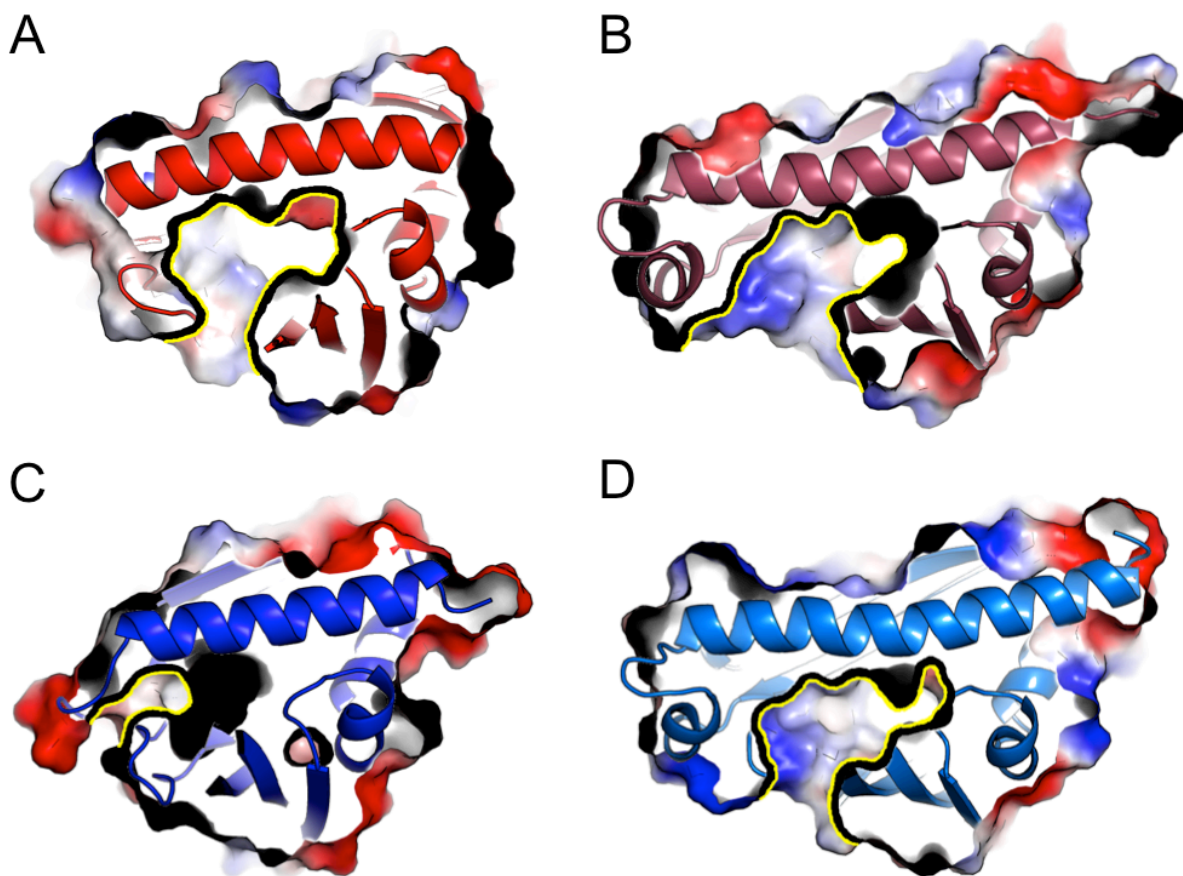
Figure 2-7. Cutaway view of BexL and StfQ N- and C-term domains for a critical comparison of pocket sizes. (A) BexL N-term pocket. (B) BexL C-term pocket. (C) StfQ N-term pocket. (D) StfQ C-term pocket.

2.3.2 StfQ and BexL N-term Comparison

The N-term of BexL has a much larger pocket than that of StfQ (Figure 2-7). The RMSD of N-term between BexL and StfQ is 1.00 Å. The N-term of BexL has a pocket length of 22.2 Å, whereas StfQ has a much shorter pocket length of 15.7 Å (Figure 2-7A and 2-7C). This is reflected in the pocket volume and area: the StfQ N-term pocket only has a volume of 523.7 $Å^3$ and an area of 500.5 $Å^2$, the BexL N-term has a much larger volume of 1722.4 $Å^3$ and an area of 885.7 $Å^2$. When comparing the BexL and StfQ N-terms, the central α-helix (Nα3) has the biggest
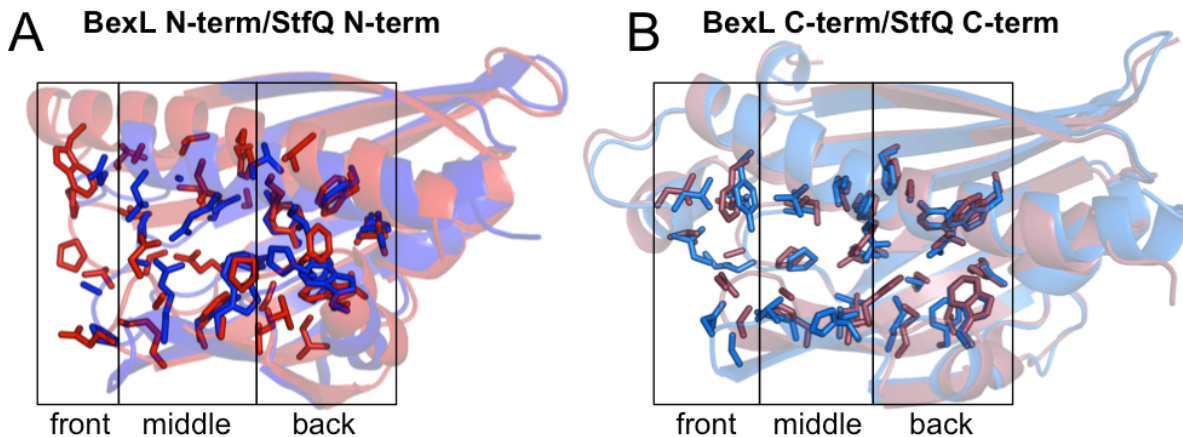
Figure 2-8. Overlay of the BexL N-term with StfQ N-term and BexL C-term with StfQ C-term. (A) Structural overlap of the N-term pockets from back, middle to front. (B) Structural overlap of the C-term pockets from back, middle to front.

shift with a 3.2 Å deviation (Figure 2-4C). This is the major element that accounts for the large difference in pocket volumes (1198.7 Å^3). Further, the loop connecting Nα3 to Nβ2 is longer in StfQ, resulting in a shorter central α-helix. In StfQ, the Nβ2 is puckered inwards, as compared to BexL, and this also pushes Nβ3 more inward towards the pocket and re-positions the Nβ2/Nβ3 loop, which effectively closes the N-term pocket of StfQ.

The N-term domains of BexL and StfQ only have a 21% sequence identity. We analyzed the pocket in detail by breaking down the pocket from the back, middle, and front. At the back of the N-term pocket, there are several bulky hydrophobic residues for both enzymes, but there are notable differences (Figure 2-8). For example, R66 of BexL overlays perfectly with the previously proposed catalytic arginine of the mono-domain ARO/CYCs TcmN, WhiE and ZhuI (Figure 2-9)(15-17). However, the StfQ N-term has a bulky W72 that perfectly aligns with R66 of BexL, resulting in the shrinking pocket size of StfQ N-term pocket. Other differences are F31 of BexL (H36 of StfQ), F77 of BexL (Y83 of StfQ), and H109 of BexL (Y115 of StfQ). The middle of the pocket for both enzymes has a number of polar residues. The only identical residue is W62 in BexL (W68 of StfQ). The front of the N-term pocket for StfQ and BexL is

Figure 2-9. Structural comparison of the N- and C-terminal domains of BexL and StfQ with other monodomain ARO/CYCs. (A) Overlay of the BexL N-term (blue), StfQ N-term (red), BexL C-term (light blue), StfQ C-term (raspberry), ZhuI (cyan), WhiE (green), and TcmN (forest green). (B) A table detailing the overlaid structures above. Residues 1, 2, and 3 correspond to the positions of important active site residues identified for the N-term of BexL, the C-term of StfQ, and the mono-domain ARO/CYCs.

	ARO/CYC class	RMSD (A)	Residue 1	Residue 2	Residue 3	Color
BexL N-term	reducing, C7-C12	0.00	R66	H109	E135	red
BexL C-term	reducing, C7-C17	2.94	R218	H259	T297	raspberry
StfQ N-term	non-reducing, C7-C12	1.01	W72	Y115	L134	blue
StfQ C-term	non-reducing, C7-C12	3.37	R218	H259	N295	light blue
ZhuI	non-reducing, C7-C12	1.51	R66	H109	D146	cyan
WhiE	non-reducing, C9-C14	1.45	R69	Q110	Q136	green
TcmN	non-reducing, C9-C14	1.57	R69	Q110	N136	forest green

flanked by a series of loops consisting of polar and nonpolar residues. In summary, it is within

the N-term domain that we observe the most difference between the two ARO/CYCs, such that

StfQ almost has no N-term pocket, while BexL has an N-term pocket comparable to the interior

pocket of mono-domain ARO/CYCs.

Figure 2-10. Sequence alignment of StfQ with other ARO/CYCs from type II PKS gene clusters that produce polyketides with a C7-C12 cyclization pattern and the lack the reduction at the C9 carbonyl group. Black dots signify the position of active site residues important for enzyme function based on the *in vitro* assay results.

```
                          1        10        20        30        40        50
                          ·        ·         ·         ·         ·         ·
StfQ_steffimycin    ....MPTLGKDTLQYTATVAAPAPLVFDLVAGLENWPQFHGPSVHAEPLG.RRDGAEEFQ
Arm2_arimetamycin   ....MLNLTKETVRHTATVTAPASLVFDLVAGLPDWSQFHGPSVHAEPLG.RRDGAELFR
DacK_dactocycline   ....MLTVDTQVVEDRCAVGAPAPDVAQVLNDVAGWPQLHRPAVHAEVRA.RGEDGDLVQ
MtmQ_mithramycin    ....MPNSQQYSVRDEILVAAPAERVRSLVDDVVGWTQLHGPAVHAHVLR.DTGDERLIR
CmmQ_chromomycin    ......MSQPHLIRHRVTVRAAAPLCFGLMADLDHWPQHFGPAVHAEVLEPDEGDGDLIR
PokC3_polyketomycin MEGTPVVNTQAPVGESVVVDAPAKVVLDLITDPDPLHYLSPSVVHADRTA.RGGAQDSVT


                          60       70        80        90        100       110
                          ·        ·         ·         ·         ·         ·
StfQ_steffimycin    HWWVIDDRTVRTWRARWRFDSQALRIGYTYDPAEPAAPAQHGEWTFRRLSDASTEVRVEY
Arm2_arimetamycin   HWWVIDDHTVRTWPARWRFDRDRLRISYEFDPAEPGAAGSRGTWTFATVSDSSTEVRVEQ
DacK_dactocycline   HWWVTSRTTVRSWLERRTVSDGGRRVDVVHEPAAAPFADVRQQWTVEDDGAGGSVIRVRH
MtmQ_mithramycin    HWAVTGRNSVRTWTARHTRSADASRISFAHVMPEAPFASLEGAWTFRE.RDGKTRVEMRH
CmmQ_chromomycin    RWALTGGTGVHVWDARRRLDRNGLRIVFDQVEPQAPLTGMRAVWSFEADGEGITTVGVTH
PokC3_polyketomycin IWTITDDERVWSRTQDRTLDADGLRVSFEDTP.ESSLAGVRGAWVLRAVSEDKTEIELRQ


                          120      130       140       150       160
                          ·        ·         ·         ·         ·
StfQ_steffimycin    ELLGDEGDAALADR.......ELRELVDCVTDAAERHEERRDLVVDFEDPLFVAGAVDDA
Arm2_arimetamycin   TPTDGDHDADRADR.......ELRQLLDSVVDAAEHDKERRDLVVDFEDPLFVAGAVEDA
DacK_dactocycline   QFTLARPDRAEAARQLELARAELAELLASIRYAAEHRAELADLVVAFEDPLYVAGPVEKA
MtmQ_mithramycin    DFTLSDGSAGPADELTARLREGSRAYLSTLGHAAENHEELDRLIISFTDSLFIAGDVKDA
CmmQ_chromomycin    EFTVTGDDPEAVRRTTDDVNARTPDQLDSLRDLAERWDEVEPLVISFEDPLFIGGRLEDV
PokC3_polyketomycin DTSALTSPEDLARAES.....ARSDLLSTVAYAAANRQELSDLVVDFRDPLFAAGSAEDA


                          170      180       190       200       210       220
                          ·        ·         ·         ·         ·         ·        ●
StfQ_steffimycin    YTYLYEADKWPERIPHVARLVMEERVPNIQFFDMDTTTPDGSAHTTRSVRVCLPGDKIVY
Arm2_arimetamycin   YAYLYEADKWPERIPHVSRLVLEEKVPNIQFFDMDTRTPDGSAHTTRSVRICLPGNKIVY
DacK_dactocycline   YTYLYEADKWPERIPHVVGLKLEQPSPDVQFFDMDTKSADGSAHTTRSVRICLDNYKIVY
MtmQ_mithramycin    YDYLWEADKWEERIPHIAKLDLEEPAPGVQFFDMLTSTPDGSSHRTRSVRVTVEPDIIVY
CmmQ_chromomycin    YEFLYAADAWPERVPHVEKLEMTEDTPGIQFFDMHTHTPDGRDHSTRSVRVCLPSHKIVY
PokC3_polyketomycin YRILYEADRWPERLAHVSRIDMTEDVPGIQFFDMDTTTSDGAPHTTRSVRICFPHSKIVY


                          230      240       250       260       270       280
                          ·     ●  ·         ·         ●         ·         ·
StfQ_steffimycin    KQIQLPKLLTGHTGHWKFTPTREGFVLGARHTATIKPSALPILGEGTTVLDARKYLRRVL
Arm2_arimetamycin   KQIHLPKLLDGHTGHWRFTATREGFILAARHAATIKPSALSILGPETSVLDARKYLRRVL
DacK_dactocycline   KQTQPPKTMTAHTGHWLITPTREGFILSARHTATVRPEGMHLLGEGTTVESARRYLRRVL
MtmQ_mithramycin    KQTELPALLDGHTGHWKFTRTPEGVIAEARHTATVKPSALHILGEGTTVQDARNYLRRVL
CmmQ_chromomycin    KQISLPALLDGHTGHWVFTETPEGVIAHARHTATIKPSALPLLGEGTTVADARKYLRRVL
PokC3_polyketomycin KQIGLPALLDAHTGHWLFTPTREGIVVESRHVAVIKPSALPILGEGTTVEDARRYLRRVL


                          290      300
                          ·    ●   ·
StfQ_steffimycin    SANSMGNLRLAKAFAEERAGV...
Arm2_arimetamycin   SANSMSNLRLAKAYTEERAGV...
DacK_dactocycline   SANSMGNLRLAKAYAEEELGW...
MtmQ_mithramycin    SANSVSNLRLAKAYAEERAGV...
CmmQ_chromomycin    STNSLQTLRYAKQFAEERVHAAHA
PokC3_polyketomycin STNSMTNLRLSKEFAEEQARA...
```

Figure 2-11. Sequence alignment of BexL with other ARO/CYCs from type II PKS gene clusters that produce polyketides with C7-C12 cyclization patterns and the C9 position is reduced/dehydrated. Black dots signify the position of active site residues important for enzyme function based on the *in vitro* assay results.

2.3.3 StfQ and BexL C-term Comparison

Unlike the N-term, the C-term of StfQ and BexL are more similar to each other, with an RMSD of 0.78 Å and 39% sequence identity. The pocket lengths are also quite similar (~24 Å). Consequently, the C-term volume and area are highly similar, with an area of 796.7 Å^2 and a

67
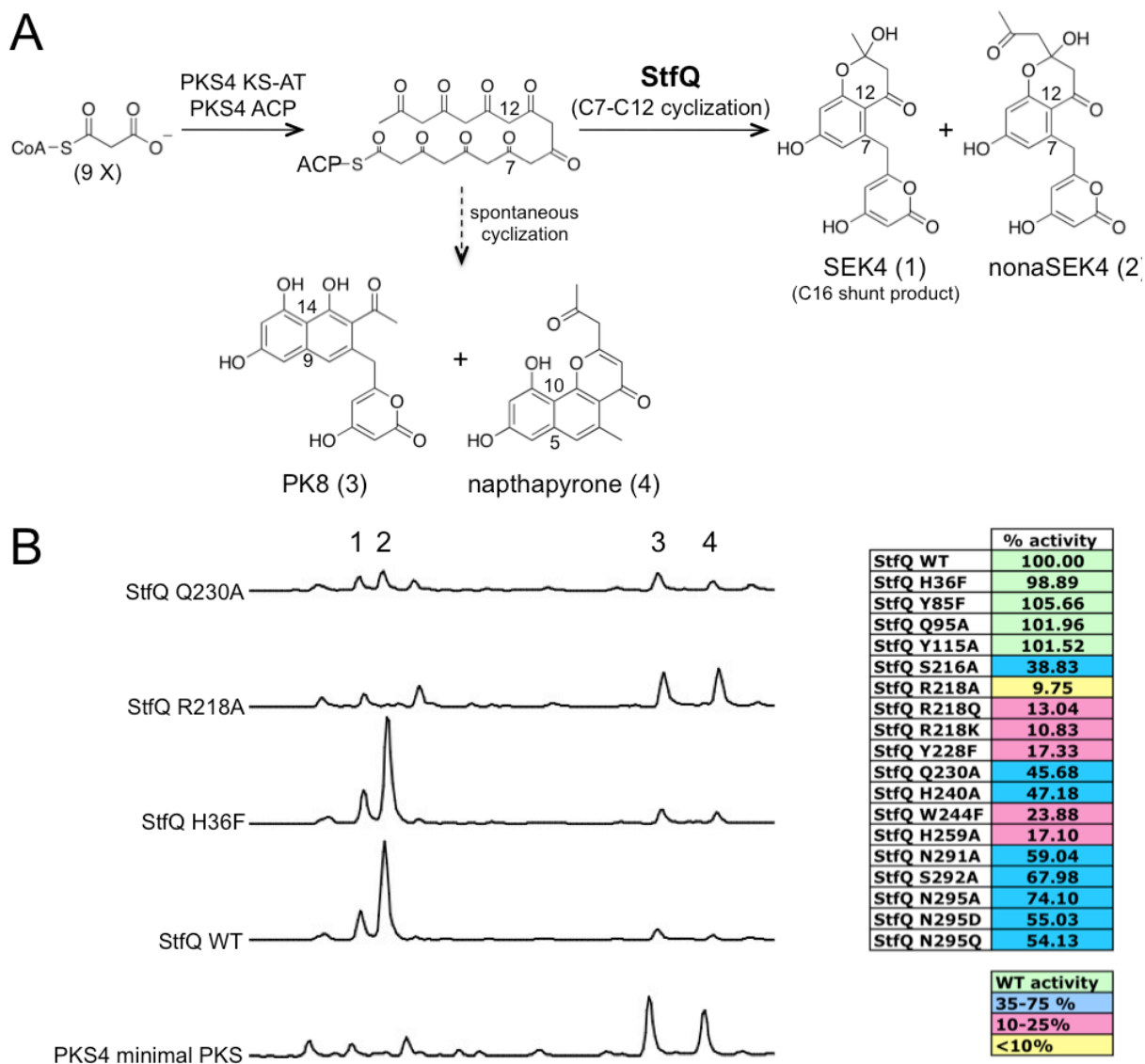
Figure 2-12. The *in vitro* assay results of StfQ. (A) Schematic diagram of the *in vitro* products of the PKS4 minimal PKS assay. (B) Typical HPLC profiles of *in vitro* reactions comparing WT StfQ to different StfQ mutants at 280 nm are displayed. The bottom trace represents the PKS4 minimal PKS, which produces PK8 (**3**), napthapyrone (**4**), and trace amounts of SEK4 (**1**). Upon addition of WT StfQ, the product profile is dominated by SEK4 (**1**) and nonaSEK4 (**2**). The % activity of each mutant StfQ was calculated by the integrated HPLC peak values for the four products (**1**, **2**, **3** and **4**), which were normalized to 100 %. The WT StfQ production of (**1** + **2**) was set as 100 % activity, and the percent activity for each mutant enzyme was calculated as follows (background production of **1** and **2** by the minimal PKS was accounted for): **% activity = (1 + 2)$_{WT}$/(1 + 2)$_{MUT}$ * 100**

volume of 1391.6 $Å^3$ for StfQ, and an area of 813.0 $Å^2$ and a volume of 1339.8 $Å^3$ for BexL (Figure 2-7B and 2-7D).

When the C-term of StfQ and BexL are overlaid, there is very little difference (Figure 2-8B). There is a slight shift in the placement of Cα4. In the reverse of what we observe in the N-term, the Cα4 of StfQ is shifted so that the C-term pocket of StfQ is more open than that of the C-term of pocket of BexL. The Cα3 of BexL C-term is also slightly moved inward when compared to the Cα3 of StfQ, which results in a wider C-term pocket entrance for StfQ.

The sequence alignments showed that the C-term of BexL and StfQ are much more similar than the N-term (Figure 2-5). The major catalytic residues that were identified in the mono-domain ARO/CYC ZhuI (R66, H109, and D146) are similar to residues in the C-term of StfQ (R218, H259, and N295), as well as both the N- (R66, H109, and E135) and C-term (R218, H260, and T292) domains of BexL. When ZhuI, StfQ (C-term), and BexL (N-term) are overlaid, the key residues match perfectly (Figure 2-9). Again, we analyzed the pocket in detail by breaking down the pocket from the back, middle, to the front (Figure 2-8). At the bottom of the pocket, there are only several residues that are different. For example, F200 in StfQ (L200 in BexL) and N295 in StfQ (T297 in BexL). Similarly, in the middle of the pocket, only one residue is different: L237 in StfQ (M237 in BexL). There are many more differences at the entrance of the pocket, which is defined by loop regions that may dictate specific interactions with their respective ACPs. In summary, the C-terms of both ARO/CYCs are highly similar to the C7-C12 mono-domain ZhuI ARO/CYC, which strongly suggests that minimally, for StfQ, the catalytic event may occur in the C-term but not N-term. In contrast, both N- and C-term of BexL may be functional.
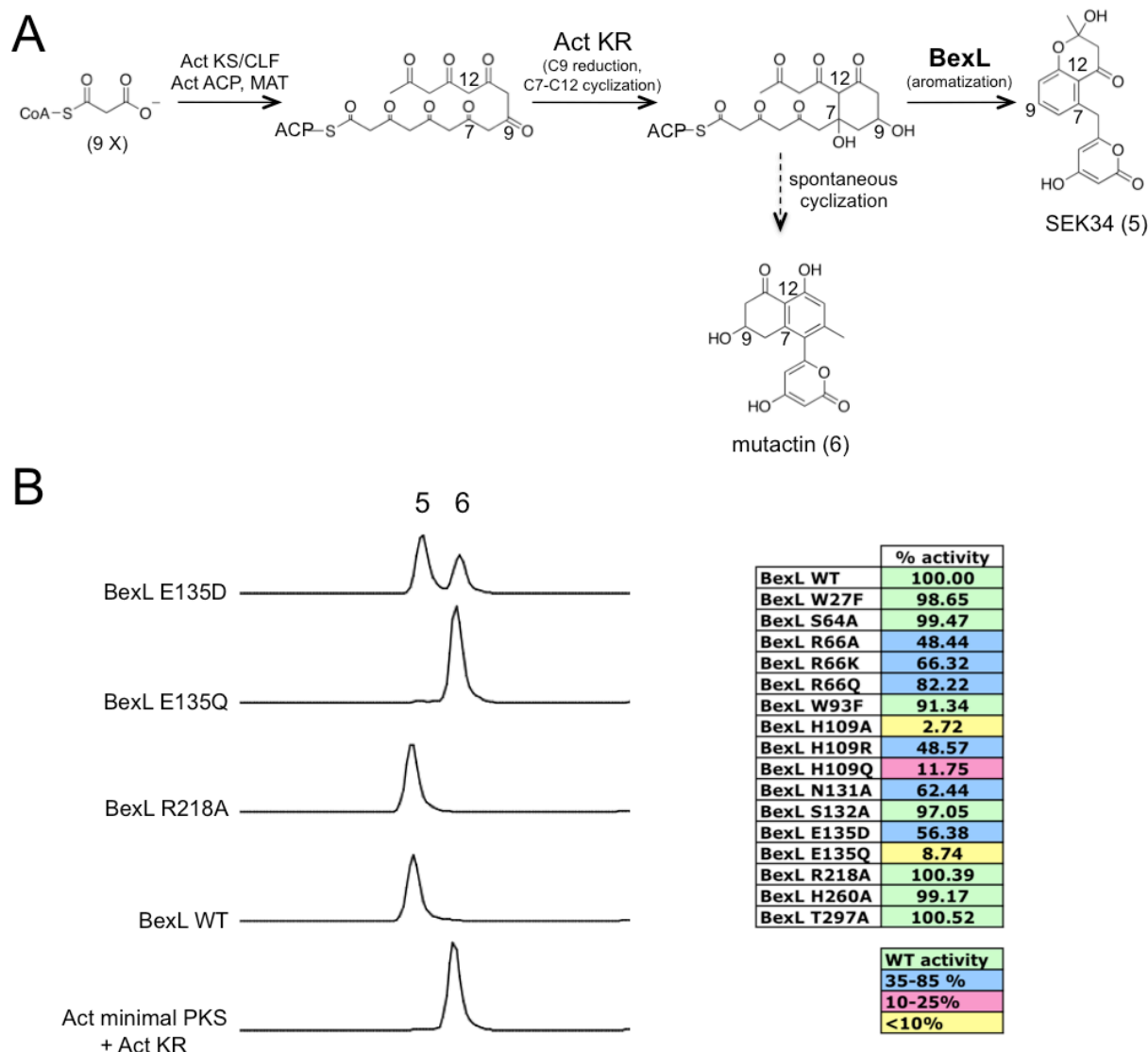
Figure 2-13. The *in vitro* assay results of BexL. (A) Schematic diagram of the *in vitro* products of the Act minimal PKS assay. (B) Typical HPLC profiles of *in vitro* reactions comparing BexL WT to different BexL mutants at 280 nm are displayed. The bottom trace represents the actinorhodin minimal PKS plus ActKR, which produces mutactin (**2**). Upon addition of WT BexL, SEK34 is produced. The % activity of each mutant BexL enzyme was calculated by the integrated HPLC peak values for **1** and **2,** which were normalized to 100 %. The WT BexL production of **1** was set as 100% activity and the percent activity for each mutant enzyme was calculated as follows: **% activity** = (**1**)$_{WT}$/(**1**)$_{MUT}$ * 100

2.3.4 Sequence Comparisons of Reducing and Non-reducing ARO/CYCs

To further explore the similarities and differences between the reducing and non-reducing ARO/CYCs, we have conducted a systematic search of reducing and non-reducing di-domain ARO/CYCs (Figure 2-11 and Figure 2-12). As mentioned above, the N-term of BexL and StfQ share a much lower sequence identity than that of the C-terms. When extending this comparison to the di-domain ARO/CYCs ActVII (reducing) and MtmQ (non-reducing), we observe a similar trend. When comparing ActVII and BexL, the N-term and C-term share 44% and 41% identity, respectively. Therefore, for the reducing di-domain ARO/CYCs, both domains are highly conserved. In contrast, for the di-domain non-reducing ARO/CYCs, the N-term and C-term of StfQ and MtmQ share a 36% and 74% identity, respectively. In summary, in the reducing ARO/CYCs both the N-term and C-term domains are moderately conserved between related enzymes. In contrast, the non-reducing ARO/CYCs display high conservation of the C-term and low conservation of the N-term (Figure 2-10 and Figure 2-11).

2.3.5 StfQ N-term Mutagenesis

To determine residues that are important for catalysis in each domain of StfQ, we conducted extensive mutagenesis of both N-term and C-term of StfQ, and evaluated the activity of StfQ using an assay that utilizes a fungal PKS (PKS4) that creates a linear poly-β-ketone. If StfQ is active, we would observe C7-C12 cyclized product. If the mutation knocks out its activity, we would observe only the shunt products (Figure 2-12, Figure 2-14C and 2-14D). Initially, residues were targeted for mutagenesis based on structural alignment with the ZhuI mono-domain ARO/CYC, which has the same cyclization pattern (C7-C12) as StfQ. The StfQ N-term contains a tryptophan (W72) at the position of the conserved bottom-pocket
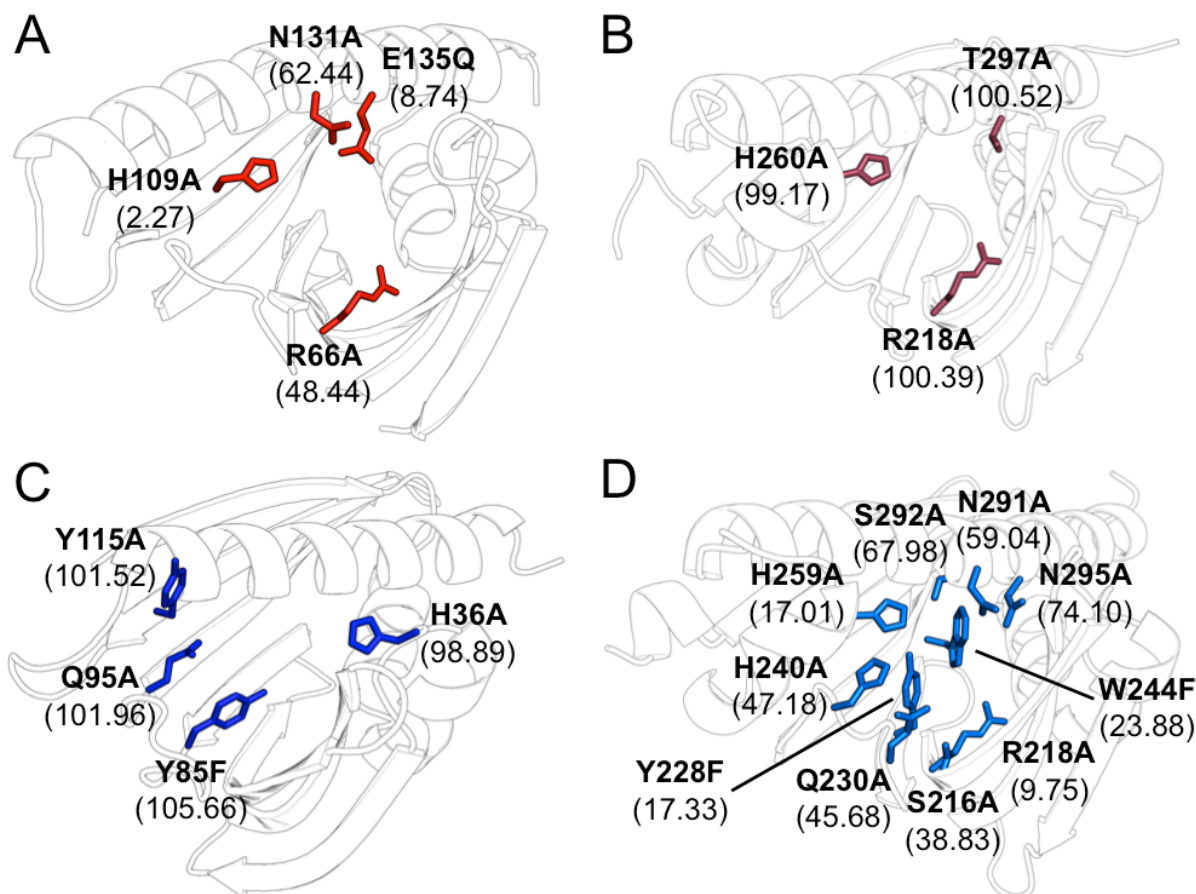
Figure 2-14. *In vitro* assay and mutagenesis results. Each residue is labeled with the corresponding mutation and the % activity compared to WT. (A) BexL N-term pocket mutagenesis reveals that R66, H109, N131, and E135 are important for activity. Mutations of H109 and E135 are especially detrimental. (B) BexL C-term pocket mutagenesis reveals that C-term pocket residues do not contribute to BexL activity. (C) StfQ N-term pocket mutagenesis reveals that N-term pocket residues do not contribute to StfQ activity. (D) StfQ C-term pocket mutagenesis reveals that S216, R218, Y228, Q230, H240, W244, H259, N291, S292, and N295 are all important for StfQ activity. Mutations of R218, H259, and Y228 are especially detrimental.

arginine of known ARO/CYCs. Mutating W72 of StfQ resulted in insoluble inclusion bodies, which suggests that W72 is important for maintaining the overall structure of StfQ. Other putative catalytic N-term residues (H36, Y85, Q95, and Y115) were also mutated, but all of these N-term mutants retain full enzyme activity. In summary, based on mutagenesis, the N-term

domain of StfQ is not important for catalysis. However, some residues (such as W72) may be important for structural integrity of the overall protein folding.

2.3.6 StfQ C-term Mutagenesis

The StfQ C-term contains many hydrophilic residues and structurally aligns well with both ZhuI and the BexL C-term. Fourteen mutants of StfQ C-term pocket residues were generated and assayed for activity (Figure 2-12). The key residues R218, Y228, W244, and H259, as predicted by the alignment with ZhuI, all showed > 70 % reduction in activity, as compared to WT StfQ. R218 occupies a position at the bottom of the StfQ C-term pocket, which is a highly conserved feature of all known ARO/CYCs. R218A led to a 90% reduction in activity and, similar to the ZhuI mutation result, could not be rescued by mutating R218 to lysine or glutamine(16). Therefore, similar to the mono-domain ZhuI, the precise location of R218 at the bottom of the pocket plays a critical role for enzyme activity, and based on the result that R218K cannot rescue its activity, the guanidinium moiety is necessary for cyclization. Therefore, the side chain of R218 may play catalytic role as the active site acid or hydrogen bond donor, similar to the mechanism previously proposed for ZhuI. Because the substrate of StfQ is an unreduced, 20-carbon poly-β-ketone, it may interact with multiple hydrogen bond donors and acceptors inside the C-term pocket, so that the substrate can be correctly oriented for cyclization inside the pocket, followed by R218-promoted C7-C12 aldol cyclization. This hypothesis corroborates our result that mutations of the above pocket residues can significantly affect activity, and is further supported by the C-term mutants Y228F, W230A, H240A, W244F, H259A, N291A, S292A, and N295A, which all show diminished activity (Figure 2-14D). In summary, contrary to the non-catalytic N-term, mutations of the C-term of StfQ significantly affect the cyclization activity in a

manner that is consistent with the cyclization mechanism proposed for that of the mono-domain C7-C12 ZhuI ARO/CYC(16).

2.3.7 BexL N-term Mutagenesis

Extensive mutagenesis coupled to an actinorhodin minimal PKS assay was conducted on both the N- and C-term of BexL based on structural alignment with ZhuI and the C-term domain of StfQ (Figure 2-13). The BexL N-term contains R66, H109, and E135, which are critical for activity, as determined by mutagenesis results (Figure 2-14A). However, R66A still retains 44% activity, which can be rescued to 66% and 82% activity with a lysine and glutamine mutation, respectively. E135A was generated, but it resulted in inclusion bodies during protein expression. Therefore, E135D and E135Q were generated, and both expressed as soluble proteins. E135Q has little activity (8.7%), but E135D rescues the activity back to 56.4%, which suggests that an acidic residue at this position is important for the cyclization activity. This matches well with ZhuI, which has an aspartic acid instead of glutamic acid, which was found to be important for activity(16). N131A also retains 60% activity, suggesting that N131 may be involved partially during the cyclization catalyzed by BexL. In summary, contrary to StfQ, mutations of the BexL N-term pocket do affect the enzyme activity.

2.3.8 BexL C-term Mutagenesis

The BexL C-term was subjected to mutagenesis based on structural alignment with ZhuI and the StfQ C-term. In sharp contrast to the BexL N-term mutagenesis results, R218A, H260A, and T297A retain full activity comparable to BexL WT (Figure 2-13 and Figure 2-14B). This result suggests that the C-term of BexL is not involved in catalysis. In summary, based on the
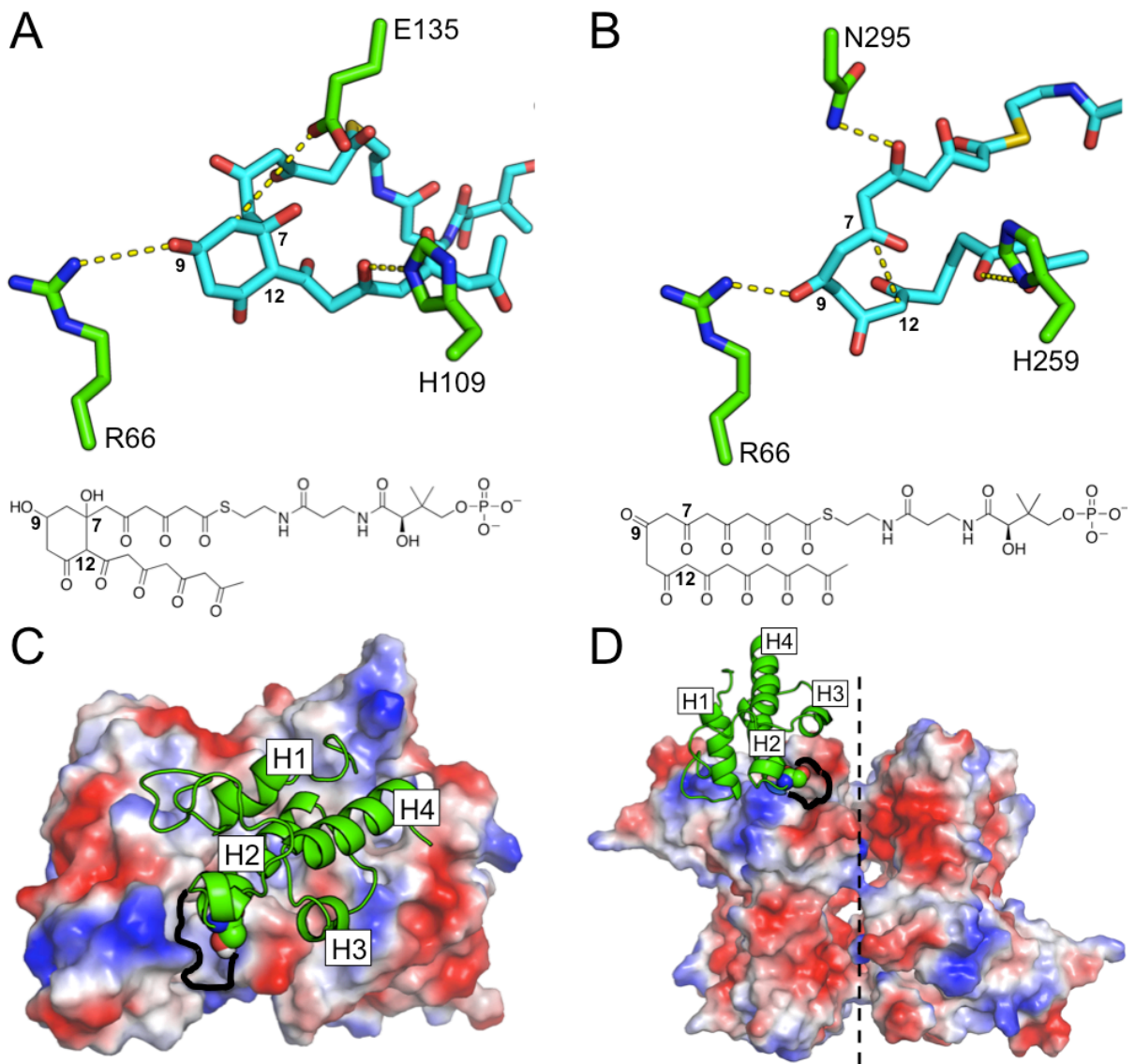
Figure 2-15. BexL and StfQ docking simulations. (A) BexL N-term with dodecaketide-pPant. (B) StfQ C-term with dodecaketide-pPant. (C) StfQ with ACP. (D) BexL with ACP. *The active site pocket entrance is outlines in black.

mutagenesis results, although StfQ and BexL both have two domains with similar folds, the C-term pocket of StfQ is catalytic but not the N-term, while the opposite is true for BexL. This significant difference reflects the fact that one aromatizes a reduced substrate, while the other cyclizes an unreduced substrate.

2.3.9 BexL Docking With Bex ACP

A homology model of the ACP from the Bex gene cluster was generated using HHPred and docked to the surface of the BexL monomer using the Hex Protein Docking server(29, 30). The Bex ACP could be docked to the BexL monomer in many different orientations, but only a small number of these docking solutions appear to be biologically relevant. During catalysis, the location of the pPant-bound serine has to be close to the entrance of the N- or C-term pockets in order to deliver the poly-β-ketone into the active site (Figure 2-15C). In this docking mode, helix 2 of ACP, which has been implicated in ACP-protein interactions for the *E. coli* FAS ACP, docks between the central α-helix and loops between Nβ1 and Nα2, and Nβ3 and Nβ4 of the N-terminal pocket entrance (31, 32). The Bex ACP model contains many negatively charged surface residues and the central α-helix in the BexL N-term has two surface arginines (R216 and R133), which can form charge-charge interactions with the Bex ACP. There are multiple positively charged residues (R283, R287, R295) on the BexL C-term central α-helix, but these residues are opposite to the C-term pocket entrance, thus preventing ACP from docking the C-term pocket. Therefore, the docking result corroborates the mutagenesis results, in that the Bex ACP docks to the N-term of BexL, where the aromatization of the first ring occurs. Moreover, our docking results prove that the BexL N-term is capable of docking the incoming Bex ACP in a biologically relevant fashion. In summary, consistent with the mutation results, docking simulation supports that the BexL N-term contains a positive patch, which can facilitate ACP binding and direct a poly-β-ketone intermediate into the active site for catalysis.

2.3.10 StfQ Docking With Stf ACP

A homology model of the ACP from the Stf gene cluster was generated using HHPred and docking was conducted using the Hex Protein Docking server(29, 30). Automated docking of the ACP the StfQ surface proved difficult, and there were no solutions where the location of the pPant-bound serine was close to the entrance of the C-term active site. Therefore, the Stf ACP was manually placed to help visualize the potential ACP binding site near the C-term pocket of StfQ (Figure 2-15D). In the StfQ dimer, the StfQ N-term pockets are facing inward towards the dimer interface, which prohibits the ACP from binding near their entrance. Therefore, docking of the Stf ACP near the N-term pocket entrance does not reveal any useful insights. In contrast, the StfQ C-term pocket is solvent accessible, and there are two argnines (R286 and R297) in the central α-helix, which can interact with the negatively charged surface of ACP. The Stf ACP can be docked to StfQ in a similar orientation as the Bex ACP with BexL, where helix 2 of ACP interacts with the central α-helix and loop regions between β-sheets. In summary, the docking simulation of Stf ACP and StfQ also agrees with the mutagenesis results, in that the Stf ACP docks to the C-term contains a positive patch, which can facilitate ACP binding and direct a poly-β-ketone intermediate into the active site for catalysis. This is the exact opposite of Bex scenario, where the ACP docks to the N-term active site entrance.

2.3.11 Docking of a 20-carbon Unreduced Poly-β-ketone With StfQ

A 20-carbon pPant-linked unreduced poly-β-ketone was docked in the N-term and C-term of StfQ to investigate if the internal pockets could accommodate the proposed substrate. Docking the StfQ substrate also reveals potential interactions between the substrate and proposed active site residues in the C-term. The 20-carbon intermediate could not be docked in the StfQ
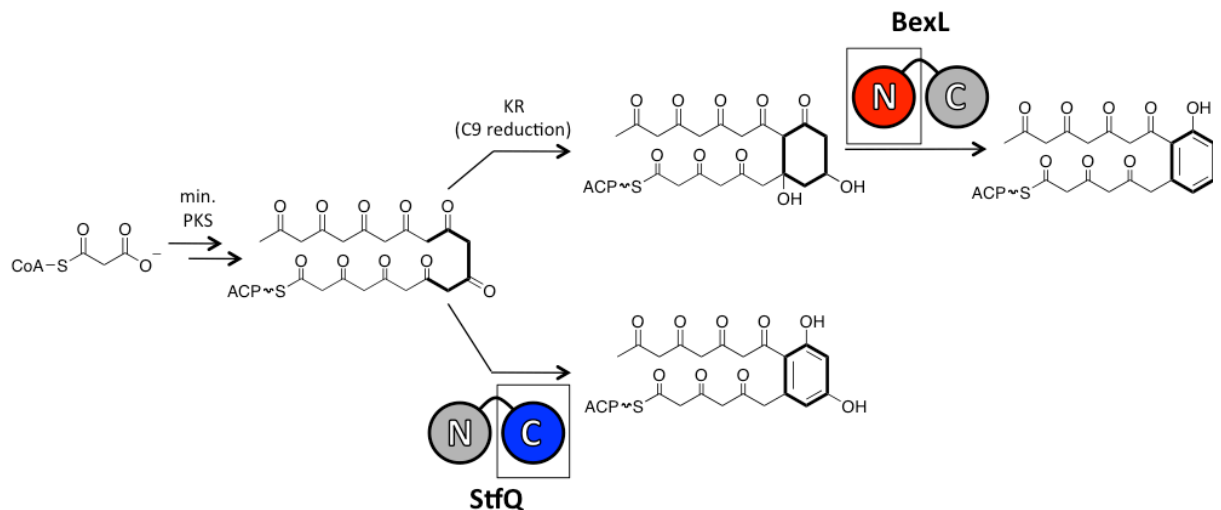
Figure 2-16. Proposed mechanism for reducing versus non-reducing di-domain ARO/CYCs. During type II PKS biosynthesis, the ACP, KS-CLF, and MAT use malonyl-CoA to generate an unreduced polyketide intermediate. In non-reducing PKSs, ketoreductase (KR) first cyclizes the unreduced polyketide with C7-C12 cyclization specificity followed my reduction of the C9 carbonyl to a hydroxyl. BexL acts on the C7-C12 cyclized, C9-reduced intermediate to catalyze aromatization of the C7-C12 cyclized ring. The N-terminal domain of BexL is responsible for catalysis. In contrast, StfQ acts directly on the unreduced polyketide intermediate to yield a C7-C12 cyclized and aromatized product. The C-terminal domain of StfQ is responsible for catalysis.

N-term due to the small pocket dimensions observed in the crystal structure (docking results not shown). It is possible that the StfQ N-term could adopt a different conformation in solution or upon substrate binding, which could bind a 20-carbon intermediate. Nevertheless, even if this is the case, our mutagenesis results suggest that the StfQ N-term pocket is not involved in catalysis. In contrast, the StfQ C-term can accommodate a 20-carbon pPant-linked intermediate, and docking solutions revealed a myriad of solutions in different conformations. Of these solutions, a subset positioned the C7 and C12 carbons within 5 Å of each other. At the same time, the C9 carbonyl group is within hydrogen bonding distance of R218 (Figure 2-15B). Although we

identified many active site residues that are important for StfQ activity *in vitro*, it is difficult to assign specific roles such as general acid or base for the promotion of the C7-C12 aldol cyclization. Residues that were found to be important for activity such as R218, H260, and N295 are located very far away from each other in the crystal structure. Therefore, when docked with the substrate, their specific roles can either be catalytic, or promote a proper orientation of the poly-β-ketone that for the C7-C12 aldol cyclization. Alternatively, the StfQ C-term may adopt a more closed conformation upon ACP/substrate binding, which may position these active site residues differently than in the crystal structure. This may account for their activity as identified by mutagenesis. Another possibility is that the putative active site residues simply act as a "mold" to help draw in and position the 20-carbon intermediate, which then spontaneously cyclizes between C7 and C12, with water serving as the general acid/base. This is an attractive hypothesis, because multiple docking solutions anchor the C9 carbonyl group near R218. This kinks the poly-β-ketone and forces C7 and C12 within 4 Å of each other, which then allows spontaneous aldol cyclization. Therefore, it is reasonable to propose that StfQ (and other non-reducing ARO/CYCs) may simply fold the poly-β-ketone using the arginine at the bottom of its pocket (R218) as an anchoring point to facilitate the proper cyclization pattern. Additional pocket residues may play a role in accelerating the kinetics of the ARO/CYC mediated aldol cyclization by providing protons, interacting with waters, and facilitating the binding and transfer of the poly-β-ketone intermediate in and out of the ARO/CYC pocket.

2.3.12 Docking of a 20-carbon C9-reduced C7-C12 Cyclized Poly-β-ketone With BexL

To investigate if the internal pockets of BexL could accommodate the proposed substrate, we docked a 20-carbon pPant-linked, C9-reduced, C7-C12 cyclized polyketide intermediate into

the N-term and C-term of BexL. The 20-carbon intermediate could be docked in both the BexL N-term and C-term due to the large pocket dimensions observed in the crystal structure. Although both pockets can accommodate the 20-carbon intermediate, mutagenesis results suggest that the BexL N-term is solely responsible for catalysis. Therefore, only BexL N-term docking results will be discussed in detail. The BexL C-term may be involved in stabilizing a growing poly-β-ketone intermediate, but our mutagenesis results do not support this hypothesis.

Similar to the StfQ C-term docking results, many solutions in multiple different conformations were identified for BexL N-term docking. Of these solutions, a subset of solutions positioned the C9 hydroxyl group in hydrogen bonding distance to R66, which was found to be necessary for activity (Figure 2-15A). Similar to StfQ, the existence of multiple solutions suggested that the active site residues such as H109, E135, and R66 may have multiple roles, ranging from general acid or general base for promoting aromatization of C7-C12 ring (such as R66), or they may simply orient the incoming substrate for the aromatization to spontaneously occur (such as H109 and E135). Additionally, the BexL N-term may adopt a more closed conformation upon substrate-ACP binding. In summary, the active site of BexL and StfQ are quite similar, and both bind C7-C12 cyclized intermediates. Therefore, the bottom pocket arginine of both enzymes may be important for anchoring intermediates in such a way that promotes either cyclization/aromatization (for StfQ) or dehydration/aromatization (for BexL).

2.3.13 Functional Considerations of BexL and StfQ in Polyketide Biosynthesis

StfQ and BexL represent di-domain ARO/CYCs from two different type II polyketide synthases. Each of these PKSs is responsible for the synthesis of aromatic polyketides with the same C7-C12 first ring cyclization pattern, but there is a key difference: the presence (Bex) or
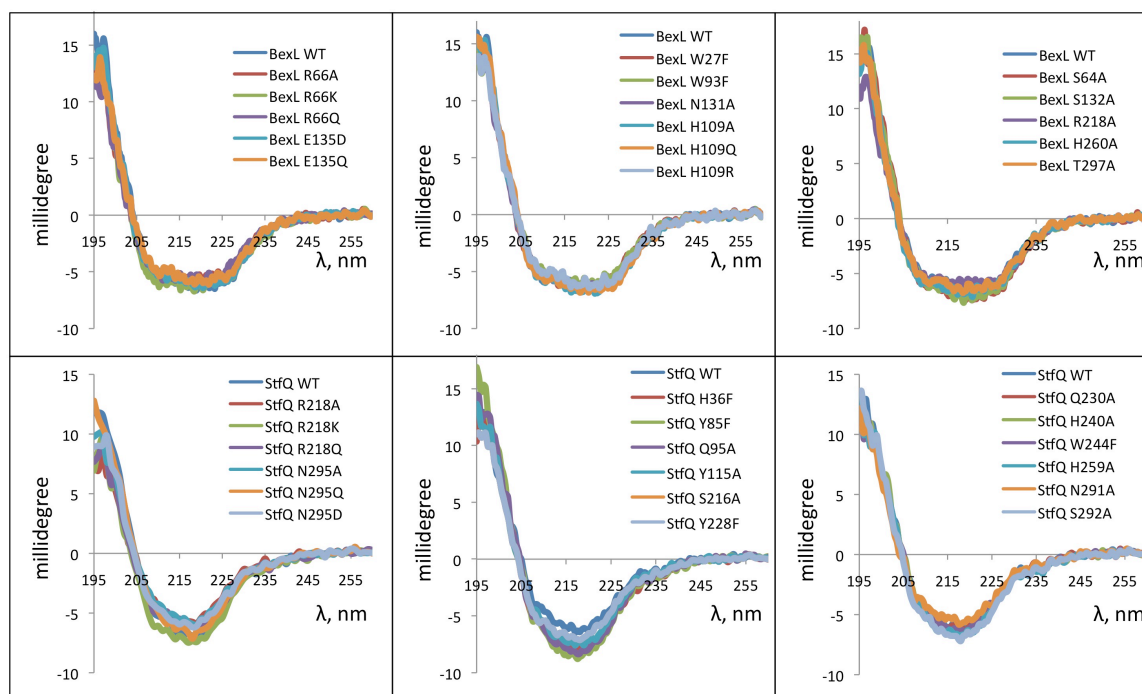
Figure 2-17. Circular dichroism spectra of BexL and StfQ mutants and wild type enzymes, showing all enzymes are properly folded. Samples were collected on a JASCO J-810 Circular Dichroism Spectrapolarimeter. All samples were prepared by diluting protein into 25 mM tris (pH 7.5) to a final concentration of 0.1 - 0.2 mg/ml.

absence (Stf) of a NADPH-dependent KR. Based on previous results(7, 25), we hypothesize that for the reducing PKS (Bex), the KR first cyclizes the elongated polyketide chain between C7 and C12, then reduces the C9 carbonyl to a hydroxyl group. Upon aromatization by BexL, the hydroxyl group is eliminated as water. In contrast, for the non-reducing PKS (Stf), which lacks a KR, StfQ receives the unreduced, linear poly-β-ketone directly after elongation, and proceeds to cyclize and aromatize the first ring specifically between C7 and C12 (Figure 2-16).

The different roles of StfQ versus BexL, as discussed above, may be related to their corresponding oligomeric state. During BE-7585A biosynthesis, the ACP is constantly shuttling between the KS-CLF for chain elongation and the KR for cyclization and reduction. BexL is competing with KR to receive the poly-β-ketone intermediate. Therefore, BexL needs to be

	BexL	StfQ	StfQ (SeMet)		
	Native	Native	Peak	Inflection	Remote
A. Crystallization	0.1 M MES pH 7, 15 % PEG 20000	0.15 M Sodium Formate, 20% PEG3350			
B. Crystallographic Data					
Wavelength (Å)	1.0246	0.99997	0.979	0.979	0.918
Space Group	$P4_12_12$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell Dimension (Å)	63.68 63.68 170.371	59.98 92.43 138.53	58.03 92.47 135.77	58.02 92.45 135.73	58.05 92.49 135.81
	$\alpha=\beta=\gamma=90°$	$\alpha=\beta=\gamma=90°$	$\alpha=\beta=\gamma=90°$	$\alpha=\beta=\gamma=90°$	$\alpha=\beta=\gamma=90°$
Resolution (Å)	50.00 - 1.46	50.00 - 1.91	50.00 - 2.76	50.00 - 2.76	50.00 - 2.76
No. of observations	193711	411186	532454	21070	304586
No. of unique observations	34205	56825	19762	14056	21025
Completeness % (last shell)	99.70 (99.22)	99.73 (98.34)	99.7(96.9)	99.6 (96.3)	99.8 (98.3)
$I/\sigma(I)$ (last shell)	20.1 (4.1)	16.38 (2.59)	9.6 (2.5)	24.2 (3.5)	10.4 (1.1)
R_{merge} %	5.9 (38.6)	7.9 (79.3)	23.2 (47.5)	12.3 (46.7)	21.5 (44.5)
Redundancy	4.1	3.8	14.6	14.6	14.6
C. Refinement					
Resolution (Å)	27.01-1.788 (1.852-1.788)	47.29-1.952 (2.022-1.952)	36.15-2.081 (2.155-2.081)		
No. of protein atoms	2700	5454	5214		
No. of cofactor atoms	25	6	0		
No. of water atoms	307	637	361		
Rfree %	0.2193 (0.2405)	0.2093 (0.2753)	0.2244 (0.3273)		
Rcrys %	0.1803 (0.2328)	0.1780 (0.2360)	0.1973 (0.2699)		
D. Geometry					
RMS bonds (Å)	0.01	0.007	0.01		
RMS angles (°)	1.33	1.05	1.19		
Ramachandran Favored (%)	99.7	97	98		
Ramachandran Generously Allowed (%)	0.3	3	2		
Ramachandran Dissallowed (%)	0	0	0		
B-factors					
Protein	31.5	31.8	25.1		
Water	37.3	47.7	0		
Ligand/ion	39.3	38.9	29.9		
Se FOM			0.4025		

Table 2-1. StfQ and BexL data collection and refinement statistics.
*Single crystals were used to collect each corresponding monochromatic or SeMet multiwavelength data set.
†Values in parentheses represent the highest resolution shell.

mobile as a monomer to compete for the substrate. In contrast, in steffimycin biosynthesis, the ACP is only shuttling between the KS-CLF and StfQ. As a dimer, StfQ can offer a larger protein surface to interact with both ACP and KS/CLF, as well as a larger protein surface for protecting the unstable poly-β-ketone intermediate that is passed from the ACP to StfQ. In summary, this study unveils the surface structures and oligomeric states of di-domain ARO/CYCs from both a reducing and non-reducing system, highlighting key differences.

2.3.14 Functional Roles of the BexL N- and C-terminal Domain Pockets

Both BexL and StfQ have two domains, and each domain has one interior pocket. In BexL, the functional role of each pocket was probed by site-directed mutagenesis. The N-term pocket contains essential catalytic residues but not the C-term pocket. Although the C-term pocket contains the putative active site residues (R218, H260, and T297) based on homology to other ARO/CYCs, mutations of these residues to alanine had no effect on enzyme activity. These results suggest that the N-terminal domain alone of BexL is responsible for dehydration/aromatization, and the C-terminal domain may be important for overall enzyme stability or protein-protein interactions with the ACP or other minimal PKS enzymes. Additionally, we were not able to express the BexL N-term domain as a soluble protein (data not show), which further supports the hypothesis that the C-term domain may aid in overall protein stability.

2.3.15 Functional Roles of the StfQ N- and C-terminal Domain Pockets

The StfQ N-term pocket contains many hydrophobic residues whose mutations resulted in very low expression and insoluble aggregates. This suggests that in contrast to BexL, residues

in the StfQ N-term are important for overall protein stability. Docking studies suggest that the StfQ N-term pocket is not large enough to accept a linear 20-carbon poly-β-ketone. However, a different conformation where the pocket volume increases may allow the binding of a cyclized intermediate. If this is the case, the hydrophobic nature of the StfQ C-term pocket may promote aromatization of a cyclized substrate. Docking studies show that the StfQ C-term domain can accept the full-length 20-carbon poly-β-ketone and allows the chain to fold appropriately for a C7-C12 cyclization. Mutational studies also strongly support that R218, H259, N295, and other hydrophilic residues are important for the cyclization of the poly-β-ketone. In summary, in contrast to BexL, residues in the StfQ N-term pocket have been determined to be structurally important, whereas the C-term domain is responsible for catalyzing the first ring C7-C12 cyclization (Figure 2-16).

2.4 Conclusion

This work allows greater understanding of how complex PKSs function and produce a diversity of natural products. The knowledge of how the individual enzymes work allows for the engineering new natural products by controlling functionality and cyclization patterns. Biosynthetic control at the enzyme level allows for the creation of potentially new pharmaceuticals and bioactive molecules.

The structures and functions of these two di-domains have opened the doors to further understand the role of the di-domains within their respective PKSs. BexL is the first model of a di-domain ARO/CYC involved in reducing PKSs that produce polyketides such as doxorubicin (anticancer), actinorhodin (antibiotic), and griseusin (anticancer)(19, 33, 34). StfQ has given us the first insight into di-domains involved in the non-reducing PKSs that biosynthesize

StfQ H36F	5' - attcTTCggcccctcggtgcatgccgaaccgc - 3'	BexL R218A	5' - gtcggtcGCGgtgtgcttcccgcaccacc - 3'
	5' - ggccGAAgaatt gaggccagtt ctccagtccg gcc - 3'		5' - gcacacCGCgaccgactcggtggtgtgcac - 3'
StfQ Q95A	5' - cgcaGCGcacggcgagtggaccttccgccgcc - 3'	BexL H260A	5' - ggagGCCaccgtcgtgctcaacaccgcgaa - 3'
	5' - cgtgCGCtgcgggcgccgccggttcgg - 3'		5' - cggtGGCctccgacgacgccgtcg - 3'
StfQ Y85F	5' - tacaccTTCgacccggccgaaccgg - 3'	BexL S64A	5' – ggaccGCGcgccgcgaactcg – 3'
	5' - gggtcGAAggtgtagccgatccgcagc - 3'		5' – cggcgCGCggtccagttcttggg – 3'
StfQ Y115A	5' - ggagGCCgagctcctcggcgacgaaggtgacg - 3'	BexL T297A	5' - ggcgGCGctcggcttcgcgaaggacc - 3'
	5' - gctcGGCctccacccgcacttcggtggacg - 3'		5' - cgagCGCcgcccggctgttggtgc - 3'
StfQ H259A	5' - cacggGCGaccgcgacgatcaagccctcg - 3'	BexL R66A	5' - gcgcGCCgaactcgacccgg - 3'
	5' - cggtCGCccgtgcgccgagcacgaatc - 3'		5' - gagttcGGCgcgcgaggtccagttc - 3'
StfQ W244F	5' - gtcacTTCaagttcacgccgacgcgcgag - 3'	BexL R66Q	5' - gcgcCAGgaactcgacccggc - 3'
	5' - gaacttGAAgtgaccggtgtgccctgtcagc - 3'		5' - gagttcCTGgcgcgaggtccagttcttgg - 3'
StfQ R218A	5' - ggtcGCCgtctgcctgcccggcgacaagatc - 3'	BexL R66K	5' - gcgcAAAgaactcgacccggcgcgg - 3'
	5' - cagacGCGgaccgaccgggtggtgtgcg - 3'		5' - gagttcTTTgcgcgaggtccagttcttgggc - 3'
StfQ R218Q	5' - ggtcCAGgtctgcctgcccggcgacaagatcgtg - 3'	BexL H109A	5' - cctgGCCgcctaccgcgcg - 3'
	5' - gcagacCTGgaccgaccgggtggtgtgcgc - 3'		5' - ggcGGCcaggagccgcaccttc - 3'
StfQ R218K	5' - ggtcAAAgtctgcctgcccggcgacaagatcg - 3'	BexL H109R	5' - cctgCGCgcctaccgcgcg - 3'
	5' - gcagacTTTgaccgaccgggtggtgtgcgc - 3'		5' - aggcGCGcaggagccgcacct - 3'
StfQ N295A	5' - gggcGCCctgcggctggcgaag - 3'	BexL H109Q	5' - cctgCAGgcctaccgcgcggtc - 3'
	5' - cgcagGGCgcccatgctgttcgcac - 3'		5' - aggcCTGcaggagccgcaccttcg - 3'
StfQ N295D	5' - gggcGACctgcggctggcgaaggcgttc - 3'	BexL E135A	5' - gtcgGCActggccgcgctcaag - 3'
	5' - cgcagGTCgcccatgctgttcgcactgagcac - 3'		5' - ccagTGCcgaccggctgttcgtg - 3'
StfQ N295Q	5' - gggcCAGctgcggctggcgaagg - 3'	BexL E135D	5' - gtcgGATctggccgcgctcaagc - 3'
	5' - cgcagCTGgcccatgctgttcgcactg - 3'		5' - ccagATCcgaccggctgttcgtgtc - 3'
StfQ H240A	5' - cagggGCCaccggtcactggaagttcacg - 3'	BexL E135Q	5' - gtcgCAActggccgcgctcaagc - 3'
	5' - ccggtGGCccctgtcagcagcttcg - 3'		5' - ccagTTGcgaccggctgttcgtgtc - 3'
StfQ Q230A	5' - acaagGCGatccagctgccgaagctgctg - 3'	BexL N131A	5' – acacgGCCagccggtcggaactggc – 3'
	5' - ctggatCGCcttgtacacgatcttgtcgccgg - 3'		5' – cggctGGCcgtgtcgaccgcc – 3'
StfQ Y228F	5' - tcgtgGCCaagcagatccagctgccgaag - 3'		
	5' - tgcttGGCcacgatcttgtcgccggg - 3'		
StfQ N291A	5' - gtgcgGCCagcatgggcaacctgc - 3'		
	5' - atgctGGCcgcactgagcacccgg - 3'		
StfQ S216A	5' - cccggGCGgtccgcgtctgcct - 3'		
	5' - cggacCGCccgggtggtgtgc - 3'		
StfQ S292A	5' - cgaacGCCatgggcaacctgcggc - 3'		
	5' - cccatGGCgttcgcactgagcacccg - 3'		

Table 2-2. BexL and StfQ mutagensis primers.

polyketides such as steffimycin (anticancer) and mithramycin (anticancer)(27, 35). StfQ mutagenesis has revealed that the residues important for the ZhuI mechanism are just as important for StfQ (in the C-term). This further supports the hypothesis that ZhuI may be a bridge between the mono-domain and the di-domain ARO/CYCs. In the reducing system, we have revealed the surprising result that the N-term domain of BexL, instead of C-term in the non-

reducing system, acts as dehydratase/aromatase catalyzing a critical reaction in the generation of bioactive polyketides.

This is the first molecular view of any di-domain ARO/CYC from either a reducing or non-reducing type II PKS system, and here we present examples of both, displaying significant differences between these two classes of ARO/CYCs. We can combine this information with what we already know about mono-domain ARO/CYCs, and we can extend this knowledge to engineer these enzymes for alternative cyclization patterns.

2.5 Materials and Methods

2.5.1 Expression and Purification of BexL and StfQ

A pET-28b(+) (Novagen) derived DNA plasmid encoding N-terminal His-tagged BexL (BexL/pET28b) was provided by Dr. Hung-Wen Liu. The BexL/pET28b plasmid was transformed into *E. coli* BL21(DE3) competent cells and plated on LB-agar plates containing kanamycin (50 µg/mL). The plates were incubated overnight at 37 °C. Positive transformants were transferred to a 5 mL starter culture of Luria-Bertani (LB) broth containing kanamycin (50 µg/mL) and grown overnight at 37 °C with shaking. The cultures were used to inoculate one liter of LB with kanamycin (50 µg/mL) and grown at 37 °C until the A_{600} reached 0.4 – 0.6. The cells were then cooled to 18 °C, and 0.1 mM IPTG was added to induce protein expression. After 12-18 hours of incubation at 18 °C, the cells were harvested by centrifugation at 5000 rpm for 15 minutes. The cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C. The frozen cell pellets were thawed on ice and re-suspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 15% glycerol, and 10 mM imidazole). The cell suspension was lysed using sonication (8 x 30 s cycles), and cellular debris was removed by centrifugation at 14000 rpm for 45 minutes. The

lysate was incubated with 5 mL Ni-IMAC resin (BioRad) at 4 °C for one hour. The resin was poured into a fritted column and the flow through fraction was collected. The resin was washed with 100 mL of lysis buffer then eluted with lysis buffer plus increasing amounts of imidazole (20-500 mM). The elutions were analyzed using SDS-PAGE, and elutions containing the protein of interest were combined and concentrated by centrifugal filtration. The theoretical molecular weight of BexL is 36148.4 Daltons, and was confirmed using SDS-PAGE.

The expression and purification of StfQ is similar to the above procedure. The plasmid pYR31 encoding N-terminal His-tagged StfQ was provided by Dr. Yi Tang at UCLA. pYR31 was transformed into *E. coli* BL21(DE3) and verified by gene sequencing. StfQ was expressed and purified using a similar procedure as described above. After Ni-IMAC purification, StfQ was buffer exchanged into the crystallization buffer (20 mM Tris pH 8.0) using a PD-10 column, and the resulting pure native protein was concentrated to 13 mg/ml.

The selenomethionine-substituted StfQ was produced in *E. coli* strain BL21(DE3) in 2 x 1 LB broth containing kanamycin at 37 °C until A_{600} reached ≈ 0.7. The cells were harvested (3,000 x g for 20 min) and resuspended 3 times with a total of 100 ml of M9 minimal medium containing the following amino acids, lysine hydrochloride, phenylalanine, threonine (5 mg each), isoleucine, leucine, valine (2.5 mg each), and L-selenomethionine (3 mg) (Sigma). The third resuspension was added to 2 x 1 L M9 minimal media plus kanamycin and the following amino acids: lysine hydrochloride, phenylalanine, threonine (50 mg each), isoleucine, leucine, valine (25 mg each); L-selenomethionine (30 mg) plus 1 mM IPTG to induce protein expression. Cells were grown for an additional 18 h at 18 °C. The cells were then harvested by centrifugation. The selenomethione-substituted StfQ protein purification procedure was completed as described above for wild type StfQ.

2.5.2 Crystallization and Data Collection

After Ni-IMAC purification, BexL was buffer exchanged into the crystallization buffer (20 mM Tris, pH 8.0) using a PD-10 column (GE). BexL was crystallized using the sitting drop vapor diffusion method at room temperature from crystal seed stocks generated using a Seed Bead (Hampton). One microliter of 4 mg/mL protein solution and one microliter of well solution (0.1 M MES (2-(*N*-morpholino)ethanesulfonic acid) pH 7.0, 15% PEG 20000) were mixed and allowed to equilibrate over 500 μL of well solution. Three-dimensional trapezoidal crystals formed overnight. Crystals were washed in well solution then flash frozen in liquid nitrogen. X-ray diffraction data using monochromatic X-rays (0.96110 Å) was collected for native BexL crystals to 1.79 Å at Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 12-2. Data was processed using HKL2000(36).

Native and selenomethionine-substituted StfQ were crystallized by hanging drop vapor diffusion method at room temperature. Two microliters of 8 - 10 mg/mL protein were mixed with 2 μl of well solution (0.2 M sodium formate, 15% PEG 3350) and equilibrated over a 500 μL well solution. The crystals were cryopreserved in 20 % glycerol, then flash-frozen in liquid nitrogen. X-ray diffraction data using monochromatic X-rays (0.96110 Å) was collected for native StfQ crystals to 2.2 Å at Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 12-2. The multi-wavelength anomalous diffraction (MAD) data of selenomethionine-substituted StfQ crystals were collected to 2.70 Å at Advanced Light Source on beamline 8.2.1. Data was processed using HKL2000.

2.5.3 StfQ Phasing, Model Building, and Refinement

StfQ was crystallized in the space group $P2_12_12_1$ with two StfQ molecules forming a dimer in the asymmetric unit. The StfQ structure was solved using heavy atom phasing by multi-wavelength anomalous diffraction (MAD) to 2.7 Å by SHELX(37). A preliminary model was built using Arp/Warp, which was used for iterative rounds of model building (COOT) and refinement (PHENIX Refine)(38-40). The resulting MAD structure was used to solve the 2.2 Å native StfQ structure by molecular replacement (Phaser)(41). This was followed by refinement (PHENIX Refine) and model building (COOT) to yield a structure with a R_{work} of 18% and R_{free} value of 21%. The following residues could not me confidently placed in the StfQ model due to missing electron density: M1-D7 (chain A), M1-D7 and D119-G122 (chain B). The StfQ structure has been deposited to the protein data bank (PDB ID: 4P2D).

2.5.4 BexL Phasing, Model Building, and Refinement

BexL was crystallized in the space group $P4_12_12$ and Matthews coefficient analysis suggested one BexL molecule in the asymmetric unit (Vm. of 2.39 $Å^3$/Da and solvent content 48.56 %). Initial phases for BexL were determined using molecular replacement (Phaser) with StfQ as a search model looking for one BexL molecule in the asymmetric unit. A preliminary model was built (PHENIX AutoBuild) and then this model used for iterative rounds of model building and refinement (PHENIX Refine)(42). This cycle was continued until R_{free} of the model was significantly reduced. At this point MES and waters were added to the model using COOT and PHENIX Refine, respectively. The model was further refined to achieve an R_{work} of 18% and an R_{free} of 22%. Data collection and refinement statistics can be found in Table 2-1. The

following residues could not be confidently place in the BexL model due to missing electron density: E44, G45, G251, and G252. The BexL structure has been deposited to the protein data bank (PDB ID: 4PSQ).

2.5.5 Site-Directed Mutagenesis

BexL and StfQ mutants were generated by site directed mutagenesis using PCR with mutagenic primers using BexL/pET28b and pYR31 as templates, respectively. Mutagenic primers are listed in Table 2-2.

2.5.6 Actinorhodin KS/CLF Expression and Purification

ActKS/CLF was expressed in *S. coelicolor* CH999 containing the pRJC006 expression vector as previously described but with slight modifications(43). Fresh spores were used to inoculate 50 mL of Super YEME containing 50 μg/mL kanamycin. The 50 mL cultures were grown for 3 days at 30 °C with shaking at 250 rpm. The 50 mL cultures were transferred to 2.5 L baffled flasks containing 500 ml of Super YEME and 50 μg/mL kanamycin. The cultures were grown for an additional 2 days under the same conditions. Protein expression was induced with 10 μg/mL thiostrepton. Cells were harvested 24 hours after induction by centrifugation (20 minutes at 5100 rpm) and the cell pellets were stored at -80 °C. Cell pellets from 2 L of cell culture were resuspended in 200 mL of lysis buffer (100 mM KP_i, 15% glycerol, 300 mM NaCl, pH 7.5, with 2 EDTA-free protease inhibitor cocktail (Roche) tablets). The cell suspension was lysed by sonication on ice (10 x 1 minute cycles). Cell debris was pelleted by centrifugation (1 hour at 14000 rpm) and DNA was precipitated using streptomycin sulfate (2% final concentration) followed by centrifugation (45 min at 14000 rpm). The supernatant was filtered

90

using a 0.45 μm filter and 30% – 50% $(NH_4)_2SO_4$ was used to precipitate ActKS/CLF overnight. Precipitated protein was pelleted by centrifugation (30 min at 14000 rpm) and redissolved in Ni-binding buffer (50 mM KP_i, 10% glycerol, 500 mM NaCl, 5 mM imidazole, pH 7.5). After resuspension, the solution was filtered using a 0.45 μm filter then bound to 2 ml of Ni-IMAC resin (Bio-Rad) pre-equilibrated with Ni-binding buffer and stirred at 4 °C for 1 hour. ActKS/CLF was eluted using Ni-binding buffer containing increasing amounts of imidazole. Fractions (2.5 mL each) containing between 375 mM and 750 mM imidazole were individually buffer exchanged into storage buffer (100 mM KP_i, 10% glycerol, pH 7.5) using a PD-10 column (GE Healthcare) and flash frozen at -80 °C.

2.5.7 *Holo*-ActACP, ActKR, and MAT Expression and Purification

The *S. coelicolor* MAT was purified from *E. coli* as previously described(44). A pET-28a derived ActACP expression plasmid was transformed into *E. coli* BAP1 cells that contain a chromosomally encoded copy of the phosphopantetheinyl transferase (Sfp) that ensures the production of holo-ACP. Expression and purification was conducted as described above for BexL. Purified ActACP was dialyzed into storage buffer (50 mM Tris, 100 mM NaCl, 10 % glycerol, 2 mM DTT, pH 8.0) and concentrated to 7 mg/ml. A pET28c derived ActKR expression plasmid was transformed into *E. coli* BL21(DE3) cells and expressed and purified as described above for BexL then dialyzed into storage buffer and concentrated to 8 mg/mL.

2.5.8 BexL *in vitro* Assays

WT and mutant BexL activity was measured using an *in vitro* assay based on the actinorhodin minimal PKS(43). The reaction volume was 250 μL containing 5 μM ActKS/CLF,

50 μM ActACP, 30 μM ActKR, 30 μM, BexL WT (or a BexL mutant), 2 μM MAT, 5 mM NADPH (Sigma) (when ActKR is present), and 5 mM malonyl-CoA (Sigma) were carried out in 100 mM KP$_i$ (pH 7.5) at room temperature for 12 hours. The reaction mixtures were extracted with 300 μL of 94 : 5 : 1 of ethyl acetate : methanol : acetic acid and the organic layer was dried using Speed Vac Concentrator. The residual oil was resuspended in DMSO (Sigma) and subject to reverse-phase HPLC analysis using a Synergi Hydro-RP column (Phenomenex). The HPLC gradient was from 5 to 50 % MeCN in a H$_2$O/0.1 % formic acid mixture over 15 minutes followed by 50 to 95 % MeCN in a H$_2$O/0.1 % formic acid mixture over 5 min. SEK4, SEK4B, mutactin, and SEK34 were identified using ESI-MS, UV-Vis absorbance, and relative retention times.

2.5.9 PKS4 KS-AT and PKS4 ACP Expression and Purification

A PKS4 KS-AT expression plasmid was provide by Dr. Yi Tang (UCLA), transformed into *E. coli* BL21(DE3), and expressed and purified as described above for BexL. Purified PKS4 KS-AT dialyzed into storage buffer (50 mM Tris (7.5), 10 % glycerol, 100 mM NaCl, and 2 mM DTT) then concentrated to 8 mg/ml. The PKS4 ACP expression plasmid was transformed into *E. coli* BAP1 cells that contain a chromsomally encoded copy of the phosphopantetheinyl transferase (Sfp) that ensures the production of holo-ACP. Expression and purification was conducted as described above for BexL. Purified PKS4 ACP was dialyzed into storage buffer and concentrated to 7 mg/ml.

2.5.10 StfQ *in vitro* Assays

WT and mutant StfQ activity was measured using an *in vitro* assay based on the fungal PKS4 minimal PKS(45). The reaction volume was 250 µL containing 10 µM PKS4-KSAT, 50 µM PKS4 ACP, 30 µM, StfQ WT (or an StfQ mutant), 2 µM MAT, and 5 mM malonyl-CoA (Sigma) were carried out in 100 mM KP_i (pH 7.5) at room temperature for 12 hours. Reactions were extracted and analyzed using the same procedure described above for the BexL assay.

2.5.11 *In silico* Docking Studies

The ligands were drawn in ChemDraw and then converted to PDB files using the NCI SMILES converter server (http://cactus.nci.nih.gov/translate/). The ligands were energy minimized and converted to .mol2 files using Chimera(46). The program GOLD was used for docking putative phosphopantetheine-linked intermediates into the C-terminal domain of StfQ and N-terminal domain of BexL(47). Docking runs for BexL and StfQ were performed by defining the ligand binding site within 20 Å of the active site arginine (R66 of BexL, or R218 of StfQ) using default settings for all parameters with 100 docking trials per docking run.

2.5.12 Circular Dichroism of BexL and StfQ Mutants

Samples were collected on a JASCO J-810 Circular Dichroism Spectrapolarimeter. All samples were prepared by diluting protein into 25 mM tris (pH 7.5) to a final concentration of 0.1 - 0.2 mg/ml (Figure 2-17).

REFERENCES

1. Marti, T., Hu, Z., Pohl, N. L., Shah, A. N., and Khosla, C. (2000) Cloning, nucleotide sequence, and heterologous expression of the biosynthetic gene cluster for R1128, a non-steroidal estrogen receptor antagonist. Insights into an unusual priming mechanism, *J Biol Chem 275*, 33443-33448.
2. Rix, U., Fischer, C., Remsing, L. L., and Rohr, J. (2002) Modification of post-PKS tailoring steps through combinatorial biosynthesis, *Nat Prod Rep 19*, 542-580.
3. Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms, *Curr Opin Chem Biol 7*, 285-295.
4. Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork, *Nat Prod Rep 24*, 162-190.
5. Das, A., and Khosla, C. (2009) Biosynthesis of aromatic polyketides in bacteria, *Acc Chem Res 42*, 631-639.
6. Zhou, H., Li, Y., and Tang, Y. (2010) Cyclization of aromatic polyketides from bacteria and fungi, *Nat Prod Rep 27*, 839-868.
7. Javidpour, P., Korman, T. P., Shakya, G., and Tsai, S. C. (2011) Structural and biochemical analyses of regio- and stereospecificities observed in a type II polyketide ketoreductase, *Biochemistry 50*, 4638-4649.
8. Bartel, P. L., Zhu, C. B., Lampel, J. S., Dosch, D. C., Connors, N. C., Strohl, W. R., Beale, J. M., Jr., and Floss, H. G. (1990) Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in streptomycetes: clarification of actinorhodin gene functions, *J Bacteriol 172*, 4816-4826.
9. Hopwood, D. A. (1990) Molecular-Genetics of Polyketides and Its Comparison to Fatty-Acid Biosynthesis, *Annu Rev Genet 24*, 37-66.
10. McDaniel, R., Ebert-Khosla, S., Fu, H., Hopwood, D. A., and Khosla, C. (1994) Engineered biosynthesis of novel polyketides: influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase, *Proc Natl Acad Sci U S A 91*, 11542-11546.
11. Alvarez, M. A., Fu, H., Khosla, C., Hopwood, D. A., and Bailey, J. E. (1996) Engineered biosynthesis of novel polyketides: properties of the whiE aromatase/cyclase, *Nat Biotechnol 14*, 335-338.
12. Shen, B., and Hutchinson, C. R. (1996) Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase, *Proc Natl Acad Sci U S A 93*, 6600-6604.
13. Zawada, R. J., and Khosla, C. (1997) Domain analysis of the molecular recognition features of aromatic polyketide synthase subunits, *J Biol Chem 272*, 16184-16188.
14. Mcdaniel, R., Ebertkhosla, S., Hopwood, D. A., and Khosla, C. (1994) Engineered Biosynthesis of Novel Polyketides - Act(Vii) and Act(Iv) Genes Encode Aromatase and Cyclase Enzymes, Respectively, *Journal of the American Chemical Society 116*, 10855-10859.
15. Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y., and Tsai, S. C. (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides, *Proc Natl Acad Sci U S A 105*, 5349-5354.

16.   Ames, B. D., Lee, M. Y., Moody, C., Zhang, W., Tang, Y., and Tsai, S. C. (2011) Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway, *Biochemistry 50*, 8392-8406.

17.   Lee, M. Y., Ames, B. D., and Tsai, S. C. (2012) Insight into the molecular basis of aromatic polyketide cyclization: crystal structure and in vitro characterization of WhiE-ORFVI, *Biochemistry 51*, 3079-3091.

18.   Gullon, S., Olano, C., Abdelfattah, M. S., Brana, A. F., Rohr, J., Mendez, C., and Salas, J. A. (2006) Isolation, characterization, and heterologous expression of the biosynthesis gene cluster for the antitumor anthracycline steffimycin, *Appl Environ Microbiol 72*, 4172-4183.

19.   Rajgarhia, V. B., and Strohl, W. R. (1997) Minimal Streptomyces sp. strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis, *J Bacteriol 179*, 2690-2696.

20.   Menendez, N., Nur-e-Alam, M., Brana, A. F., Rohr, J., Salas, J. A., and Mendez, C. (2004) Biosynthesis of the antitumor chromomycin A3 in Streptomyces griseus: analysis of the gene cluster and rational design of novel chromomycin analogs, *Chem Biol 11*, 21-32.

21.   Daum, M., Peintner, I., Linnenbrink, A., Frerich, A., Weber, M., Paululat, T., and Bechthold, A. (2009) Organisation of the biosynthetic gene cluster and tailoring enzymes in the biosynthesis of the tetracyclic quinone glycoside antibiotic polyketomycin, *Chembiochem 10*, 1073-1083.

22.   Trefzer, A., Pelzer, S., Schimana, J., Stockert, S., Bihlmaier, C., Fiedler, H. P., Welzel, K., Vente, A., and Bechthold, A. (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic, *Antimicrob Agents Chemother 46*, 1174-1182.

23.   Xu, Z., Jakobi, K., Welzel, K., and Hertweck, C. (2005) Biosynthesis of the antitumor agent chartreusin involves the oxidative rearrangement of an anthracyclic polyketide, *Chem Biol 12*, 579-588.

24.   Sasaki, E., Ogasawara, Y., and Liu, H. W. (2010) A biosynthetic pathway for BE-7585A, a 2-thiosugar-containing angucycline-type natural product, *J Am Chem Soc 132*, 7405-7417.

25.   Javidpour, P., Bruegger, J., Srithahan, S., Korman, T. P., Crump, M. P., Crosby, J., Burkart, M. D., and Tsai, S. C. (2013) The determinants of activity and specificity in actinorhodin type II polyketide ketoreductase, *Chem Biol 20*, 1225-1234.

26.   Sriram, M., Liaw, Y. C., Gao, Y. G., and Wang, A. H. (1991) Molecular structure of antitumor drug steffimycin and modelling of its binding to DNA, *J Biomol Struct Dyn 9*, 251-269.

27.   Brodasky, T. F., Mizsak, S., and Hoffstetter, J. R. (1985) Steffimycin C, a new member of the steffimycin anthracyclines. Isolation and structural characterization, *J Antibiot (Tokyo) 38*, 849-855.

28.   Olano, C., Abdelfattah, M. S., Gullon, S., Brana, A. F., Rohr, J., Mendez, C., and Salas, J. A. (2008) Glycosylated derivatives of steffimycin: insights into the role of the sugar moieties for the biological activity, *Chembiochem 9*, 624-633.

29.   Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res 33*, W244-W248.

30.   Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D., and Ritchie, D. W. (2010) HexServer: an FFT-based protein docking server powered by graphics processors, *Nucleic Acids Res 38*, W445-449.

31.   Masoudi, A., Raetz, C. R., Zhou, P., and Pemble, C. W. t. (2014) Chasing acyl carrier protein through a catalytic cycle of lipid A production, *Nature 505*, 422-426.

32.   Nguyen, C., Haushalter, R. W., Lee, D. J., Markwick, P. R., Bruegger, J., Caldara-Festin, G., Finzel, K., Jackson, D. R., Ishikawa, F., O'Dowd, B., McCammon, J. A., Opella, S. J., Tsai, S. C., and Burkart, M. D. (2014) Trapping the dynamic acyl carrier protein in fatty acid biosynthesis, *Nature 505*, 427-431.

33.   Rudd, B. A., and Hopwood, D. A. (1979) Genetics of actinorhodin biosynthesis by Streptomyces coelicolor A3(2), *J Gen Microbiol 114*, 35-43.

34.   Yu, T. W., Bibb, M. J., Revill, W. P., and Hopwood, D. A. (1994) Cloning, sequencing, and analysis of the griseusin polyketide synthase gene cluster from Streptomyces griseus, *J Bacteriol 176*, 2627-2634.

35.   Lombo, F., Brana, A. F., Mendez, C., and Salas, J. A. (1999) The mithramycin gene cluster of Streptomyces argillaceus contains a positive regulatory gene and two repeated DNA sequences that are located at both ends of the cluster, *J Bacteriol 181*, 642-647.

36.   (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D Biol Crystallogr 50*, 760-763.

37.   Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD, *Acta Crystallogr D Biol Crystallogr 58*, 1772-1779.

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

39.   Adams, P. D., Afonine, P. V., Bunkoczi, 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.

40.   Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7, *Nat Protoc 3*, 1171-1179.

41.   Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography 40*, 658-674.

42.   Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D 64*, 61-69.

43.   Matharu, A. L., Cox, R. J., Crosby, J., Byrom, K. J., and Simpson, T. J. (1998) MCAT is not required for in vitro polyketide synthesis in a minimal actinorhodin polyketide synthase from Streptomyces coelicolor, *Chem Biol 5*, 699-711.

44.   Kumar, P., Koppisch, A. T., Cane, D. E., and Khosla, C. (2003) Enhancing the modularity of the modular polyketide synthases: transacylation in modular polyketide synthases catalyzed by malonyl-CoA:ACP transacylase, *J Am Chem Soc 125*, 14307-14312.

45. Zhang, W., Li, Y., and Tang, Y. (2008) Engineered biosynthesis of bacterial aromatic polyketides in Escherichia coli, *Proc Natl Acad Sci U S A 105*, 20683-20688.
46. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.
47. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD, *Proteins 52*, 609-623.

Chapter 3

Determining the Molecular Basis For Starter Unit Selection During Daunorubicin Biosynthesis

3.1 Abstract

Type II polyketides are a large class of polyaromatic natural products with anticancer, antibiotic, and antiviral activity. Type II polyketides are biosynthesized from malonyl-CoA building blocks to form an aglycone core, which is further modified by different tailoring enzymes. In rare cases, biosynthesis is initiated by the incorporation of a non-malonyl derived starter unit, which adds molecular diversity to the poly-β-ketone backbone. Starter unit incorporation can be mediated by one or more enzymes, which are involved in the generation and transfer of the starter unit onto the acyl carrier protein (ACP) that subsequently is fed into the downstream PKS domains. How novel starter units are primed onto PKSs remains poorly understood. Daunorubicin is a type II polyketide that is widely used to treat many cancers. Daunorubicin biosynthesis employs a unique propionyl starter unit, which is selected for by the enzyme DpsC. Here we present the crystal structures of apo-DpsC, acyl-DpsC, and a DpsC co-crystal structure with a bound substrate analogue. The DpsC crystal structures reveal the molecular determinants for catalytic activity and propionyl substrate specificity as well as provide a structural basis for rational engineering of starter unit selection during daunorubicin biosynthesis.

3.2 Introduction

Type II polyketides are a diverse group of natural products that have been a historically important source for pharmaceutical development(1-4). Type II polyketides include bioactive
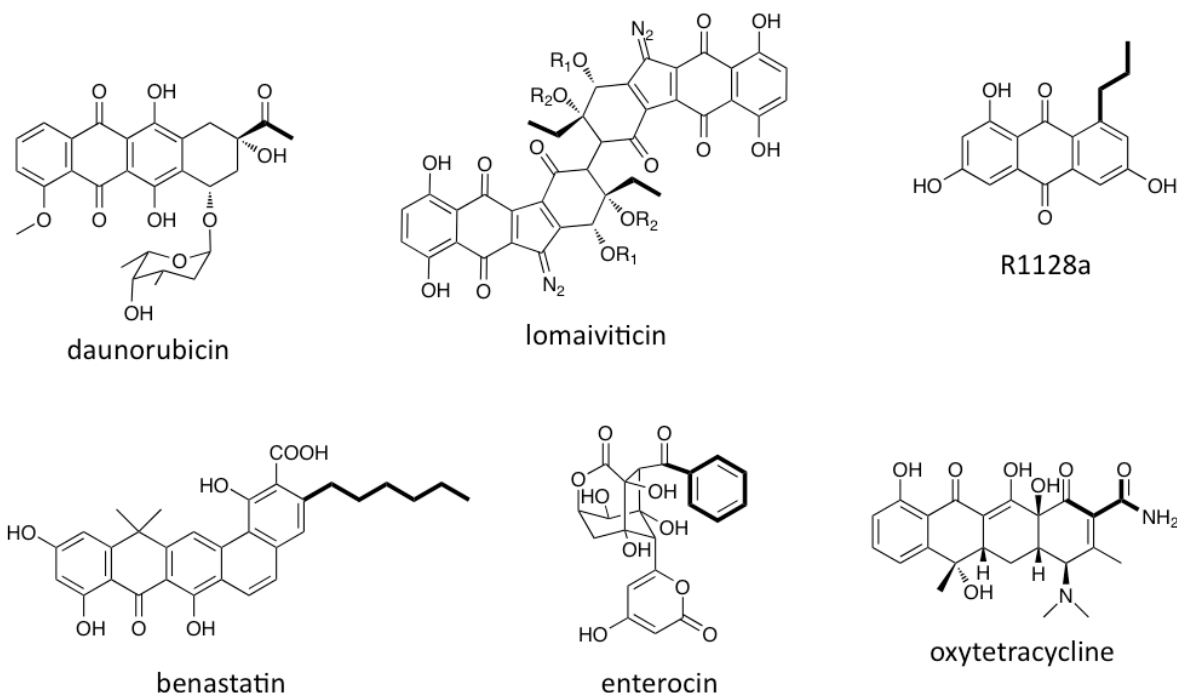
Figure 3-1. The diversity of starter units in type II polyketides. Daunorubicin, lomaiviticin R1128a, benastatin, enterocin, and oxytetracycline are all type II PKS products with novel starter units. Starter units are shown in bold.

natural products such as daunorubicin and tetracycline, which have anticancer and antibiotic properties, respectively(5, 6). Daunorubicin has been used as a chemotherapeutic agent since its discovery in the 1950s and contains an atypical propionyl starter unit (Figure 3-1) (5, 7-9).

Type II polyketide biosynthesis is initiated by priming the active site cysteine of the ketosynthase (KS) with an acetyl starter unit(10). To achieve KS priming, the majority of type II polyketide synthases (PKSs) decarboxylate a malonyl group in the active site of KS to form an acetyl-cysteine acyl-enzyme intermediate(11). Once primed, the elongation of the acyl-enzyme

intermediate occurs in the KS active site by decarboxylative Claisen condensations(12). Although KS priming by malonate is the most common method to initiate type II polyketide biosynthesis, alternative initiation mechanisms have been discovered, such as propionate incorporation in daunorubicin biosynthesis(8).

The biosynthesis and incorporation of unique starter units in type II PKSs is an important area of investigation, because starter unit diversity allows for unique opportunities to expand the chemical space of bioactive natural products(13). Atypical starter units have been identified in the multiple type II PKS natural products. For example, daunorubicin (anticancer), R1128 (anticancer), enterocin (antibiotic), benastatin (anticancer), lomaiviticin (anticancer), and oxytetracycline (antibiotic) all have non-acetate starter units(9, 14-19) (Figure 3-1). The genes responsible for starter unit synthesis and incorporation into these type II polyketides have been identified, and they were studied to different extents *in vitro* and *in vivo*. In the cases of enterocin, benastatin, and oxytetracycline, the knowledge of starter unit biochemistry was applied to biosynthesize new, engineered polyketides(16, 19-21). These studies highlight the importance of starter unit diversity in type II PKSs and the potential to generate new natural product analogues with improved bioactivity.

The enzyme DpsC uses propionyl-CoA as a substrate and is responsible for incorporating a propionyl starter unit into daunorubicin biosynthesis(8, 22). Early *in vitro* studies of DpsC revealed that DpsC can catalyze the reactions of both a priming KS and an acyltransferase (AT) (Figures 3-2, 3-3A-B)(8). However, the molecular basis for the substrate specificity and catalytic promiscuity of DpsC is unknown. Based on protein sequence homology, DpsC is similar to known priming KSs (KS IIIs) such as FabH (*Ec*FabH) from the *E. coli* type II fatty acid synthase (FAS), and a PKS KS III ZhuH from *Streptomyces R1128* (12, 23). DpsC also shares high

sequence homology to the recently characterized enzyme CerJ from *Streptomyces tendae*, which transfers malonate and malonate-derivatives onto a type II PKS natural product(24). Because DpsC has unique catalytic activities and the ability to incorporate an atypical starter unit in biosynthesis, DpsC is an attractive target for understanding and engineering PKSs. Here, we
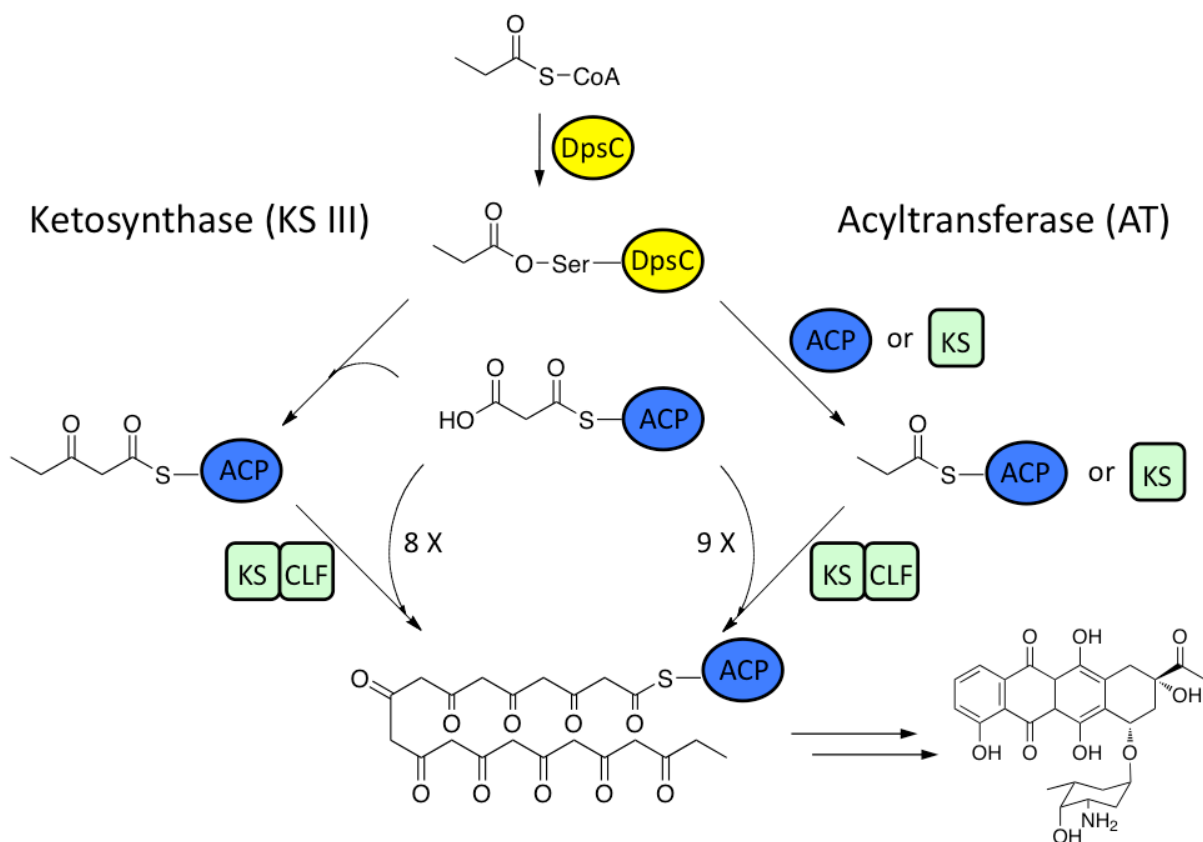
Figure 3-2. The role of DpsC in daunorubicin biosynthesis. DpsC uses propionyl-CoA as a substrate to form an acyl-enzyme intermediate. The acylated active site serine of DpsC then reacts with malonyl-ACP (ketosynthase activity) or the PPT thiol of holo-ACP (acyltransferase activity). Both enzymatic activities ultimately generate a 21-carbon poly-β-ketone intermediate, which is processed to yield daunorubicin.

present five crystal structures of DpsC: apo-DpsC, three acyl-ezymes (acetyl-, propionyl-, and butanoyl-DpsC), and a co-crystal structure with a substrate analogue. Our results provide a structural basis to explain the acyltransferase activity of DpsC, and reveal residues important for its substrate specificity. This information can be used to engineer DpsC to incorporate different

starter units into polyketide biosynthesis. Additionally, we developed a phosphopantheine analogue that can be loaded onto the acyl carrier protein (ACP) and stabilize ACP/DpsC interactions.
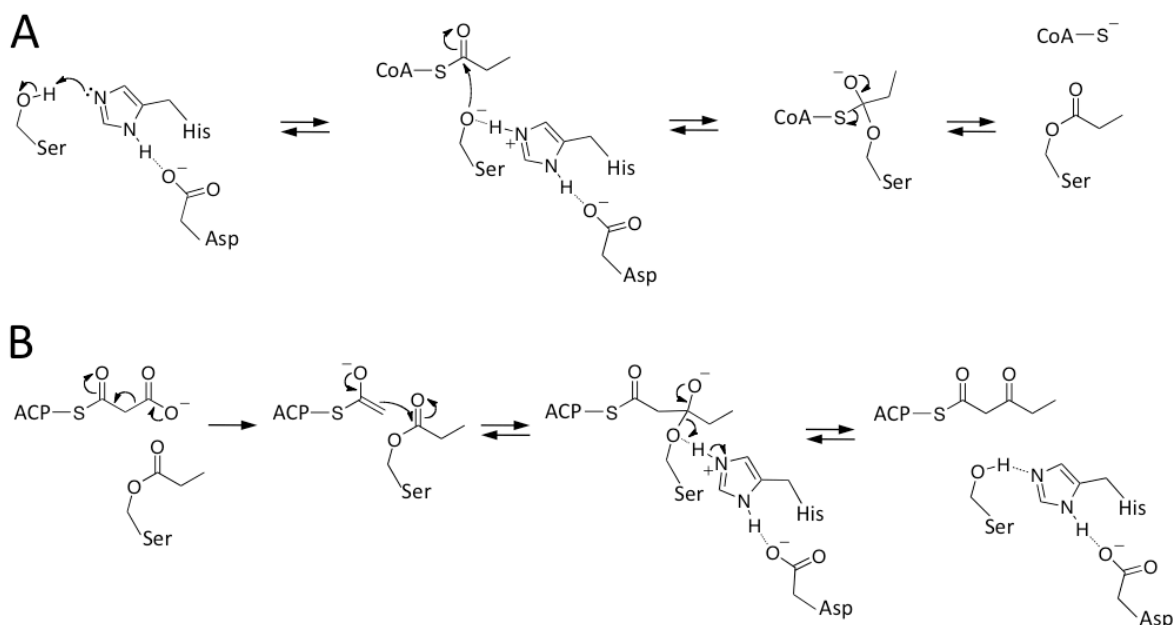
Figure 3-3. The active site residues and proposed mechanisms of DpsC AT and KS activity. (A) The DpsC active site contains a Ser-His-Asp catalytic triad that is responsible for acyl transfer. In the proposed mechanism of DpsC acyl transfer, Ser118 is activated as a nucleophile for acylation because of a neighboring His297 that forms a hydrogen bond with the hydroxyl proton on Ser118. The His297 is further stabilized by a neighboring aspartate. (B) DpsC also catalyzes decarboxylation of malonyl-ACP, which elongates the acyl-enzyme intermediate.

3.3 Results and Discussion

3.3.1 Overall Structure of DpsC

The crystal structure of DpsC reveals a thiolase fold common to ketosynthases and the recently reported acyltransferase CerJ(24) (Figure 3-4A and 4B). DpsC crystallized in the space group $P6_522$ with two DpsC molecules per asymmetric unit, and the crystals diffracted to a
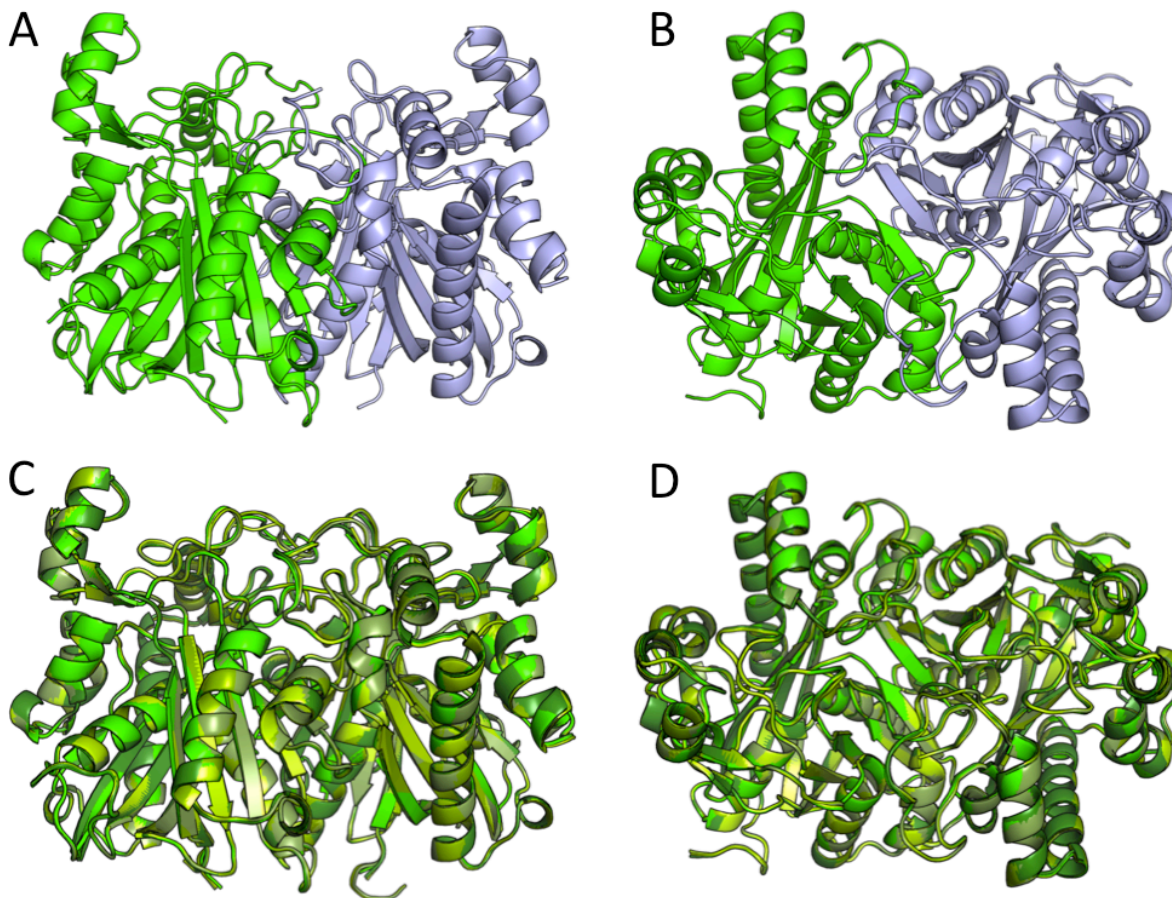
Figure 3-4. The crystal structures of DpsC, acetyl-DpsC, propionyl-DpsC, butanoyl-DpsC, and DpsC substrate analogue (**4**) co-crystal. (A) Front view of the DpsC homodimer with monomer A in green and monomer B in blue. (B) Top view of the DpsC homodimer. (C) Front view of the overlaid crystal structures of DpsC, acetyl-DpsC, propionyl-DpsC, butanoyl-DpsC, and DpsC substrate analogue (**4**) co-crystal. Each structure is a different shade of green. (D) Top view of the overlaid crystal structures of DpsC, acetyl-DpsC, propionyl-DpsC, butanoyl-DpsC, and DpsC substrate analogue (**4**) co-crystal in different shades of green.

resolution of 2.5 Å. The two DpsC monomers form a tight homodimer, with each monomer related by a 2-fold rotational axis of non-crystallographic symmetry. The extensive DpsC dimer interface area is estimated to be 2745.3 Å^2 (PISA)(25). When overlaid, the DpsC monomers are nearly identical with an overall RMSD of 0.18 Å. The DpsC core secondary structure consists of two β-α-β-α-(β)-α-β-β motifs connected by a series of loops, which is similar to other known KS

structures and CerJ(12, 23, 24). The DpsC structure can be further split into two β-α-β-α-(β)-α-β-β motifs, as defined by the N- and C-terminal sections. The N-terminal core motif consists of Nα1-Nα3, Nβ1-Nβ5, and loops L1-L8. L1 contains two α-helices and two β-sheets designated L1(α1), L1(α2), L1(β1), and L1(β2). The C-terminal core motif consists of Cα1-Cα3, Cβ1-Cβ4,
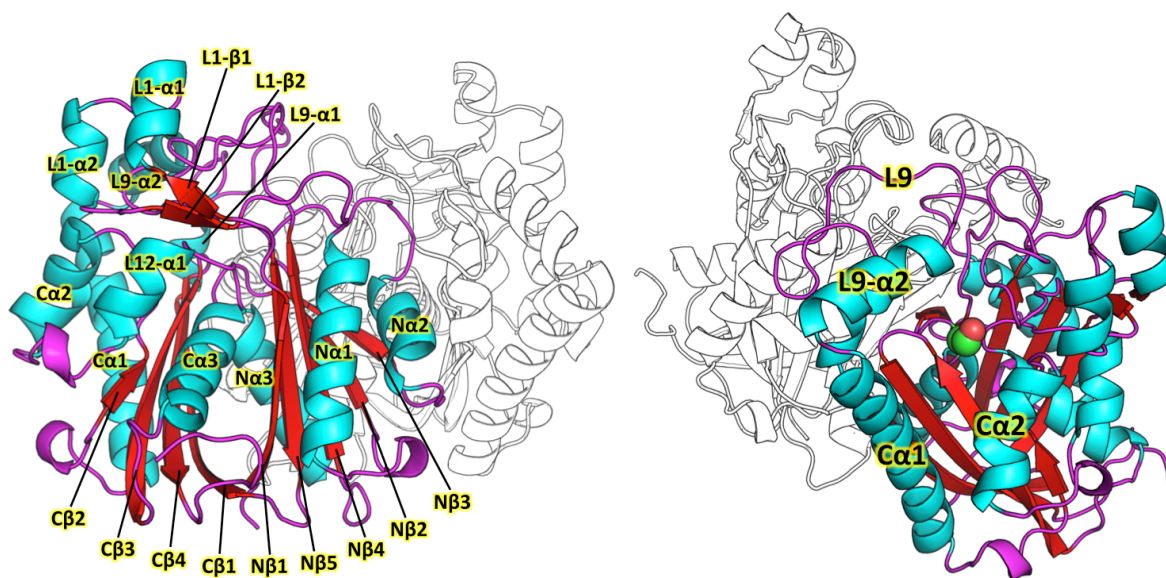
Figure 3-5. The secondary structural elements of DpsC. The DpsC dimer is shown with monomer A colored by secondary structure and monomer B outlined in black and white. α-Helices are in cyan, β-sheets are in red, and loop regions are in magenta. The double β-α-β-α-(β)-α-β-β core motifs are displayed on the left. The active site serine is shown in spheres and the secondary structure elements of the active site pocket entrance are displayed on the right.

and loops L9-L15. L9 contains two α-helices designated L9(α1), and L9(α2). L12 contains one α-helix designated L12(α1). In DpsC, the missing β-strand in the second core motif is a small loop between Cα2 and the L12(α1) region (Figure 3-5A-5B). In summary, DpsC contains a thiolase fold similar to other KSs and the acyltransferase CerJ.

3.3.2 Comparison of DpsC With Other KS IIIs and ATs

The DpsC Ser-His-Asp active site is unique among known KS III enzymes, which have a conserved Cys-His-Asn catalytic triad(12, 23, 26). In KS IIIs, the active site cysteine is acylated
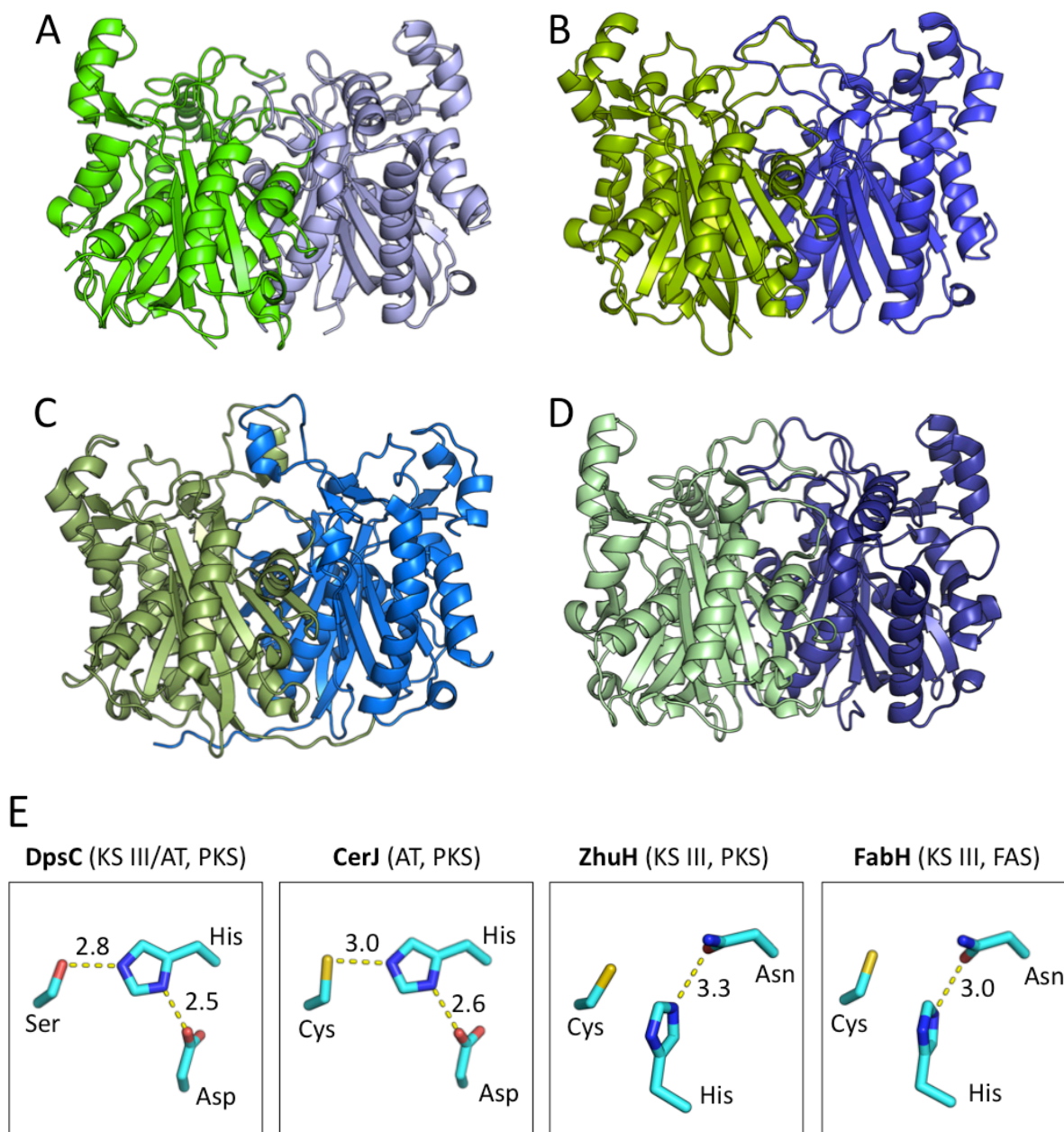
Figure 3-6. The crystal structures and active sites of DpsC, CerJ, ZhuH, and FabH with monomers A and B in different shades of green and blue, respectively. (A) The crystal structure of DpsC. (B) The crystal structure of CerJ. (C) The crystal structure of ZhuH. (D) The crystal structure of FabH. (E) The actives sites of DpsC, CerJ, ZhuH, and FabH. The DpsC and CerJ active sites have the same overall orientation of the catalytic triad. In comparison, DpsC has serine and CerJ has cysteine as the active site nucleophile. The orientation of active site residues of ZhuH and FabH are different than DpsC and CerJ. ZhuH and FabH both contain cysteine as the active site nucleophile, but their active site histidine is not engaged in hydrogen bonding with the cysteine. In ZhuH and FabH, the histidine and asparagine are important for decarboxylation of malonyl-ACP, not acyl-transfer.

with an acyl-CoA derived starter unit, and histidine and asparagine facilitate decarboxylative condensation with malonyl-ACP(12). In contrast to KS IIIs, DpsC contains a Ser-His-Asp catalytic triad (Figures 3-6E). Namely, in DpsC, the highly conserved cysteine is replaced with a serine (Ser118), which is located on the L5 loop at the N-terminal end of Nα3 in both DpsC and other KS IIIs. The location of the active site nucleophile (Cys in all known KSs except DpsC) has increased nucleophilic character because of the helix dipole effect, which lowers the pKa of the active site cysteine (or serine)(27, 28). The active site histidine (His297) and aspartic acid (Asp302) of DpsC are located on helix Cα3 and loop L12, respectively. This positioning differs significantly from the locations of His243 and Asn274 in the *Ec*FabH crystal structure (PDB ID: 1EBL), as well as the equivalent active site residues of ZhuH (PDB ID: 1MZJ), which is a polyketide KS III (12, 23) (Figure 3-7). His243 and Asn274 of *Ec*FabH are located on loop L11 and loop L12, respectively (Figure 3-7A). Asn274 of *Ec*FabH occupies the same position as His297 of DpsC but when overlaid the Asn274 side chain is positioned orthogonally to His297 of DpsC. Moreover, Asn274 of *Ec*FabH does not form a hydrogen bond with the active site cysteine (C112)(Figures 3-6E and 3-7). In DpsC, His297 of DpsC is positioned at the back of the active site pocket below the catalytic Ser118 and is less accessible to the substrates when compared to Asn274 of *Ec*FabH. Further, His297 of DpsC forms a tight hydrogen bond with Ser118, which increases the nucleophilicity of Ser118. His297 of DpsC also forms a hydrogen bond with Asp302, which is located on helix Cα3 in the interior of DpsC below the active site Ser118 (Figure 3-7). Asp302 further increases the nucleophilicity of Ser118 by polarizing His297. Interestingly, the Ser-His-Asp catalytic triad of DpsC is more closely related to serine protease Ser-His-Asp catalytic triads or acyltransferase (ATs) Ser-His catalytic dyads than typical KS active site triads (Cys-His-His or Cys-His-Asn)(29, 30).
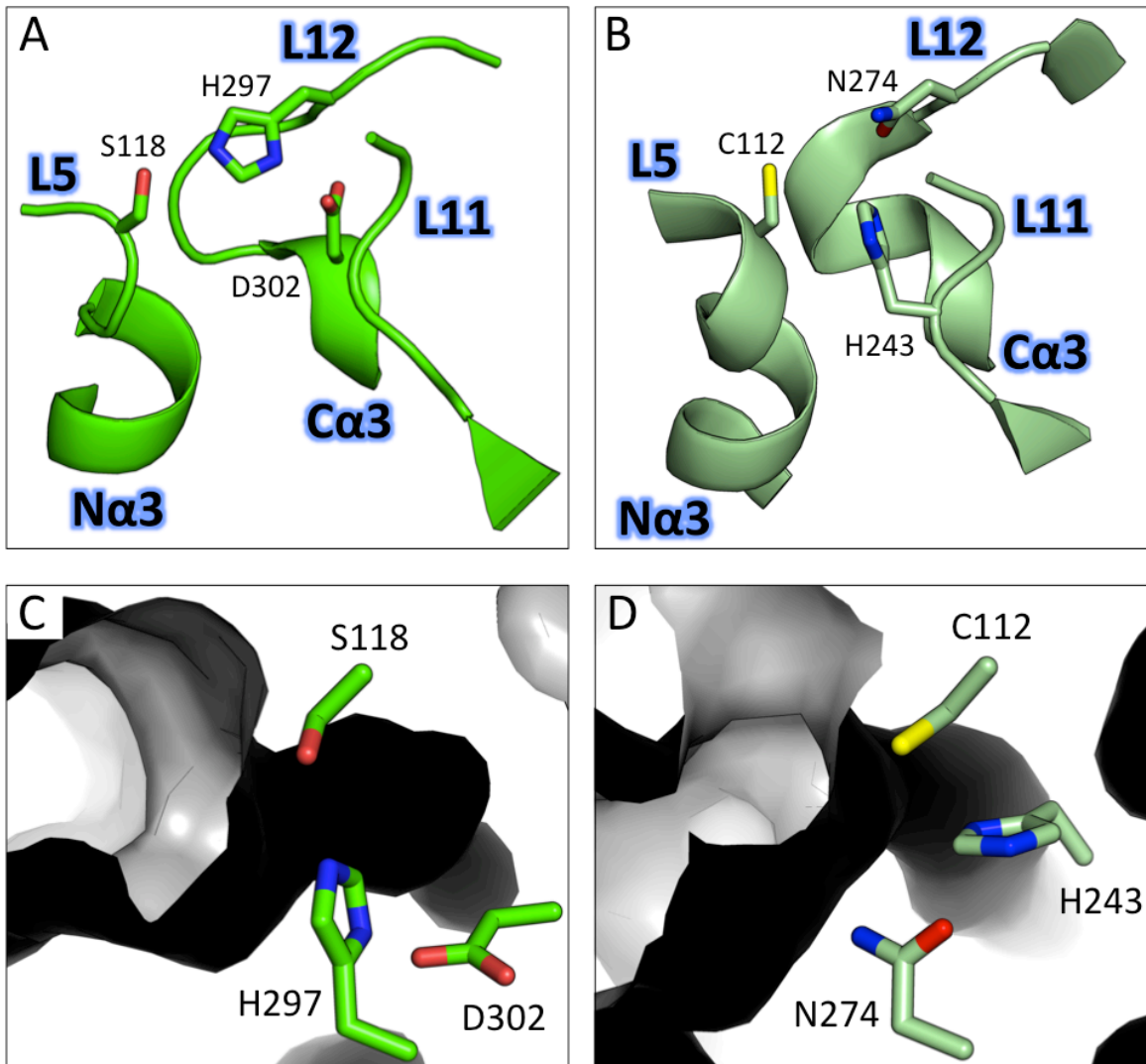
Figure 3-7. The differences in spatial orientation between the DpsC and *Ec*FabH active site residues. (A) In DpsC, Ser118, His297, and Asp302, are located on L5, L12, and Cα3, respectively. (B) In *Ec*FabH, Cys112, His243, and Asn274, are located on L5, L12, and L11, respectively. (C) His297 and Asp302 of DpsC are behind the active site pocket and His297 is positioned so that the aromatic ring stacks against the side of the pocket. (D) His243 and Asn274 of *Ec*FabH are positioned to the side of the active site pocket with both of their side chains able to form hydrogen bonds with substrates in the active site.

The acyltransferase CerJ has an overall sequence identity of 41 % with DpsC and also a similar active site. The CerJ active site consists of Cys-His-Asp, which only differs from DpsC by the replacement of cysteine with serine. The DpsC and CerJ active sites residues are

positioned at the same secondary structural elements and in the same orientation (Figure 3-6E). CerJ, DpsC, and other KS IIIs all have the same fold, but CerJ does not catalyze decarboxylative elongation(24). Instead, CerJ catalyzes the transfer of malonyl, methylmalonyl, and dimethylmalonyl groups onto a sugar moiety of the type II polyketide cervimycin(24). Based on the active site triad analysis above, DpsC displays acyltransferase activity because it contains a similar active site arrangement to CerJ and other ATs. However, the ability of DpsC to catalyze decarboxylative condenstation is surprising, because the DpsC active site differs from known KS IIIs. In *Ec*FabH, His243 is essential for decarboxylation. In contrast, the structurally equivalent position on the L11 loop of DpsC is occupied by P225. The next residue on loop L11 in DpsC is histidine (His226), but it is buried deep behind the active site and is not structurally accessible for decarboxylation. Near the DpsC active site, there is another histidine (His198), which is located in the active site tunnel 7.8 Å away from Ser118. Although His198 is quite far from Ser118, molecular docking of a decarboxylation substrate mimic, malonyl-pPant, revealed that His198 can form a hydrogen bond with a substrate carbonyl and may facilitate decarboxylation (Figure 3-8). The DpsC catalyzed decarboxylative condensation mechanism is unknown. However, based on the structural comparisons with *Ec*FabH and molecular docking with malonyl-pPant, DpsC may employ a previously uncharacterized His198-dependent decarboxylation mechanism.

3.3.3 Acetyl-, Propionyl-, and Butanoyl-DpsC Crystal Structures

DpsC has been shown to accept acetyl-CoA, propionyl-CoA, and butanoyl-CoA as its substrates. However, propionyl-CoA is considered the biological substrate of DpsC, because acetyl and butanoyl starter units are not incorporated into daunorubicin biosynthesis(8). To gain
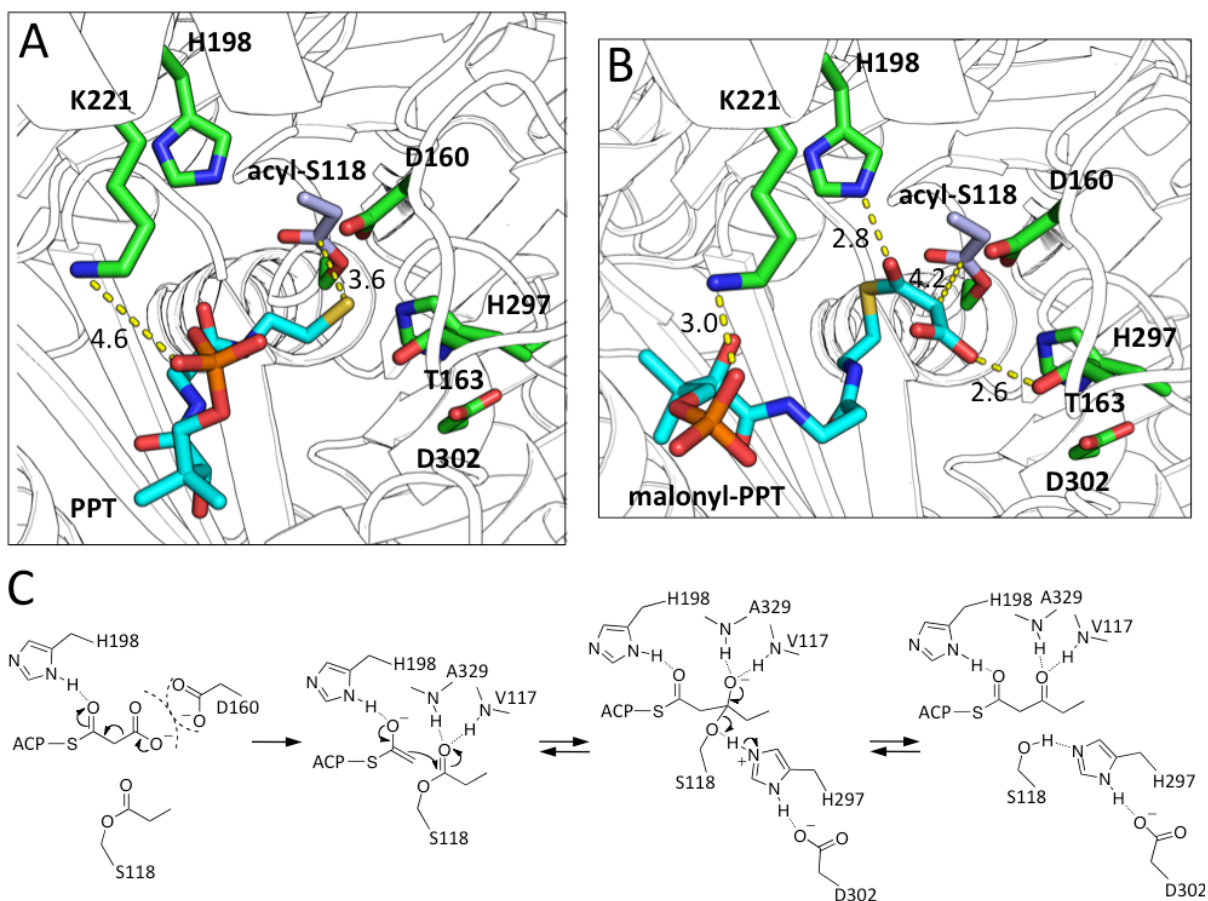
Figure 3-8. The results of molecular docking of pPant and malonyl-pPant into the propionyl-DpsC active site. (A) To understand the mechanism of acyl transfer from Ser118 of DpsC to *holo*-ACP, pPant was docked into the propionyl-DpsC active site crystal structure. Lys221 interacts with the terminal phosphate and at the opposite end of the pPant the thiol comes in close contact with the ester carbon of the acyl-enzyme intermediate. (B) Lys221 interacts with the terminal phosphate and acts as an anchoring point for pPant binding. The carboxylate is pushed into the active site pocket and forms a hydrogen bond with Thr163. The thioester carbonyl of the malonyl moiety forms a hydrogen bond with His198, which may facilitate decarboxylation by stabilizing the resulting enolate. Additionally, electrostatic repulsion between the malonyl carboxylate and Asp160 may be a driving force for decarboxylation. (C) The proposed mechanism of DpsC catalyzed decarboxylation involving His198 and Asp160.

a better understanding of how DpsC recognizes acyl-CoAs, we solved three co-crystal structures

of DpsC bound with different acyl groups. Apo-DpsC crystals were soaked with acetyl-CoA,

propionyl-CoA, and butanoyl-CoA to generate the acyl-enzyme structures acetyl-DpsC,

propionyl-DpsC, and butanoyl-DpsC, respectively (Figures 3-4C-D and 3-9). The acyl-enzyme
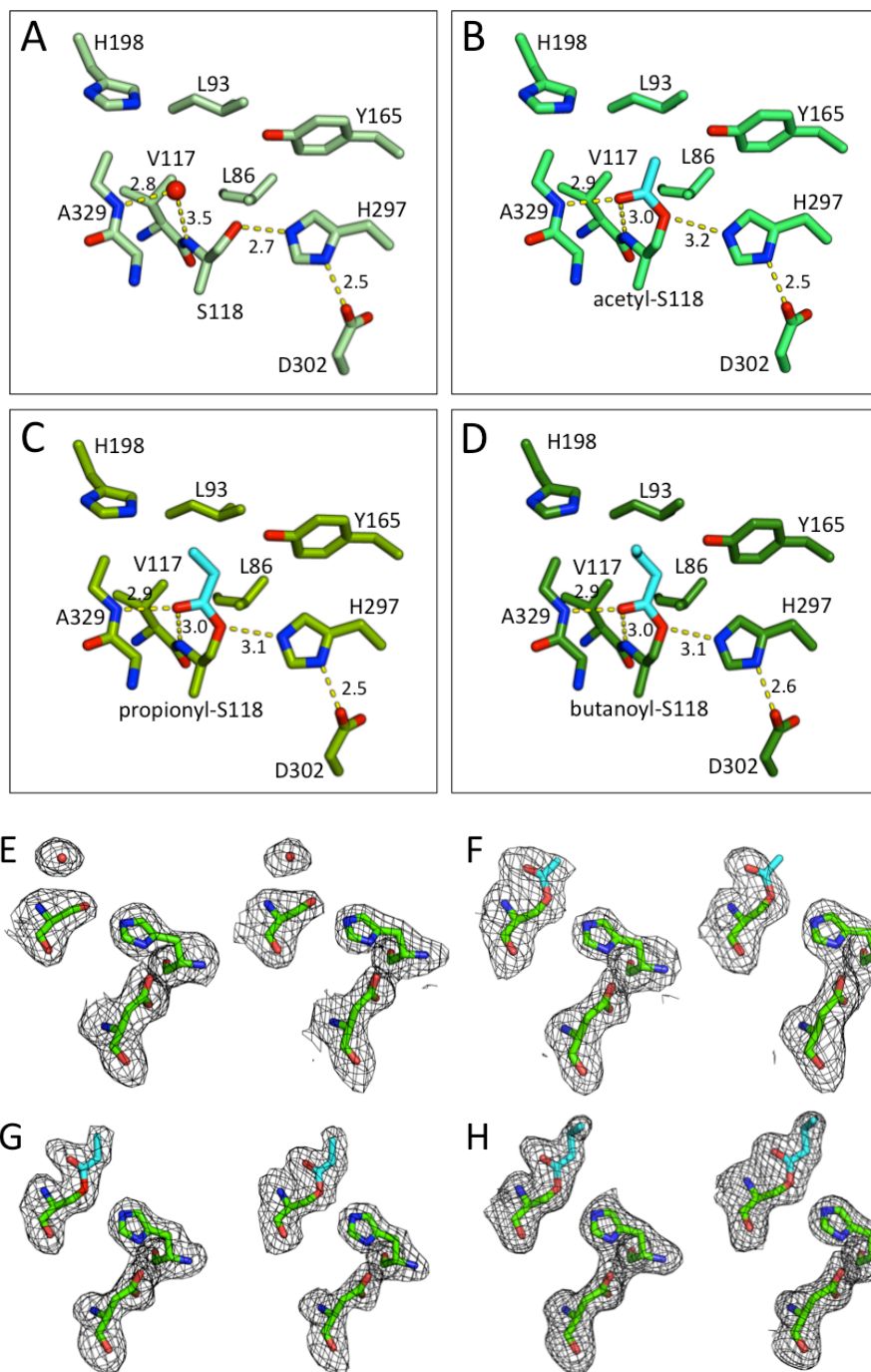
Figure 3-9. The active sites and SA omit maps of the DpsC, acetyl-DpsC, propionyl-DpsC, and butanoyl-DpsC. DpsC active site (A), acetyl-DpsC (B), propionyl-DpsC (C), and butanoyl-DpsC (D). SA omit maps of the DpsC active site (E), acetyl-DpsC (F), propionyl-DpsC (G), and butanoyl-DpsC (H)*.

*For SA omit maps in E-H, the active site of monomer A is on the left and the active site of monomer B is on the right. Acyl groups are colored in cyan.

structures overlay very well with the apo-DpsC structure, with overall RMSDs of less than 0.18 Å. In all DpsC structures, the L9 loop has elevated B-factors for both monomers (Figure 3-10). Residues Gln212-Arg213 in the apo-DpsC structure and residue Arg213 in the L9 loop region of the acetyl-DpsC structure could not be modeled because of a lack of electron density. In the apo-DpsC structure, the active site Ser118 hydroxyl is 2.7 Å from His297. In the acetyl-, propionyl, and butanoyl-DpsC structures, the Ser118 hydroxyls are 3.2 Å, 3.1 Å, and 3.1 Å from His297, respectively. In all acyl-enzyme DpsC structures the Ser118 ester carbonyl sits in an oxyanion hole formed by the amide protons of Val117 and Ala329 (Figure 3-9B-D). In the apo-DpsC structure, water is bound in the oxyanion hole formed by the amide NH of Val117 and Ala329 (Figure 3-9A). The binding of acyl-Ser118 in the oxyanion hole pulls Ser118 away from His297 and facilities the attack of a nucleophile (*holo*-ACP or decarboxylated malonyl-ACP). When the acyl-enzyme intermediate reacts with *holo*-ACP or decarboxylated malonyl-ACP, a tetrahedral intermediate is formed in the oxyanion hole. For propionyl-ACP to leave the active site, the tetrahedral intermediate must collapse and transfer the negative charge to the Ser118 oxygen, which is stabilized by a protonated His297. The active sites of acetyl-, propionyl-, and butanoyl-DpsC are all very similar with respect to distances between the active site residues and acyl-serine carbonyl group with the oxyanion hole. The acyl binding region is the same in all DpsC acyl-enzyme crystal structures and consists of Leu93, Leu86, Val117 (Figure 3-9). Leu93 is located on loop L3 from the opposite monomer and is a major determinant of pocket length. It is currently not known how only the propionyl starter unit is incorporated into biosynthesis.

There are multiple hypotheses regarding the above phenomenon. In the PKSs, the proof-reading thioesterases have been shown to remove incorrect starter units from ACP, which ensures the biosynthesis of correct products(31, 32). Additionally, an AT was shown to hydrolyze
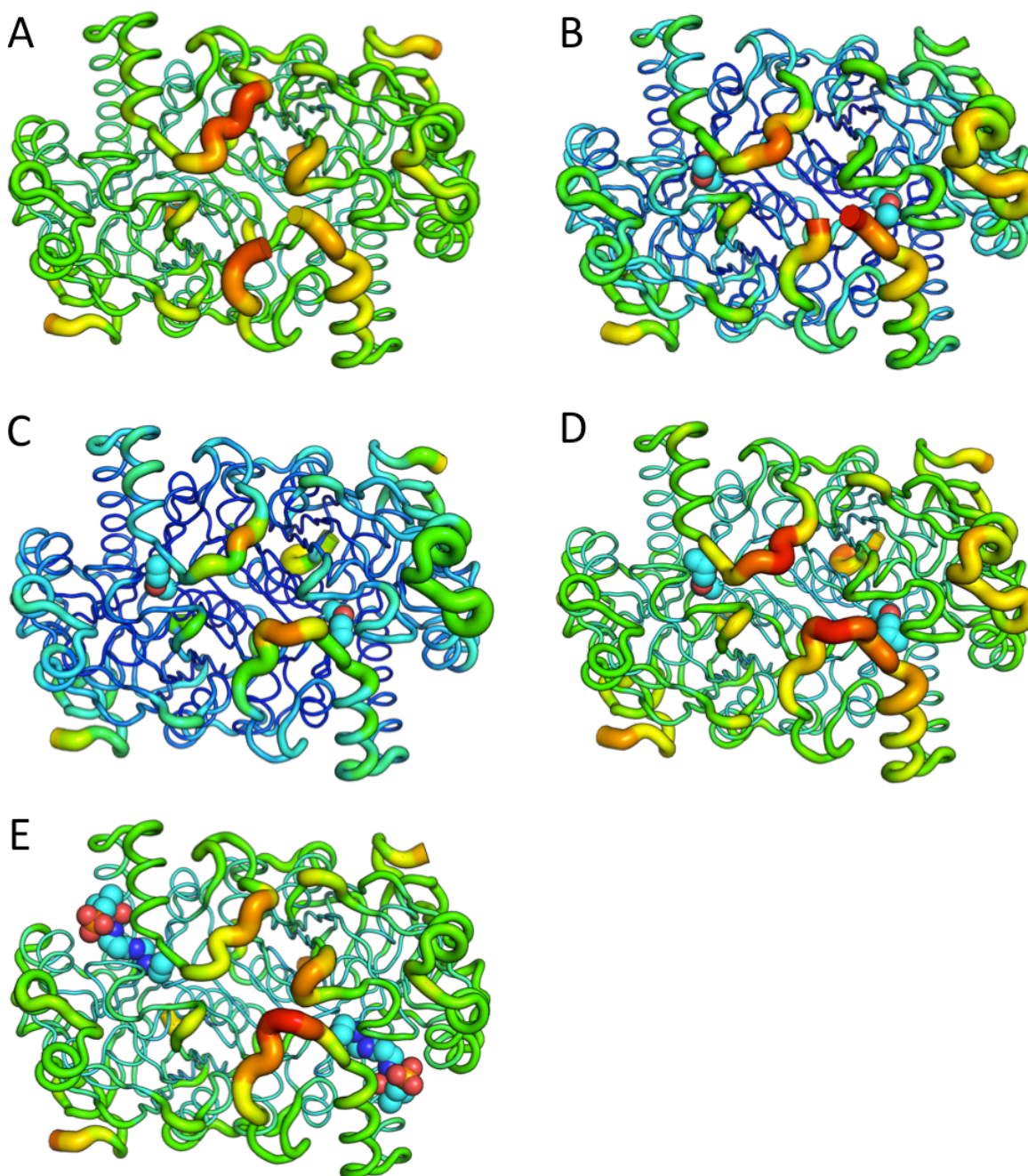
Figure 3-10. B-factor representations of DpsC, acetyl-DpsC, propionyl-DpsC, butanoyl-DpsC, and DpsC substrate analogue (**4**) co-crystal. (A) B-factor representation of DpsC. (B) B-factor representation of acetyl-DpsC. (C) B-factor representation of propionyl-DpsC. (D) B-factor representation of butanoyl-DpsC. (E) B-factor representation of DpsC substrate analogue (**4**) co-crystal.

incorrectly-loaded acyl-ACPs in a type I PKS(33). These results demonstrate that PKSs are capable of using different enzymes to ensure that ACP is loaded with the correct acyl starter unit. Biochemical experiments are currently underway to understand the role DpsC might play in a starter unit editing mechanism during daunorubicin biosynthesis.

3.3.4 DpsC Co-crystal Structure With a Substrate Analogue

Studying protein-protein interactions between an ACP and its partner enzymes is difficult because of the transient nature of these interactions(34). The idea of mixing and matching enzymes and domains from different biosynthetic pathways to generate new products is an attractive idea. Despite the potential in this area, it is difficult to successfully mix and match enzymes from different pathways because we lack specific knowledge of protein-protein interactions (35). Understanding the role of protein-protein interactions between biosynthetic enzymes is necessary for effective combinatorial biosynthesis (Figure 3-11). To study ACP/DpsC protein-protein interactions, we developed a propyl-amide pantetheine analogue (**3**) that could be loaded onto the ACP and induce stable complex formation with DpsC (Figures 3-12 and 3-13). Compound **3** was synthesized by coupling a protected pantetheine analogue (**1**) with propanoic acid, and deprotection of the coupled product yielded **3**, which mimics both the propionyl-phopsopanthetheine of the propionyl-CoA substrate and the product propionyl-ACP. **3** can be loaded onto ACP using enzymes involved in CoA biosynthesis and Sfp, a phosphopantetheinyl transferase, to yield **4** (Figure 3-13). Results of DpsC complex formation using **4** will be discussed in section 3.3.5. Mimic **3** can also be chemoezymatically phosphorylated using the enzyme CoaA at the terminal hydroxyl position to yield propionyl-phopsopanthetheine (**5**), which was co-crystallized with DpsC to reveal the pantetheine binding

113

tunnel and determine if the propionyl group binds in the same location and orientation as in the acyl-DpsC structures.

In the DpsC co-crystal structure, **5** is located in the tunnel leading to the active site of DpsC in both monomers, although the electron density for **5** is slightly better defined in monomer A (Figure 3-14E). Lys221 (L9-α2) and Arg271 (Cα2) interact strongly with the terminal phosphate of **5,** demonstrating the importance of the pPant phosphate for binding (Figure 3-14A-B). Moving towards the active site, the dimethyl moiety of **5** binds a
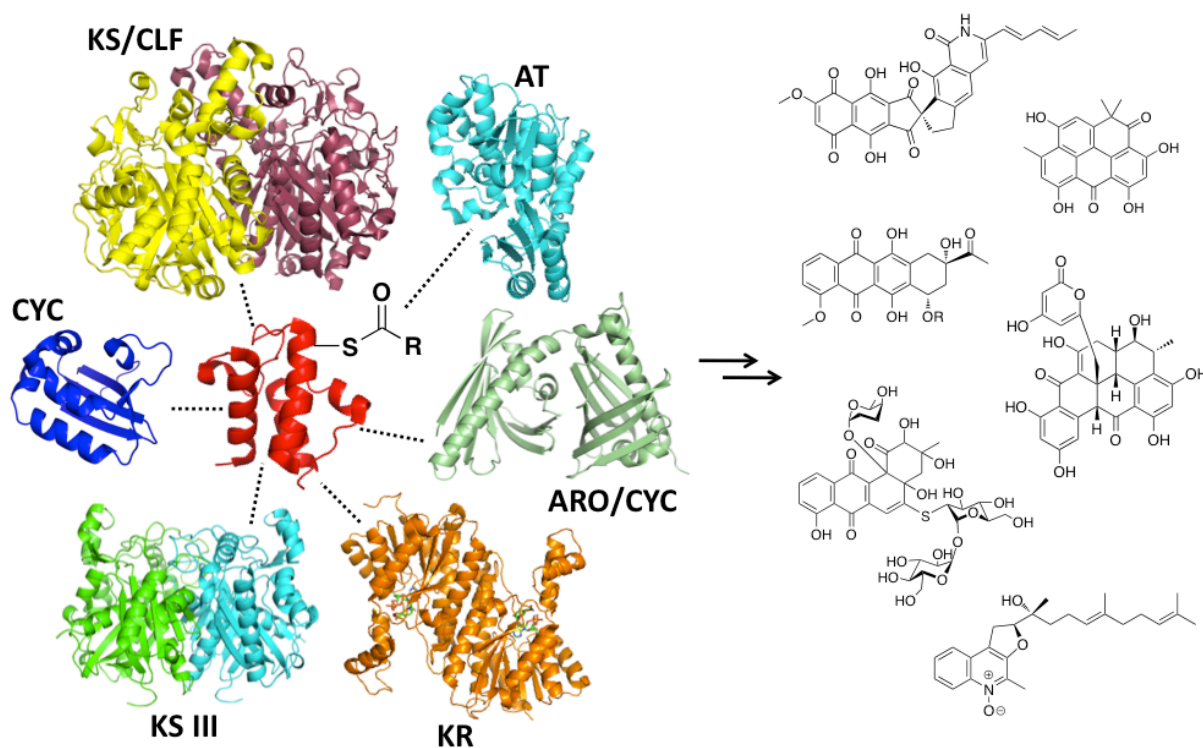
Figure 3-11. The role of protein-protein interactions in type II polyketide biosynthesis. Protein-protein interactions between the ACP and other enzymes in type II PKSs produce a variety of structurally complex natural products.

hydrophobic pocket between Cα1, L9-α2, and L9-α1. The hydroxyl group of **5** does not have any direct hydrogen bonds with DpsC; however, a water molecule is tightly coordinated between this hydroxyl group, the terminal phosphate, and an amide proton of **5** (Figure 3-14C). The two

central carbonyl groups of **5** interact with the same water molecule and also do not have direct interactions with DpsC. The ethylene group, on the other hand, has hydrophobic interactions with Val327 (Cβ3). Finally, the propyl-amide binds the active site and is poised for acyl transfer onto Ser118, with the carbonyl group 3.0 Å away from Ser118 (Figure 3-14B). The amide carbonyl group forms hydrogen bonds with the backbone amide protons of Ala329 and Ser118, which is very similar to the interaction observed for the ester carbonyl groups in all other acyl-DpsC co-crystal structures. The hydrophobic propionyl moiety binds a hydrophobic cleft formed
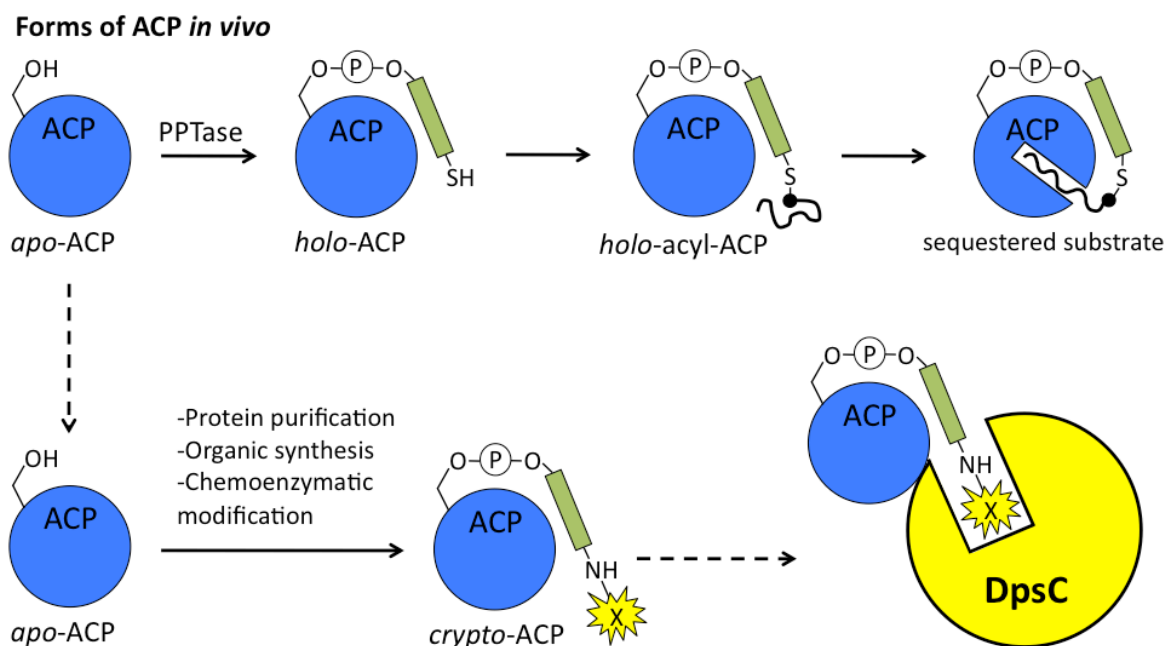
Figure 3-12. Forms of ACP *in vivo* and a schematic diagram of loading ACP with chemical probes to stabilze ACP/protein complexes *in vitro*. ACP is post-translationally modified with phosphopantetheine (pPant) to form *holo*-ACP. The growing polyketide is linked via a thioester bond to the thiol of pPant. ACP has also been shown to sequester, or protect, the growing intermediate during biosynthesis. *Apo*-ACP can be purified and chemoenzymatically loaded with pPant analogues that contain specific chemical groups that stabilize protein-protein interactions between ACP and target enzymes.

by Leu217, Leu93, Leu86, and Val117 (Figure 3-14B). When overlaid with the propionyl-DpsC structure, the DpsC co-crystal structure with **5** shows that the propionyl group of Ser118 is in almost the exact same orientation as the propionyl group of **5** (Figure 3-14D). In summary, the

115
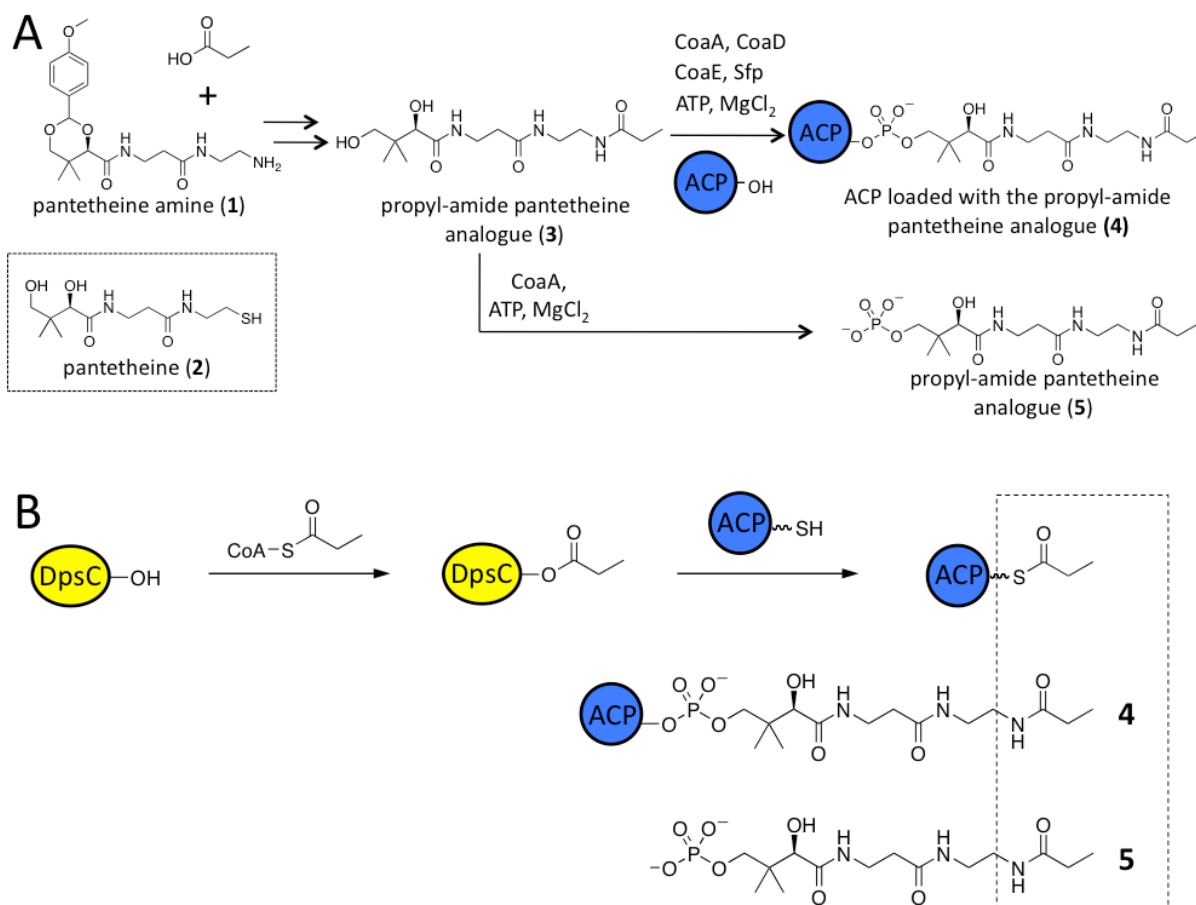
Figure 3-13. Chemical modification of ACP using mechanism-based pantetheine analogue probes. Pantetheine amine (**1**) is a protected analogue of pantetheine (**2**) that can be coupled with a carboxylic acid, and after deprotection it yields pantetheine analogues with stable amide linkages. These analogues can be loaded onto ACP *in vitro* using enzymes from the CoA biosynthetic pathway, or it can be directly phosphorylated to form substrate mimics for co-crystallization such as (**4**). (B) The DpsC acyltransfer reaction yielding propionyl-ACP. Propionyl-ACP with an amide linkage (**3**) can be generated using this methodology. The propionyl-ACP containing the amide linkage can then be used to stabilize the ACP/DpsC complex *in vitro*. *Dr. Gaurav Shakya synthesized the propyl-pantetheine analogue 3.

DpsC co-crystal structure shows that **5** extends deep into the active site, and when linked to the

ACP, both the propionyl and pPant groups may stabilize the interaction between the ACP and

DpsC.

116
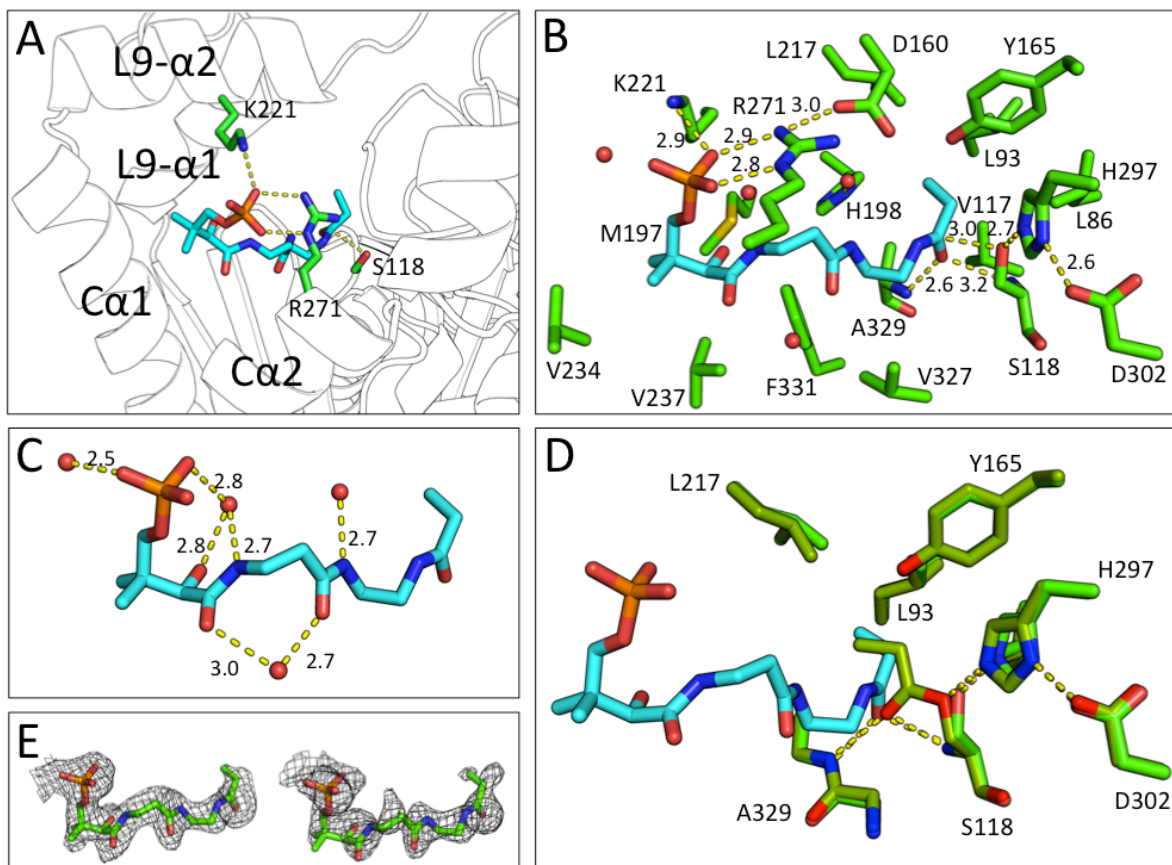
Figure 3-14. The DpsC active site and residues that interact with phosphopantetheine analogue (**5**). (A) A view looking from the solvent towards the active site tunnel of DpsC. Lys221 (α-7) and Arg271 (α-10) anchor the phosphate of **5** and the dimethyl moiety of **5** sits between α-7 and α-8. (B) The substrate analogue **4** is bound in the active site of DpsC by a series of hydrogen bonds, hydrophobic interactions, and charge-charge interactions. Lys221 and Arg217 form salt bridges to the phosphate moiety of the phosphopantetheine analogue at the solvent exposed face. The hydrophobic groups of the pantetheine portion of the analogue interact with Met197, Val234, Phe331, and Val327. The propionyl moiety sits in the active site with the amide carbonyl forming hydrogen bonds the amide hydrogens of the backbone residues Ala329 and Ala118. The hydrophobic portion of the propionyl moiety forms hydrophobic contacts in the acyl-binding region with Leu86, Leu93, Val117, and Leu217. (C) A series of water molecules forms hydrogen bonds with the phosphopantetheine moiety. (D) An overlay of the propionyl-DpsC and DpsC co-crystal with **5** showing similar orientations of the propionyl group. (E) SA omit maps of **5** bound to DpsC (monomer A, top, and monomer B, bottom).

3.3.5 ACP/DpsC Complex Formation Studies

To generate a stable ACP/DpsC complex for crystallization studies, **3** was chemoenzymatically loaded onto DpsG, the ACP from the daunorubicin biosynthesis pathway,
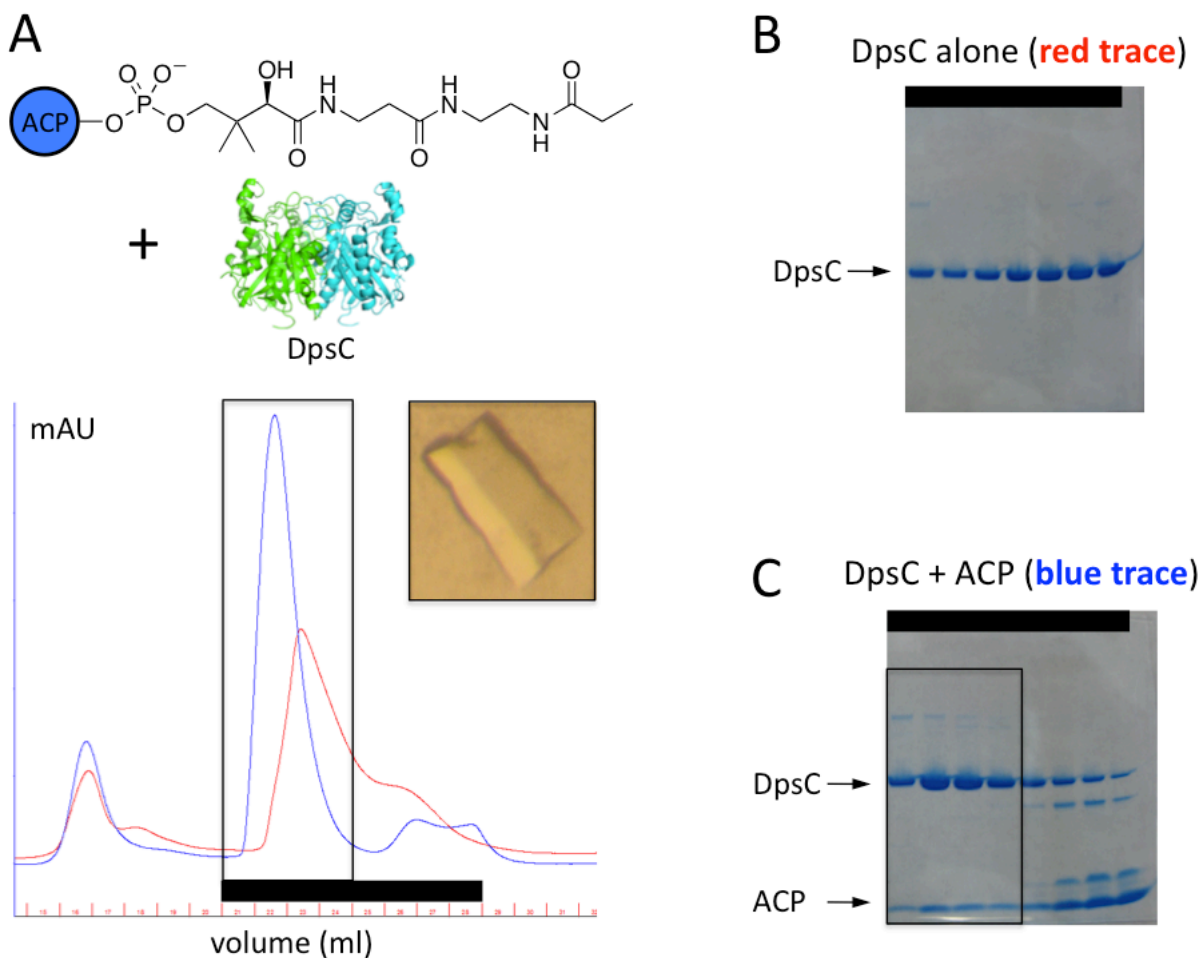
Figure 3-15. Formation of the ACP/DpsC complex and analysis by size exclusion chromatography. (A) The ACP/DpsC complex can be purified using gel filtration chromatography (blue). DpsC alone (red trace) elutes after the ACP/DpsC complex suggesting a smaller size in solution than the ACP/DpsC complex. Crystals were also generated using the purified ACP/DpsC complex but were found to only contain DpsC. (B) SDS-PAGE analysis of DpsC alone after gel filtration. The black bar above the gel lanes corresponds to fractions collected under the black bar on the gel filtration chromatogram. (C) SDS-PAGE analysis of the ACP/DpsC complex after gel filtration. ACP co-elutes with DpsC suggesting the formation of a stable ACP/DpsC complex.

to form **4** (Figure 3-13A). **4** was then incubated with purified DpsC and found to induce complex formation in solution and the ACP/DpsC complex was purified by gel filtration chromatography (Figure 3-15). The ACP/DpsC complex was crystallized in a different crystallization condition than DpsC. Intensity data was collected and the structure solved to 2.6 Å, but the ACP was not present in the electron density. To determine if the ACP was present in the crystals, individual

Figure 3-16. Glycosylated natural products containing an acylated hydroxyl or thiol and putative acyltransferase enzymes identified from a protein BLAST search of DpsC. (A) Natural products with DpsC homologues present in their gene clusters. Possible acylation sites are highlighted in bold. (B) A table describing DpsC homologues identified by a protein BLAST from known gene clusters that produce products that contain acylated substructures.

crystals were extensively washed and then dissolved in acetonitrile for MALDI analysis. MALDI

revealed a clear mass signal for **4** present in the crystals; therefore, it was surprising when the

ACP was not present in the electron density. It is possible that ACP was tightly associated to the outside of the crystal lattice but not present in the internal repeating asymmetric unit. The stable ACP/DpsC complex may have dissociated during crystal formation and excluded the ACP during lattice formation. This may explain why the ACP/DpsC complex grew in a different crystal condition that DpsC alone.

3.3.6 Bioinformatic Analysis Linking DpsC to AT/KSs

The active site of DpsC contains an atypical Ser-His-Asp catalytic triad, which has elements of a KS active site and also an AT active site. DpsC contains a thiolase fold, which is characteristic of KSs, and DpsC can act as both a KS and an AT. Taken together, these observations suggest that DpsC sits at the crossroads between a KS and an AT. A protein BLAST search using the amino acid sequence of DpsC revealed many uncharacterized bacterial enzymes annotated as KS III-like enzymes. Although many of the enzymes identified by BLAST have no verified function, a small set of enzymes from known biosynthetic gene clusters were found. The AT CerJ, which is discussed above, shares the highest sequence identity with DpsC among BLAST hits with known structures. A series of other putative KS IIIs were identified along with the natural product produced from their biosynthetic gene clusters and aligned in Figures 3-16 and 3-17. Interestingly, most of the putative KS III enzymes are from gene clusters whose products contain acylated glycosides, such as the gene cluster containing CerJ. By visual inspection, we identified acylation sites in the natural products from the biosynthetic pathways with DpsC homologues (Figure 3-16). The DpsC homologues in these gene clusters may use acyl-CoAs as substrates to acylate glycosides similar to CerJ; therefore, their active sites are expected to have residues and sizes that reflect their corresponding substrate specificity.

```
         1        10        20        30        40        50
DpsC    MSVPQGAPGDLYVAGCGVWLPPPVTTEQALAAGHCDRRLASSTRMLSVAVA...DKETPAE
CosE    .....MNDGALYIRATGTWLPPAERVADAVADGRCEAAVARSTRMESVTVS...AGEPAPE
CerJ    .....MRWENLFVSGVAAWLPPLSTAQDAVMAGLLDPARSKLRGIESVTVASDAEEDAPPR
SsfG    .......MYIRSVASYLPPRTSTQQAVDDGLYDAKDLGFMQLTGALFAG...DTSPAE
PokM2   MQGQDLYVRSLGVHLPEPVSTAQAVAEGRLDAEDVADGELTGVRVAG...DVPAPE
PlaP2   .....MRIRDIYLAGIGVFLPTAQSIGSAVKQGLVPAGTAAEYGLSGVVVAG...RTAAPE
ChlB3   .....MRTPDIFIKATGRFLPETVSVEWAVEQGHYSAEDAELHELGGAAVAG...DTPAPD
Cal04   .....MRTAGTYIRGIGAYLPETVTVEEAVAQGLYPQEDIETHGLGGAAIAG...ELPAPD
AviI    .....MQIADLYIGALGVFVPPCVSVEWAVERGLYPAEEAEAHELGGVAIAG...DVPPPE

         60        70        80        90       100       110
DpsC    MAALAAQTALDRSGVAPAHVDLVLHASLYFQGHHLWAPSSYVQRVAVGNRCPAMEVRQVSN
CosE    MAARAARTALERAAAEPGDIDLVLHASFYYQGFDLWAPASYVQRAAVGNQCPAMEIRQVSN
CerJ    MAARAARAALGRGDVDPADVSLVLHSSLWFQGIDLWPAASYVAHEAVCRHVPAFGLAQRCN
SsfG    MATAAARTALERGGVTSQELDYLIHSNVYQPGPHAWSLPGYILRELGGGKASVLELRTGCS
PokM2   LAVRAARQLFERVGADPSVVDLILYASVWHQGPDGWGPQSYVQRHLGG.EALAVELRHGCT
PlaP2   MALAAAQEALESSGTSAEDIGLLLYTGVWHQGPDGWPQAYLQRHLLGDDLLAMEVRNGCN
ChlB3   MALWAAQQAVKRCGHRPEDLGLLLYVDSWHQGPDGWPQYYLQRHLVGGDVLAVEIQQGCN
Cal04   MALRAAQDALKESELGRGDIDLLLYAAAWHQGPEGWLAHSYIQHYLLGGVPRATEIRQGCN
AviI    MALRAAQQAVKRWGGSPTEFDLLLYASTWHQGPDGWPPQAYLQRHLVGGDMLALEIRQGCN

        120       130       140       150       160       170
DpsC    GGMAALELARAYLLAAP.DRVAALITTGDRMHPPGFDRWSSDPG.TVYADGGTALVLSRQG
CosE    GGMAALDLAWAYLAADP.ARVAALLTTGDRFCPPGFDRWRSDPG.TVYGDGGTALVVSRHS
CerJ    GGMGAIELAGAYLGSGIGAGHAALLTTGDRFAGPRIDRWNSVDV.TMYGDGAAALVLSTRD
SsfG    GMLSSVELAVAQLTGAQ.RYSNVLLGAENFDPWMERWKSDVY..ITGDAASALLLSSTS
PokM2   GVFSGLQLAAGFLRGNP.QSSDALLGAENFGTPLIDRWRATGGMLVGDGASALLLGKEP
PlaP2   GTFGALELAAGYLKAFP.HVPAALVTASDNFGTPLIDRWNPGDGVAYLGDGASAVVVSTTP
ChlB3   GMFSALELAAAHLRAGP.RPGSALVVAADNFGTPLFDRWTTGPG.YIAGDGAAALVVLTTEP
Cal04   GMFTMLELAASYLKAAP.ERKAAMLVAADNYGTPLLDRWRTNLG.FILGDAASAVVLSTES
AviI    GVFSALELAVGYLQADP.NRTSALIVAADNYGTPLIDRWRMGPG.FIGGDALPPWCCK.RP

        180       190       200       210       220       230
DpsC    GFARLRSLVTVSEPVLEGMHRGGHPFGPPSPEEQRAVDLDAH....KRAYVAEAGSSFSVA
CosE    GFARVRSIATVSAPDLEGMHRGDDPFGPAPFSHRDVVDLDAH....KRDFLARHGVAPTVA
CerJ    GFARVLSTATGVDNSLEILARGDEPFAPHPVEPSPVADLGTR....TVRGAELADLPDLTH
SsfG    GFARVVSAASGSSPELEKMNRGSEPLHPPGETA..VVDLRVR....HAVVTDTMGDREAAN
PokM2   GFAKLLWVDCVAVPELEAMHRSGEPMFPPPGATIGRRLDFSER...NQHFLKYEGEQMRPLL
PlaP2   GLAELRSLCTATFSAMEEAHRGGEPLFPPGATTATALDYGVRGAAFQRVAEEDGSWVRLLL
ChlB3   GFARLLAVRSLAVPEAEQMHRGAEPLFPPGATIGRPLNFTSRNAAFRELSLTEGLGTGALM
Cal04   GFVELMSVCSITVPEAEEVHRGGEPMFPPPGATLAKELDFGAR...LFYHITEQTPVLAVLG
AviI    GFARLCSVASKGLPEIESLHRGDEPLFPPSITRGRATDFSAR...IGQQFATRSPAFVAMA

        240       250       260       270       280       290
DpsC    RVSAGQEEALTGALEAAGAGLDDISRVVLPHMGWRRLSAAYFNKWHIQPERTTWEFGRRTG
CosE    RVAAAQEAVVKQALADAGLGLADIDRFVLPHLGHRRLHVAYFAKFGIDPARSTWSWSRTIG
CerJ    RYIDLLVAAKTQALEDAGTAIEDIAHAVIP.VSRRGTGHELHDLLGLPDERTSWAYGRTTG
SsfG    IMAAQYVELAERAAADGGIKLADVDKLVYHNMAG.FIVGFFAEQMGLSLEHSTWEFGRTLG
PokM2   ..MKAHQDLYDRWTAATGLGVEDVTFLSYMNTSRTTIEKRLLPLLGLPAERTTWEVGRDIG
PlaP2   GHQRHNVVCTEQALEEAGVTADDITHVLIHGMPRRAAAS.YLRILGFPLERSTWDFSRTVG
ChlB3   RVHQRTLEVVEKTLSEAGITLGDITRVAYMNFSREIVEQRCMAALGLPMSASTWEFGRKLG
Cal04   EAQETMTTVAEQALAEAGIGTGDLAKVSFMNYSREVVEQRCMAPLGLGMEKSTWDFGRMIG
AviI    EIQDHMMDVAERALAGAGIGMADVARVSFMNYHAKWSSSAEWRPGGCRCPGPRGNSAGRSA

        300       310       320       330       340       350
DpsC    HLGGGDPIAGFDHLVGSGRLAPGELCLLVSVGAGFSWSCAVVELLERPSWAAAPAAR...
CosE    HLGAGDPFAGLDHLATAGELGPGDTCLLVGAGFDPASPSWRDR.......
CerJ    HVGAGDQYAGLAHLVENALVQPGDRVLLFGGGAGYTCTAAVVEILRMPAQ.......
SsfG    HLGCSDQAVGLEHLLLTGELSPGDRVMLLAASAGFATSCILLEILDVPDWPAPAPIAG..
PokM2   HLGVSDQLVALDRLLCDGRLKPGDHYLMYGLGPGMTVAGAMFEITEVPSWTR.......
PlaP2   HLGASDHMAALHHLLATERLHAGDHVLLCGFSPGVTYKAAVVRVLGTDPTRKREAQASGV
ChlB3   HLGASDQVVALDELVTTGELGPGDHLLMLGMGPGVTLSCAVVKVLTPAPWSD.......
Cal04   HCGASDHLLALHHSLRAGEVAAGDHVLWLAMGPGVEFTAAVLRVLDNPYVER.......
AviI    TCGASDHLLSMEHIVRTGELAPGDHVLQLATAPGLVVSSAVLQVLESPDWDE.......
```

Figure 3-17. Sequence alignment of putative DpsC homologues. The active site residues are marked with a red dot and residues in the acyl binding region are marked with a blue dot.

	Active Site Residues			Middle Pocket Residues			Active Site Loop	Relative Substrate Size
DpsC	S118	H297	D302	V117	Y165	L217	L93	small (propionyl)
CosE	S113	H292	D297	V112	Y160	L212	L88	
CerJ	C116	H295	D300	R115	Y164	L216	L91	
SsfG	C108	H282	D287	G107	T153	L210	A83	
PokM2	C112	H291	D296	G111	V160	F212	G88	
PlaP2	C113	H296	D301	G112	Y160	Y213	G88	
ChlB3	C113	H296	D301	G112	A160	F212	G88	
CalO4	C113	H293	D298	G112	L160	F212	G88	large (substituted benzoyl)
AviN	C113	T292	D297	G112	G160	F211	G88	
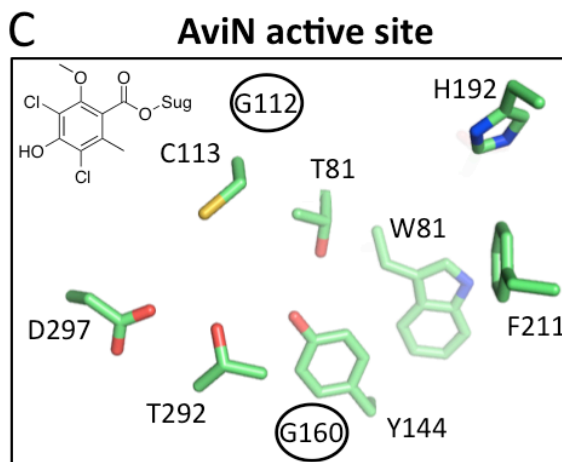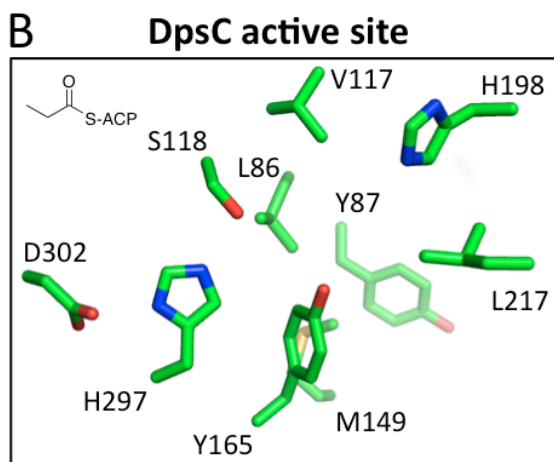
Figure 3-18. Analysis of active site and acyl binding region residues of DpsC homologues. (A) A table comparing the active site, middle pocket, and active site loop residues of DpsC homologues. (B) DpsC active site displaying residues in the acyl binding region. (C) The active site of the AviN homology model showing the effect of smaller middle pocket residues on pocket size. The AviN active site is much larger than the DpsC active site and also contains F211, which may form π-π stacking interaction with an aromatic substrate.

Homology models of DpsC homologues identified by BLAST were generated and overlaid with DpsC structure to determine their putative active site residues, especially with regard to substrate binding. A sequence alignment of DpsC homologues was also generated (Figure 3-16).

The table describing residues surrounding the active sites of the homology models is detailed in Figure 3-18. There is a clear trend that the size of the putative substrate is directly correlated with the size of the substrate pocket, as reflected by decreasing size of the side chains of residues that define the substrate pocket. For example, residues corresponding to Leu93 of

DpsC invariably change to glycine as the putative substrate size increases. There is also a clear switch in the middle pocket residue corresponding to Leu217 of DpsC as the substrate goes from aliphatic to aromatic. Further, the residue Leu217 of DpsC is located at the side of the active site pocket and changes to an aromatic residue, phenylalanine or tyrosine, as the substrate becomes aromatic. This trend at the Leu217 position may occur because of favorable π-π stacking between phenylalanine or tyrosine and an aromatic substrate. Finally, residues at the top and bottom of the active site pocket, corresponding to Val117 and Tyr165 of DpsC, also change to smaller residues as the substrate becomes larger (Figure 3-18). In summary, the above analysis of DpsC homologues with different substrate specificities not only helps determine residues important for substrate specificity, but it also has revealed insights into how the DpsC active site could be engineered to introduce new starter units into biosynthesis.

3.3.7 Conclusion

Five different DpsC crystal structures were reported in this work: apo-DpsC, three acyl-DpsC structures, and a DpsC co-crystal structure with the substrate analogue **5**. Collectively, these structures reveal that the DpsC Ser-His-Asp active site catalytic triad differs from other KS III active site triads (Cys-His-Asn). Rather, the geometry of the DpsC active site more closely resembles the active site of the acyltransferase CerJ, a first observation for a type II priming KS. All three acyl-DpsC crystal structures display the acyl group bound in a hydrophobic cleft at the end of the active site pocket, with the Ser118 acyl ester carbonyl bound in the oxyanion hole formed by the amide protons of Ala329 and Ser118. The Ser118 hydroxyl oxygen is stabilized by His297, which is further stabilized by Asp302. This increases the nucleophlicity of Ser118 and helps explain the acyltransferase activity of DpsC. The DpsC co-crystal structure with **5**

	Apo-DpsC	Acetyl-DpsC	Propionyl-DpsC	Butanoyl-DpsC	DpsC co-crystal with 5
Crystallization	0.18 M sodium citrate, 26 % PEG 3350	0.18 M sodium citrate, 26 % PEG 3350	0.18 M sodium citrate, 26 % PEG 3350	0.18 M sodium citrate, 26 % PEG 3350	0.18 M sodium citrate, 26 % PEG 3350
Crystallographic Data					
Wavelength (Å)	0.9999	0.9775	1.0000	0.9999	0.9795
Space Group	$P6_522$	$P6_522$	$P6_522$	$P6_522$	$P6_522$
Cell Dimension a, b, c (Å)	90.033, 90.033, 304.689	91.213, 91.213, 315.643	91.477, 91.477, 316.833	91.382, 91.382, 315.783	89.702, 89.702, 302.704
	$\alpha=\beta=90°, \gamma=120°$	$\alpha=\beta=90°, \gamma=120°$	$\alpha=\beta=90°, \gamma=120°$	$\alpha=\beta=90°, \gamma=120°$	$\alpha=\beta=90°, \gamma=120°$
Resolution (Å)	50.00 - 2.50	50.00-2.40	50.00 - 2.40	50.00 - 2.20	50.00-2.30
No. of observations	557086	500940	857439	411366	479239
No. of unique observations	26403	31516	40753	40709	33191
Completeness %	99.3 (98.8)	100.00 (100.00)	99.6 (99.2)	99.9 (100)	100.0 (100.0)
$I/\sigma(I)$	24.1 (6.0)	18.6 (4.3)	25.0 (5.3)	18.4 (2.6)	19.8 (6.0)
R_{merge} %	14.2 (65.0)	13.9 (79.9)	16.5 (75.9)	11.3 (81.9)	17.8 (44.3)
Redundancy	21.1	15.9	22.5	10.1	14.4
Refinement					
Resolution (Å)	50.00 - 2.50 (2.59 - 2.50)	50.00 - 2.40 (2.49 - 2.40)	50.00 - 2.4 (2.49 - 2.40)	50.00 - 2.20 (2.28 - 2.20)	50.00 - 2.30 (2.38 - 2.30)
No. of protein atoms	5058	5068	5102	5104	5073
No. of ligand atoms	0	0	0	0	52
No. of water atoms	173	248	572	431	316
R_{free} %	20.88 (23.47)	21.57 (29.88)	20.29 (24.10)	22.49 (31.98)	19.76 (23.02)
R_{crys} %	16.18 (18.96)	17.70 (20.80)	16.64 (18.10)	17.79 (25.03)	15.93 (17.81)
Geometry					
RMS bonds (Å)	0.008	0.023	0.009	0.008	0.008
RMS angles (°)	1.09	1.40	1.17	1.16	1.10
Ramachandran Favored (%)	98.00	96.00	97.00	96.00	98.00
Ramachandran Allowed (%)	1.55	3.55	2.56	3.41	1.70
Ramachandran Dissallowed (%)	0.45	0.45	0.44	0.59	0.30
Average B-factors (Å²)					
Protein	32.40	34.30	17.50	29.90	29.00
Water	33.80	35.30	26.40	35.70	40.20
Ligands	N/A	N/A	N/A	N/A	33.10

Table 3-1. DpsC crystallographic data collection and refinement statistics.
*Numbers in parentheses denote the highest resolution shell.

revealed important binding interactions between the pPant and DpsC and demonstrated how the acyl binding region orients the substrate. In an attempt to crystallize an ACP/DpsC complex, **5** was loaded onto the ACP (DpsG) and incubated with DpsC to form a complex. The ACP/DpsC complex was purified using gel filtration chromatography and crystallization was attempted; however, this did not yield an ACP/DpsC complex crystal structure. Nevertheless, this methodology proves that relatively simple pantetheine analogues can be used to stabilize non-covalent protein-protein interactions between an ACP and target enzymes.

3.4 Materials and Methods

3.4.1 Expression and Purification of DpsC

The pET-28b(+) (Novagen) derived DNA plasmid encoding N-terminal His-tagged DpsC (DpsC/pET28b) was generously provided by Dr. John Crosby. The DpsC/pET28b plasmid was transformed into *E. coli* BL21(DE3) competent cells and plated on LB-agar plates containing kanamycin (50 µg/mL). The plates were incubated overnight at 37 °C. Positive transformants were transferred to a 5 mL starter culture of Luria-Bertani (LB) broth containing kanamycin (50 µg/mL) and grown overnight at 37 °C with shaking, which was then used to inoculate one liter of LB with kanamycin (50 µg/mL). Cultures were grown at 37 °C until the A_{600} reached $0.4 - 0.6$. The cells were then cooled to 18 °C, and 0.1 mM IPTG was added to induce protein expression. After 12-18 hours of incubation at 18 °C, the cells were harvested by centrifugation at 5000 rpm for 10 minutes. The cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C. The frozen cell pellets were thawed on ice and resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10% glycerol, 10 mM imidazole). The cell suspension was lysed using sonication (8 x 30 s cycles), and cellular debris was removed by centrifugation at 14000 rpm for 45 minutes. The

lysate was incubated with 5 mL Ni-IMAC resin (BioRad) at 4 °C for one hour. The resin was poured into a fritted column and the flow-through fraction was collected. The resin was washed with 100 mL of lysis buffer then eluted with lysis buffer plus increasing amounts of imidazole (20-500 mM). The elutions were analyzed using SDS-PAGE, and the elutions containing the protein of interest were combined and dialyzed overnight into storage buffer (50 mM Tris pH 8.0, 100 mM NaCl, 10% glycerol). One unit of bovine thrombin (Sigma) per mg of DpsC was added to the combined protein elutions prior to dialysis to cleave the N-terminal His-tag. Following the dialysis, DpsC was incubated with Ni-IMAC resin and passed over a fritted column to remove any uncleaved DpsC and the flow through was analyzed by MALDI-TOF MS to confirm cleavage of the His-tag (data not shown). The cleaved DpsC protein was then concentrated to 5 mg/ml, buffer exchanged into crystallization buffer (25 mM Tris, pH 7.5), and flash frozen in liquid nitrogen.

3.4.2 Crystallization of DpsC, Acetyl-DpsC, Propionyl-DpsC, Butanoyl-DpsC, and DpsC + Propyl-pPant Analogue

Frozen aliquots of DpsC were thawed on ice then filtered with a 0.22 µM filter prior to crystallization. DpsC was crystallized using the sitting drop vapor diffusion method at room temperature from crystal seed stocks generated using a Seed Bead (Hampton). Crystals were grown by adding one microliter of 5 mg/mL protein solution to one microliter of well solution (0.18 M sodium citrate, 26% PEG 3350) containing crystal seeds and allowed to equilibrate with 500 µL of well solution. Three-dimensional hexagonal crystals formed over one week. For acetyl-DpsC, propionyl-DpsC, and butanoyl-DpsC, seeded wild type DpsC crystals were soaked with 5 mM acetyl-CoA, 5 mM butanoyl-CoA, or 5 mM propionyl-CoA 12-16 hours before being

frozen in liquid nitrogen. For co-crystallization with **5**, DpsC was incubated on ice for one hour with 5 mM of **5,** and diffraction quality crystallization was achieved with multiple rounds of seeding, starting with wild type DpsC crystals and subsequent rounds with co-crystal seeds. X-ray diffraction data was collected for apo-DpsC and all acyl-DpsC crystals at the Advanced Light Source (ALS) using either Beamline 8.2.1 or 8.2.2. The resolution of the data sets were 2.50 Å, 2.40 Å, 2.40 Å, and 2.20 Å, for apo-, acetyl-, propionyl-, and butanoyl-DpsC, respectively. Diffraction data for the DpsC-propyl pantetheine analogue **5** co-crystals was collected at the Stanford Synchrotron Radiation Laboratory (SSRL), Beamline 12-2, to a resolution of 2.30 Å. All data were processed using indexed, integrated, and scaled using HKL2000(36). Crystallographic data and refinement statistics are presented in Table 3-1.

3.4.3 DpsC Phasing, Model Building, and Refinement

DpsC crystallized in the space group $P6_522$ with two DpsC monomers forming a dimer in the asymmetric unit. The apo-DpsC structure was solved using molecular replacement (Phaser) with CerJ (PDB ID: 3S3L) as a starting model(37). A preliminary model was built using (PHENIX AutoBuild), which was then used for iterative rounds of model building (COOT) and refinement (PHENIX Refine)(38-40). The model was refined to an R_{work} of 16.2% and R_{free} of 20.9%. Acetyl-DpsC, propionyl-DpsC, butanoyl-DpsC, and DpsC co-crystal with **5** were all solved by molecular replacement using the apo-DpsC structure as a search model. Model building and refinement were carried out in a similar manner as described to DpsC above with a few exceptions. For acetyl-, propionyl-, and butanoyl-DpsC, after multiple rounds of refinement, there is well-defined electron density for the acyl groups attached to Ser118. The acyl-serine moiety was added using COOT and further refined. For the DpsC co-crystal with **5**, after

multiple rounds of refinement, there is well-defined electron density for **5**. The ligand was drawn in ChemDraw and converted to a PDB using the NCI SMILES converter server (http://cactus.nci.nih.gov/translate/). The ligand was manually docked in the DpsC active site using COOT with restraints generated using PHENIX Elbow(41). Data collection and refinement statistics for all structures can be found in Table 3-1.

3.4.4 DpsG Expression and Loading Reaction

A pET-28b(+) (Novagen) derived DNA plasmid encoding N-terminal His-tagged DpsG (DpsG/pET28b) was generously provided by Dr. John Crosby. The expression and purification of DpsG was accomplished as described above to DpsC with a few minor differences. To obtain apo-DpsG, cells were grown at 37 °C until the A_{600} reached 1, which was then induced with 0.1 mM IPTG and grown for an additional 4 hours at 37 °C. Cells were harvested, and DpsG was purified as described above for DpsC. Purified DpsG (400 µM) was incubated with CoaA (5 µM), CoaE (5 µM), CoaD (5 µM), Sfp (5 µM), $MgCl_2$ (5 mM), ATP (8 mM), and propyl-pantetheine amide **3** (800 µM) in 50 mM sodium phosphate (pH 7.5) for 2 hours. Loading of the pantetheine analogue **3** was confirmed using MALDI-TOF MS (data not shown) to generate the propyl amide-linked ACP **4**. The final reaction volume was 5 mL. **4** was further purified using a Hi Trap Q column (GE Healthcare). The DpsG loading reaction was directly applied to the Hi Trap Q column, and **4** was eluted with 25 mM Tris (pH 7.5), 300 mM NaCl, and 5 % glycerol. **4** was concentrated to 0.8 mM, flash frozen in liquid nitrogen, and stored at -80 °C.

3.4.5 Chemoenzymatic Phosphorylation of 3

Compound **3** was incubated with the pantetheine kinase CoaA (5 μM) in the phoshporylation buffer (25 mM potassium phosphate, 10 mM $MgCl_2$, 8 mM ATP, pH 7.5) for 3 hours at 37 °C to generate **5**. The final product **5** was purified by HPLC using a semi-prep C18 column with a linear gradient of water/acetonitrile containing 0.1 % (v/v) formic acid. Fractions containing **5** were collected and verified by ESI-MS then lyophilized, resuspended in water, and stored at -20 °C.

3.4.6 ACP/DpsC Complex Purification and Gel Filtration Studies

4 (200 μM) was incubated overnight at 4 °C with DpsC (100 μM) in 25 mM Tris (pH 7.5), 100 mM NaCl, and 5 % glycerol. The complex was purified using a Superdex 200 gel filtration column (GE Healthcare), and the protein fractions were analyzed by SDS-PAGE. Fractions containing both ACP and DpsC were pooled, concentrated to 8 mg/ml, and screened for crystallization.

3.4.7 Sequence Alignments and the Generation of Homology Models of DpsC Homologues

A search for proteins similar to DpsC was conducted using BLAST that searched for non-redundant protein sequences. Genes from known biosynthetic gene clusters were analyzed for products containing an acylated sugar moiety, and they were chosen for homology model generation using HHPred(42). Sequence alignments were conducted using ClustalW and graphics were generated using Espript (http://espript.ibcp.fr)(43, 44).

References

1.	Cragg, G. M., and Newman, D. J. (2001) Natural product drug discovery in the next millennium, *Pharm Biol 39 Suppl 1*, 8-17.
2.	Zhang, M. Q., and Wilkinson, B. (2007) Drug discovery beyond the 'rule-of-five', *Curr Opin Biotechnol 18*, 478-488.
3.	Harvey, A. L. (2008) Natural products in drug discovery, *Drug Discov Today 13*, 894-901.
4.	Weber, T., Welzel, K., Pelzer, S., Vente, A., and Wohlleben, W. (2003) Exploiting the genetic potential of polyketide producing streptomycetes, *J Biotechnol 106*, 221-232.
5.	Di Marco, A., Cassinelli, G., and Arcamone, F. (1981) The discovery of daunorubicin, *Cancer Treat Rep 65 Suppl 4*, 3-8.
6.	Nelson, M. L., and Levy, S. B. (2011) The history of the tetracyclines, *Ann N Y Acad Sci 1241*, 17-32.
7.	Lichtman, M. A. (2013) A historical perspective on the development of the cytarabine (7days) and daunorubicin (3days) treatment regimen for acute myelogenous leukemia: 2013 the 40th anniversary of 7+3, *Blood Cells Mol Dis 50*, 119-130.
8.	Bao, W., Sheldon, P. J., Wendt-Pienkowski, E., and Hutchinson, C. R. (1999) The Streptomyces peucetius dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide, *J Bacteriol 181*, 4690-4695.
9.	Rajgarhia, V. B., and Strohl, W. R. (1997) Minimal Streptomyces sp. strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis, *J Bacteriol 179*, 2690-2696.
10.	Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork, *Nat Prod Rep 24*, 162-190.
11.	Staunton, J., and Weissman, K. J. (2001) Polyketide biosynthesis: a millennium review, *Nat Prod Rep 18*, 380-416.
12.	Davies, C., Heath, R. J., White, S. W., and Rock, C. O. (2000) The 1.8 A crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from escherichia coli, *Structure 8*, 185-195.
13.	Moore, B. S., and Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units, *Nat Prod Rep 19*, 70-99.
14.	Piel, J., Hertweck, C., Shipley, P. R., Hunt, D. M., Newman, M. S., and Moore, B. S. (2000) Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate 'Streptomyces maritimus': evidence for the derailment of an aromatic polyketide synthase, *Chem Biol 7*, 943-955.
15.	Marti, T., Hu, Z., Pohl, N. L., Shah, A. N., and Khosla, C. (2000) Cloning, nucleotide sequence, and heterologous expression of the biosynthetic gene cluster for R1128, a non-steroidal estrogen receptor antagonist. Insights into an unusual priming mechanism, *J Biol Chem 275*, 33443-33448.
16.	Xu, Z., Schenk, A., and Hertweck, C. (2007) Molecular analysis of the benastatin biosynthetic pathway and genetic engineering of altered fatty acid-polyketide hybrids, *J Am Chem Soc 129*, 6022-6030.
17.	Kersten, R. D., Lane, A. L., Nett, M., Richter, T. K., Duggan, B. M., Dorrestein, P. C., and Moore, B. S. (2013) Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in Salinispora tropica, *Chembiochem 14*, 955-962.

18.   Zhang, W., Watanabe, K., Wang, C. C., and Tang, Y. (2007) Investigation of early tailoring reactions in the oxytetracycline biosynthetic pathway, *J Biol Chem 282*, 25717-25725.

19.   Kalaitzis, J. A., Cheng, Q., Thomas, P. M., Kelleher, N. L., and Moore, B. S. (2009) In vitro biosynthesis of unnatural enterocin and wailupemycin polyketides, *J Nat Prod 72*, 469-472.

20.   Zhang, W., Ames, B. D., Tsai, S. C., and Tang, Y. (2006) Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module from the oxytetracycline polyketide synthase, *Appl Environ Microbiol 72*, 2573-2580.

21.   Xu, Z., Metsa-Ketela, M., and Hertweck, C. (2009) Ketosynthase III as a gateway to engineering the biosynthesis of antitumoral benastatin derivatives, *J Biotechnol 140*, 107-113.

22.   Bao, W., Sheldon, P. J., and Hutchinson, C. R. (1999) Purification and properties of the Streptomyces peucetius DpsC beta-ketoacyl:acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis, *Biochemistry 38*, 9752-9757.

23.   Pan, H., Tsai, S., Meadows, E. S., Miercke, L. J., Keatinge-Clay, A. T., O'Connell, J., Khosla, C., and Stroud, R. M. (2002) Crystal structure of the priming beta-ketosynthase from the R1128 polyketide biosynthetic pathway, *Structure 10*, 1559-1568.

24.   Bretschneider, T., Zocher, G., Unger, M., Scherlach, K., Stehle, T., and Hertweck, C. (2012) A ketosynthase homolog uses malonyl units to form esters in cervimycin biosynthesis, *Nat Chem Biol 8*, 154-161.

25.   Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J Mol Biol 372*, 774-797.

26.   Sachdeva, S., Musayev, F., Alhamadsheh, M. M., Neel Scarsdale, J., Tonie Wright, H., and Reynolds, K. A. (2008) Probing reactivity and substrate specificity of both subunits of the dimeric Mycobacterium tuberculosis FabH using alkyl-CoA disulfide inhibitors and acyl-CoA substrates, *Bioorg Chem 36*, 85-90.

27.   Kortemme, T., and Creighton, T. E. (1995) Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family, *J Mol Biol 253*, 799-812.

28.   Hol, W. G., van Duijnen, P. T., and Berendsen, H. J. (1978) The alpha-helix dipole and the properties of proteins, *Nature 273*, 443-446.

29.   Serre, L., Verbree, E. C., Dauter, Z., Stuitje, A. R., and Derewenda, Z. S. (1995) The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-A resolution. Crystal structure of a fatty acid synthase component, *J Biol Chem 270*, 12961-12964.

30.   Polgar, L. (2005) The catalytic triad of serine peptidases, *Cell Mol Life Sci 62*, 2161-2172.

31.   Kalaitzis, J. A., Cheng, Q., Meluzzi, D., Xiang, L., Izumikawa, M., Dorrestein, P. C., and Moore, B. S. (2011) Policing starter unit selection of the enterocin type II polyketide synthase by the type II thioesterase EncL, *Bioorg Med Chem 19*, 6633-6638.

32.   Hu, Z., Pfeifer, B. A., Chao, E., Murli, S., Kealey, J., Carney, J. R., Ashley, G., Khosla, C., and Hutchinson, C. R. (2003) A specific role of the Saccharopolyspora erythraea thioesterase II gene in the function of modular polyketide synthases, *Microbiology 149*, 2213-2225.

33. Jensen, K., Niederkruger, H., Zimmermann, K., Vagstad, A. L., Moldenhauer, J., Brendel, N., Frank, S., Poplau, P., Kohlhaas, C., Townsend, C. A., Oldiges, M., Hertweck, C., and Piel, J. (2012) Polyketide proofreading by an acyltransferase-like enzyme, *Chem Biol 19*, 329-339.

34. Bruegger, J., Haushalter, B., Vagstad, A., Shakya, G., Mih, N., Townsend, C. A., Burkart, M. D., and Tsai, S. C. (2013) Probing the selectivity and protein.protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers, *Chem Biol 20*, 1135-1146.

35. Menzella, H. G., and Reeves, C. D. (2007) Combinatorial biosynthesis for drug development, *Curr Opin Microbiol 10*, 238-245.

36. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol 276*, 307-326.

37. Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography 40*, 658-674.

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

39. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D 64*, 61-69.

40. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D 68*, 352-367.

41. Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Crystallogr D 65*, 1074-1080.

42. Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res 33*, W244-W248.

43. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res 42*, W320-324.

44. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Res 31*, 3497-3500.

Chapter 4

Structural Insights Into Anthranilate Priming During Type II Polyketide Biosynthesis

4.1 Abstract

The incorporation of non-acetate starter unit during type II polyketide biosynthesis helps to diversify natural products. Currently, there are few enzymatic strategies for the incorporation of non-acetate starter units in type II PKS pathways. Here we report the crystal structure of AuaEII, the anthranilate:CoA ligase responsible for the generation of anthranoyl-CoA, which is used as a starter unit by a type II PKS during aurachin biosynthesis. We present structural and protein sequence comparisons to other known aryl:CoA ligases. We also compare the AuaEII crystal structure to a model of a CoA ligase homologue, AuaE, which is present in the same gene cluster. AuaE is predicted to have the same CoA ligase fold as AuaEII, but instead of CoA ligation, AuaE catalyzes acyl transfer of anthranilate from anthranoyl-CoA to the acyl carrier protein (ACP). Together, this work provides insight into the molecular basis for starter unit selection of anthranilate during type II PKS biosynthesis.

4.2 Introduction

Type II polyketide biosynthesis typically begins by loading an acetyl starter unit onto the terminal thiol of 4'-phosphopantethine (pPant) linked to an acyl carrier protein (ACP), which is next transferred to the active site cysteine of ketosynthase (KS) and then undergoes iterative elongation using malonyl-CoA as a building block(1). In some cases, different (non-acetyl) starter units are employed in polyketide biosynthesis, thus yielding more structurally diverse natural products(2). For example, priming ketosynthases use an acyl-CoA as a substrate for one
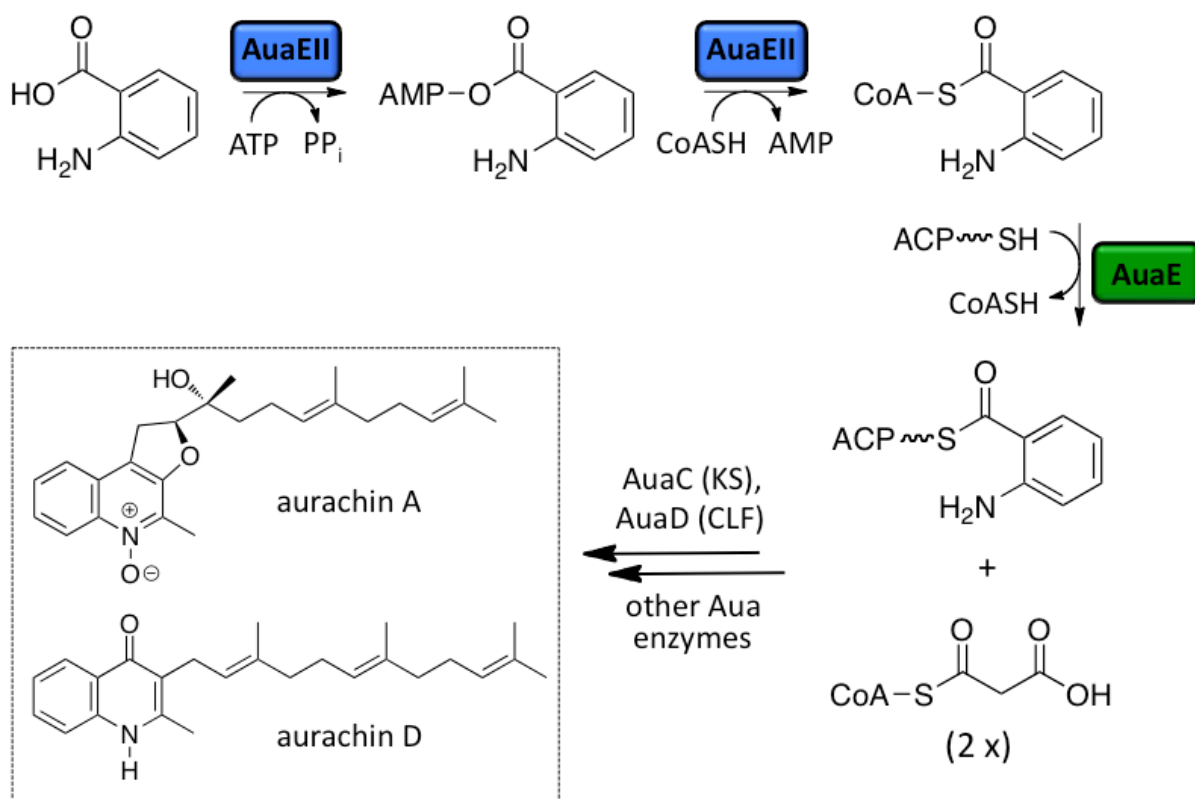
Figure 4-1. The roles of AuaEII and AuaE in aurachin biosynthesis. Aurachin biosynthesis begins with the activation of anthranilic acid into a CoA ester that is catalyzed by AuaEII. AuaEII first activates anthranilic acid using ATP to form anthranoyl-AMP, which is susceptible to nucleophilic attack by CoA to yield anthranoyl-CoA. AuaE uses anthranoyl-CoA as a substrate and transfers the anthranoyl group onto ACP (AuaB) that is then elongated by a type II PKS (AuaB, AuaC, and AuaD). Subsequent elongation and cyclization yields a quinoline core, which is modified by farnesyl group. Additional reactions yield a variety of aurachin natural products (only aurachin A and aurachin D are pictured).

round of extension using malonyl-CoA and then the elongated substrate is transferred to the ketosynthase active site cysteine. The growing polyketide is then further extended by the ketosynthase/chainlength factor (KS/CLF)(3, 4). Non-acetate starter units have been observed in the daunorubicin (propionyl-CoA) and R1128(propionyl-CoA and isobutyryl-CoA) biosynthetic pathways(4, 5). Further, other strategies for diversifying starter unit incorporation have been discovered. For example, a bifunctional decarboxylase/acyltransferase (AT) LomC was identified in the lomaiviticin gene cluster and found to decarboxylate methylmalonyl-CoA to form propionyl-CoA, which is then loaded onto the ACP and incorporated into lomaiviticin biosynthesis(6). Additionally, the enterocin pathway harbors the enzyme EncN, which uses ATP to activate and transfer benzoic acid onto the ACP that is then elongated by KS/CLF(7). The incorporation of non-acetate starter units is important for the generation of molecular complexity in polyketide biosynthesis. Therefore, there is a need to understand the structural and mechanistic basis for the enzymatic reactions responsible for starter unit selection. This work extends our current understanding of novel starter incorporation by a comprehensive analysis of the crystal structure of the anthranilate:CoA ligase AuaEII, which is involved in anthranilate priming during aurachin biosynthesis.

Aurachins are a class of quinilone alkaloids produced by *Stigmatella aurantiaca* Sg a15 that inhibit the electron transport chain(8, 9). A gene cluster responsible for aurachin biosynthesis was previously identified and sequenced(10). Two CoA ligase homologues, AuaEII and AuaE, were found to be responsible for a unique anthranilate priming mechanism that generates and incorporates the anthranilate starter unit into aurachin biosynthesis (Figure 4-1)(11). In *vitro* experiments revealed that AuaEII is an anthranilate:CoA ligase and AuaE is an anthranoyl-CoA:ACP acyltransferase (AT)(11). Surprisingly, both AuaEII and AuaE are
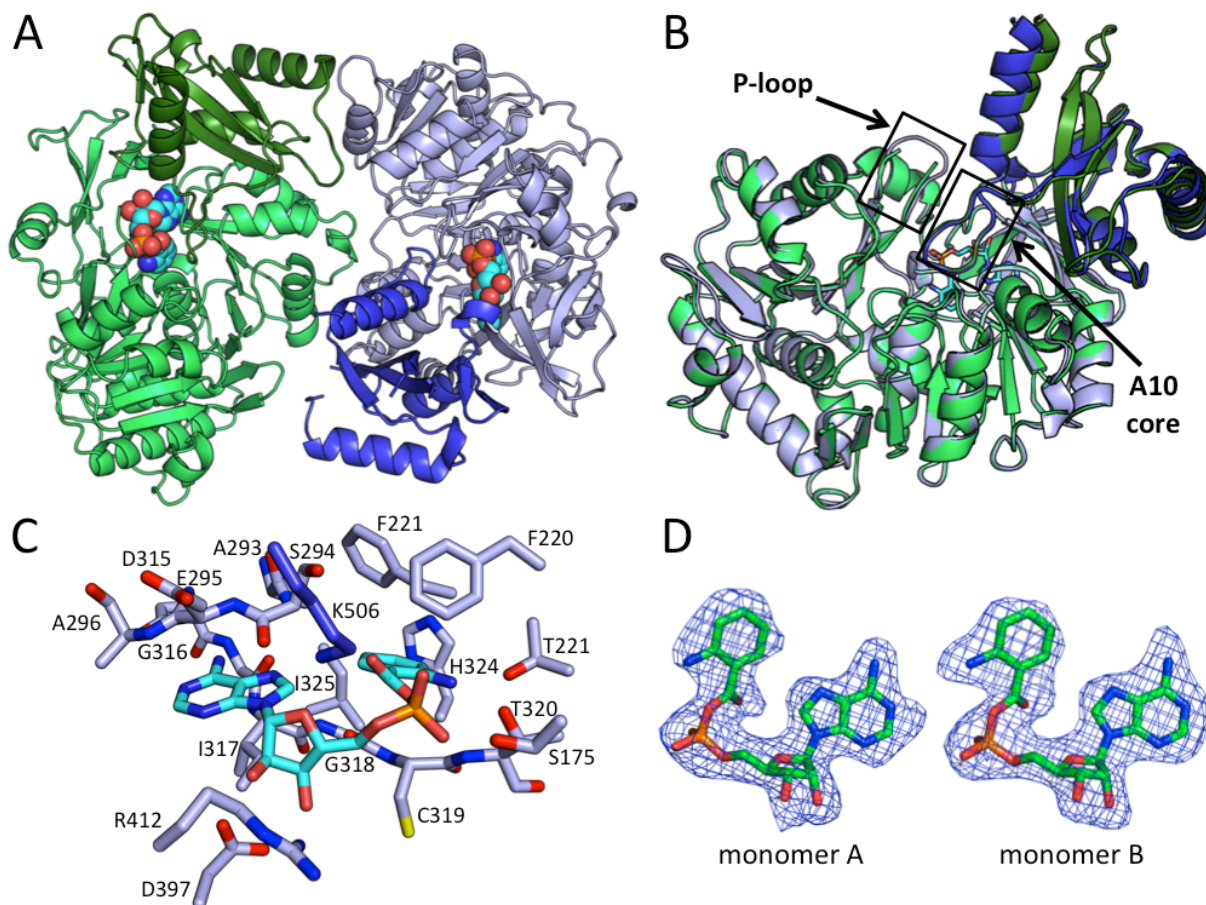
Figure 4-2. The crystal structure of AuaEII bound to anthranoyl-AMP. (A) The crystal structure of AuaEII is composed of two monomers per asymmetric unit that are related by non-crystallographic symmetry. The N-terminal domains of monomers A and B are colored light green and light blue respectively, and the C-terminal domains of monomers A and B are colored green and blue, respectively. Anthranoyl-AMP is represented in spheres and lies between the N- and C- terminal domain of both monomers. (B) AuaEII monomers A and B overlay well with minor structural differences. Conserved regions of the CoA ligase family (P-loop and A10 core) are labeled. The P-loop of AuaEII monomer A could not be completely modeled in the crystal structure due to weak electron density. (C) The anthranoyl-AMP intermediate is locked into place by a series of interactions with both protein side chains and backbone amide nitrogens and carbonyl oxygens. The adenine ring is largely held into place between two stretches of the protein backbone. The ribose ring forms hydrogen bonds with Lys506, Arg412, and Asp397. The aryl-binding region is largely hydrophobic except for Thr221 which hydrogen bonds with the 2-amino group of the anthranoyl moiety. (D) The SA-omit maps of anthranoyl-AMP intermediates in each monomer showing well-defined electron density for the ligands.

members of the CoA ligase family based on protein sequence analysis, but AuaE only functions

as an AT and has no CoA ligase activity. Our current work establishes a structural basis for the

substrate specificity of AuaEII and provides multiple hypotheses for the acyltransferase mechanism of AuaE.

CoA ligases are members of the adenylate-forming enzymes and catalyze the transformation of a carboxylic acid substrate into an acyl-CoA using ATP, Mg^{2+} and CoA(12). The adenylate-forming enzymes are composed of three classes: type I (non-ribosomal peptide synthase adenylation domains, acyl- and aryl-CoA synthetases, and oxidoreductases), type II (aminoacyl-tRNA synthetases), and type III (NRPS independent synthetases)(12). AuaEII is an aryl:CoA ligase and belongs to the type I ligase class. Other related aryl:CoA ligases include benzoate:CoA ligase, coumarate:CoA ligase, and 4-chlorobenzoate:CoA ligase(13-15). In comparison, because the functional divergence of AuaE as an acyltransferase, it cannot be defined by any of the above categories(11). How AuaEII and AuaE can have such different functions remains a mystery.

Here, we describe the crystal structure of AuaEII and a comprehensive comparison of AuaE and AuaEII. The AuaEII structure agrees with the previously described structural characteristics of aryl:CoA ligases, thus giving us an excellent opportunity to systematically evaluate the molecular basis of substrate specificity for the CoA ligases. Further, the AuaEII structure represents the first crystal structure of a dedicated anthranilate:CoA ligase. Below, we present the structural analysis of AuaEII in comparison to other aryl:CoA ligase crystal structures and an AuaE homology model. These comparisons reveal the molecular details of substrate specificity of the aryl:CoA ligases and provide insight into the novel type II PKS priming mechanism during aurachin biosynthesis.
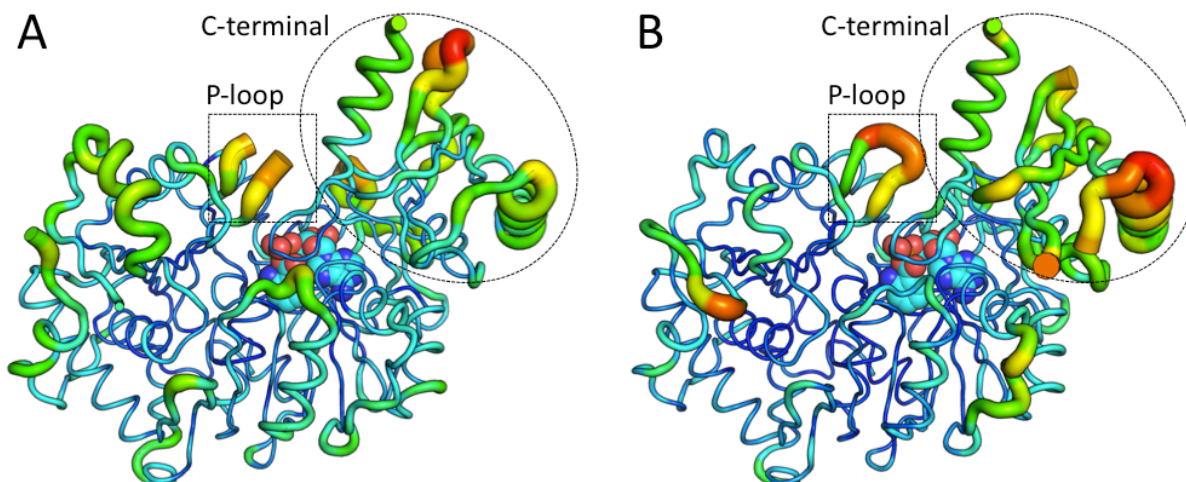
Figure 4-3. B-factor putty representation of AuaEII. High B-factors are represented by thick red tubes, low B-factors are represented by thin blue tubes. A. AuaEII monomer A B-factor representation displaying higher B-factors in the C-terminal region as well as a partially modeled P-loop. B. AuaEII monomer B B-factor representation displaying high B-factors in the C-terminal region as well as the P-loop.

4.3 Results and Discussion

4.3.1 Overall AuaEII Structure

AuaEII crystallized in the space group $P2_12_12_1$ with two AuaEII molecules per asymmetric unit. The AuaEII dimer is composed of two monomers that are related by a 2-fold rotational axis of symmetry. The protein-protein interface between monomers is not extensive, and is largely mediated by solvent accessible hydrophilic contacts (Figure 4-2A). Each monomer contains a large N-terminal and small C-terminal domain, which is in accordance with other structurally characterized CoA ligases. A typical aryl:CoA ligase contains two domains: a large N-terminal domain and smaller C-terminal domain(16). The N-terminal domain is responsible for substrate specificity, and the flexible C-terminal domain adopts different conformations, depending on whether adenylation or CoA acylation occurs(17). The N-terminal domain has been described as a combination of a β-barrel and a five layered α-β-α-β-α structure, where two β-sheets are flanked by a series of short α-helices(16). The C-terminal domain sits on top of the

Figure 4-4. Structural comparison of the active sites of the AuaEII crystal structure and AuaE homology model with other aryl-CoA ligases. (A) AuaEII (anthranoyl CoA ligase) bound to anthranoyl-AMP. (B) AuaE homology model (no ligand). (C) Benzoate CoA ligase (*B. xenovorans*) bound to benzoate. (D) Benzoate CoA ligase (*R. palustris*) bound to benzoate. (E) 4-chlorobenzoate ligase bound to 4-chlorobenzoyl-AMP. (F) 4-coumarate CoA ligase bound to 4-coumaroyl-AMP. (G) PaaK1 (phenylacetate CoA ligase) bound to phenyacetyl-AMP. (H) PaaK2 (phenylacetate CoA ligase) bound to phenyacetyl-AMP. (I) The N-terminal domain of 6 aryl CoA ligases and the AuaE homology model were aligned to the N-terminal domain of AuaEII for comparison of the aryl binding region in order to gain insight into substrate specificity of these aryl:CoA ligases. The overall % identity, residues in the aryl binding region, and 2-amino binding region are listed for comparison with AuaEII and AuaE.

N-terminal domain and consists of a core of three antiparallel β-sheets surrounded by three α-helices(18).

In AuaEII, the structures of both monomers are highly similar with an overall RMSD of 0.217 Å. The largest structural deviations occur in two α-helices in the C-terminal domain (Figure 4-2B). The N-terminal domain consists of 12 α-helices and 19 β-sheets connected by 28 loops. The core structure of the N-terminal domain contains a series of short central α-helices, which are flanked on both sides by β-sheets. The β-sheets on both sides are further surrounded by α-helices, which define the solvent accessible surface of the N-terminal domain. The N-terminal domain also contains a pseudo β-barrel structure formed by three anti-parallel two-stranded β-sheets, which is situated directly below the C-terminal domain.

The C-terminal domain is structurally distinct from the N-terminal domain. Starting with a short β-hairpin followed three anti-parallel β-strands surrounded by three α-helicies, the C-terminal domain also contains a long, highly conserved loop, termed the A10 core, which leads into the final C-terminal α-helix. Anthranoyl-AMP is located in the active site between the N- and C-terminal domains, and all atoms of both ligands were modeled with well-defined electron density (Figure 4-2D). In summary, the crystal structure of AuaEII matches well with other members of the aryl:CoA ligase family and the flexible C-terminal domains of both monomers have elevated B-factors relative to the large N-terminal domains.

4.3.2 The C-terminal Conformation of AuaEII

CoA ligases typically crystallize with the C-terminal domain in conformation 1 or conformation 2. Conformation 1 has the A10 core facing the active site of the N-terminal domain, and this conformation facilitates the adenylation reaction(19). Conformation 2

represents the thioester forming conformation. In conformation 2, the C-terminal domain is rotated ~140 degrees relative to conformation 1, resulting in the movement of A10 core motif away from the N-terminal active site(17). In the AuaEII crystal structure, the C-terminal domain is in conformation 1, and the conserved lysine Lys506 of the A10 core interacts with the phosphate of the anthranoyl-AMP intermediate (Figure 4-2C). The charge-charge interaction between Lys506 and the anthranoyl-AMP phosphate may be important for stabilizing conformation 1 of AuaEII and promoting crystallization.

Multiple aryl:CoA ligase crystal structures (benzoate:CoA ligase, 4-chlorobenzoate:CoA ligase, 4-coumarate:CoA ligase, and phenylacetyl:CoA ligase) have been solved with an aryl-AMP intermediate bound in the N-terminal pocket, yet the C-terminal domains of these structures do not all adopt the same conformation(13, 18, 20, 21). It has been hypothesized that in different CoA ligase enzymes, there is an equilibrium between conformations 1 and 2, and this equilibrium may be different for different CoA ligases during crystallization(17). In the AuaEII crystal structure, the C-terminal domains of both monomers have elevated B-factors for the majority of residues except in the A10 core region, where Lys506 interacts with the anthranoyl-AMP intermediate (Figure 4-3). In summary, anthranoyl-AMP is located in the active site of the AuaEII crystal structure and the C-terminal domain is in conformation 1 with Lys506 of the A10 core interacting with the phosphate of anthranoyl-AMP.

4.3.3 Anthranoyl-AMP Interactions in the AuaE Active Site

The AuaEII active site, similar to those of other aryl:CoA ligases, lies in a cleft between the N- and C-terminal domains. The anthranoyl portion of the anthranoyl-AMP intermediate is bound in an internal pocket of the N-terminal domain, whereas the AMP portion interacts with

Accession ID	Enzyme	Organism	Identity
CCA65703.1	anthranilate-CoA ligase	Stigmatella aurantiaca Sg a15	100%
KFE69504.1	long chain fatty acid-CoA ligase	Hylangium minutum	76%
ADJ94058.1	putative benzoate-CoA ligase BzlA	Clostridia bacterium	49%
ETX03435.1	hypothetical protein ETSY1 00060	Candidatus entotheonella sp. TSY1	63%
ADJ94003.1	putative benzoate-CoA ligase BzlA	Clostridia bacterium	48%
ADJ94007.1	putative benzoate-CoA ligase BzlA	Clostridia bacterium	47%
YP_006721104.1	benzoate-CoA ligase	Geobacter metallireducens GS-15	48%
WP_029133166.1	4-hydroxybenzoate-CoA ligase	Sedimenticola selenatireducens	48%
YP_002535674.1	benzoate-CoA ligase	Geobacter daltonii FRC-32	48%
YP_007944607.1	benzoate-CoA ligase	Desulfotomaculum gibsoniae DSM 7213	45%
BAE91925.1	benzoate-CoA ligase	Magnetospirillum sp. TS-6	47%

Figure 4-5. Top 10 non-redundant protein BLAST results of AuaEII.

both the N- and C-terminal domains (Figure 4-2C). The adenine portion of the anthranoyl-AMP is situated between the backbone amides of Ser294, Glu295, Ala296 from the top, and Ile317 on the bottom, and forms multiple hydrogen bonds with other residues. The adenine C6 amine forms hydrogen bonds to the carboxylate of Asp315 and the backbone carbonyl group of Gly316. The N1 of adenine ring also forms a hydrogen bond to a solvent-accessible water molecule. The ribose portion of the anthranoyl-AMP intermediate interacts with two residues with opposing charges, Lys506 and Asp397. Lys506 is located on the in a highly conserved A10 core of the C-terminal domain and forms hydrogen bonds with the 5' hydroxyl and oxygen in the ribose ring. The Asp397 carboxylate forms tight hydrogen bonds with the 2' and 3' hydroxyls of the ribose ring and Arg412 is also in hydrogen bond distance to the 3' hydroxyl. The phosphate forms weak hydrogen bonds with the amide hydrogen of Thr320 and the hydroxyl of Ser175. Additionally, three solvent linked water molecules form hydrogen bonds to the phosphate. The anthranoyl aromatic ring is located under the end of an α-helix where three consecutive residues (Phe220, Thr221, and Phe222) interact with the anthranoyl ring. First, Phe222 has hydrophobic interactions with the edge of the anthranoyl group and defines the bottom of the substrate pocket. Second, Thr221 forms a hydrogen bond with the amine of the anthranoyl aromatic ring that locks the intermediate into the observed orientation. Third, Phe220 has hydrophobic interactions with
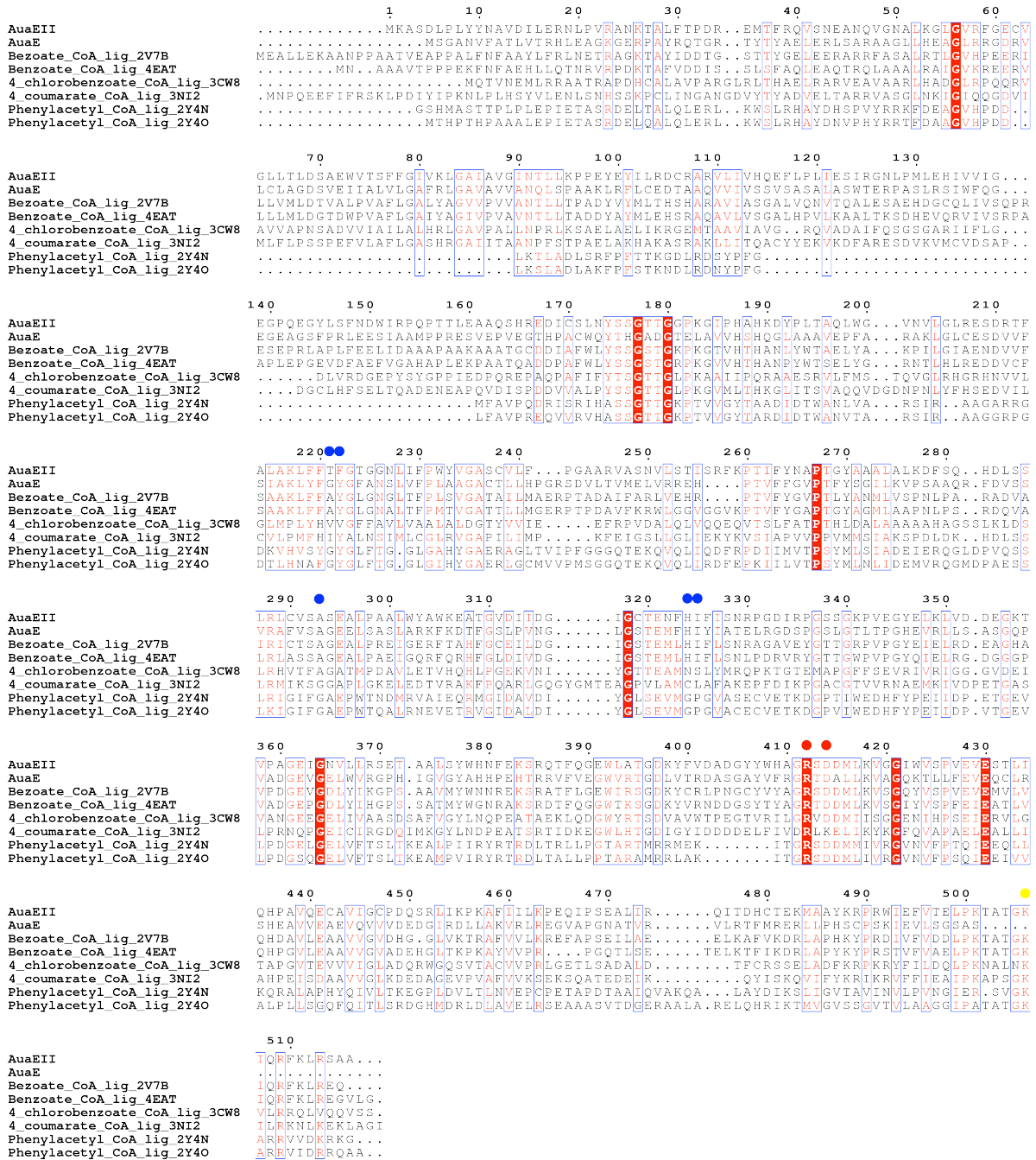
Figure 4-6. Sequence alignment of AuaEII and AuaE with aryl:CoA ligases with known crystal structure. Aryl binding residues are colored with blue dots. Conserved hinge region residues are colored with red dots. The conserved lysine of the A10 region is colored with a yellow dot.

the top of the anthranoyl aromatic ring, and the edge of the Phe220 aromatic ring is perpendicular to the carbonyl group of the anthronyl ester. The back edge of the anthranoyl aromatic ring is bound in a pocket defined by His324, Ile325, and Ala293. His324 is engaged in face-on-edge π-π interactions with the anthranoyl aromatic ring from one side, while the adenine ring of the anthranoyl-AMP intermediate sits edge to edge with the anthranoyl aromatic ring. The bottom of the anthranoyl aromatic ring undergoes π-π stacking with the amide bond between Gly318 and Cys319 (Figures 4-2C). In summary, the anthranoyl-AMP intermediate sits between the N-terminal and C-terminal domains and reveals residues in the N-terminal domain that are important for substrate specificity.

4.3.4 Comparisons Between the Aryl Binding Regions of AuaEII and AuaE

The selectivity of aryl:CoA ligases is determined entirely by residues in the aryl binding region of the N-terminal domain. The aryl-binding region is where the carboxylic acid substrate binds. This region contains specific residues that interact with the aryl substituent. To date, many CoA ligase crystal structures have been solved bound with an acid substrate or aryl-adenylate intermediate, and these structures reveal residues that are important protein-substrate interactions(13, 17, 18, 21). An in depth comparison of the AuaEII crystal structure to the AuaE homology model and six previously determined aryl:CoA ligase crystal structures reveal insights into aryl substrate binding (Figure 4-4). The aryl-binding region of AuaEII consists of histidine (His234), isoleucine (Ile325), alanine (Ala293), and phenyalanine (Phe222). In the case of AuaEII, a threonine (Thr221) is located near the aryl binding region and forms a hydrogen bond with the 2-amino group of the anthranilate moiety of the bound anthranoyl-AMP. Therefore, the

A

Accession	Enzyme	Organism	Identity
CAL48957.1	anthranilate-CoA-ACP transferase	Stigmatella aurantiaca Sg a15	100%
WP_025582066.1	4-hydroxybenzoate-CoA ligase	Cupriavidus taiwanensis	38%
ADJ93951.1	putative benzoate-CoA ligase BzlA	Clostridia bacterium	35%
YP_002005350.1	benzoate-CoA ligase	Cupriavidus taiwanensis LMG 19424	38%
WP_016906742.1	hypothetical protein	Streptomyces sulphureus	39%
YP_488250.1	benzoate-CoA ligase	Rhodopseudomonas palustris HaA2	37%
YP_568017.1	benzoate-CoA ligase	Rhodopseudomonas palustris BisB5	37%
WP_007912907.1	4-hydroxybenzoate-CoA ligase	Ktedonobacter racemifer	35%
WP_008143446.1	4-hydroxybenzoate-CoA ligase	Bradyrhizobium sp. YR681	36%
NP_767717.1	acetyl-CoA synthetase	Bradyrhizobium diazoefficiens	36%
YP_001261334.1	benzoate-CoA ligase	Sphingomonas wittichii RW1	37%

B

```
                 465        475        485        493          *
                  .          .          .          .
AuaE           GNATVRVLRTFMRERLLPHSCPSKIEVLS........GSAS........
WP_025582066.1 HDGMAAELQAFVKSRLAPYKYPRQIECVPELPKTATGKIQRFRLRQREQAARG..
ADJ93951.1     GEEMVKELQLYVKNKIAPYKYPRWIEFINELPKTATGKYRDLS...........
YP_002005350.1 HDGMAAELQAFVKSRLAPYKYPRQIECVPELPKTATGKIQRFRLRQREQAARG..
WP_016906742.1 .QQLEERVQAALRRRLPPFKRPRTLHLVDALPLTSTGKTARFLLRGQEKERES..
YP_488250.1    PDGLFEALKDHVQKVGPWKYPRWIEVVPSIPKTATGKIQRFKLREGAQ......
YP_568017.1    ADGLFEQLKDHVKHVGPWKYPRWIEVVPSIPKTATGKIQRFKLRDGASG.....
WP_007912907.1 SESLEEELKEFVKGRLAPHKYPRYIEFRAELPKTATGKIQRFLLRGTGETVPLDI
WP_008143446.1 TTDLQEMLKEHVKQKIGPWKYPRWIDVVDSIPKTATGKFKLRDGAN......
NP_767717.1    TANLQEMLKEHVKQKIGPWKYPRWIDVVDSIPKTATGKIQRFKLREGAN......
YP_001261334.1 SDAIEAALKAHVKQLLAPYKYPRWIEFLDALPKTATGKVERYKLRQPPAA.....
                                              A10 core
```

C

Accession	Enzyme	Organism	identity
WP_029001793.1	4-hydroxybenzoate-CoA ligase	Azohydromonas australica	42%
AHG64790.1	benzoate-CoA ligase	Advenella mimigardefordensis	45%
WP_014750022.1	4-hydroxybenzoate-CoA ligase	Advenella kashmirensis	44%
EYF01798.1	benzoate-CoA ligase	Chondromyces apiculatus	44%
WP_027299446.1	4-hydroxybenzoate-CoA ligase	Rhodospiralles bacterium URHD0088	41%
WP_024006624.1	4-hydroxybenzoate-CoA ligase	Advenella kashmirensis	45%
WP_017775601.1	4-hydroxybenzoate-CoA ligase	Burkholderia kururiensis	44%
WP_019939360.1	4-hydroxybenzoate-CoA ligase	Bordetella sp. FB-8	40%
WP_0261546600.1	CoA ligase	Rhizobium leguminosarum	38%
KCV10937.1	AMP-binding enzyme C-terminal domain	Bordetella holmesii CDC-H785-BH	40%

D

```
                  465        475        485        495    *
                   .          .          .          .
AuaE_Cterminal  GNATVRVLRTFMRERLLPHSCPSKIEVLSG...SAS...............
WP_029001793.1  GEALAEALKSFVKEQLAPRHIEFVAELPKTATGKIQRFKLREREAHQ.....
AHG64790.1      DDTLRDALKAFVKSRLAPHKYPRSIEFLEELPKTATGKIQRFRLRELNQHEVA...
WP_014750022.1  DDTLKDTLKAFVKSRLAPHKYPRSIEFLDEIPKTATGKIQRFRLREMNQHEVA...
EYF01798.1      GDHLVLALQAFVKGRLAPHKYPREIAFVTELPKTATGKIQRFRLREGG........
WP_027299446   GSALAEALQRHVKERLAPYKYPRWIEFIAELPKTATGKIQRFKLRERAGKTG....
WP_024006624.1  DDALKEVLKAFVKSRLAPHKYPRSIEFLDELPKTATGKIQRFRLRELNQHEPA...
WP_017775601.1  LDRLVDELKAFVKERLAPHKYPRDIQFVDDIPKTATGKIQRFKLRELL........
WP_019939360.1  GSATCASLQAYVKTRLAPFKYPREIDFIEELPKTATGKIQRFRLRQREGAAEPAKV
WP_026154660.1  GDVQTRILRDYVKSTLMPHKYPRIIQYVAELPKTGTGKIDRQSLLKLPVPVD....
KCV10937.1      DASTGAALQDYVKSHLAPFKYPRQINFTEELPKTATGKIQRFRLRQMEEASL....
                                              A10 core
```

Figure 4-7. AuaE protein BLAST search results and sequence alignments. (A) Top 10 non-redundant BLAST results for AuaE. (B) Alignment of the top 10 AuaE BLAST results focusing on the C-terminal region. The A10 core is boxed and the position of conserved lysine is labeled with an asterisk. (C) Top 10 non-redundant BLAST results for the predicted C-terminal domain (R408 – S497) of AuaE. (D) Alignment of the top 10 AuaE C-terminal BLAST results. The A10 core is boxed and the position of conserved lysine is labeled with an asterisk.

aryl binding region of AuaEII is most similar to benzoate CoA-ligases except for the residue Thr221, which is important for anthranilate specificity.

AuaE has an overall sequence identity of 28.6% with AuaEII. The aryl binding site of the AuaE homology model consists of histidine (His320), isoleucine (Ile321), alanine (Ala289), and tyrosine (Tyr218). The aryl binding region of AuaE is very similar to that of AuaEII. The only difference is that AuaE has a tyrosine (Tyr218) in place of phenylalanine (Phe222) of AuaEII. AuaE has a glycine (Gly217) at the position that corresponds to Thr221 of AuaEII. Thr221 of AuaEII is responsible for forming a hydrogen bond with the anthranoyl 2-amino group. Therefore, it is surprising that the AuaE model would have glycine at this position, because the aryl binding region of both enzymes is proposed to process the 2-amino bearing anthranoyl moiety. The small size of glycine would not exclude anthranoyl-CoA binding, but it would lack a hydrogen bond donating/accepting side chain present in threonine. It is possible that the anthranoyl group of AuaE's substrate, anthranoyl-CoA, binds in a different orientation. Alternatively, the cavity of the aryl binding region alone may be sufficient for productive binding of the aromatic substrate, and additional hydrogen bonding to the 2-amino moiety is not necessary.

4.3.5 The Nature of Low Sequence Similarity Among Aryl:CoA Ligases

Among aryl:CoA ligases with very similar substrates (benzoate, 4-chlorobenzoate, 4-coumarate, phenylacetate, and anthranilate), the overall sequence identity can be very low (Figure 4-4). For example, the substrate differences of AuaEII and 4-chlorobenzoate-CoA ligase are minor (2-amino vs. 4-choro substituent), yet the two enzymes only share 21.6 % of sequence identity. CoA ligases are very common in primary metabolism, and diversification of substrate

specificity is easily achievable by a small number of mutations in the N-terminal substrate-binding region. Therefore, it is not surprising that low sequence identity is observed for different CoA ligases that act on similar substrates. This is especially intriguing in the case of AuaE and AuaEII, because AuaE only has an overall sequence identity of 28.6 % with AuaEII, despite its close proximity in the same genomic region and their joint roles in the same biosynthetic pathway. These observations suggest that AuaE did not arise from duplication of AuaEII (or vice versa). The situation is quite different in the active site. Even though the overall sequence identity is only 28.6 % between AuaE and AuaEII, residues in the aryl binding region that directly interact with the substrates (based on the AuaEII crystal structure and AuaE homology model) are nearly identical (His324/His320, Ile325/Ile321, Ala293/Ala289, Phe222/Tyr218, for AuaEII/AuaE). AuaE and AuaEII are most closely related to benzoate CoA ligases (Figure 4-5). Therefore, they could have both independently evolved from different benzoate (or other aryl) CoA ligases.

4.3.6 Rational for Low Overall Sequence Conservation of Aryl:CoA Ligases

Aryl:CoA ligases are often employed by bacteria in detoxify environmental contaminants such as 4-chlorobenzoate and benzoate(22-25). The driving force for activation of aromatic acids as CoA thioesters may have directed the evolution of CoA ligases, which results in a group of aryl:CoA ligases that share conserved active site residues but low overall sequence identity (such as AuaEII and benzoate:CoA ligases). When analyzing the aryl:CoA ligase crystal structures, it is clear that the aryl binding regions of AuaEII, AuaE, and both benzoate-CoA ligases are similar. However, the aryl binding region is different from that of 4-chlorobenzoate-CoA ligase, despite the structural similarities between benzoate and 4-chlorobenzoate. The aryl binding
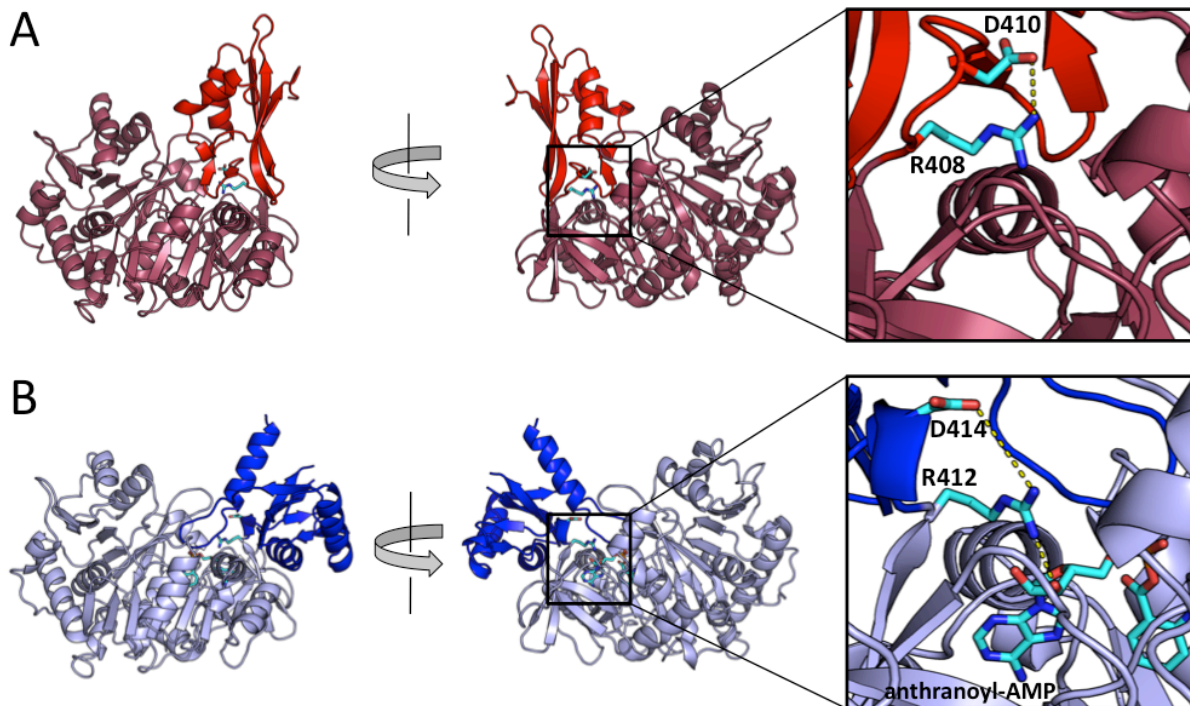
Figure 4-8. Aspartic acid/arginine pair of the AuaEII and AuaE hinge regions. (A) The AuaE homology model in conformation 2 reveals the conserved hinge region residues Asp410 and Arg408 form a salt bridge, which stabilizes this conformation. (B) The AuaEII crystal structure in conformation 1 showing the conserved hinge region residues Asp414 and Arg412 do not form a salt bridge and Arg412 is in hydrogen bonding distance with the a hydroxyl of the anthranoyl-AMP ribose ring.

regions of 4-coumarate:CoA ligase and both phenylacetyl:CoA ligases are also distinct from the AuaEII, AuaE, the benzoate:CoA ligases, and 4-chlorbenzoate:CoA ligase. Nevertheless, several features are still retained, such as an aromatic amino acid at the position corresponding to F220 of AuaEII (Figure 4-4). In summary, our analysis concludes that the need for activation of aromatic acids as CoA thioesters, especially in bacteria, has led to the emergence of a wide variety of aryl:CoA ligases that often share low overall sequence identity but with highly conserved residues in the aryl binding region.

4.3.7 Overall Motif Differences of AuaE and Other Aryl:CoA Ligases

Previously, we completed an in depth sequence analysis of AuaE and identified two regions, the A3 (also known as the P-loop) and A10 cores, which vary greatly in sequence when compared to other CoA ligases(11). The A3 region is responsible for binding and positioning the β and γ-phosphates of ATP during the adenylation reaction, whereas the A10 region is mobile and contains a conserved lysine residue that stabilizes the negative charge formed on the α-phosphate during the adenylation reaction. The A10 region then undergoes a conformational change that allows CoA to enter the active site and complete the CoA-ligation (17, 26-30). In contrast, the function of AuaE is to catalyze acyltransfer from anthranoyl-CoA to ACP, which does not involve ATP binding or adenylation. For this reason, in AuaE, a high conservation of the P-loop and A10 regions is not critical for its function. In AuaEII, these regions are conserved. However, the P-loop could not be completely modeled in both AuaEII monomers, and only monomer B has well defined electron density for this region. The B-factors of the P-loop region in AuaEII are elevated, and this has also been observed in other CoA ligase crystal structures in which ATP is not bound, suggesting high flexibility of the P-loop in the absence of ATP. In summary, AuaE lacks conservation of the important A10 and P-loop regions of aryl:CoA ligases.

4.3.8 Absence of Conserved C-terminal Residues in AuaE

The N-terminal domain of AuaE shares a high degree of sequence identity to other CoA ligases. However, the C-terminal domain is truncated, and lacks the highly conserved region A10 core important for the ligase activity (Figures 4-6 and 4-7). A BLAST search followed by sequence alignment of the AuaE C-terminal domain revealed ~20 residues are missing when
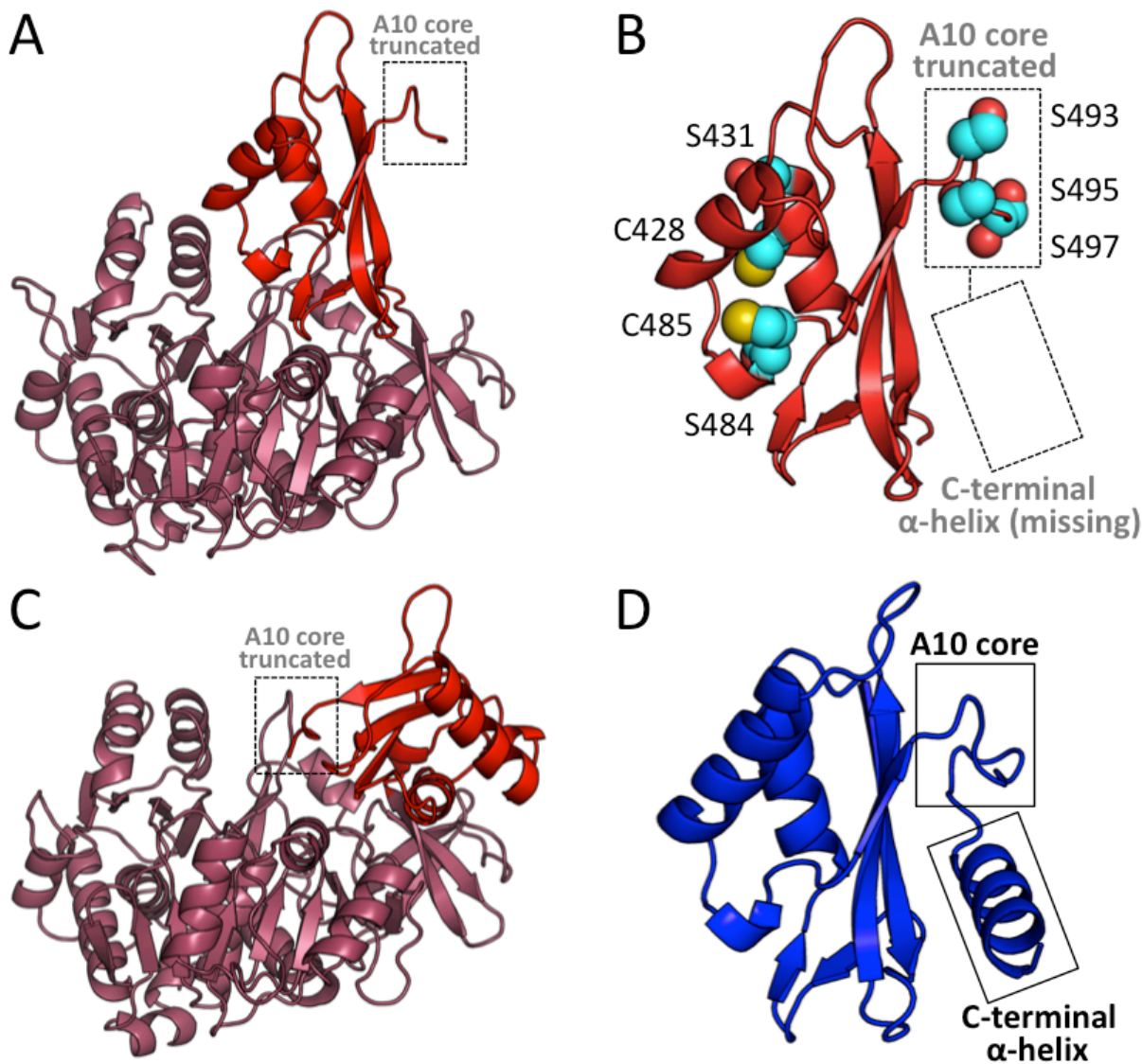
Figure 4-9. AuaE homology model in both conformation 1 and conformation 2 and comparison of the C-terminal domains of AuaEII and AuaE. (A) Cartoon representation of an AuaE homology model generated using the structure of benzoate:CoA ligase (PDB ID: 4EAT) with the C-terminal domain (colored in red) in conformation 1 (adenylate forming conformation). The truncated A10 core is labeled and located away from the N-terminal (colored in raspberry) active site. (B) Close up view of the C-terminal domain of the AuaE homology model in conformation 1, showing the truncated A10 core and missing C-terminal helix. Residues (serines and cysteines) that may be involved in acyltransfer are represented in spheres. (C) Cartoon representation of the AuaE homology model generated using the structure of benzoate:CoA ligase (PDB ID: 2V7B) with the C-terminal domain (colored in red) in conformation 2 (thioester forming conformation). (D) C-terminal domain of AuaEII displaying the conserved A10 core and C-terminal helix, for comparison with the AuaE homology model.

compared to the C-terminal domain of other CoA ligases. The A10 core is comprised of a highly conserved loop that leads into a C-terminal α-helix in AuaEII. The conserved loop contains a lysine residue, which has been shown to stabilize the negative charge of the phosphate pentahedral intermediate formed when the acyl-adenylate is generated (26, 27). AuaE does not catalyze a typical CoA ligase reaction; therefore, it is likely that the highly conserved A10 core is missing, because it does not need the A10 core. A second BLAST search using only the AuaE C-terminal domain as a query sequence showed that the most similar proteins are CoA ligases and, when aligned, all the sequences have a full length C-terminal domain (Figure 4-7). This result suggests that AuaE is a unique member of the CoA ligase where the loss of CoA ligase function has occurred at the cost of losing the A10 core. However, AuaE somehow gained the AT activity.

4.3.9 Conserved Aspartic Acid of the Hinge Region

The 140° degree rotation of the C-terminal domain between adenylation and thioester forming reactions has been well studied in 4-chlorobenzoate:CoA ligase, which revealed that a conserved aspartic acid in the hinge region between the N- and C-terminal domains is crucial for the conformational change(28). In the AuaE homology model, the aspartic acid (Asp410) is located in the hinge region at the same position as the conserved aspartic acid of 4-chlorobenzoate:CoA ligase (Figure 4-8). Because of the missing A10 motif of AuaE, it is unclear about the influence that such a conformational change may have on the acyltransferase activity of AuaE. However, because of the presence of the conserved aspartic of the hinge region, AuaE may undergo conformational changes during catalysis. Furthermore, in conformation 2 (thioester forming conformation), the aspartic acid of the hinge region forms a salt bridge with a nearby
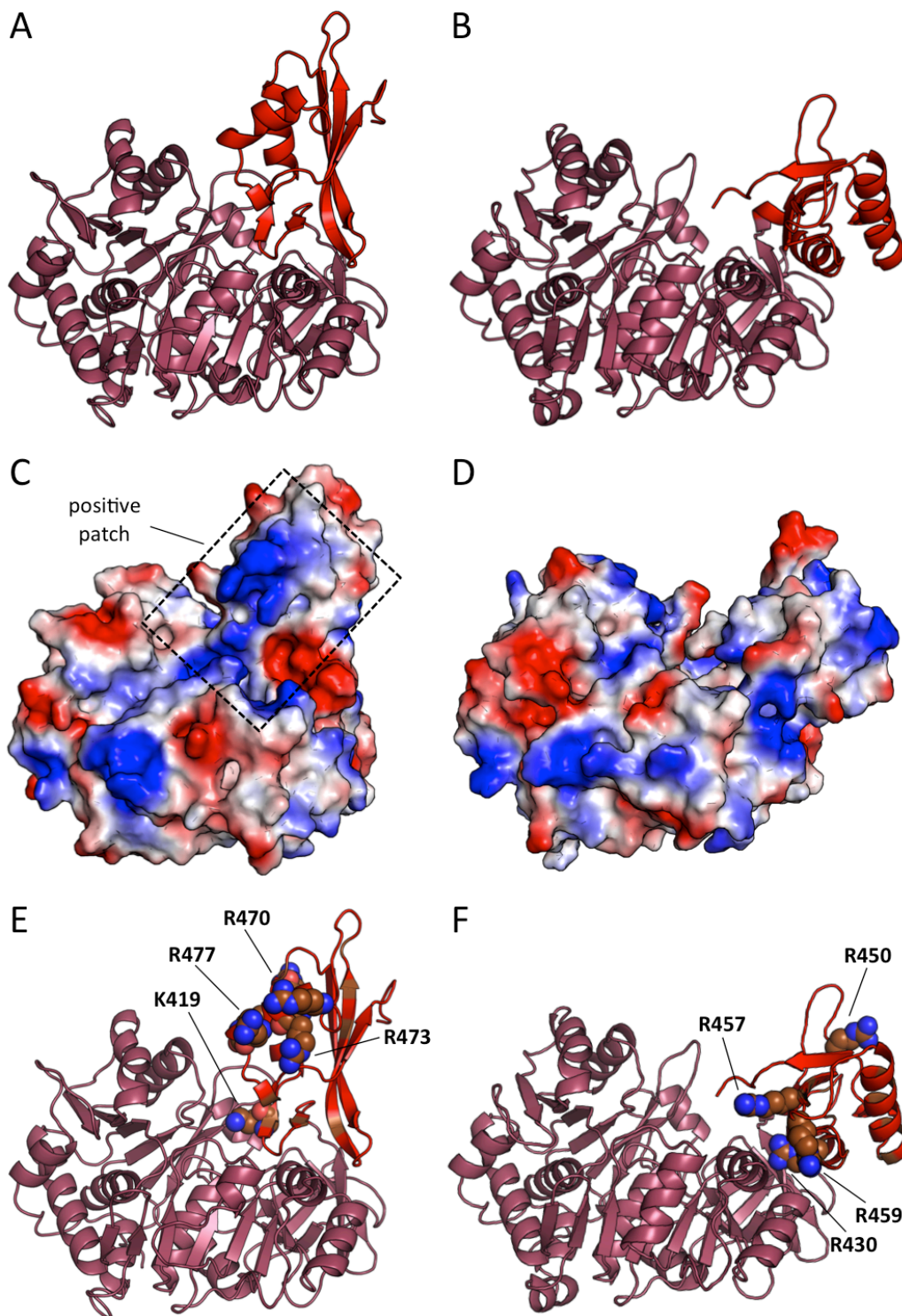
Figure 4-10. AuaE homology models in conformation 1 and 2 with electrostatic surface analysis. (A) AuaE model of conformation 2. (B) AuaE model of conformation 1. (C) AuaE conformation 2 electrostatic surface*. (D) AuaE conformation 1 electrostatic surface*. (E) Positively charged residues facing the N-terminal of AuaE in conformation 2. (F) Positively charged residues facing the N-terminal of AuaE in conformation 1. *Red represents negative and blue is positive surface charge.

arginine, which stabilizes conformation 2. AuaE also has arginine (Arg408) at this position, which further supports the possibility of a conformational change mechanism, or possibly that conformation 2 may be important for the acyltransferase activity of AuaEII. AuaEII also has the conserved aspartic acid/arginine pair in the hinge region as expected for the CoA ligases, However, the two residues are not engaged in a charge-charge interaction, because the C-terminal domain of AuaEII is in conformation 1 (Figure 4-8). In summary, AuaE contains a conserved aspartic acid/arginine pair, which has been shown to stabilize conformation 2 in aryl-CoA ligases. Therefore, conformation 2 may be important for the acyltransferase activity of AuaE.

4.3.10 Possible Mechanisms of AuaE Acyltransferase (AT) Activity

CoA ligases have a CoA binding site, which allows the phosphopantetheine of CoA to extend into the N- and C-terminal domain interface and react with the acyl-adenylate intermediate. *In vitro*, it has been demonstrated that the substrate for AuaE is anthranoyl-CoA and the product is anthranoyl-ACP, but the mechanism of this reaction is unknown. There are two likely possibilities for how AuaE catalyzes the acyl transfer from anthranoyl-CoA to *holo*-ACP: 1) A classic AT-like mechanism or 2) AuaE acts simply as a reaction surface.

The classic AT mechanism is used by enzymes such as malonyl-CoA:ACP transacylase (MCAT) that has an active site serine, with a multistep mechanism beginning with an acyl-CoA as the substrate. The acyl group is trans-esterified onto an active site serine to form an acyl-intermediate, which then reacts with *holo*-ACP to yield an acyl-ACP product. Other ATs have also been identified that use cysteine as the active site residue to produce the covalent acyl intermediate(31). Enzymes with a CoA ligase fold, such as AuaE, are able to produce and bind

	AuaEII
Crystallization	0.1 M sodium acetate pH 4.5, 0.1 M magnesium acetate, 8% PEG 8000
Crystallographic Data	
Wavelength (Å)	0.9537
Space Group	$P2_12_12_1$
Cell Dimensions (a, b, c) (Å)	50.054, 132.219, 160.118
	$\alpha=\beta=\gamma=90$ °
Resolution (Å)	50.00 - 2.60
No. of observations	215077
No. of unique observations	33543
Completeness %	99.7 (97.4)
I/σ(I)	15.1 (5.1)
R_{merge} %	11.2 (32.2)
Redundancy	6.4
Refinement	
Resolution (Å)	46.81-2.60 (2.69-2.60)
No. of protein atoms	8062
No. of ligand atoms	64
No. of water atoms	488
R_{free} %	0.224 (0.267)
R_{crys} %	0.174 (0.203)
Geometry	
RMS bonds (Å)	0.017
RMS angles (°)	0.99
Ramachandran Favored (%)	98
Ramachandran Allowed (%)	1.8
Ramachandran Dissallowed (%)	0.2
Average B-factors ($Å^2$)	
Protein	21.90
Water	22.30
Ligands	15.00

Table 4-1. AuaE crystallographic data collection and refinement statistics.
*Numbers in parentheses denote the highest resolution shell.

acyl-CoAs; therefore, AuaE may bind anthranoyl-CoA and undergo trans-esterification to form a

stable enzyme intermediate on a serine or cysteine. Once the AuaE enzyme acyl-intermediate is

formed, *holo*-ACP can form a complex with AuaE and allow the pPant to react with the acyl-

intermediate and complete the AT reaction. The classic AT mechanism is an attractive hypothesis, because there are multiple serine and cysteine residues in the C-terminal domain, which may be the site of acylation. In the AuaE homology model Cys426, Ser431, Ser484, Cys485, and Ser487 are all located on the active site side of the C-terminal domain while in the thioester forming conformation 2. Ser493, Ser495, and Ser497 are all located in the truncated A10 region and would be close to the active site, if AuaE adopts conformation 1 during catalysis (Figure 4-9). Both conformations could potentially facilitate acyl transfer. Additionally, in conformation 2, a series of positively charged residues are located on the C-terminal domain facing the aryl-binding region, and these residues form a positive patch that could bind the ACP.

Alternatively, AuaE may simply act as a surface to bring anthranoyl-CoA and *holo*-ACP together and stabilize the formation of a tetrahedral intermediate that results from a nucleophilic attack of the ACP-linked pPant thiol onto the thioester carbonyl group of anthranoyl-CoA. In this scenario, anthranoyl-CoA would be stabilized in a favorable conformation by AuaE for acyl transfer. This would be achieved by placing the thioester carbonyl oxygen of anthranoyl-CoA into an oxyanion hole. Once anthranoyl-CoA is bound to AuaE, the positive surface of the AuaE C-terminal domain can dock with ACP to facilitate acyl transfer (Figure 4-10).

4.4 Conclusion

We have solved the crystal structure the CoA ligase AuaEII that is responsible for the generation of anthranoyl-CoA from anthranilic acid. AuaEII activity is essential for the production of the aurachin class of natural products and is part of a novel starter unit activation strategy to incorporate anthranilic acid into type II polyketide biosynthesis. The crystal structure of AuaEII bound to anthranoyl-AMP reveals specific residues involved in substrate selectivity

and adds to a growing body of work on this diverse enzyme class. AuaE is a novel enzyme that is predicted to have the same overall fold as a CoA ligase, except that the C-terminal domain is truncated, and instead of CoA ligase activity, AuaE is an acyltransferase. We generated multiple homology models of AuaE and used a combination of structural comparisons and sequence alignments to visualize how AuaE might function as an AT instead of a CoA ligase. Future structural studies of AuaE are needed to fully understand the novel acyltransferase activity of this enzyme. Taken together, these results represent the first crystal structure of an anthranoyl-CoA ligase, which paves the foundation for understanding the priming mechanism of a novel starter unit in type II polyketide biosynthesis.

4.5 Materials and Methods

4.5.1 Protein Expression and Purification

A pET-29b (+) (Novagen) derived DNA plasmid encoding a C-terminal non-cleavable His-tagged AuaEII (AuaEII/pET29) was transformed into *E. coli* BL21(DE3) competent cells and plated on LB-agar plates containing kanamycin (50 µg/mL). The plates were incubated overnight at 37 °C. Positive transformants were transferred to a 5 mL starter culture of Luria-Bertani (LB) broth containing kanamycin (50 µg/mL) and grown overnight at 37 °C with shaking, which was then used to inoculate two liters of LB with kanamycin (50 µg/mL) and grown at 37 °C until the A_{600} reached 0.4 – 0.6. The cells were then cooled to 18 °C, and 0.1 mM IPTG was added to induce protein expression. After 12-18 hours of incubation at 18 °C, the cells were harvested by centrifugation at 5000 rpm for 15 minutes. The cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C. The frozen cell pellets were thawed on ice and re-suspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 % glycerol, and 10 mM

imidazole). The cell suspension was lysed using sonication (8 x 30 s cycles), and the cell debris was removed by centrifugation at 14000 rpm for 45 minutes. The clarified lysate was incubated with 5 mL Ni-IMAC resin (BioRad) at 4 °C for one hour. The resin was poured into a fritted column and the flow through fraction was collected. The resin was washed with 100 mL of lysis buffer then eluted with lysis buffer plus increasing amounts of imidazole (20-500 mM). The elutions were analyzed using SDS-PAGE, and elutions containing the protein of interest were pooled, and dialyzed over night into storage buffer (50 mM Tris 8.0, 300 mM NaCl, 10 % glycerol, and 2 mM DTT). The dialyzed protein was concentrated to 8-10 mg/ml by centrifugal filtration using a 30 kDa MWCO concentrator (Millipore) then flash frozen in liquid nitrogen, and stored at -80 C° for further use. The theoretical molecular weight of AuaEII is 58477 daltons, which was confirmed using MALDI-TOF MS (data not shown).

4.5.2 *In Vitro* Production of Anthranoyl-AMP and AuaEII Crystallization

AuaEII was thawed on ice and buffer exchanged into crystallization buffer (25 mM Tris 8.0, and 2 mM DTT) using a PD-10 column (GE Healthcare). In order to produce crystals that diffract, AuaEII was incubated with different combinations of anthranilic acid, ATP, AMP, CoA, and $MgCl_2$ before co-crystallization screening. The most promising co-crystallization condition was found by incubating AuaEII (95 µM) with anthranilic acid (2 mM), ATP (5 mM), and $MgCl_2$ (2 mM) at 37 °C for one hour to form the anthranoyl-AMP intermediate. The reaction was then filtered with a 0.22 µM spin filter and screened for crystallization. For crystallization optimization, 1.7 µL of the incubated AuaEII reaction mixture was mixed with 1.7 µL of well solution (8 % PEG 8000, 0.1 M magnesium acetate, and 0.1 sodium acetate, pH 4.5) and allowed to equilibrate over 500 µL of well solution using the sitting drop vapor diffusion method. Small

rod-shaped crystals grew over 1-3 days. Crystals also formed when AuaEII was incubated with AMP + anthranilic acid and incubated on ice for 45 minutes prior to crystallization, but the optimized crystals from this co-crystallization did not diffract beyond ~10 Å. Despite many different co-crystallization strategies, we were unable to obtain any diffraction-quality crystals of AuaE.

4.5.3 X-ray Data Collection, Processing, and Refinement

Crystals of AuaEII were flash frozen directly in liquid nitrogen before data collection. X-ray diffraction data using monochromatic X-rays (0.9537 Å) was collected for an AuaEII crystal to a resolution of 2.6 Å at Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 12-2. The data was indexed, integrated, and scaled using HKL2000(32). AuaEII crystallized as one dimer per asymmetric unit in the space group $P2_12_12_1$ and initial phases were determined using molecular replacement (PHENIX Phaser-MR) with the structure of benzoate CoA ligase (PDB ID: 4EAT) as a search model(33). A preliminary model was built (PHENIX AutoBuild) and then this model was used for iterative rounds of manual model building (COOT) and refinement (PHENIX Refine)(34-36). Well-defined electron density for anthranoyl-AMP was observed in each AuaEII monomer after multiple rounds of refinement. The anthranoyl-AMP ligand was drawn in Chemdraw and then converted to a PDB using the NCI SMILES converter server (http://cactus.nci.nih.gov/translate/), and ligand restraints were generated using PHENIX Elbow(37). Anthranoyl-AMP was placed in each monomer manually using COOT. Waters were added with PHENIX Refine. Refinement was continued until R_{free} and R_{work} reached values of 22.4 % and 17.4 %, respectively. Data collection and refinement statistics can be found in Table 4-1. The following residues could not modeled confidently in the AuaEII crystal structure due to

missing or weak electron density: Pro141, Gln142, G;u143, Gly144, Tyr145, and Thr179 of monomer A and Ser451, Arg452, and Leu435 of monomer B.

4.5.4 AuaE Homology Model Generation

HHPred was used to generate two homology models of AuaE using benzoate:CoA ligases (PDB IDs: 4EAT and 2V7B) as templates with two different C-terminal conformations (38). The C-terminal domains of CoA ligases have been observed in multiple conformations as determined by the crystal structures of several enzymes(19). Even when two different enzymes are bound to a similar intermediate (such as an acid substrate), the C-terminal domains could be in different conformations. The template for the first AuaE homology model, benzoate:CoA ligase (from *Rhodopseudomanas palustris*) (PDB ID: 4EAT), has its C-terminal domain in the thioester conformation (conformation 2), where A10 core rotates away from the active site. The template for the second AuaE homology model, benzoate:CoA ligase (from Burkholderia xenovorans LB400) (PDB ID: 2V7B), has its C-terminal domain in the adenylation conformation (conformation 1), where the A10 core rotates towards the active site. Because of the AuaE C-terminal truncation, it is difficult to predict the exact conformation that the C-terminal domain will adopt during catalysis; therefore, we have chosen to generate two models, which represent the two most common conformations observed in the crystal structures of CoA ligases(19).

4.5.6 AuaEII and AuaE Sequence Alignment and Structural Comparison

A search for proteins similar to AuaEII and AuaE was conducted using BLAST searching both non-reduntant protein sequences and also Protein Data Bank (PDB) proteins. Sequence

159

alignments were conducted using ClustalW(39) and graphics were generated using Espript (http://espript.ibcp.fr)(40).

References

1. Das, A., and Khosla, C. (2009) Biosynthesis of aromatic polyketides in bacteria, *Acc Chem Res 42*, 631-639.
2. Moore, B. S., and Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units, *Nat Prod Rep 19*, 70-99.
3. Das, A., and Khosla, C. (2009) In vivo and in vitro analysis of the hedamycin polyketide synthase, *Chem Biol 16*, 1197-1207.
4. Meadows, E. S., and Khosla, C. (2001) In vitro reconstitution and analysis of the chain initiating enzymes of the R1128 polyketide synthase, *Biochemistry 40*, 14855-14861.
5. Bao, W., Sheldon, P. J., Wendt-Pienkowski, E., and Hutchinson, C. R. (1999) The Streptomyces peucetius dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide, *J Bacteriol 181*, 4690-4695.
6. Waldman, A. J., and Balskus, E. P. (2014) Lomaiviticin biosynthesis employs a new strategy for starter unit generation, *Org Lett 16*, 640-643.
7. Kalaitzis, J. A., Cheng, Q., Thomas, P. M., Kelleher, N. L., and Moore, B. S. (2009) In vitro biosynthesis of unnatural enterocin and wailupemycin polyketides, *J Nat Prod 72*, 469-472.
8. Friedrich, T., van Heek, P., Leif, H., Ohnishi, T., Forche, E., Kunze, B., Jansen, R., Trowitzsch-Kienast, W., Hofle, G., Reichenbach, H., and et al. (1994) Two binding sites of inhibitors in NADH: ubiquinone oxidoreductase (complex I). Relationship of one site with the ubiquinone-binding site of bacterial glucose:ubiquinone oxidoreductase, *Eur J Biochem 219*, 691-698.
9. Meunier, B., Madgwick, S. A., Reil, E., Oettmeier, W., and Rich, P. R. (1995) New inhibitors of the quinol oxidation sites of bacterial cytochromes bo and bd, *Biochemistry 34*, 1076-1083.
10. Sandmann, A., Dickschat, J., Jenke-Kodama, H., Kunze, B., Dittmann, E., and Muller, R. (2007) A Type II polyketide synthase from the gram-negative Bacterium Stigmatella aurantiaca is involved in Aurachin alkaloid biosynthesis, *Angew Chem Int Ed Engl 46*, 2712-2716.
11. Pistorius, D., Li, Y., Mann, S., and Muller, R. (2011) Unprecedented anthranilate priming involving two enzymes of the acyl adenylating superfamily in aurachin biosynthesis, *J Am Chem Soc 133*, 12362-12365.
12. Schmelz, S., and Naismith, J. H. (2009) Adenylate-forming enzymes, *Curr Opin Struct Biol 19*, 666-671.
13. Bains, J., and Boulanger, M. J. (2007) Biochemical and structural characterization of the paralogous benzoate CoA ligases from Burkholderia xenovorans LB400: defining the entry point into the novel benzoate oxidation (box) pathway, *J Mol Biol 373*, 965-977.
14. Hu, Y., Gai, Y., Yin, L., Wang, X., Feng, C., Feng, L., Li, D., Jiang, X. N., and Wang, D. C. (2010) Crystal structures of a Populus tomentosa 4-coumarate:CoA ligase shed light on its enzymatic mechanisms, *Plant Cell 22*, 3093-3104.
15. Coschigano, P. W., Haggblom, M. M., and Young, L. Y. (1994) Metabolism of both 4-chlorobenzoate and toluene under denitrifying conditions by a constructed bacterial strain, *Appl Environ Microbiol 60*, 989-995.
16. Conti, E., Franks, N. P., and Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes, *Structure 4*, 287-298.

17. Reger, A. S., Wu, R., Dunaway-Mariano, D., and Gulick, A. M. (2008) Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase, *Biochemistry 47*, 8016-8025.

18. Gulick, A. M., Lu, X., and Dunaway-Mariano, D. (2004) Crystal structure of 4-chlorobenzoate:CoA ligase/synthetase in the unliganded and aryl substrate-bound states, *Biochemistry 43*, 8670-8679.

19. Gulick, A. M. (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase, *ACS Chem Biol 4*, 811-827.

20. Wu, R., Reger, A. S., Cao, J., Gulick, A. M., and Dunaway-Mariano, D. (2007) Rational redesign of the 4-chlorobenzoate binding site of 4-chlorobenzoate: coenzyme a ligase for expanded substrate range, *Biochemistry 46*, 14487-14499.

21. Law, A., and Boulanger, M. J. (2011) Defining a structural and kinetic rationale for paralogous copies of phenylacetate-CoA ligases from the cystic fibrosis pathogen Burkholderia cenocepacia J2315, *J Biol Chem 286*, 15577-15585.

22. Goris, J., De Vos, P., Caballero-Mellado, J., Park, J., Falsen, E., Quensen, J. F., 3rd, Tiedje, J. M., and Vandamme, P. (2004) Classification of the biphenyl- and polychlorinated biphenyl-degrading strain LB400T and relatives as Burkholderia xenovorans sp. nov, *Int J Syst Evol Microbiol 54*, 1677-1681.

23. Denef, V. J., Klappenbach, J. A., Patrauchan, M. A., Florizone, C., Rodrigues, J. L., Tsoi, T. V., Verstraete, W., Eltis, L. D., and Tiedje, J. M. (2006) Genetic and genomic insights into the role of benzoate-catabolic pathway redundancy in Burkholderia xenovorans LB400, *Appl Environ Microbiol 72*, 585-595.

24. Chang, K. H., Liang, P. H., Beck, W., Scholten, J. D., and Dunaway-Mariano, D. (1992) Isolation and characterization of the three polypeptide components of 4-chlorobenzoate dehalogenase from Pseudomonas sp. strain CBS-3, *Biochemistry 31*, 5605-5610.

25. Dunaway-Mariano, D., and Babbitt, P. C. (1994) On the origins and functions of the enzymes of the 4-chlorobenzoate to 4-hydroxybenzoate converting pathway, *Biodegradation 5*, 259-276.

26. Horswill, A. R., and Escalante-Semerena, J. C. (2002) Characterization of the propionyl-CoA synthetase (PrpE) enzyme of Salmonella enterica: residue Lys592 is required for propionyl-AMP synthesis, *Biochemistry 41*, 2379-2387.

27. Branchini, B. R., Murtiashaw, M. H., Magyar, R. A., and Anderson, S. M. (2000) The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase, *Biochemistry 39*, 5433-5440.

28. Wu, R., Reger, A. S., Lu, X., Gulick, A. M., and Dunaway-Mariano, D. (2009) The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase, *Biochemistry 48*, 4115-4125.

29. Osman, K. T., Du, L., He, Y., and Luo, Y. (2009) Crystal structure of Bacillus cereus D-alanyl carrier protein ligase (DltA) in complex with ATP, *J Mol Biol 388*, 345-355.

30. Kochan, G., Pilka, E. S., von Delft, F., Oppermann, U., and Yue, W. W. (2009) Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A, *J Mol Biol 388*, 997-1008.

31. Bretschneider, T., Zocher, G., Unger, M., Scherlach, K., Stehle, T., and Hertweck, C. (2012) A ketosynthase homolog uses malonyl units to form esters in cervimycin biosynthesis, *Nat Chem Biol 8*, 154-161.

32. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol 276*, 307-326.
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, *Journal of Applied Crystallography 40*, 658-674.
34. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D 64*, 61-69.
35. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D 68*, 352-367.
36. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D 60*, 2126-2132.
37. Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Crystallogr D 65*, 1074-1080.
38. Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res 33*, W244-W248.
39. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Res 31*, 3497-3500.
40. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res 42*, W320-324.

Chapter 5

Insights Into Complex Oxidation During BE-7585A Biosynthesis: Structural

Determination and Analysis of the Polyketide Monooxygenase BexE

5.1 Abstract

Type II polyketides can be categorized by their core cyclic structures. The core is typically composed of multiple fused rings that are described as linear (anthracycline), angular (angucycline), or other more complex patterns (discoid). The core structures of type II polyketides are essential for bioactivity. Type II polyketide synthases (PKSs) use common poly-β-ketone intermediates to generate a diverse sets of bioactive molecules with vastly different structures. A critical step in this diversification process is to generate different core structures. Biosynthesis of core structures requires multiple enzymes such as ketoreductases (KRs), aromatase/cyclases (ARO/CYCs), cyclases (CYCs), and oxygenases (OXYs). Recently, a set of oxygenases (BexI, BexE, and BexM) was discovered in the biosynthetic gene cluster for BE-7585A, an angucyclic type II polyketide with anticancer properties. BE-7585A has an angucyclic core structure; however, C^{13} isotope labeling experiments suggest that the angucycline core may arise from an oxidative rearrangement of a linear anthracycline. Here we present the crystal structure and analysis of BexE, the oxygenase that is proposed to catalyze a key step in the oxidative rearrangement to generate the angucycline core. Through structural analysis and docking simulations, we provide insights into the role of BexE in BE-7585A biosynthesis and lay the groundwork for engineering oxidation in type II PKSs.

5.2 Introduction

Angucyclines are a diverse group of bioactive natural products produced by type II polyketide synthases (PKSs) that contain modified benz[*a*]anthraquinone cores(1-3). Many angucyclines have been discovered with anticancer and antibiotic properties such as the urdamycins, landomycins, gaudimycins, oviedomycins, jadomycins, and azicemicins (Figure 5-1)(4-9). The biosynthesis of angucylines has been studied extensively due to their important biological activity and pharmaceutical potential(3).

Type II polyketides are biosynthesized two carbons at a time from malonyl-CoA building blocks to yield reactive poly-β-ketones(2, 3). The reactive poly-β-ketone intermediates are regiospecifically reduced, cyclized, and aromatized by ketoreductases (KRs), cyclases (CYCs), and aromatase/cyclases (ARO/CYCs), respectively(3, 10). These initial enzymatic transformations yield polycyclic aromatic intermediates, termed the "aglycones". The aglycones then undergo a variety of tailoring reactions that include, but are not limited to oxidation, methylation, reduction, dehydration, and glycosylation(11). Tailoring enzymes are nature's tools to diversify type II polyketide biosynthesis. Many type II PKS products share common aglycone cores, and different combinations of tailoring enzymes modify the aglycone core to yield a diverse molecules with anticancer, antibiotic, and antiviral activities(10, 11). For example, the widely used anticancer drugs daunorubicin and doxorubicin are both hydroxylated during biosynthesis, which is important for bioactivity(12). Tailoring enzymes often use structurally complex and unstable substrates, which are formed transiently during biosynthesis(6). Therefore, structural and functional characterization of these enzymes, their substrates, and ensuing products is extremely challenging. Here, we report the structural and functional studies of the
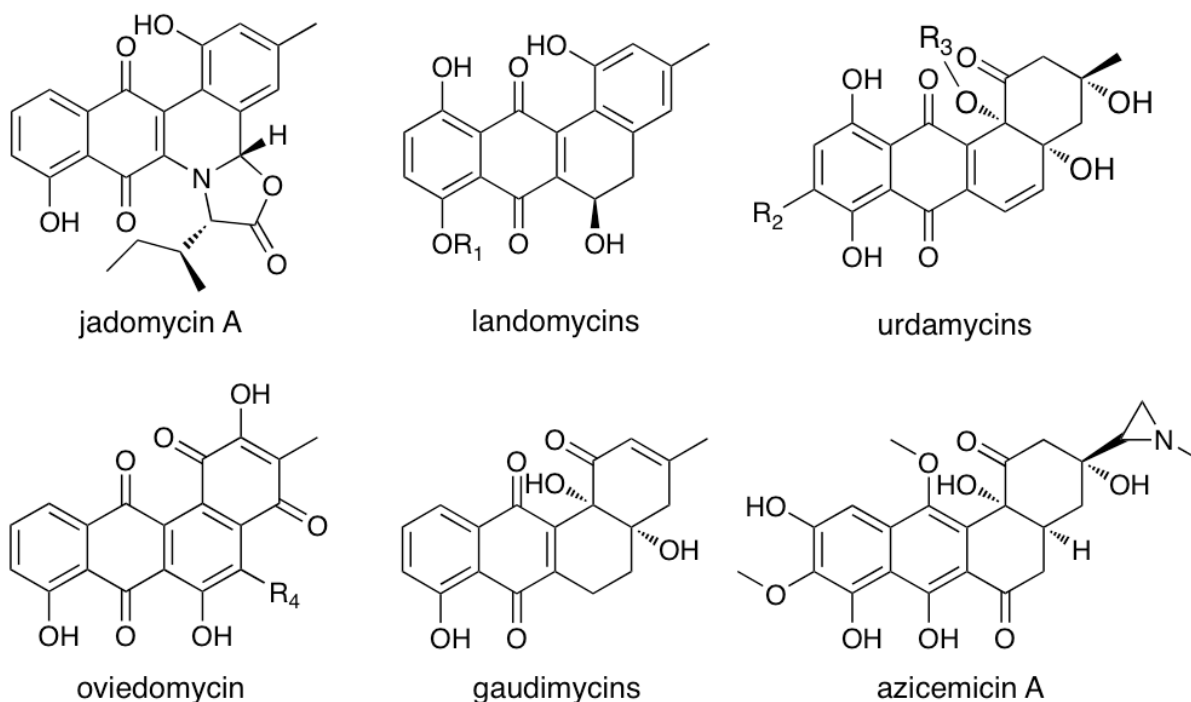
Figure 5-1. A representative set of angucycline type II polyketides.

oxygenase BexE, which is involved in an uncharacterized oxidation step during BE-7585A biosynthesis(13).

BE-7585A is an angucyclic polyketide with anticancer properties that was isolated from *Amycolotopsis orientalis* sp. *vinerea*(13). Because of its complex structure, BE-7585A has inspired multiple biosynthetic studies(13-15). C^{13} isotope labeling suggests that BE-7585A is initially biosynthesized as an anthracycline then undergoes an oxidative rearrangement to yield an angucycline (Figure 5-2)(13). The BE-7585A gene cluster contains three candidate oxygenase genes, BexE, BexI, and BexM, which are proposed to catalyze oxidative tailoring of the BE-7585A aglycone core(13). Based on protein sequence similarity, BexE is most similar to aromatic hydroxylases that oxidize angucyclines. Elucidating the structure and function of BexE should help us understand how the angucycline core is produced. Here, we present the crystal structure of BexE bound to FAD, molecular docking of proposed substrates, and preliminary *in*
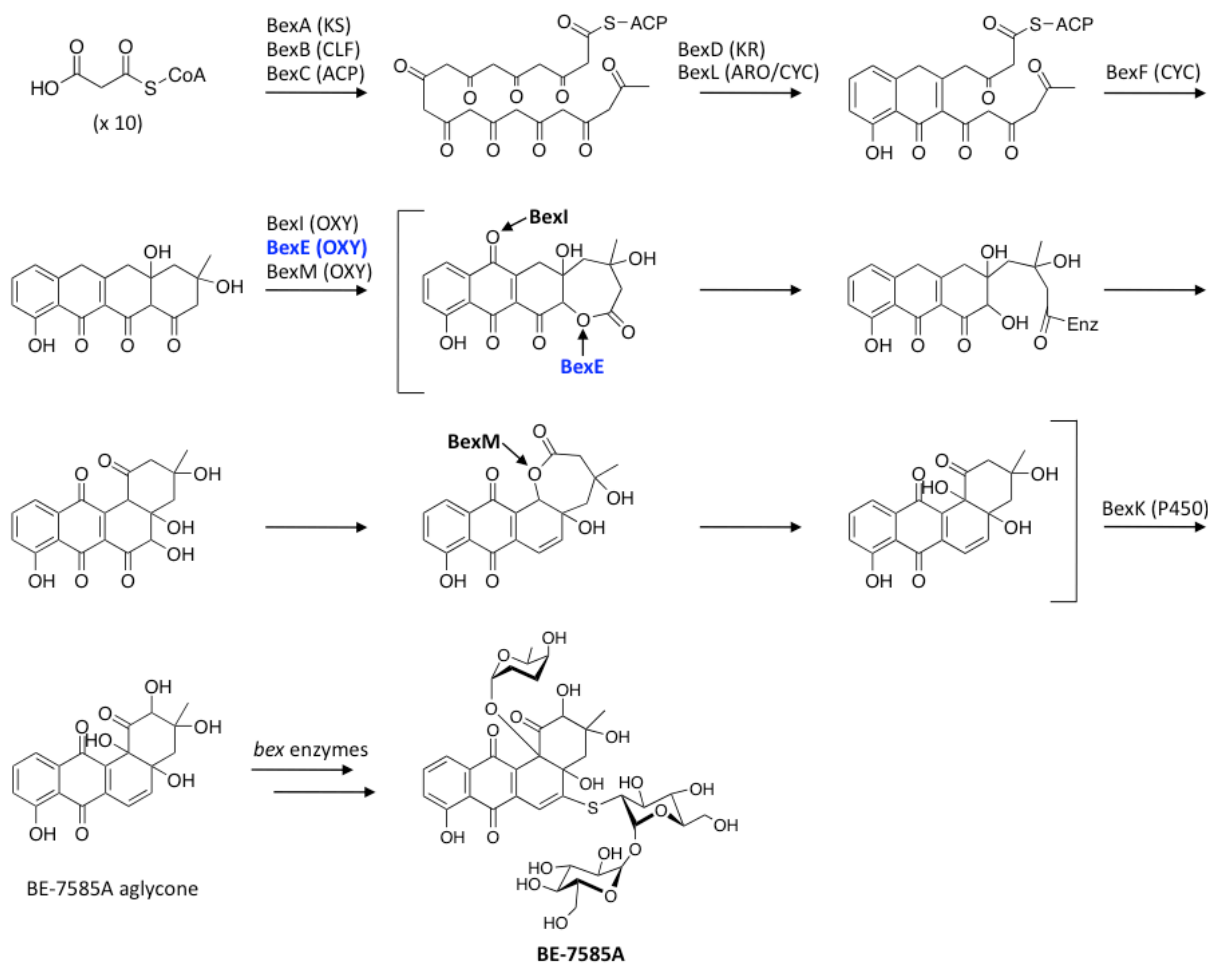
Figure 5-2. The proposed biosynthesis of BE-7585A with an emphasis on the oxidation reactions. The minimal PKS produces a 20-carbon poly-β-ketone that is reduced at the C9 position and cyclized. BexF is proposed to catalyze 4th ring cyclization to produce the tetracyclic intermediate that is the substrate for a series of putative oxidation steps. BexE, BexI, and BexM are proposed to form the BE-7585A aglycone core, which is further modified by hydroxylation and glycosylation to yield the final product.

vitro functional studies. A critical comparison of BexE with RdmE, a biosynthetic FAD-dependent oxygenase is also presented. These results offer the first insight into oxidative tailoring during BE-7585A biosynthesis and provide a basis for engineering oxidation during type II polyketide biosynthesis to yield new bioactive molecules.

5.3 Results and Discussion

5.3.1 Overall Structure of BexE

BexE crystallized in the space group C2 with two BexE molecules per asymmetric unit. The BexE dimer is composed of two monomers related by a non-crystallographic 2-fold rotational axis of symmetry (Figure 5-3A). The protein-protein interface between monomers is moderate with a surface area estimated to be 922.7 Å^2 (PISA)(16). Both monomers are nearly identical and when overlaid the overall RMSD is 0.241 Å. Each monomer contains an FAD binding domain (residues Met1-Thr169 and Thr259-Gly372), a middle domain (residues Ala170-Ala258), and a C-terminal domain (Leu373-Arg487) (Figure 5-3B). This domain arrangement is in agreement with other structurally characterized members of the para-hydroxybenzoate hydroxylase pHBH superfamily, including related type II polyketide and alkaloid monoxygenases (Figure 5-4 and Figure 5-5)(17-22). The FAD binding domain consists of ten α-helices (α1-α6 and α8-α10) and ten β-sheets (β1-β9 and β14) that are connected by a series of 18 loops (L1-L13 and L18-L22). FAD is bound almost entirely in the FAD binding domain, while only contacting the middle domain by hydrophobic interactions with the 1,2 dimethyl moiety of the isoalloxazine ring (Figure 5-3B and 5-3C). The middle domain is inserted between α6 and β14 of the FAD binding domain. The middle domain consists of one α-helix (α7) and four β-sheets (β10-β13) that are connected by a series of four loops (L14-L17). The middle domain has been implicated in substrate binding and, in some cases, undergoes structural rearrangements during substrate binding and product release(19, 21-25). In BexE, a large pocket leads into the isoalloxazine ring of FAD, which is formed between the FAD binding domain and the middle domain. This pocket is hypothesized to be the substrate binding pocket. The C-terminal domain has a thioredoxin fold but lacks the conserved active site cysteines for thioredoxin
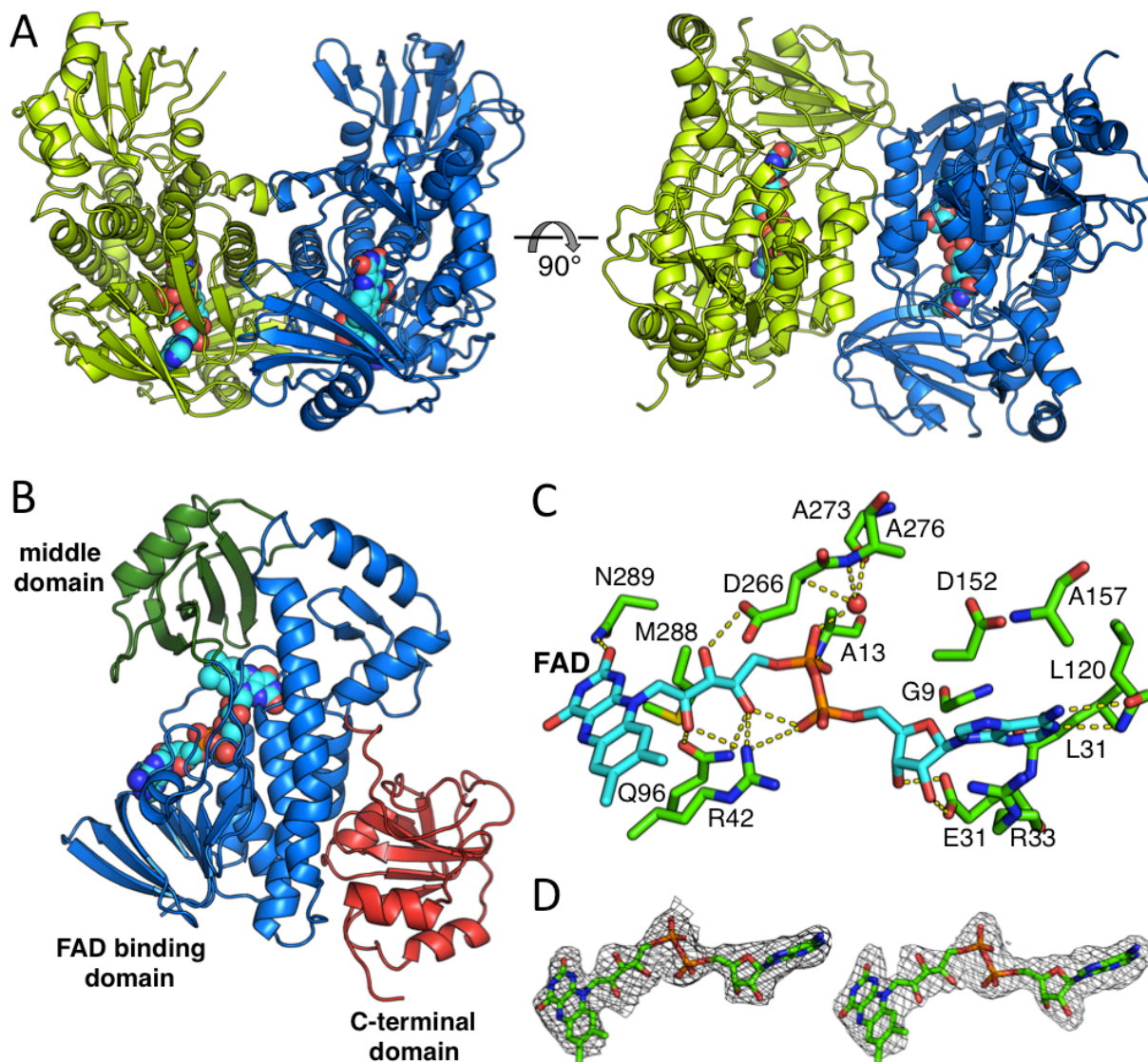
Figure 5-3. The overall structure of the BexE dimer and domain organization of the BexE monomer. (A) The BexE dimer (monomer A in green and monomer B in blue). (B) The BexE monomer is composed of three domains: the FAD binding domain (residues 1-169 and 259-372, colored in blue), the middle domain (residues 170-258, colored in green), and the C-terminal domain (373-487, colored in red). (C) A molecular view of the BexE FAD binding site. (D) SA-omit electron density maps displaying clear density for FAD in monomer A (left) and monomer B (right).

activity(26). Therefore, the C-terminal domain may be important for structural stability of BexE, but not for catalytic activity.
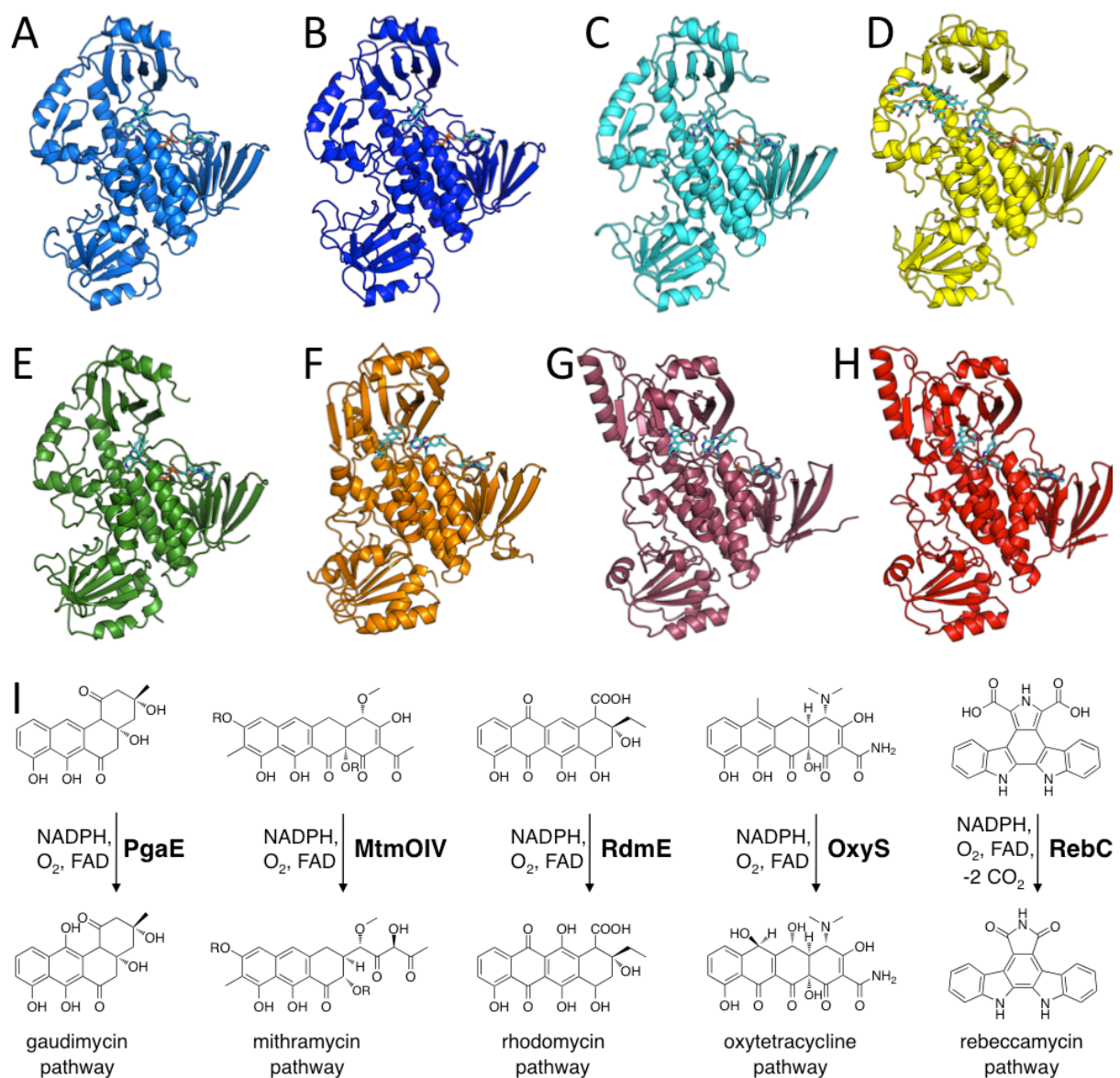
Figure 5-4. The crystal structures of oxygenases from polyketide and alkaloid biosynthetic pathways. The middle domain is located at the top of each structure, the FAD binding domain in the middle, and the C-terminal domain is at the bottom. (A) BexE (polyketide oxygenase) bound to FAD. (B) PgaE (polyketide oxygenase) bound to FAD. (C) CabE (polyketide oxygenase) bound to FAD. (D) MtmOIV (polyketide oxygenase) bound to FAD and premithramycin B. (E) OxyS (polyketide oxygenase) bound to FAD. (F) RdmE (polyketide oxygenase) bound to FAD and aklavinone. (G) RebC (alkaloid oxygenase) bound to FAD and K252c. (H) RebC (alkaloid oxygenase) bound to FAD and K252c. (I) Reactions catalyzed by oxygenases that are related structurally related to BexE.

170

5.3.2 FAD Binding in BexE

FAD consists of three parts: the adenosine (adenine and ribose), riboflavin, and the bridging diphosphates. The adenine ring of FAD is positioned between Arg33 (L3) and Asp152 (L12), which interact with opposite faces of the adenine ring (Figure 5-3C). Gly9 (L1), Ala157 (α6), and Leu31 (β2) define the adenine-binding site by interacting with the edges of the adenine ring. Arg33 forms a hydrogen bond with N7 of adenine while the L120 (β6) amide NH and carbonyl form hydrogen bonds with the N1 of adenine and amine group at the C6 position. The ribose ring of FAD is anchored tightly in place by Glu32 (β2), which forms hydrogen bonds to the 2′ and 3′ hydroxyl groups. The first phosphate, moving from the adenosine towards the isoalloxazine ring, forms an intramolecular hydrogen bond with the riboflavin hydroxyl group. This phosphate also forms a weak hydrogen bond with Arg42. The second phosphate forms a hydrogen bond with the amide NH of Ala13 (α1) and also a water molecule, which is held in place the Ala273 (β14) carbonyl oxygen and Ala276 (L19) amide NH. The three hydroxyl groups of the riboflavin form hydrogen bonds with Arg42 (L3), Gln96 (α5), and Asp275 (L19). The isoalloxazine is positioned between the C-terminal end of α8, loops L3 and L9, and β10 and β13. L3 and L9 interact with opposite faces of the isoalloxazine ring, while β10 and β13 make hydrophobic contacts with the dimethyl moiety. The C2 carbonyl oxygen of the isoalloxazine ring forms hydrogen bonds with the Asn289 (α8) side chain and M288 (α8) amide NH. The extensive interactions between FAD and BexE support that the cofactor is tightly bound in the structure.

Enzyme	Organism	Natural product	Type	PDB ID	% Identity	RMSD (A)	Ligand(s)	FAD
BexE	*Amycolatopsis orientalis*	BE-7585A	type II PKS	N/A	100	0.00	FAD	in
PgaE	*Streptomyces PGA64*	gaudimycin	type II PKS	2AQ1	74	0.70	FAD	in
PgaE	*Streptomyces PGA64*	gaudimycin	type II PKS	4ICY	74	0.60	FAD	in
CabE	*Streptomyces H201*	unknown	type II PKS	2QA2	46	0.53	FAD	in
MtmOIV	*Streptomyces argillaceus*	mithramycin	type II PKS	4K5S	48	1.04	FAD, premithramycin B	in
MtmOIV	*Streptomyces argillaceus*	mithramycin	type II PKS	3FMW	48	1.02	FAD	in
OxyS	*Streptomyces rimosus*	oxytetracycline	type II PKS	4K2X	45	0.82	FAD	in
RdmE	*Streptomyces purpurascens*	rhodomycin	type II PKS	3IHG	31	3.10	FAD, aklavinone	out
RebC	Lechevalieria aerocolonigenes	rebbecamycin	alkaloid	2R0C	28	4.35	FAD	out
RebC	Lechevalieria aerocolonigenes	rebbecamycin	alkaloid	4EIP	28	4.04	FAD, K252c	in
RebC	Lechevalieria aerocolonigenes	rebbecamycin	alkaloid	4EIQ	28	3.95	KCT	N/A
RebC	Lechevalieria aerocolonigenes	rebbecamycin	alkaloid	2R0G	28	5.59	FAD, chromo + 7-K252c	in
RebC	Lechevalieria aerocolonigenes	rebbecamycin	alkaloid	2R0P	28	3.01	FAD, K252c soaked	out
RebC	Lechevalieria aerocolonigenes	rebbecamycin	alkaloid	3EPT	28	4.37	FAD (reduced)	in

Table 5-1. A table listing the crystal structures of biosynthetic oxygenases that are related to BexE. The % identity and RMSD are relative to the BexE structure. The last column, "FAD" corresponds to the conformation of FAD.

5.3.3 The Role of FAD in BexE and Related Oxygenases

Many oxygenase enzymes, such as BexE, employ FAD for redox-based catalysis(27). For NAD(P)H dependent enzymes, the FAD cofactor is reduced by NAD(P)H and then reacts with molecular oxygen to form a highly reactive oxygen species at the C4a position of the isoalloxazine ring. FAD reduction is proposed to occur when the isoalloxazine ring of FAD is flipped away from the active site and towards the solvent. This outward conformation of FAD is term the "out" conformation(28). Once FAD is reduced by NAD(P)H, the isoalloxazine moves back into the active site pocket, where oxidation of the cofactor and subsequent oxidation of the
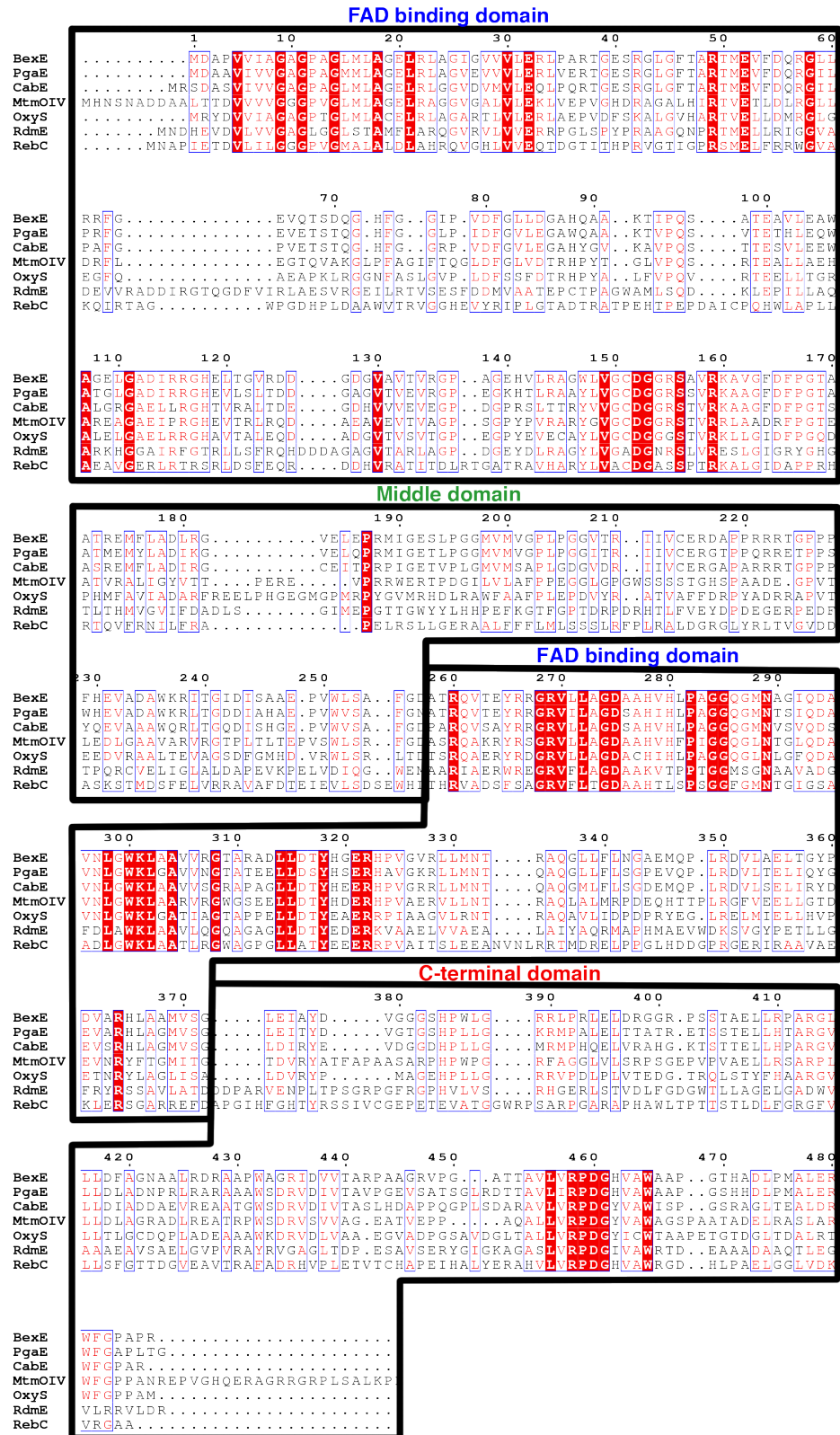
Figure 5-5. Sequence alignment of BexE with other oxygenases from biosynthetic pathways.

substrate is proposed to occur. When the FAD moiety is positioned inwards towards the active site pocket, the conformation is the "in" conformation. Multiple crystal structures of FAD/NAD(P)H-dependent oxygenases have been solved with FAD in either the "in" or "out" conformations (Figure 5-4 and Table 5-1)(6, 17, 19, 21-25). A subset of these structures have a substrate bound that allows the identification of substrate binding pockets, as well as the relative locations between FAD and the bound substrate(19, 22). In summary, while homologous oxygenases have been crystallized with FAD in the "in" conformation or "out" conformations, the BexE FAD is in the "in" conformation with a large open active site pocket (Table 5-1). The BexE co-factor conformation and BexE pocket with be discussed in detail in section 5.3.10.

5.3.4 Identifying the BexE Substrate Using *In Vitro* Functional Assays

To determine the substrate and product of the reaction catalyzed by BexE, our co-workers in Jurgen Rohr's laboratory at the University of Kentucky tested the ability of BexE to oxidize linear and angucyclic tetracyclic polyketides (Figure 5-6). *In vitro* assays were conducted with an NADPH regeneration system and included BexF, which is a putative 4th ring cyclase. Initial BexE assays in the presence of NADPH using possible linear substrates such as presteffimycinone, premithramycinone, aklaviketone, and anthracyclinone did not lead to product formation (Figure 5-6). Additionally, assays testing the reversibility of BexE using possible angucycline substrates such as landomycinone, 11-deoxy-landomycinone, tetrangulol, and tetrangomycin in the presence of NADP[+] showed no product formation (Figure 5-6). However, new products were observed when BexE was incubated in the presence of NADPH with the crude lysate from *Streptomyces galilaeus* (ATCC 31615) mutant H036, which produces aklaviketone and anthracyclinone (Figure5-7)(29). The crude lysate predominantly contains
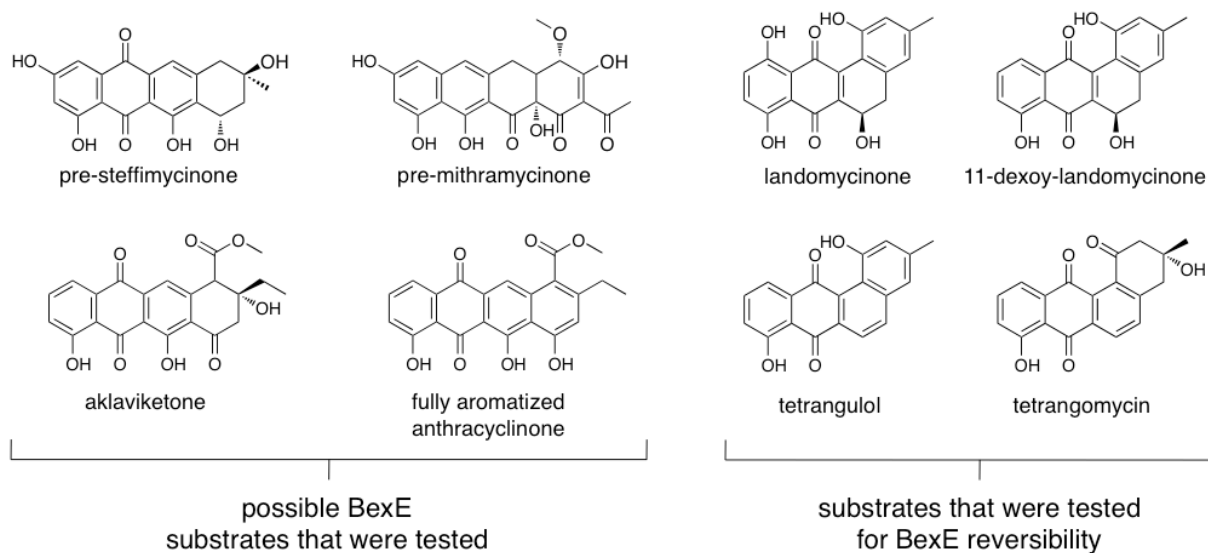
Figure 5-6. Possible substrates tested for BexE activity.

aklaviketone and anthracyclinone; however, there are also many unstable and uncharacterized biosynthetic intermediates en route to these products in the lysate. To identify the true BexE substrate, the crude lysate was fractionated, and individual fractions were tested for activity with BexE (Figure 5-8). A single peak present in "fraction 2" was converted into two new products when incubated with BexE. Interestingly, the unknown BexE substrate accumulates in the crude lysate over time and reaches a maximum amount after 5 hours at room temperature. We are currently determining the chemical structure of the BexE substrate and products. In summary, we have developed an *in vitro* functional assay for BexE using the crude lysate of *Streptomyces galilaeus* (ATCC 31615) mutant H036, which produces aklaviketone. The lysate contains an uncharacterized product that is most likely a shunt product or biosynthetic intermediate, which is produced during aklaviketone biosynthesis.
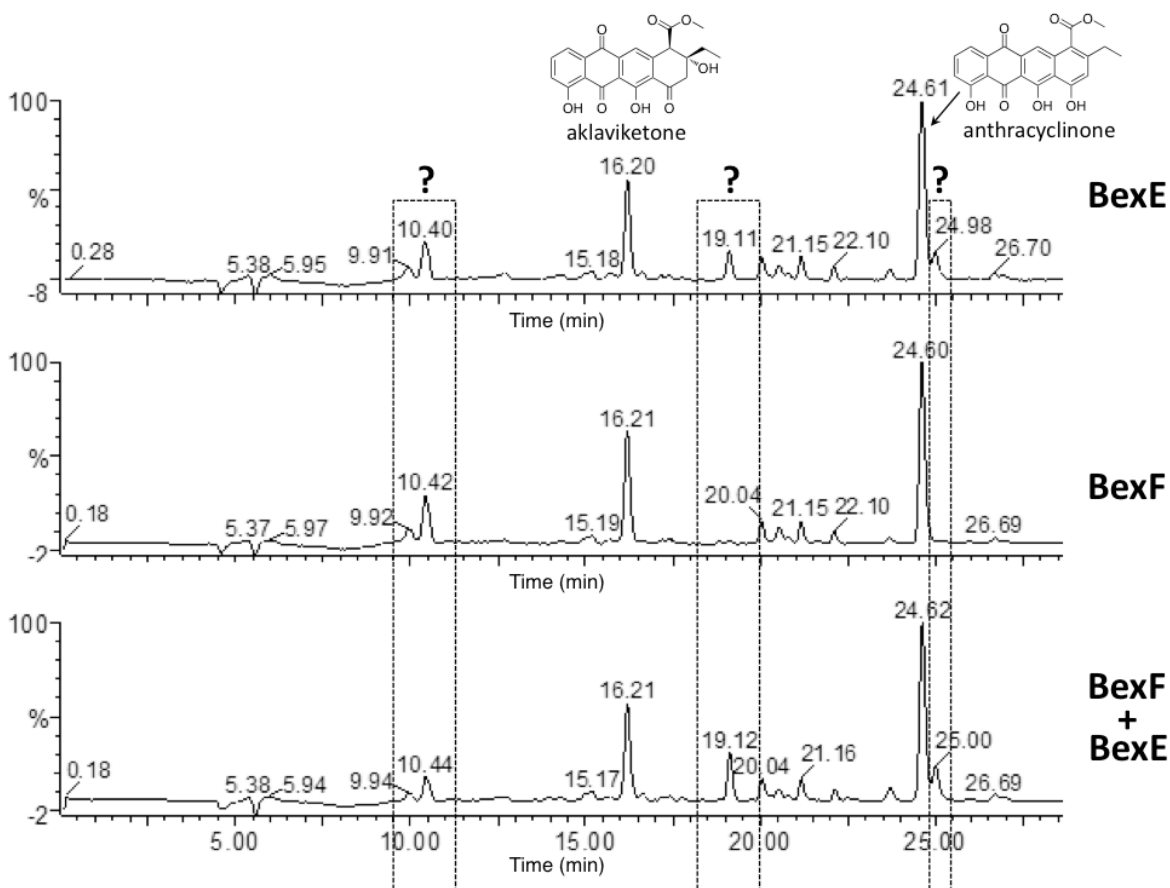
Figure 5-7. HPLC analysis of reactions using the *Streptomyces galilaeus* (ATCC 31615) lysate with BexE and BexF. The main products in the lysate are aklaviketone and anthracyclinone, shown at 16.20 and 24.61 minutes, respectively. The addition of BexE to the *Streptomyces galilaeus* (ATCC 31615) lysate results in the appearance of two new peaks at 19.11 and 24.98 minutes. The addition of BexF to the lysate resulted in an increase in the size of the peak at 10.42 minutes. The addition of BexE and BexF to the lysate resulted in the appearance of the same two new peaks at 19.11 and 24.98 minutes, which were present with BexE alone. However, when BexF and BexE are present, the peaks at 19.11 and 24.98 minutes are increased relative to BexE alone . Therefore, the presence of BexF increases BexE product formation. *Dr. Xia Yu in Dr. Jurgen Rohr's group conducted the *in vitro* BexE studies.

5.3.5 The Effect of BexF During the BexE *In Vitro* Assay

BexF is the proposed 4th cyclase during BE-7885A biosynthesis and shares 39% identify to TcmI, the 4th ring cyclase during tetracenomycin biosynthesis (Figure 5-9A)(30, 31). BexF is proposed to catalyze cyclization(s) immediately before the BexE-catalyzed oxidation during BE-7585A biosynthesis; however, the substrate and product of BexF are unknown. When BexF was

176
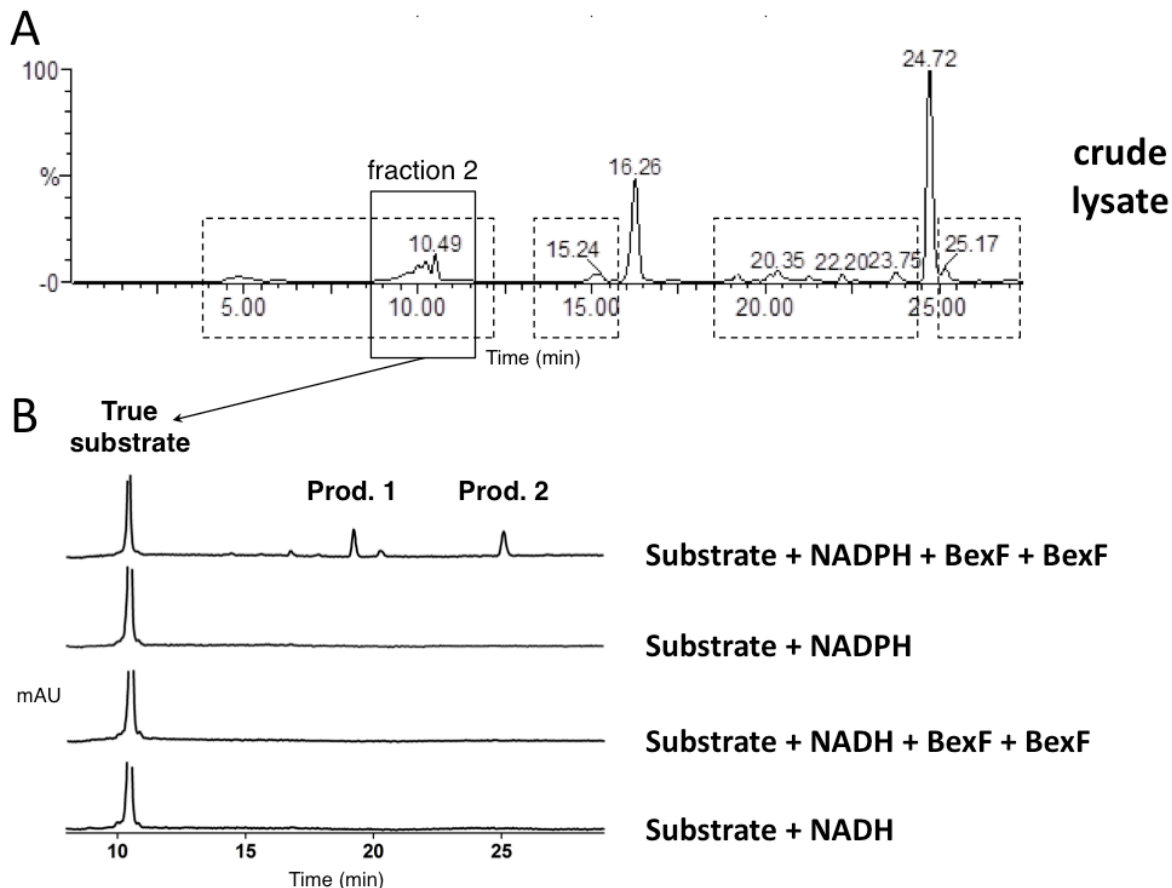
Figure 5-8. The fractionation of the *Streptomyces galilaeus* (ATCC 31615) lysate to identify the BexE substrate. (A) BexE activity was only observed in lysate that was incubated at room temperature for 5 hours. Therefore, HPLC was used to fractionate areas with increased peak sizes in the lysate after 5 hours of incubation to identify the BexE substrate (shown in dotted boxes). Fraction 2 contained a peak at 10.49 minutes, which contains the true BexE substrate. (B) Fraction 2 was assayed with BexE and BexF in the presence of NADPH or NADH. Assays with the purified fraction 2 confirmed that the BexE reaction is NADPH dependent. *Dr. Xia Yu in Dr. Jurgen Rohr's group conducted fractionation and *in vitro* BexE studies.

incubated with BexE and the *Streptomyces galilaeus* (ATCC 31615) mutant H036 crude lysate, BexE product formation increased (Figure 5-7). Therefore, BexF may affect BexE product formation through a number of possible mechanisms. For example, we hypothesize that BexF is involved in the generation or stabilization of the BexE substrate. Alternatively, BexF may form a complex with BexE and alter the kinetics of the BexE reaction, or help shuttle a reactive substrate into the BexE active site. There are previous reports that different oxygenases may

177

associate together, but never a cyclase (BexF) and oxygenase (BexE). The observation that the inclusion of BexF increases BexE product formation strongly supports that these two enzymes, a cyclase and an oxygenase, associate together. This would be the first observation of such association between these types of tailoring enzymes.

5.3.6 Possible BexF and BexE Substrates

Based on our *in vitro* assay results, the BexE substrate in the crude lysate of *Streptomyces galilaeus* (ATCC 31615) mutant H036 is not aklaviketone or anthracyclinone. Therefore, the BexF or BexE substrate is possibly a tricyclic compound such as aklanonic acid, aklanonic acid methyl ester, 12-deoxy-aklanonic acid, or a tetracyclic substrate such as aklaviketonic acid or 12-deoxy-aklaviketonic acid (Figure 5-9B). Because of the unavailability or instability of the proposed substrates, the unaromatized intermediates shown in Figure 5-9B could not be tested *in vitro*. However, we are exploring compounds such as aklanonic acid and aklanonic acid methyl ester as BexF or BexF substrates. 12-deoxy-aklanonic acid and 12-deoxy-aklanonic acid methyl ester are prone to oxidation at the C12 position; therefore, these compounds may not be stable over long periods of time. However, during biosynthesis, 12-deoxy-aklanonic acid derivatives may be stabilized by sequestration into an enzyme active site or binding pocket, which could protect the C12 position from oxidation. The protected 12-deoxy-aklanonic acid could then be shuttled directly into the correct active site for catalysis. BexF may catalyze 4th ring formation of 12-deoxy-aklanonic acid to form 12-deoxy-aklaviketonic acid, which we hypothesize to be the BexE substrate.

The presence or absence of the methyl ester at C10 may also influence activity. The C10 carboxy methyl ester is important for the cyclization of aklanonic acid methyl ester to form

178
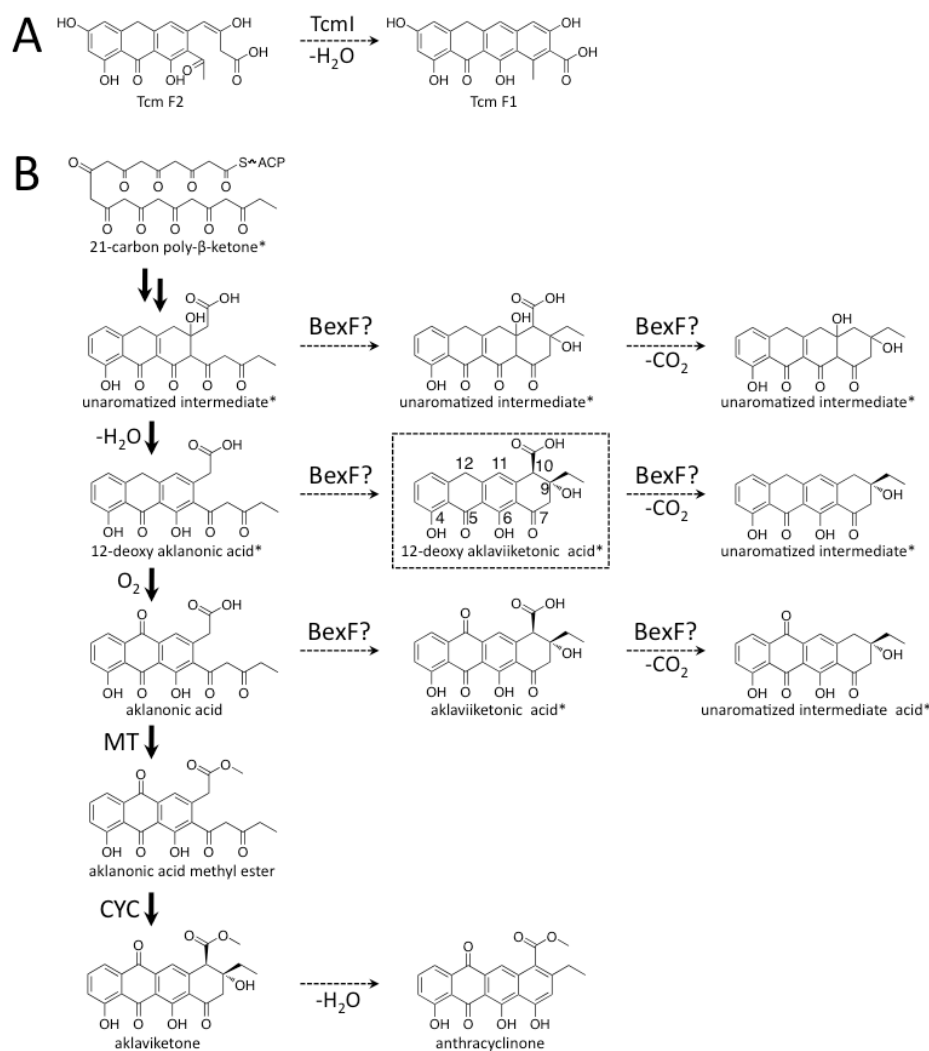
Figure 5-9. The reaction catalyzed by TcmI and type II polyketide biosynthesis of aklaviketone in *Streptomyces galilaeus* (ATCC 31615). (A) TcmI is a 4th ring cyclase from the tetracenomycin biosynthesis pathway that catalyzes closure of the 4th ring of Tcm F2 to form Tcm F1. (B) Aklaviketone biosynthesis in *Streptomyces galilaeus* (ATCC 31615) follows a linear pathway starting from a 21-carbon intermediate (shown in bold arrows). Many putative intermediates occur transiently during aklaviketone biosynthesis, and we hypothesize that BexF may cyclize an intermediate to form the BexE substrate. Based on molecular docking studies, we propose that BexF generates 12-deoxy-aklaviketonic acid, which is then used by BexE as a substrate. *Intermediates or shunt products that have not been isolated due to low yield or instability.

179
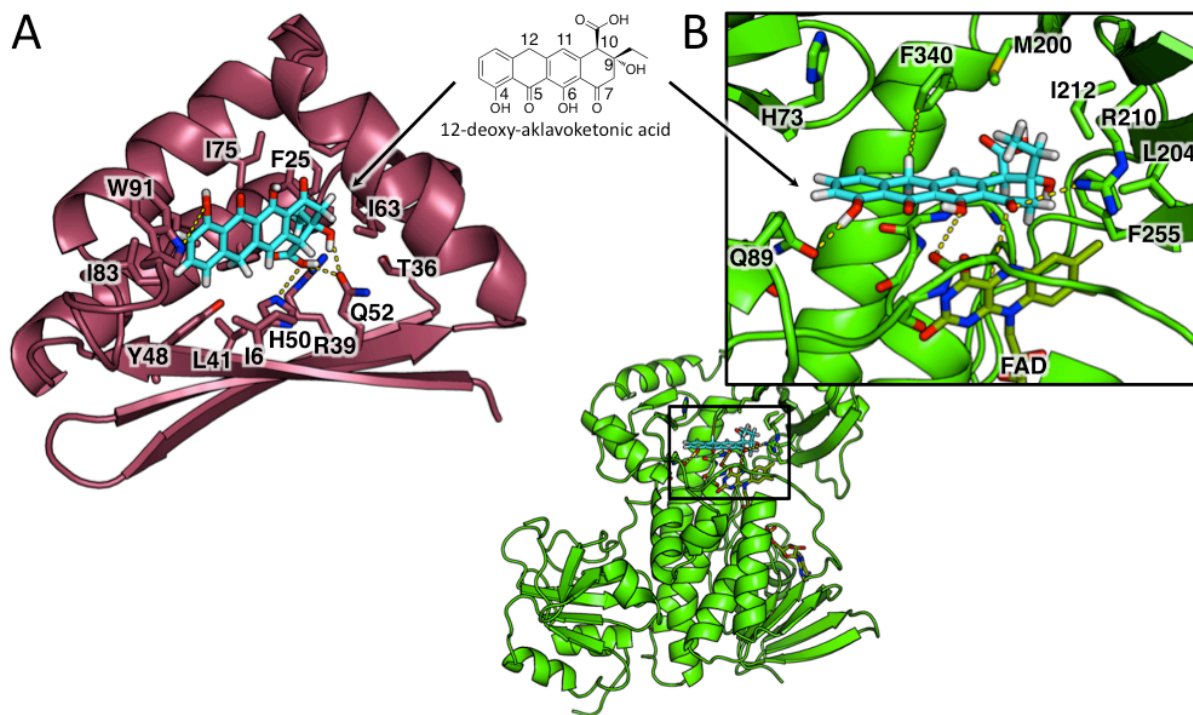
Figure 5-10. The docking of 12-deoxy-aklaviketonic acid into the active sites of the BexF homology model and BexE crystal structure. (A) 12-deoxy-aklaviketonic acid was docked into the active site pocket of the BexF homology model to determine if this product would engage in favorable interactions with active site residues. 12-deoxy-aklaviketonic acid docks in the BexF active site, with favorable hydrophobic and hydrophilic interactions with active site residues. Specifically, the 4th ring of 12-deoxy-aklaviketonic acid is docked deep in the active site and makes interactions with His50, Arg39, and Gln52. The tetracyclic core its in a hydrophobic cavity and is anchored by a hydrogen bond between the C4 hydroxyl group and Trp91. (B) 12-deoxy-aklaviketonic acid was docked into the active site pocket of the BexE crystal structure to determine if this product would engage in favorable interactions with active site residues. 12-deoxy-aklaviketonic acid docks with the 4th ring close to the C4a carbon of FAD, the site of peroxy-flavin generation. The C4 hydroxyl group of 12-deoxy-aklaviketonic acid is anchored by a hydrogen bond with Gln89. Hydrophobic interactions occur between the hydrophobic side 12-deoxy-aklaviketonic acid and F340. Additionally, the propyl moiety of 12-deoxy-aklaviketonic acid forms hydrophobic contacts with Met200, Ile212, and Phe255. Arg210 forms a hydrogen bond with the C7 carbonyl, which may stabilize an oxyanion if this carbonyl undergoes nucleophilic attack by the peroxy-flavin intermediate.

aklaviketone by the cyclase DnrD in daunorubicin biosynthesis(32). Therefore, the C10 carboxylate may be an important determinant of substrate specificity for BexE.

Another possible determinant of substrate specificity is the stereochemistry at the C9 position. During aklaviketone biosynthesis in *Streptomyces galilaeus* (ATCC 31615) the

stereochemistry at the C9 position is set during the enzyme-catalyzed cyclization of aklanonic acid methyl ester(32). However, aklanonic or 12-deoxy-aklanonic acid may undergo spontaneous decarboxylation, resulting in the formation of a carbanion at the C10 position. The reactive C10 carbanion is poised to form a six-membered ring by attacking the C9 carbonyl. This reaction would produce equal mixture of stereoisomers at the C9 position. It is possible that spontaneous decarboxylation occurs, which leads to the formation of the BexE substrate with different stereochemistry than aklaviketone. BexF may increase the rate of decarboxylation/cyclization of aklanonic or 12-deoxy-aklanonic acid, and thus it would favor one stereoisomer over the other. This scenario could explain the increase in product formation when BexF is present in the assay with crude lysate.

In summary, we have found that BexE can convert a to-be-determined substrate from the crude lysate of *Streptomyces galilaeus* (ATCC 31615) into two unknown products in an NADPH-dependent reaction. *Streptomyces galilaeus* (ATCC 31615) produces aklaviketone; therefore, we proposed a series of candidate BexE substrates that are expected to be intermediates during aklaviketone biosynthesis.

5.3.7 BexF Active Site Docking

BexF is likely to produce the BexE substrate. To gain more information about the possible structure of the BexE substrate, we investigated possible substrates and products of BexF using molecular docking with the BexF homology model (Figure 5-10A). A homology model of BexF was generated using HHPred based on the TcmI crystal structure(33). Linear tricyclic and tetracyclic polyketide intermediates were docked into the BexF model to reveal the impact of C9 stereochemistry, C12 oxidation, and the C10 carboxyl group on substrate binding
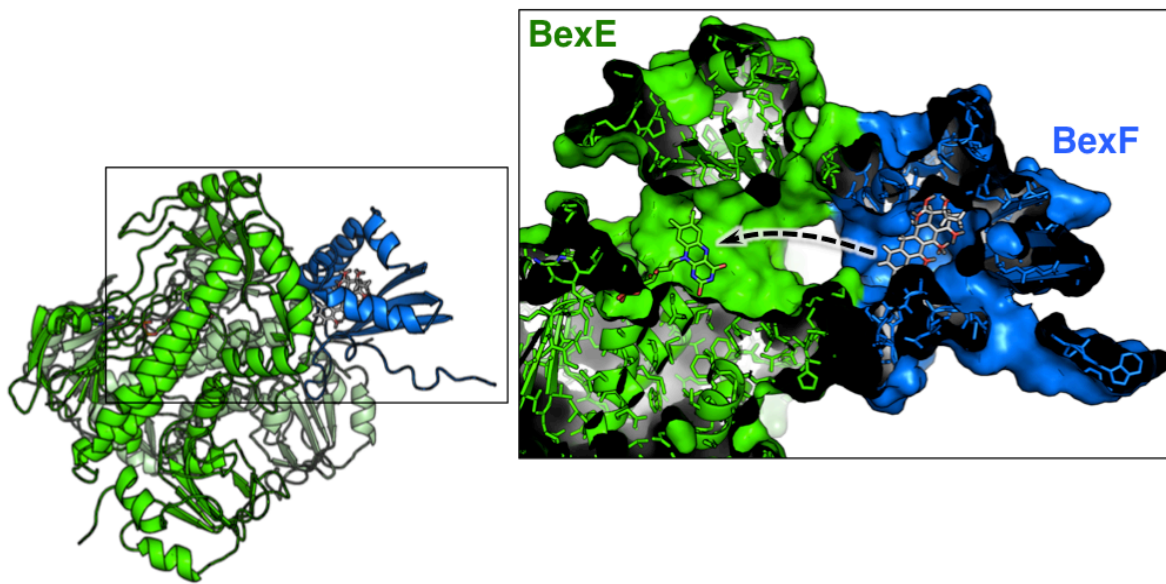
Figure 5-11. A cartoon representation of the BexF homology model bound to a tetracyclic substrate and docked to the surface of BexE. The BexF model contains a large pocket and active site entrance. When BexF is docked with BexE, a clear interface between the two enzyme active site entrance forms to reveal a tunnel, which we hypothesize is responsible for transferring a reactive substrate from the BexF active site to the BexE active site.

in the active site. In total, over 500 docking solutions were analyzed for different possible BexF substrates and products. 12-deoxy-aklaviketonic acid, a putative BexF product, showed the most favorable interactions with BexF (Figure 5-10A). These preliminary docking results suggest that 12-deoxy-aklaviketonic acid is a possible BexF product; however, aklaviketonic acid or a different tetracyclic intermediate may be the true product. Structural elucidation of the BexE substrate from the *Streptomyces galilaeus* (ATCC 31615) lysate will greatly enhance our understanding of the reaction catalyzed by BexF.

5.3.8 BexE Active Site Docking

To determine the potential substrate for BexE, multiple tetracyclic intermediates were docked into the BexE active site near FAD. Similar to BexF, 12-deoxy-aklaviketonic acid docked well into the BexE active site (Figure 5-10B). The tetracyclic structure of 12-deoxy-

aklaviketonic acid stacks on the backbone of L3 and forms a series of hydrogen bonds and hydrophobic interactions with BexE (Figure 5-10B). The BexE substrate pocket is sufficiently large to accommodate a range of tetracyclic molecules. However, upon substrate binding, we hypothesize that the middle domain undergoes a conformational change, which seals the substrate in the active site. We are currently scaling up the fractionation of the *Streptomyces galilaeus* (ATCC 31615) lysate to generate enough BexE substrate for co-crystallization studies. A BexE co-crystal structure with the true substrate will shed light on the nature of conformational changes upon substrate binding. In summary, docking results of BexE with 12-deoxy-aklaviketonic acid support our hypothesis that 12-deoxy-aklaviketonic acid may be the BexE substrate.

5.3.9 Docking a BexF Homology Model with BexE

To visualize a possible BexE/BexF complex, a homology model of BexF was generated using HHPred based on the TcmI crystal structure and docked with BexE using COOT(33, 34). The docked BexE/BexF complex was further energy minimized using Chimera(35). The BexF homology model has a large open active site and wide entrance capable of binding tetracyclic linear and angucycline polyketides. The BexF model was docked with the BexE structure in an orientation that positions their active site entrances in close contact (Figure 5-11). This orientation mimics a possible binding interaction between BexE and BexF that allows BexF to pass an unstable substrate into the BexE active site. The BexF model binds to the middle domain BexE, which in other related oxygenases can undergo structural rearrangements upon substrate binding(23). Additionally, there is a large central cavity between the two BexE monomers that may facilitate BexF docking by providing a large interface for BexF docking. It is possible that
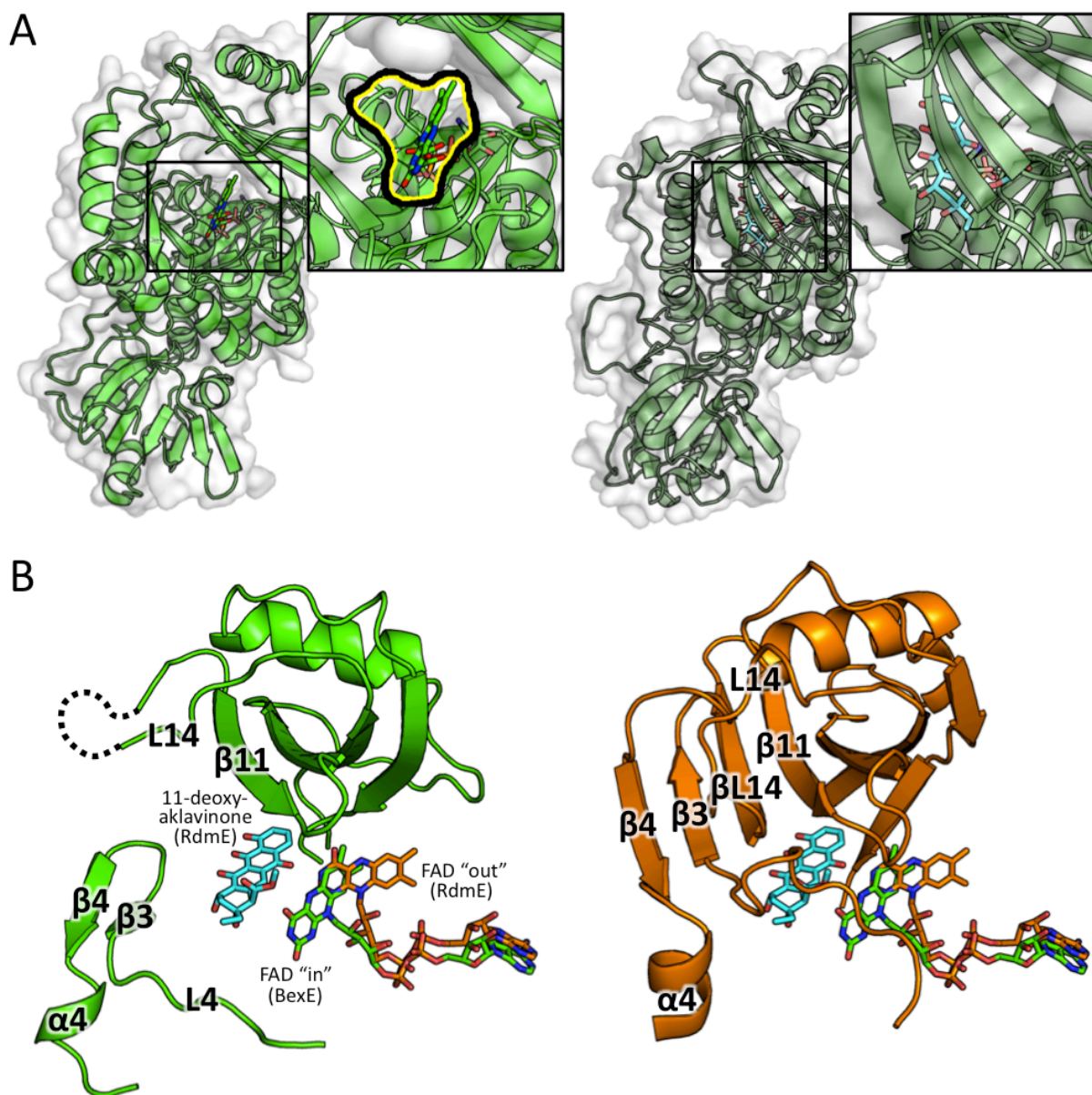
Figure 5-12. A structural comparison between the middle domains of BexE with FAD in the "in" position and RdmE bound to aklaviketone and FAD in the "out" position. (A) BexE with FAD in the "in" position with a zoomed in view of the putative active site entrance (left). RdmE bound to aklaviketone with FAD in the "out" position with a zoomed in view of the area corresponding to the BexE active site entrance (right). In RdmE, there is no active site entrance in this area where the substrate is bound. (B) In BexE (left, light green), FAD is in the "in" position with an open active site. The β-hairpin consisting of β3 and β4 is relaxed and shifted downward compared to the same region in RdmE. The L14 region of BexE could not be modeled because of lack of electron density; however, in RdmE this loop forms a β-strand (βL14), which forms a beta sheet with β3, β4, and β11. *Aklaviketone, and FAD from both structures are overlaid in both BexE and RdmE to highlight the FAD motion with respect to substrate position.

BexF protects and shuttles a reactive intermediate to the BexE active site, and structural rearrangements in the middle domain of BexE occur during complex formation with BexF. In summary, we propose that a BexE/BexF complex may form that connects the active site entrances of BexF and BexE to facilitate transfer of a reactive product from BexF to BexE.

5.3.10 Comparison with RdmE

BexE is structurally similar to polyketide hydroxylase RdmE with an RMSD 3.10 Å. Their overall sequence identity is 31 %, and we hypothesize that the BexE substrate is similar to the RdmE substrate, aklaviketone. RdmE is an FAD/NADPH-dependent polyketide monooxygenase that is responsible for hydroxylation at the C11 position of aklavinone to form rhodomycinone, which is precursor to daunorubicin and many other type II PKSs natural products(19). Aklaviketone, the substrate of RdmE, is structurally related to the proposed BexE substrate (Figure 5-2 and Figure 5-4). The co-crystal structure of RdmE bound to its substrate aklaviketone and FAD was determined (Figure 5-2 and Figure 5-12)(19). The RdmE co-crystal structure reveals the aklaviketone binding site with FAD in the "out" conformation. The middle domain of RdmE closes down over the substrate and effectively seals the substrate into a cavity. BexE has a large open active site leading directly into the FAD binding pocket. In contrast, the same region of RdmE, which corresponds to the BexE active site entrance, is completely blocked while the substrate is enclosed inside the pocket (Figure 5-12). BexE and RdmE may have different active site pocket entrances; however, the middle domain forms both pockets. This allows us to examine the relationship between substrate binding in the middle domain and the "in" and "out" conformations of FAD more comprehensively.

The RdmE crystal structure suggests that upon substrate binding, the FAD moves from the "in" position to the "out" position. In other related oxygenase structures in the absence of

185
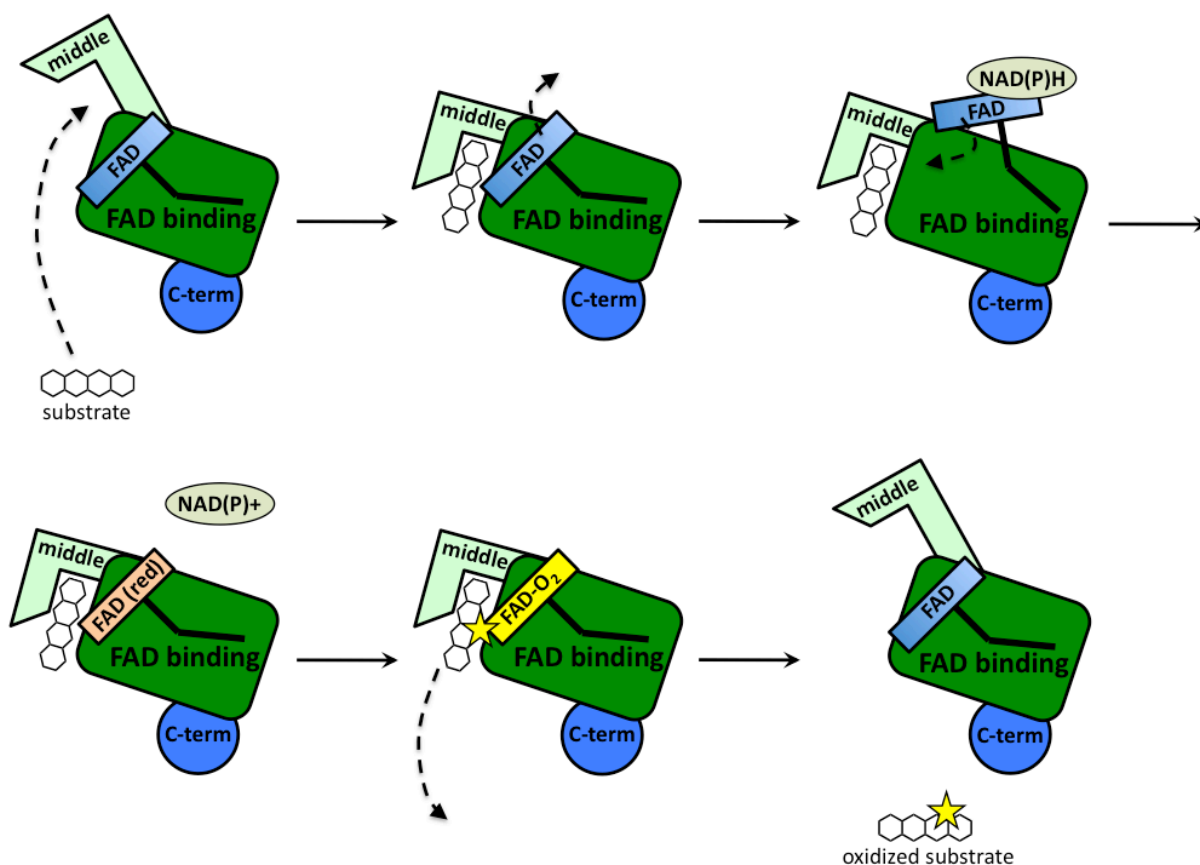
Figure 5-13. The proposed steps during substrate binding and conformational changes in the middle domain during catalysis. Substrate binding induces a conformational change in the middle domain of BexE, which causes FAD to move from the "in" to the "out" position. When FAD is in the "out" position, NAPDH reduction occurs and FAD returns to the "in" position. The reduced FAD reacts with molecular oxygen to form a reactive C4a peroxy-flavin intermediate. The reactive C4a peroxy-flavin intermediate delivers oxygen to the substrate and the product is released.

substrate FAD is in the "in" position(6, 25). This is a reasonable hypothesis, because when FAD is reduced in the "out" position, the reduced FAD reacts with molecular oxygen to generate a peroxy-flavin intermediate. Peroxy-flavin intermediates are unstable, and without a substrate to oxidize, the peroxy-flavin will decay. If substrate binding does not control the position of FAD and peroxy-flavin generation, NADPH is needlessly consumed. A model for substrate binding, peroxy-flavin binding, and conformation changes in the middle domain is proposed in Figure 5-13. In summary, BexE and RdmE may have similar substrates. Based on the RdmE crystal

	BexE
Crystallization	0.1 M MES pH 7.0, 0.2 M ammonium sulfate, 30% PEG 3350
Crystallographic Data	
Wavelength (Å)	1.0000
Space Group	C2
Cell Dimensions (a, b, c) (Å)	150.672, 80.379, 105.079
	$\alpha=\gamma=90°$ $\beta=126.191°$
Resolution (Å)	50.00 - 2.65
No. of observations	218698
No. of unique observations	29488
Completeness %	100.00 (99.8)
I/σ(I)	21.6 (3.9)
R_{merge} %	8.4 (46.9)
Redundancy	7.4
Refinement	
Resolution (Å)	50.00 - 2.65 (2.74 - 2.65)
No. of protein atoms	7162
No. of ligand atoms	121
No. of water atoms	58
R_{free} %	24.56 (35.29)
R_{crys} %	19.62 (29.89)
Geometry	
RMS bonds (Å)	0.005
RMS angles (°)	0.960
Ramachandran Favored (%)	96
Ramachandran Allowed (%)	3.58
Ramachandran Dissallowed (%)	0.42
Average B-factors ($Å^2$)	
Protein	58.20
Water	55.80
Ligands	48.10

Table 5-2. BexE crystallographic data collection and refinement statistics.
*Numbers in parentheses denote the highest resolution shell.

structure bound to aklaviketone, we hypothesize that the BexE substrate will bind in an active site pocket formed by the middle domain. Additionally, upon substrate binding, the FAD will move from the "in" position to the "out" position for reduction and subsequent peroxy-flavin generation.

5.4 Conclusion

The crystal structure of the oxygenase BexE was solved to 2.65 Å. It reveals a large active site pocket adjacent to the FAD cofactor, which is positioned in the "in" conformation that has been observed for other FAD/NADPH oxygenase crystal structures in the absence of substrate. Because BexE is part of a large number of enzymes involved in BE-7585A biosynthesis, the identification of its true substrate is a challenging task. Attempts at developing a BexE functional assay with late-stage tetracyclic type II PKS intermediates eliminated many potential BexE substrates and helped establish the substrate tolerance of BexE. Additionally, contrary to other oxygenases, we found that BexE cannot catalyze a reverse reaction using angucyclic polyketides, which supports the unique activity of BexE. We did observe NADPH-dependent BexE activity when BexE was incubated with the lysate of *Streptomyces galilaeus* (ATCC 31615), which produces aklaviketone. The *Streptomyces galilaeus* (ATCC 31615) lysate was fractionated and a single peak was identified as the BexE substrate. Importantly, BexF, the putative 4th ring cyclase that is hypothesized to provide the BexE substrate during BE-7585A biosynthesis, was found to increase BexE product formation in our assay. To gain insight into the structure of the BexE substrate, we conducted molecular docking studies on possible biosynthetic intermediates using a homology model of BexF and the crystal structure of BexE. Our docking results suggest that the substrate of BexE may be 12-deoxy-aklaviketonic acid. We are actively pursuing structural determination of the BexE substrate and products. When BexE was compared with the related type II PKS oxygenase RdmE, we gained insights into how substrate binding may induce conformation changes in the middle domain and regulate catalysis. Together, these results offer an initial glimpse into the complex oxidation reactions that occur during BE-7585A biosynthesis and lay a framework for engineering oxidations in type II PKSs.

5.5 Materials and Methods

5.5.1 Expression and Purification of BexE

The pET-28b (Novagen) derived DNA plasmid encoding a cleavable N-terminal His-tagged BexE (BexE/pET28b) was provided by Dr. Hung-Wen Liu. BexE/pET28b was transformed into *E. coli* BL21(DE3) competent cells and plated on LB-agar plates containing kanamycin (50 μg/mL). The plates were incubated overnight at 37 °C. A positive transformant was transferred to a 5 mL starter culture of Luria-Bertani (LB) broth containing kanamycin (50 μg/mL). The starter culture was grown overnight at 37 °C with shaking. The culture was used to inoculate two liters of LB with kanamycin (50 μg/mL). The culture was grown at 37 °C until the A_{600} reached 0.4 − 0.6. The cells were cooled to 18 °C and 0.1 mM IPTG was added to induce protein expression. After 12-18 hours of incubation at 18 °C, the cells were harvested by centrifugation at 5000 rpm for 15 minutes. The cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C. The frozen cell pellets were thawed on ice and re-suspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 15% glycerol, 10 mM imidazole). The cell suspension was lysed using sonication (8 x 30 s cycles), and cell debris was removed by centrifugation at 14000 rpm for 45 minutes. The lysate was incubated with 5 mL Ni-IMAC resin (BioRad) at 4 °C for one hour. The resin was poured into a fritted column and the flow-through fraction was collected. The resin was washed with 100 mL of lysis buffer then eluted with lysis buffer plus increasing amounts of imidazole (20-500 mM). The elutions were analyzed using SDS-PAGE, and elutions containing the protein of interest were combined and dialyzed overnight into crystallization buffer (50 mM Tris pH 8.0, 300 mM NaCl, 15% glycerol). The dialyzed protein was concentrated to 5 mg/ml and frozen in liquid nitrogen. The appearance of the elutions was bright yellow due to the presence of the FAD cofactor bound to BexE. The theoretical molecular

weight of BexE (including the His-tag) is 53589 Daltons, which was confirmed using SDS-PAGE.

5.5.2 BexE Crystallization

Frozen aliquots of BexE were thawed on ice and filtered with a 0.22 µM spin filter prior to crystallization. 1.7 µL of BexE (5 mg/ml) was mixed with 1.7 µL of well solution (30 % PEG 3350, 0.2 M ammonium sulfate, 0.1 M MES, pH 4.5) and allowed to equilibrate with 500 µL of well solution using the sitting drop vapor diffusion method. Medium sized clusters containing 3D blades formed over 1 week at 4 °C. The clusters were broken and single blades were used for diffraction experiments.

5.5.3 BexE Data Collection, Model building, and Refinement

Crystals of BexE were flash frozen in liquid nitrogen before data collection. X-ray diffraction data using monochromatic X-rays (1.0 Å) were collected for BexE crystals to a resolution of 2.65 Å at the Advanced Light Source (ALS) on Beamline 8.2.1. The diffraction data was processed using HKL2000(36). BexE crystallized as one homodimer per asymmetric unit in the space group C2. Initial phases were determined using molecular replacement (PHENIX Phaser) with the structure of the aromatic hydoxylase PgaE (PDB ID: 2QA1) as a search model(37). A preliminary model was built (PHENIX AutoBuild) and this model used for iterative rounds of model building (COOT) and refinement (PHENIX refine)(34, 38, 39). After one round of refinement, clear electron density for the FAD cofactor was observed in both BexE monomers. Once all side chains were correctly placed, FAD was inserted into each monomer and included for all subsequent rounds of model building and refinement. Clear electron density was

located near multiple arginines and three sulfates were modeled and refined in these areas. BexE was refined to yield an R_{work} of 19.62% and an R_{free} of 24.56%. Data collection and refinement statistics can be found in Table 5-1. The following residues could not me confidently place in the BexE model due to missing electron density: Ile190-Glu192, and Ala445 (monomer A), M189-Glu192, and Gly399-Arg401 (monomer B).

5.5.4 BexE Molecular Docking

Putative substrates for oxidation were drawn using ChemDraw and converted to PDBs using the NCI SMILES converter server (http://cactus.nci.nih.gov/translate/). The putative substrate PDB files were energy minimized and converted to .mol2 files using Chimera(35). The BexF homology model was generated using HHPred based on the crystal structure of TcmI. The BexE and BexF PDBs were converted to .mol2 files and waters were removed using Chimera. GOLD docking was used to dock the putative substrates in the BexE active site within 15 Å of the FAD C4a carbon(40). For BexF, substrates were docked within 15 Å of the bottom pocket arginine Arg39.

References

1. Metsa-Ketela, M., Palmu, K., Kunnari, T., Ylihonko, K., and Mantsala, P. (2003) Engineering anthracycline biosynthesis toward angucyclines, *Antimicrob Agents Chemother 47*, 1291-1296.
2. Krohn, K., and Rohr, J. (1997) Angucyclines: Total syntheses, new structures, and biosynthetic studies of an emerging new class of antibiotics, *Top Curr Chem 188*, 127-195.
3. Kharel, M. K., Pahari, P., Shepherd, M. D., Tibrewal, N., Nybo, S. E., Shaaban, K. A., and Rohr, J. (2012) Angucyclines: Biosynthesis, mode-of-action, new natural products, and synthesis, *Nat Prod Rep 29*, 264-325.
4. Han, L., Yang, K., Ramalingam, E., Mosher, R. H., and Vining, L. C. (1994) Cloning and characterization of polyketide synthase genes for jadomycin B biosynthesis in Streptomyces venezuelae ISP5230, *Microbiology 140 (Pt 12)*, 3379-3389.
5. Henkel, T., Rohr, J., Beale, J. M., and Schwenen, L. (1990) Landomycins, new angucycline antibiotics from Streptomyces sp. I. Structural studies on landomycins A-D, *J Antibiot (Tokyo) 43*, 492-503.
6. Kallio, P., Liu, Z., Mantsala, P., Niemi, J., and Metsa-Ketela, M. (2008) Sequential action of two flavoenzymes, PgaE and PgaM, in angucycline biosynthesis: chemoenzymatic synthesis of gaudimycin C, *Chem Biol 15*, 157-166.
7. Lombo, F., Brana, A. F., Salas, J. A., and Mendez, C. (2004) Genetic organization of the biosynthetic gene cluster for the antitumor angucycline oviedomycin in Streptomyces antibioticus ATCC 11891, *Chembiochem 5*, 1181-1187.
8. Ostash, B., Korynevska, A., Stoika, R., and Fedorenko, V. (2009) Chemistry and biology of landomycins, an expanding family of polyketide natural products, *Mini Rev Med Chem 9*, 1040-1051.
9. Tsuchida, T., Sawa, R., Takahashi, Y., Iinuma, H., Sawa, T., Naganawa, H., and Takeuchi, T. (1995) Azicemicins A and B, new antimicrobial agents produced by Amycolatopsis. II. Structure determination, *J Antibiot (Tokyo) 48*, 1148-1152.
10. Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork, *Nat Prod Rep 24*, 162-190.
11. Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms, *Curr Opin Chem Biol 7*, 285-295.
12. Di Marco, A., Cassinelli, G., and Arcamone, F. (1981) The discovery of daunorubicin, *Cancer Treat Rep 65 Suppl 4*, 3-8.
13. Sasaki, E., Ogasawara, Y., and Liu, H. W. (2010) A biosynthetic pathway for BE-7585A, a 2-thiosugar-containing angucycline-type natural product, *J Am Chem Soc 132*, 7405-7417.
14. Sasaki, E., and Liu, H. W. (2010) Mechanistic studies of the biosynthesis of 2-thiosugar: evidence for the formation of an enzyme-bound 2-ketohexose intermediate in BexX-catalyzed reaction, *J Am Chem Soc 132*, 15544-15546.
15. Sasaki, E., Zhang, X., Sun, H. G., Lu, M. Y., Liu, T. L., Ou, A., Li, J. Y., Chen, Y. H., Ealick, S. E., and Liu, H. W. (2014) Co-opting sulphur-carrier proteins from primary metabolic pathways for 2-thiosugar biosynthesis, *Nature 510*, 427-431.
16. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J Mol Biol 372*, 774-797.

17. Koskiniemi, H., Metsa-Ketela, M., Dobritzsch, D., Kallio, P., Korhonen, H., Mantsala, P., Schneider, G., and Niemi, J. (2007) Crystal structures of two aromatic hydroxylases involved in the early tailoring steps of angucycline biosynthesis, *J Mol Biol 372*, 633-648.

18. Kallio, P., Patrikainen, P., Belogurov, G. A., Mantsala, P., Yang, K., Niemi, J., and Metsa-Ketela, M. (2013) Tracing the evolution of angucyclinone monooxygenases: structural determinants for C-12b hydroxylation and substrate inhibition in PgaE, *Biochemistry 52*, 4507-4516.

19. Lindqvist, Y., Koskiniemi, H., Jansson, A., Sandalova, T., Schnell, R., Liu, Z., Mantsala, P., Niemi, J., and Schneider, G. (2009) Structural basis for substrate recognition and specificity in aklavinone-11-hydroxylase from rhodomycin biosynthesis, *J Mol Biol 393*, 966-977.

20. Wang, P., Bashiri, G., Gao, X., Sawaya, M. R., and Tang, Y. (2013) Uncovering the enzymes that catalyze the final steps in oxytetracycline biosynthesis, *J Am Chem Soc 135*, 7138-7141.

21. Ryan, K. S., Howard-Jones, A. R., Hamill, M. J., Elliott, S. J., Walsh, C. T., and Drennan, C. L. (2007) Crystallographic trapping in the rebeccamycin biosynthetic enzyme RebC, *Proc Natl Acad Sci U S A 104*, 15311-15316.

22. Bosserman, M. A., Downey, T., Noinaj, N., Buchanan, S. K., and Rohr, J. (2013) Molecular insight into substrate recognition and catalysis of Baeyer-Villiger monooxygenase MtmOIV, the key frame-modifying enzyme in the biosynthesis of anticancer agent mithramycin, *ACS Chem Biol 8*, 2466-2477.

23. Ryan, K. S., Chakraborty, S., Howard-Jones, A. R., Walsh, C. T., Ballou, D. P., and Drennan, C. L. (2008) The FAD cofactor of RebC shifts to an IN conformation upon flavin reduction, *Biochemistry 47*, 13506-13513.

24. Goldman, P. J., Ryan, K. S., Hamill, M. J., Howard-Jones, A. R., Walsh, C. T., Elliott, S. J., and Drennan, C. L. (2012) An unusual role for a mobile flavin in StaC-like indolocarbazole biosynthetic enzymes, *Chem Biol 19*, 855-865.

25. Beam, M. P., Bosserman, M. A., Noinaj, N., Wehenkel, M., and Rohr, J. (2009) Crystal structure of Baeyer-Villiger monooxygenase MtmOIV, the key enzyme of the mithramycin biosynthetic pathway, *Biochemistry 48*, 4476-4487.

26. Martin, J. L. (1995) Thioredoxin--a fold for all reasons, *Structure 3*, 245-250.

27. Torres Pazmino, D. E., Winkler, M., Glieder, A., and Fraaije, M. W. (2010) Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications, *J Biotechnol 146*, 9-24.

28. Brender, J. R., Dertouzos, J., Ballou, D. P., Massey, V., Palfey, B. A., Entsch, B., Steel, D. G., and Gafni, A. (2005) Conformational dynamics of the isoalloxazine in substrate-free p-hydroxybenzoate hydroxylase: single-molecule studies, *J Am Chem Soc 127*, 18171-18178.

29. Ylihonko, K., Hakala, J., Niemi, J., Lundell, J., and Mantsala, P. (1994) Isolation and characterization of aclacinomycin A-non-producing Streptomyces galilaeus (ATCC 31615) mutants, *Microbiology 140 (Pt 6)*, 1359-1365.

30. Thompson, T. B., Katayama, K., Watanabe, K., Hutchinson, C. R., and Rayment, I. (2004) Structural and functional analysis of tetracenomycin F2 cyclase from Streptomyces glaucescens. A type II polyketide cyclase, *J Biol Chem 279*, 37956-37963.

31.	Shen, B., and Hutchinson, C. R. (1993) Tetracenomycin F2 cyclase: intramolecular aldol condensation in the biosynthesis of tetracenomycin C in Streptomyces glaucescens, *Biochemistry 32*, 11149-11154.

32.	Kendrew, S. G., Katayama, K., Deutsch, E., Madduri, K., and Hutchinson, C. R. (1999) DnrD cyclase involved in the biosynthesis of doxorubicin: purification and characterization of the recombinant enzyme, *Biochemistry 38*, 4794-4799.

33.	Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res 33*, W244-W248.

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

35.	Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.

36.	Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol 276*, 307-326.

37.	Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography 40*, 658-674.

38.	Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D 64*, 61-69.

39.	Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D 68*, 352-367.

40.	Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD, *Proteins 52*, 609-623.

Chapter 6

Structural Studies of Protein-Protein Interactions in the *E. coli* Type II Fatty Acid Synthase Using Mechanism-based Crosslinking and X-ray Crystallography

6.1 Abstract

The acyl carrier protein (ACP) plays a central role in both primary and secondary metabolism. In *E. coli*, the type II ACP (AcpP) is responsible for transporting a nascent fatty acyl chain during biosynthesis. In different biosynthetic pathways including fatty acid (FA) biosynthesis, AcpP interacts with at least 12 enzymes while it delivers intermediates into different active sites. The interactions between AcpP and its partner enzymes are transient, which has hampered efforts to gain structural knowledge of how AcpP interacts with its partner enzymes. In order to gain a better understanding of how AcpP interacts with the ketosynthase (KS) FabB and dehydratase (DH) FabA during *E. coli* FA biosynthesis, enzyme-specific mechanism-based crosslinking strategies were used to isolate stable, crosslinked AcpP=FabB and AcpP=FabA complexes. These complexes were purified, crystallized, and their crystal structures solved to 2.4 Å and 1.9 Å, respectively*. The AcpP=FabB structure is the first structure of an ACP/KS complex and reveals important residues for complex formation. This work represents a major breakthrough in understanding type II FA biosynthesis and provides a structural basis for engineering the production of pharmaceuticals and biofuels in *E. coli.*

*The AcpP=FabA crystal structure was published in *Nature* in 2014, and this study was primarily completed by Dr. Chi Nguyen and our collaborators in Dr. Michael Burkart's laboratory. I am a co-author on this publication and aided in crystallizing the AcpP=FabA

195

complex and in solving the crystal structure of the complex. This structure will be briefly reviewed in section 6.3.1 by adapting a portion of the text and multiple images from our published work. However, experimental details are not included in the material and methods of this thesis. For a more detailed discussion of the AcpP=FabA crystal structure, NMR studies, and molecular dynamics simulations, please refer to reference 14.

6.2 Introduction

The type II acyl carrier protein (AcpP) plays a central role in transporting starting materials and intermediates throughout the fatty acid biosynthetic pathway before incorporation into lipogenesis (Figure 6-1A)(1-5). In *E. coli*, AcpP interacts with at least 12 enzymes involved in fatty acid biosynthesis, plus seven other enzymes from disparate biosynthetic pathways (Figure 6-1A)(6-10). AcpP has been demonstrated to sequester growing metabolites in an interior hydrophobic cavity, presumably to protect these intermediates from non-selective reactivity, and selective protein-protein interactions are believed to be a prerequisite for the delivery of an ACP-bound substrate into the appropriate catalytic partner(11). Given the importance of ACP-protein interactions in both metabolic and regulatory processes, such a "switchblade mechanism" is a highly attractive hypothesis, though conclusive evidence remains elusive due to the inherently transient nature of ACP-partner complexes (Figure 6-1C)(12, 13).

In order to gain a better understanding of protein-protein interactions in the *E. coli* type II FAS, we applied mechanism-based crosslinking to generate stable protein complexes for structural and spectroscopic studies. Here we present the crystal structures of the AcpP=FabA and AcpP=FabB complexes, which reveal molecular details about the specificity of protein-protein interactions between these enzymes(14). We hypothesize that knowledge of the differential AcpP interactions between distinct enzymes in type II fatty acid biosynthesis can be used to engineer this pathway to produce altered fatty acid profiles *in vitro* and *in vivo*. By manipulating FAS protein-protein interactions, we aim to control the chain length and saturation level of fatty acids, which will advance the field of biofuel production. Our initial studies on the AcpP=FabA complex will be presented, followed by the analysis of the AcpP=FabB complex, and concluded with a structural comparison between both complexes.
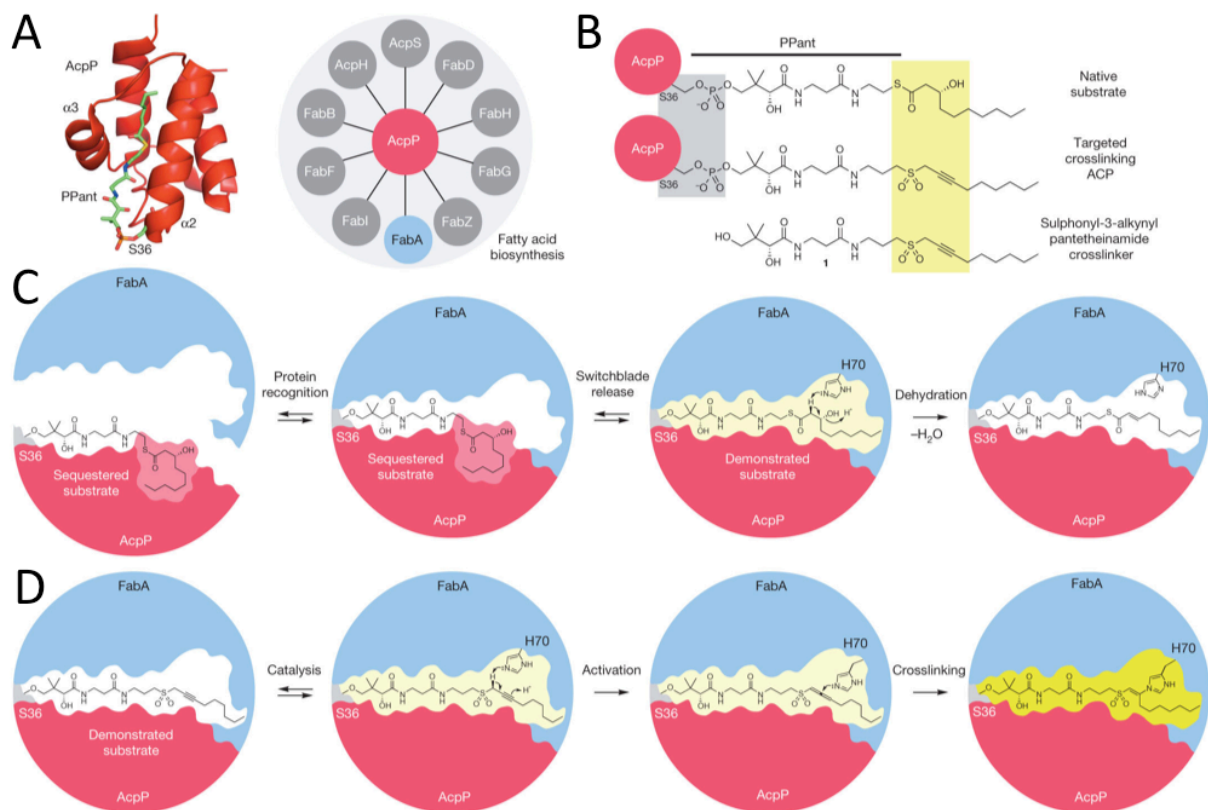
Figure 6-1. *E. coli* AcpP and crosslinking strategy. (A) AcpP is a small, acidic protein comprised of four α-helices that interacts with at least 21 catalytic enzymes, 12 of which belong to the type II FAS (10 shown here). The apolar interior of helix II (α2) and helix III (α3) form a hydrophobic cavity that sequesters the growing metabolite attached to the phosphopantetheine (pPant) prosthetic group. (B) A native substrate of FabA (top) and modified AcpP with targeted sulphonyl-3-alkynyl crosslinking probe (middle), derived from the crosslinking pantetheinamide analogue **1** (bottom). (C) Proposed mechanism of FabA. Protein–protein interactions between AcpP and FabA induce release of the sequestered substrate from AcpP into the active site of FabA, where dehydration is catalyzed. (D) Crosslinking strategy to form the AcpP=FabA complex with mechanism-based crosslinking probe **1**.

6.3 Results and Discussion

6.3.1 AcpP=FabA Structural Determination and Analysis

We recently deployed synthetic probes to study AcpP-protein interactions, including a sulfonyl-3-alkyne based probe (**1**) designed to capture AcpP in a functionally relevant association with FabA (3-hydroxyacyl-ACP dehydratase) with demonstrated specificity between the *E. coli* AcpP and FabA (Figure 6-1B-D)(15-17). Probe **1**, when loaded onto the AcpP in place of the natural pPant, creates a uniformly crosslinked species (AcpP=FabA) that forms reproducible crystals in the tag-free form. No crystals form without **1**, demonstrating the necessity of applying probes such as **1** to capture the ACP in its association with partner enzymes.

The AcpP=FabA crystals were diffracted to 1.9 Å, and the AcpP=FabA crystal structure was solved by molecular replacement using the apo-FabA dimer (PDB ID: 1MKA) and a butyryl-AcpP (PDB ID: 2K94) as search templates. Final refinement revealed the AcpP$_2$=FabA$_2$ complex structure (Figure 6-2A), consistent with protein sizing data in solution. The dimeric FabA forms a "double hotdog" topology, with two anti-parallel "hotdog" helices surrounded by the combined 14-stranded β-sheet (Figure 6-2A)(18-20). The two AcpP monomers adopt the four-helix bundle fold and dock helices II and III (α2 and α3) with the β5-6 loop of FabA (Figure 6-2C). The interface area is small (597 Å^2 and 512 Å^2 for the two interfaces), consistent with the transient nature of AcpP-partner interactions. AcpP has a negatively charged helical cleft between helices II and III, which interacts with a positively-charged arginine-rich patch (the "positive patch,") (Figure 6-2C)(3). AcpP=FabA interactions are mainly electrostatic but also include conserved, hydrophobic residues (Figure 6-2D). The high sequence conservation of
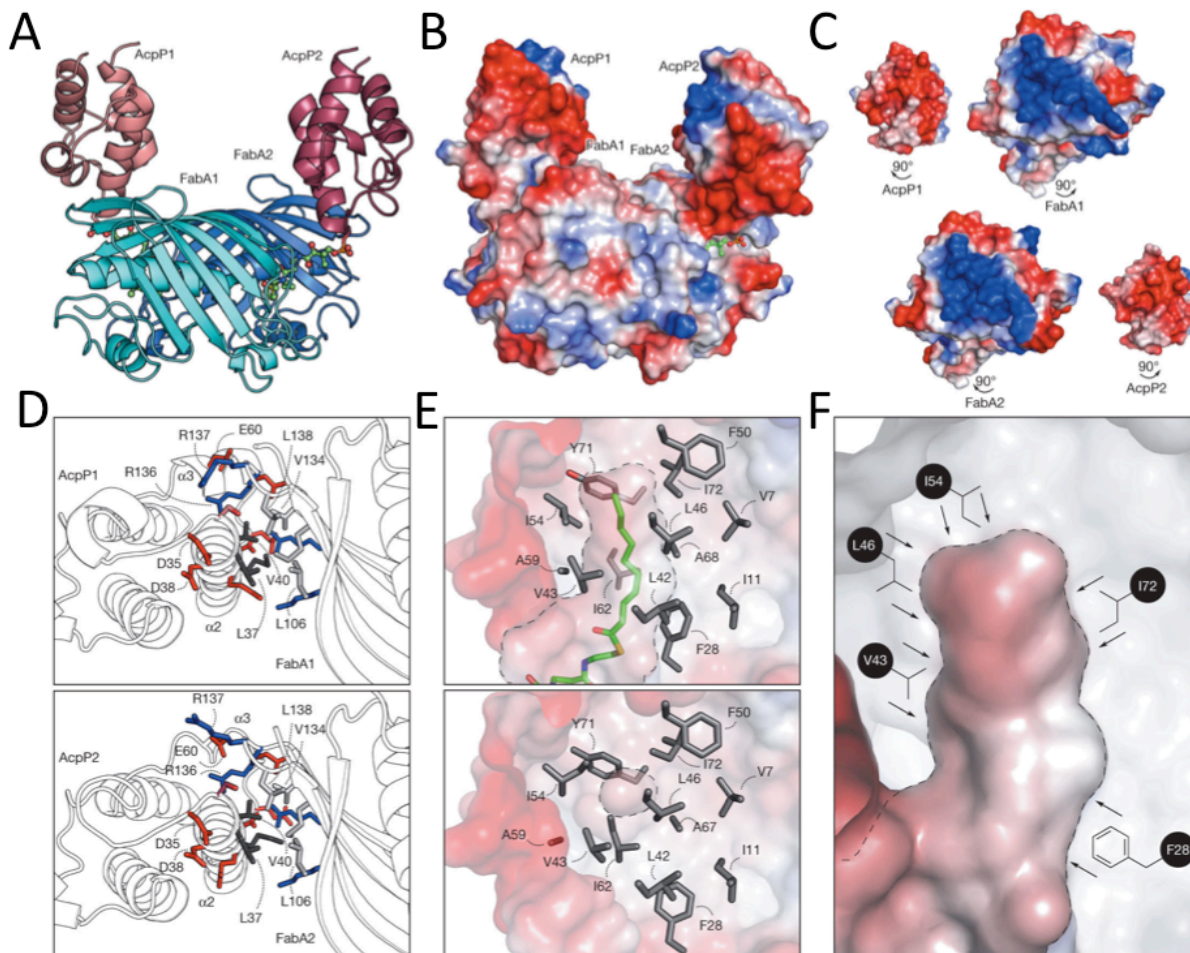
Figure 6-2. The structure of crosslinked AcpP=FabA. (A) X-ray crystal structure of AcpP=FabA at 1.9 Å. (B) The molecular surface mapped with calculated vacuum electrostatic potential of AcpP=FabA. Blue shading indicates electropositive and red shading indicates electronegative protein surfaces. (C) Rotation of (B) by 90° at the interfaces between each AcpP=FabA to visualize electrostatic pairing. (D) Expanded view of both interfaces in AcpP=FabA, indicating salt bridges and hydrophobic interactions between helix II (α2) and helix III (α3) of AcpP and the "positive patch" of FabA. (E) Comparison between hydrophobic cleft of AcpP with sequestered substrate (top, from PDB 2FAE, with long interior hydrophobic cavity outlined with dashed line) and AcpP1 in AcpP=FabA (bottom, reduced interior cavity). (F) The interior cavity of 2FAE labeled with the hydrophobic residues labeled. The contraction of these hydrophobic residues collapses the interior cavity in AcpP=FabA.

residues at the AcpP=FabA interface is consistent with previous reports of ACP-partner complex structures, strongly supporting the conserved nature of the positive patch in ACP partner proteins(7, 21-24).

While only Arg137 of FabA shifts between the two FabAs, many residues of helix III move extensively between AcpP1 and AcpP2 (Figure 6-2D), resulting in different topology near the contact interface between helices II and III (Figure 6-2C). Thus, the first AcpP=FabA interaction represents a snapshot when AcpP completes its docking with FabA, resulting in less movement of AcpP1. The second AcpP=FabA interaction represents a snapshot of AcpP in transition, where extensive movement of helix III is necessary in order for AcpP2 to bind or dissociate from FabA.

The natural FAS substrates contain both the phosphopanthetheine (pPant) and acyl chain moieties (Figure 6-1B). The application of probe **1** helps visualize both pPant and substrate binding. Probe **1** covalently connects the active site Ser36 of AcpP and His70 of FabA, and it binds a highly conserved cavity of FabA. Unlike acyl-AcpP structures that contain a hydrophobic interior pocket to sequester the acyl chain, the AcpP in the AcpP=FabA complex structure contains no interior pocket (Figure 6-2B) and closely resembles apo-AcpP, because five conserved hydrophobic residues between helices II-III move inward and collapse the interior pocket (Figure 6-2F)(13, 25). This drastic change reflects a dynamic AcpP moving from the sequestered-substrate state to the open switchblade state to position the substrate within partner enzyme FabA (Figure 6-1C-D). In summary, the crystal structure of the AcpP=FabA complex was solved and elucidates how AcpP binds and releases a sequestered substrate into the FabA active site. Additionally, residues at the AcpP/FabA binding interface were identified and can now be compared to other AcpP/target enzyme interfaces for use in the rational engineering of protein-protein interaction in the *E. coli* type II FAS. A comparison of the AcpP/FabA interface to the AcpP/FabB interface will be discussed in section 6.3.6.
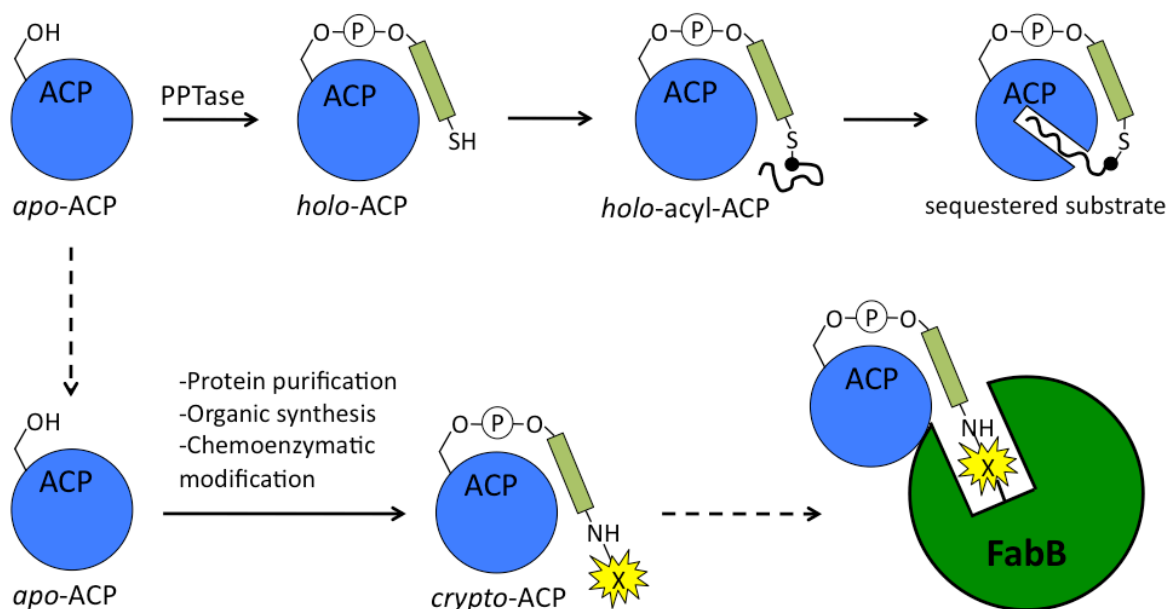
Figure 6-3. The forms of AcpP *in vivo* and FabB crosslinking strategy. AcpP is post-translationally modified with phosphopantetheine (pPant) to form *holo*-AcpP. The growing polyketide is linked via a thioester bond to the thiol of pPant. AcpP has been shown to sequester, or protect, the growing intermediate during biosynthesis. *Apo*-AcpP can be purified and chemoenzymatically loaded with pPant analogues that contain specific chemical groups that crosslink AcpP and target enzymes.

6.3.2 AcpP=FabB Crosslinking

To visualize the protein-protein interactions between AcpP and FabB, we sought to obtain a high-resolution crystal structure of an AcpP=FabB complex. In this work, we applied a similar strategy as in the case of the AcpP=FabA crosslinking and structure determination described above. A previously described chloroacryl pantetheine analogue (**2**) has been loaded onto various acyl carrier proteins and shown to be a mechanism-based crosslinker that targets the active site cysteine of KS enzymes (Figure 6-3 and Figure 6-4)(26-34). The crosslinker **2** was previously used by our lab to investigate ACP-KS interactions in a fungal type I polyketide synthase; however, a high-resolution crystal structure could not be obtained in this study(34). In the current study, crosslinker **2** was loaded onto AcpP from the *E. coli* type II FAS and incubated
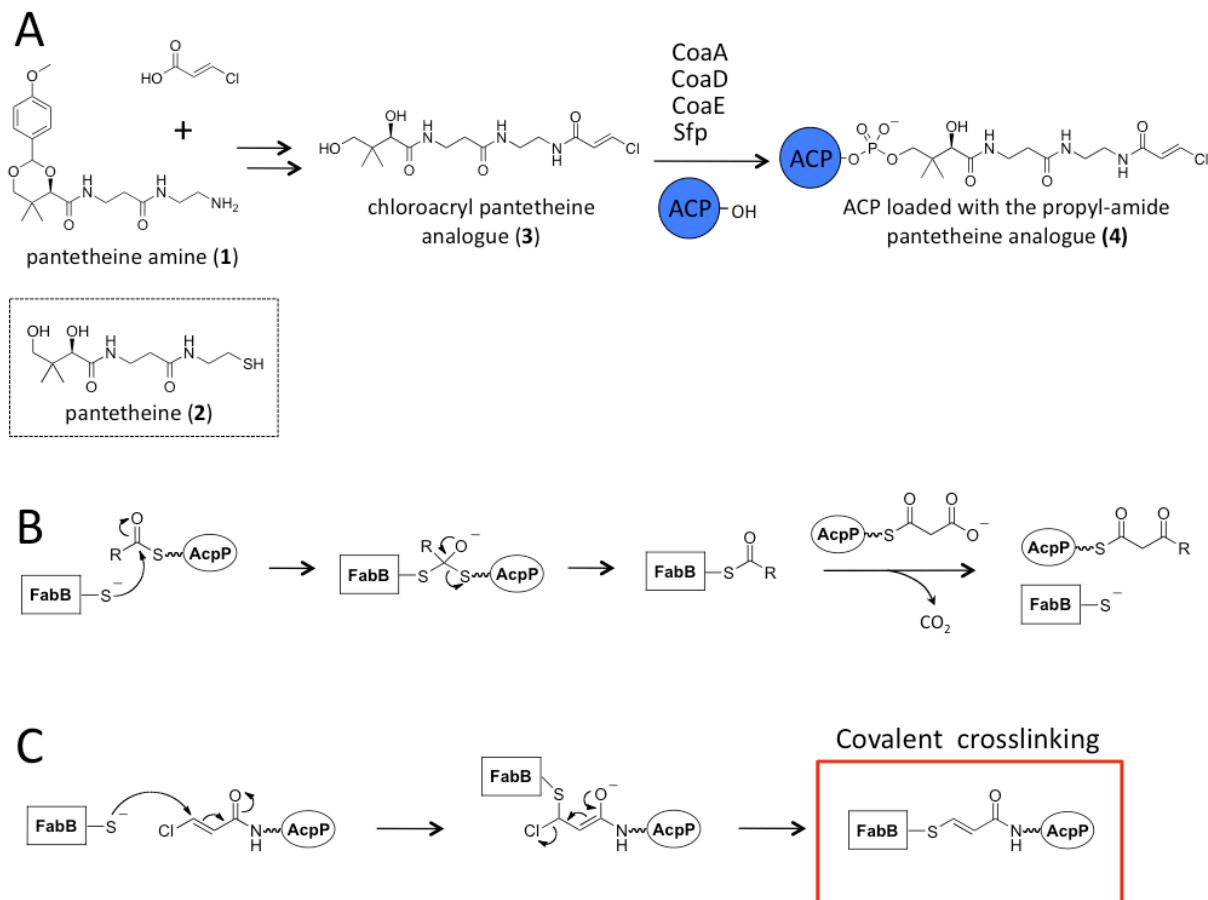
Figure 6-4. An overview of crosslinker synthesis and the FabB crosslinking mechanism. (A) Protected pantetheine amine is coupled with chloroacryl acid and the deprotected product yields the chloroacryl pantetheine analogue (**3**), which can be chemoenzymatically loaded onto AcpP using enzymes from the CoA biosynthesis pathway (CoaA, CoaD, and CoaE), and the phosphopantetheinyl transferase Sfp. (B) The active site cysteine of FabB attacks the thioester of acyl-AcpP to form a tetrahedral intermediate, which collapses to transfer the acyl group to the active site cysteine of FabB. The FabB acyl-enzyme intermediate is then elongated by decarboxylative addition using malonyl-AcpP as a substrate. (C) The active site cysteine of FabB is proposed to attack the alkene of the chloroacryl-based crosslinker resulting in the formation of an enolate. The oxyanion of the enolate collapses to cause double bond migration and expels a chloride anion, which results in a stable crosslinked species.

with FabB to generate a stable $AcpP_2=FabB_2$ complex that readily crystallized (Figure 6-5). A high-resolution crystal structure of the AcpP=FabB complex was solved, revealing important residues for protein-protein interactions between AcpP and FabB (Figure 6-6).

6.3.3 AcpP=FabB Overall Structure

The AcpP=FabB complex is composed of two AcpP molecules and two FabB molecules (Figure 6-6). The AcpPs (AcpP1 and AcpP2) are both covalently bound to the active sites of the two FabB monomers (FabB1 and FabB2) (Figure 6-7E). The structures of FabB1 and FabB2 in the AcpP=FabB complex are very similar to previously determined FabB crystal structures(35-40). FabB1 and FabB2 retain the same thiolase fold in the AcpP=FabB complex when apo-FabB (PDB ID: 2VB7) is overlaid with the AcpP=FabB complex. The structures are nearly identical with an RMSD of 0.245 Å. In the AcpP=FabB complex, both AcpPs are highly dynamic, displaying elevated B-factors and many side chains with poorly defined electron density (Figure 6-6B-C and Figure 6-7A-B). However, helix III of AcpP1 is slightly more disordered, and FabB1 has elevated B-factors relative to FabB2. Therefore, the interface between AcpP1 and FabB1 (interface 1) is termed the "destabilized interface" and the interface between AcpP2 and FabB2 is termed the "stabilized" interface. AcpP1 interacts with both FabB1 and FabB2 with interaction surface areas of 139.8 Å^2 and 451.1 Å^2, respectively. AcpP2 also interacts with both FabB1 and FabB2 with interaction surface areas of 445.8 Å^2 and 149.5 Å^2, respectively. Therefore the total interaction surface areas are very similar with values of 590.9 Å^2 for AcpP1 and 595.3 Å^2 for AcpP2. The differences between these two interfaces will be discussed in detail in the following section.

6.3.4 Comparison of the Crosslinked Active Sites of FabB1 and FabB2 in the AcpP=FabB Crystal Structure

The monomers FabB1 and FabB2 in the AcpP=FabB complex overlay well with an RMSD of 0.132 Å. The active site of FabB consists of Cys163, His297, and His333. When
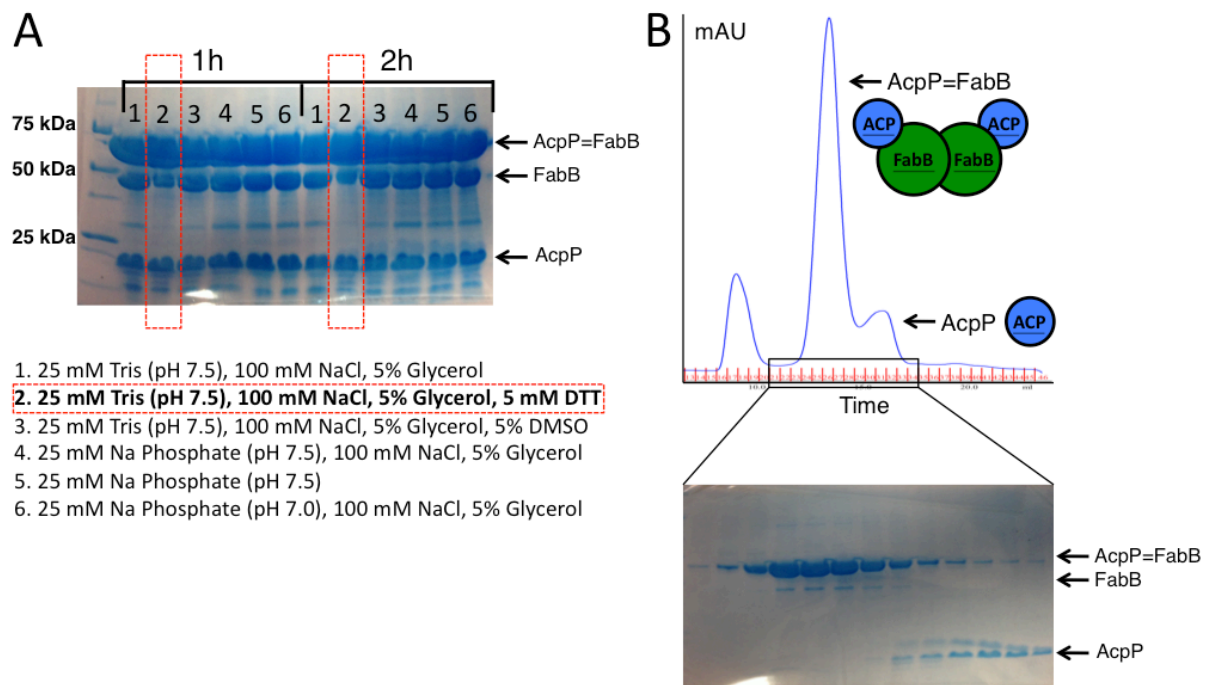
Figure 6-5. AcpP=FabB crosslinking optimization and purification. (A) SDS-PAGE gel analysis of a time course crosslinking experiment with different buffer conditions and additives such as DMSO and DTT. Condition 2 gave the highest amount of crosslinking therefore this condition was further optimized and scaled up to produce the AcpP=FabB complex. (B) Gel filtration chromatogram of a scaled up AcpP=FabB crosslinking reaction showing clear separation of excess AcpP and the AcpP=FabB complex (top). SDS-PAGE analysis showing ~95 % pure AcpP=FabB complex is obtained after gel filtration.

FabB1/AcpP1 and FabB2/AcpP2 are overlaid, the active site residues overlay with little deviation. The crosslinker is bound in the same orientation with respect to Cys163, His297, and His333. Additionally, Thr300 and Thr302 make hydrogen bonds to an amide carbonyl of the crosslinker (Figure 6-7E). However, moving towards the AcpPs, the crosslinkers adopt different conformations. Significantly, the dimethyl moiety from AcpP1 is shifted upwards moving out of the hydrophobic pocket formed by Pro272, Met204, Val240, and Phe392 of FabB1, and Leu37 in AcpP1. Although this slight difference is present in our model, the electron density in this area can accommodate multiple conformations of the crosslinker. Therefore, conclusions drawn from analyzing the conformations of the crosslinker must be corroborated by other methods. However,
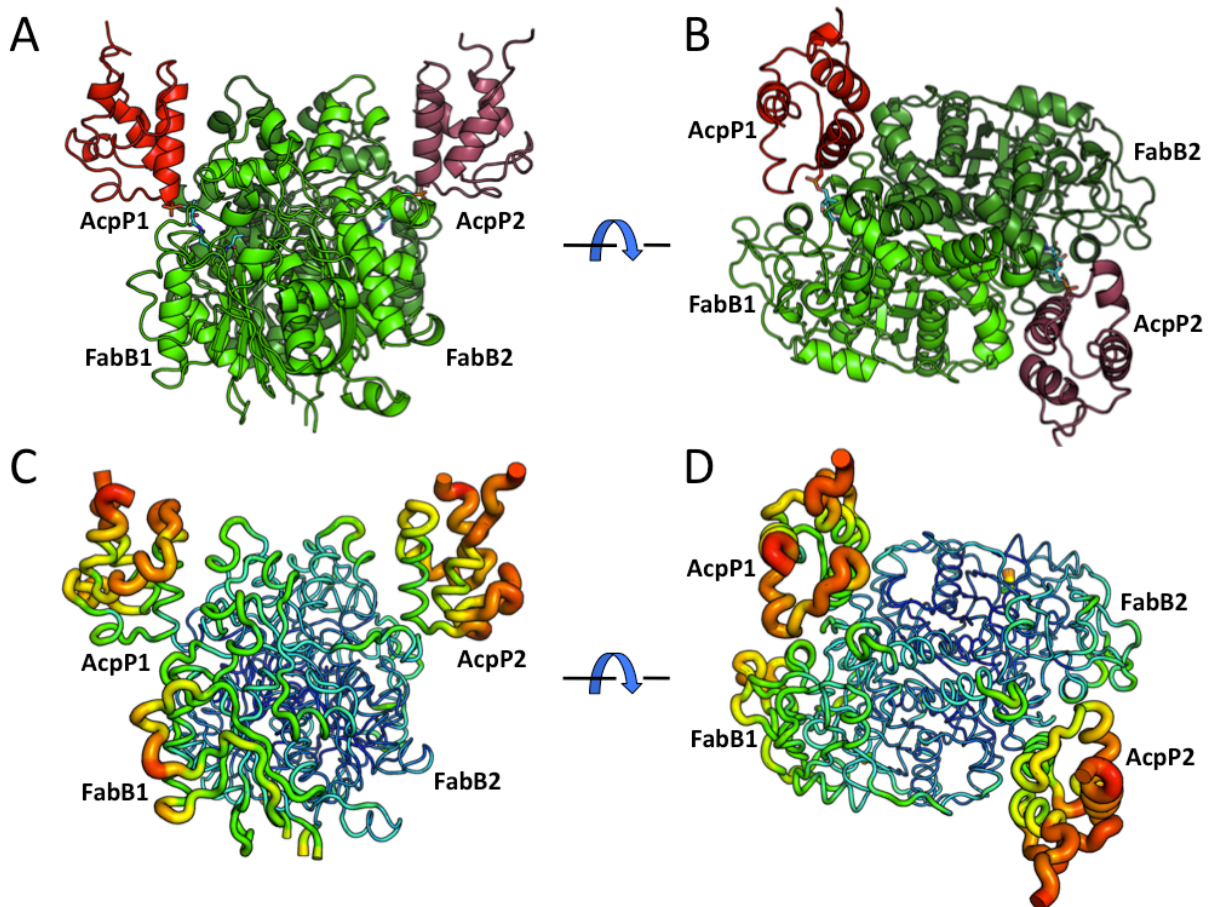
Figure 6-6. The overall structure of the AcpP=FabB complex in cartoon (A-B) and B-factor putty representation (C-D).
*Jake Milligan assisted in crystallization optimization, data collection, and refinement of the AcpP=FabB complex.

the electron density for the middle section of the crosslinker is weaker in AcpP1 than in AcpP2 (Figure 6-7C-D). This suggests that there is more disorder in the AcpP1 crosslinker and that the two AcpPs may be in different interaction states with the FabB dimer.

6.3.5 Comparison of the AcpP1/FabB1 and AcpP2/FabB2 Interfaces

The two AcpPs (AcpP1 and AcpP2) are not identical in the AcpP=FabB complex crystal structure. However, the majority of protein-protein interactions are between negatively charged residues on AcpP1 and positively charged residues on FabB2 (Figure 6-8A-B and Table 6-1).
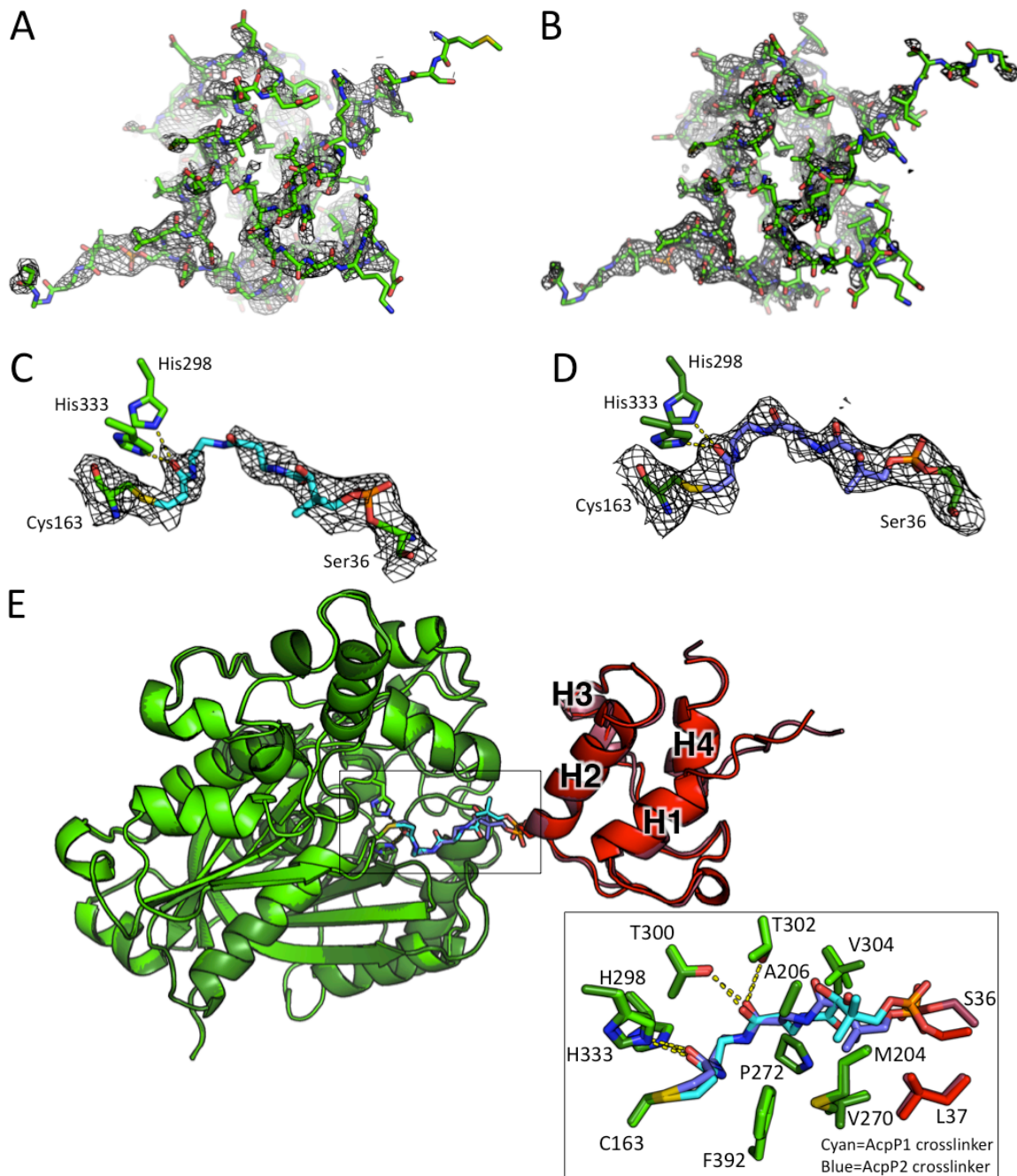
Figure 6-7. The SA-omit maps of AcpP1, AcpP2, and the crosslinkers in each FabB active site. (A) The SA-omit map of AcpP1 showing clear electron density for the protein backbone with missing side chain density for some residues. (B) The SA-omit map of AcpP2 showing clear electron density for the protein backbone with missing side chain density for some residues. (C) The SA-omit map of the crosslinker attached to AcpP1 extending into the FabB active site. (D) The SA-omit map of the crosslinker attached to AcpP2 extending into the FabB active site. (E) An overlay of AcpP1/FabB1 with AcpP2/FabB2 (top left) and an overlay of the crosslinkers in both FabB1 and FabB2 actives sites showing molecular interactions with specific side chains.
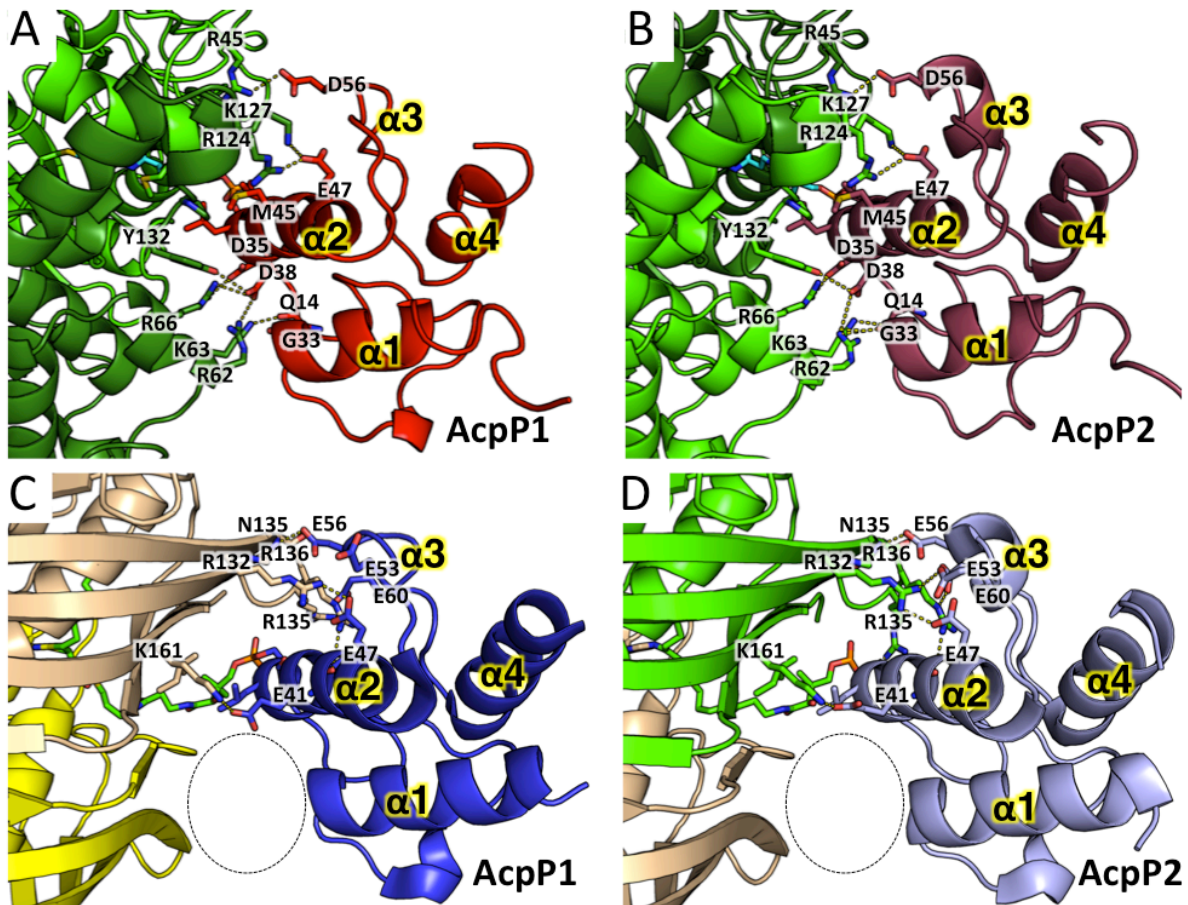
Figure 6-8. The AcpP interactions in the AcpP=FabA and AcpP=FabB complexes.
(A) Interactions between AcpP1 and FabA1 in the AcpP=FabB complex crystal structure.
(B) Interactions between AcpP2 and FabA2 in the AcpP=FabB complex crystal structure.
(C) Interactions between AcpP1 and FabA1 in the AcpP=FabA complex crystal structure.
(D) Interactions between AcpP1 and FabA1 in the AcpP=FabA complex crystal structure.

These interactions are the same for AcpP2 and FabB1. AcpP primarily interacts with the surface

of the FabB dimer via its helix II and helix III. However, Lys63 and Arg62 form hydrogen bonds

with two main chain carbonyl groups of helix I of AcpP. The interactions between the two

AcpPs with the FabB dimer are very similar; however, helix III of AcpP1 is disordered, whereas

helix 3 of AcpP2 is ordered (Figure 6-8A-B). Helix III has been hypothesized to be involved in

recognition of AcpP interaction partners and helping to deliver the AcpP-sequestered substrate

into the active site of the target enzyme(3, 14). Therefore, the presence of a disordered helix III

in AcpP1 and ordered helix III in AcpP2 suggests that AcpP1 represents a more stable binding mode than AcpP2. This phenomenon is also observed in the AcpP=FabA complex crystal structure. Interestingly, no residues from either FabB monomer interact with the negatively charged phosphate on the phosphopantetheine moiety. In summary, the majority of interactions between AcpP1 and AcpP2 with the FabB dimer are very similar; however, the conformation of helix III is different between the two AcpPs. Therefore, helix III may play a role in enzyme recognition and in transfering a growing intermediate from AcpP to the FabB active site.

6.3.6 Comparison the AcpP Residues at the AcpP/FabA and AcpP/FabB Interfaces

The residues at the interface between AcpP1 and AcpP2 in the AcpP=FabB crystal structure are similar. However, when comparing the AcpP interacting residues between the AcpPs in the AcpP=FabA and AcpP=FabB structures, clear differences become apparent. In the AcpP=FabB structure, helix I of both AcpPs interacts with FabB. This interaction is completely absent in both AcpPs in the AcpP=FabA crystal structure (Figure 6-8). There are a common set of residues that AcpP uses to interact with both FabA and FabB: Asp56, Glu47, and Glu41. However, residues Asp35, Asp38, Gln14, and Gly33 are unique for AcpP binding to FabB, whereas Glu60, Thr39, and Glu53 are unique for AcpP binding to FabB (Table 6-1). Additionally, Arg137 in FabA1 of the AcpP=FabA structure contacts the phosphate moiety of the phosphopantetheine crosslinker. No residues in the AcpP=FabB complex interact with the corresponding phosphate. In summary, the AcpP has a subset of residues that are used for interacting with both FabA and FabB, which can be visualized in the AcpP=FabA and AcpP=FabB complexes. However, the AcpP also has distinct interacting residues for each target protein. The interactions between AcpP and FabA are different that the interactions between
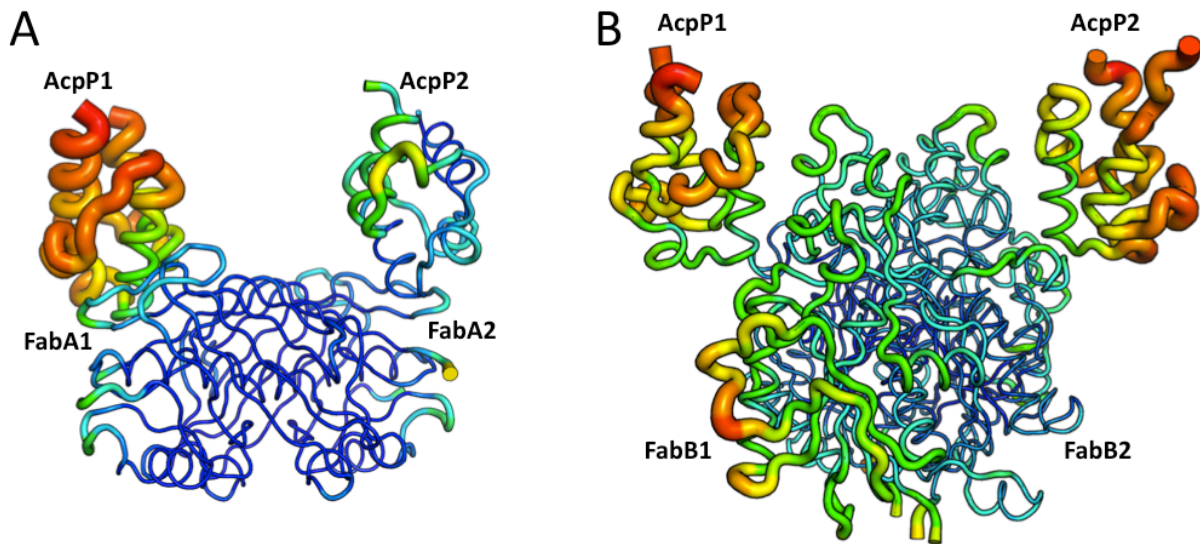
Figure 6-9. Side by side comparison of B-factor putty representations of the AcpP=FabA and AcpP=FabB crystal structures. (A) AcpP=FabA B-factory putty representation. (B) AcpP=FabB B-factory putty representation.

AcpP and FabB. Therefore, rational mutagenesis can be applied to alter AcpP interactions favoring or disfavoring specific interaction partners.

6.3.7 Comparison and Analysis of B-factors in AcpP=FabA and AcpP=FabB Crystal Structures

In the AcpP=FabA crystal structure, AcpP1 has higher B-factors than those of AcpP2 (Figure 6-9A). This suggests that AcpP1 and AcpP2 are dynamically different in the AcpP=FabA crystal structure. We hypothesize that in the AcpP=FabA crystal structure, we have caught a snapshot of two different binding states of AcpP. The ordered AcpP2 is in a stable binding mode, while AcpP1 is disordered and may represent an associating AcpP or dissociating AcpP binding mode. In the AcpP=FabA crystal structure, two AcpPs are covalently tethered to the two FabA active sites. This is possibly not a natural state of the FabA dimer *in vivo;* therefore, one active site may be able to transmit information to the other active site and

AcpP1	FabB1	FabB2	Distance (Å)
D56	R45		3.1
E47		K127	3.1
E47		R124	2.4
D35		R66	2.9
D38		R66	3.3
D38		R62	2.6
D38		K63	3.6
Q14 (C=O)		R62	3.3
G33		K63	2.7

AcpP2	FabB1	FabB2	Distance (Å)
D56		R45	3.0
E47	K127		3.1
E47	R124		3.2
D35	R66		2.9
D38	R66		3.2
D38	R62		4.1
D38	K63		3.0
Q14 (C=O)	R62		3.7
G33 (C=O)	K63		3.3

AcpP1	FabA1	FabB2	Distance (Å)
D56		N135 (NH)	2.9
E60		R136	2.8
T39 (C=O)		R136	3.3
E41		K161	3.7
E47		R132	2.8

AcpP2	FabA1	FabB2	Distance (Å)
D56	N135 (NH)		2.9
E60	R136		2.9
T39 (C=O)	R136		3.3
E41	K161		2.7
E47	R132		2.8
E53	R132		3.5
pPant (P-O)	R137		3.4

Table 6-1. Residues that interact at the interfaces of the AcpP=FabA (bottom) and AcpP=FabB (top) complexes. "C=O" corresponds to protein backbone carbonyls, and "NH" corresponds to protein backbone amide protons. "P-O" corresponds to the phosphate oxygen of the AcpP-linked pPant analogue.

influence catalytic turnover by forcing an AcpP-linked product out of the opposing active site once catalysis is complete. This scenario matches well with the observation that AcpP1 is in a disordered state while AcpP2 is in an ordered state.

In the AcpP=FabB crystal structure, both AcpPs have elevated B-factors. However, FabB1 has elevated B-factors relative to FabB2. Similar to the AcpP=FabA complex, the AcpP=FabB contains two AcpPs that are covalently linked to the FabB active sites, and this is most likely not a natural state *in vivo*. The two AcpPs in the AcpP=FabB crystal structure have almost twice as many molecular interactions when compared to the AcpPs in the AcpP=FabA crystal structure (Table 6-1). Therefore, the interactions between the AcpPs and FabB may be stronger than the AcpPs and FabA. Similar to our hypothesis for the AcpP=FabA binding

	AcpP=FabB
A. Crystallization	0.1 M sodium acetate pH 5.6, 0.2 M ammonium acetate, 18% PEG 4000
B. Crystallographic Data	
Wavelength (Å)	0.9999
Space Group	$P2_1$
Cell Dimensions (a, b, c) (Å)	59.33, 103.92, 83.40
	$\alpha=\gamma=90°$, $\beta=110.61$
Resolution (Å)	50.00-2.40
No. of observations	134614
No. of unique observations	37095
Completeness % (last shell)	99.95 (99.97)
I/σ(I) (last shell)	7.02 (1.64)
R_{merge} %	9.8 (43.0)
Redundancy	3.6
C. Refinement	
Resolution (Å)	50.00 - 2.40
No. of protein atoms	7131
No. of ligand atoms	50
No. of water atoms	109
R_{free} %	21.70 (32.88)
R_{crys} %	18.61 (29.65)
D. Geometry	
RMS bonds (Å)	0.005
RMS angles (°)	1.05
Ramachandran Favored (%)	93
Ramachandran Allowed (%)	4.9
Ramachandran Dissallowed (%)	2.1
Average B-factors ($Å^2$)	
Protein	62.00
Water	54.50
Ligands	75.60

Table 6-2. AcpP=FabB X-ray data collection and refinement statistics. *Values in parentheses represent values from the highest resolution shell.

interactions above, we hypothesize that in the AcpP=FabB crystal structure, AcpP1, which has a disordered helix III and weaker electron density for the crosslinker **2**, represents an unstable binding mode. This mode may represent an associating or dissociating AcpP, while AcpP2 represents a stable binding mode. Additionally, the elevated B-factors in FabB1 may be the result

of a long-range interaction that is relayed from FabB2, because the occupation of the FabB2 active site with the covalently tethered AcpP2.

6.3.8 Proposed "Switch Blade" Mechanism for AcpP-Mediated Substrate Delivery and Catalysis by FabA and FabB

The molecular details of how AcpP interacts with and delivers substrates to the correct partner enzymes at the right times during FA biosynthesis are unknown. AcpP is capable of carrying a variety of substrates that differ in chain length and levels of saturation, and it guides these intermediates to specific active sites of different enzymes. Our working hypothesis is termed the "switch blade" mechanism, where AcpP carries the growing substrate inside its internal hydrophobic cavity and releases the substrate into a target enzyme active site upon AcpP binding. This process may be influenced by the substrate itself. Specifically, we hypothesize that the sequestered substrate relays structural information to surface residues of AcpP, and this structural relay modulates the affinity of AcpP for different target enzymes. This mechanism allows AcpP to regulate the timing of catalysis based on the substrate inside its internal cavity. We predict that helix II and helix III of AcpP are the most important structural elements for this process. However, other areas of AcpP, such as helix I, are also involved in protein-protein interactions. Based on these results, we are starting to develop a structural picture of how AcpP interacts with different enzymes in the type II FAS pathway.

6.4 Conclusion

We have solved the crystal structures of the AcpP=FabA and AcpP=FabB complexes from the type II FA biosynthesis pathway in *E. coli*. This work proves that mechanism-based

crosslinking can be used as an effective tool to crystallize transient protein complexes and gain insight into the molecular details of how AcpP interacts with different partner enzymes. The AcpP=FabA structural studies reveal two dynamically different AcpPs representing two different binding modes. The two observed binding modes reveal insights into the involvement of helix III in the proposed "switch blade" mechanism of substrate delivery during dehydration in fatty acid biosynthesis. The AcpP=FabB crystal structure reveals, for the first time, a snapshot of an ACP in complex with a KS and allows for critical comparison between two AcpP/target enzyme complexes from the type II FA biosynthesis pathway. In the AcpP=FabB structure, the AcpPs have more molecular interactions with the target enzyme FabB than the AcpPs in the AcpP=FabA crystal structure. The differential interaction between AcpP and these target enzymes provides a snapshot of how AcpP mediates specific interactions with different targets. We are currently using this knowledge to alter residues on the AcpP surface with the aim of modulating the affinity of AcpP for target enzymes.

6.5 Materials and Methods

*The description of materials and methods corresponding to the AcpP=FabA crystal structure, NMR studies, and molecular dynamics simulations can be found in reference 14.

6.5.1 FabB Protein Expression and Purification

A pET-28b (+) (Novagen) derived DNA plasmid encoding a N-terminal cleavable His-tagged FabB (FabB/pET-28b) was transformed into *E. coli* BL21(DE3) competent cells and plated on LB-agar plates containing kanamycin (50 μg/mL). The plates were incubated overnight at 37° C. Positive transformants were transferred to a 5 mL starter culture of Luria-Bertani (LB)

broth containing kanamycin (50 µg/mL) and grown overnight at 37° C with shaking. The 5 ml starter culture was then used to inoculate two liters of LB with kanamycin (50 µg/mL) and grown at 37 °C until the A_{600} reached 0.4 – 0.6. The cells were then cooled to 18 °C, and 0.1 mM IPTG was added to induce protein expression. After 12-18 hours of incubation at 18 °C, the cells were harvested by centrifugation at 5000 rpm for 15 minutes. The cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C. The frozen cell pellets were thawed on ice and re-suspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10% glycerol, and 10 mM imidazole). The cell suspension was lysed using sonication (8 x 30 s cycles), and the cell debris was removed by centrifugation at 14000 rpm for 45 minutes. The supernatant was incubated with 5 mL Ni-IMAC resin (BioRad) at 4 °C for one hour. The resin was poured into a fritted column and the flow through fraction was collected. The resin was washed with 100 mL of lysis buffer then eluted with lysis buffer plus increasing amounts of imidazole (20 - 500 mM). The elutions were analyzed using SDS-PAGE, and elutions containing the protein of interest were combined and dialyzed over night into storage buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10% glycerol, and 2 mM DTT). The dialyzed protein was concentrated to 10 mg/ml by centrifugal filtration using a 30 kDa MWCO concentrator (Millipore) then flash frozen in liquid nitrogen and stored at -80 °C for further use. The theoretical molecular weight of N-terminal His-tagged FabB is 44835 Daltons, which was confirmed using MALDI-TOF mass spectrometry (data not shown).

6.5.2 AcpP Protein Expression and Purification

A pET-28b (Novagen) derived DNA plasmid encoding N-terminal His-tagged AcpP (AcpP/pET-28b) was transformed into *E. coli* BL21(DE3) competent cells and plated on LB-agar plates containing kanamycin (50 µg/mL). The expression and purification of AcpP was

accomplished as described above for FabB with a few minor differences. To obtain apo-AcpP, cells were grown at 37 °C until the A_{600} reached 1 and then induced with 0.1 mM IPTG and grown for an additional 4 hours at 37 °C. Cells were harvested, and AcpP was purified as described above for FabB with minor differences. After Ni-IMAC purification, the elutions were analyzed using SDS-PAGE. The elutions containing the protein of interest were pooled and dialyzed overnight into storage buffer (50 mM Tris pH 8.0, 100 mM NaCl, 10% glycerol). 1 unit of bovine thrombin (Sigma) per mg of AcpP was added to the combined protein elutions prior to dialysis to cleave the N-terminal His-tag. Following the dialysis, AcpP was incubated with Ni-IMAC resin and passed over a fritted column to remove any uncleaved AcpP, and the flow through was analyzed by MALDI-TOF MS to confirm cleavage of the His-tag (data not shown). The flowthrough was directly applied to a Hi Trap Q column (GE Healthcare), and AcpP was eluted with 25 mM Tris pH 7.5, 300 mM NaCl, and 5% glycerol. AcpP was concentrated to 10 mg/mL, flash frozen in liquid nitrogen, and stored at -80 °C.

6.5.3 AcpP Crosslinker Loading and Purification of Chloroacryl-AcpP

Purified apo-AcpP (400 µM) was incubated with CoaA (5 µM), CoaE (5 µM), CoaD (5 µM), Sfp (5 µM), MgCl₂ (5 mM), ATP (8 mM), and the chloroacryl pantetheine analogue **2** (800 µM) in 50 mM sodium phosphate (pH 7.5) for 2 hours. Loading of the pantetheine analogue was confirmed using MALDI-TOF MS (data not shown) to generate the chloroacryl-AcpP (**3**). The final reaction volume was 5 mL. Chloroacryl-AcpP was further purified using a Hi Trap Q column. The AcpP loading reaction was directly applied to the Hi Trap Q column, and chloroacryl-AcpP was eluted with 25 mM Tris pH 7.5, 300 mM NaCl, and 5 % glycerol.

Chloroacryl-AcpP was concentrated to 0.8 mM, flash frozen in liquid nitrogen, and stored at -80 °C.

6.5.4 AcpP=FabB Complex Generation and Purification

The chloroacryl-AcpP (400 μM) was incubated with FabB (100 μM) for 16 hours at 37 °C for the generation of FabB=AcpP complex. Reaction completion was verified by SDS-PAGE. The AcpP=FabB complex was purified and buffer exchanged into crystallization buffer (25 mM Tris, pH 7.5) using a Superdex 200 gel filtration column (GE Healthcare). Fractions were analyzed by SDS-PAGE, and fractions containing the pure FabB=AcpP complex were combined, concentrated using a 30 kDa MWCO concentrator (Millipore) to 12 mg/ml, then flash frozen in liquid nitrogen, and stored at -80 °C for further use.

6.5.5 AcpP=FabB Complex Crystallization

Frozen aliquots of the AcpP=FabB complex were thawed on ice and filtered with a 0.22 μM spin filter prior to crystallization. 1 μL of AcpP=FabB (12 mg/ml) was mixed with 1 μL of well solution (18 % PEG 4000, .2 M ammonium acetate, 0.1 M sodium acetate, pH 5.6) and allowed to equilibrate with 500 μL of well solution using the sitting drop vapor diffusion method. Large, rod-shaped crystals formed over 1-2 days at room temperature.

6.5.6 X-ray Data Collection, Processing, and Refinement

Crystals of AcpP=FabB were flash frozen directly in liquid nitrogen before data collection. X-ray diffraction data using monochromatic X-rays (0.9999 Å) was collected for an AcpP=FabB crystal to a resolution of 2.4 Å at the Advanced Light Source (ALS) at beamline

8.2.1. The data was indexed, integrated, and scaled using HKL2000(41). AcpP=FabB crystallized as one dimer per asymmetric unit in the space group P2$_1$, and the initial phases were determined using molecular replacement (PHENIX Phaser-MR) using the structure of apo-FabB (PDB ID: 2VB7) as a search model(42). A preliminary model was built (PHENIX AutoBuild), which was then used for iterative rounds of manual model building (COOT) and refinement (PHENIX Refine)(43-45). Well-defined electron density of the crosslinker was observed in the active site tunnel of each monomer that was connected to Cys163, corresponding to the covalent linkage of the AcpP-linked chloroacryl moiety. Structural parameters and restraints for the chloroacryl ligand were generated using PHENIX Elbow(46). The crosslinker was manually built in the electron density in the active site tunnels of both monomers. A parameter file was used to link the terminal alkene of the phosphopantetheine crosslinker to the FabB active site Cys163, and the terminal phosphate was linked to Ser36 of AcpP

Well-defined electron density for helix II of AcpP was observed near the entrance to the active site in both FabB monomers. Attempts at ensemble molecular replacement using models of FabB and AcpP failed. Therefore, the AcpP from the AcpP=FabA crystal structure (PDB ID: 4KEH) was extracted and manually placed as a rigid body in using the electron density for helix II as an initial starting point using COOT. After AcpP was placed at the active site entrance of both FabB monomers, iterative cycles of modeling and refinement were carried out to yield a final model with R$_{free}$ and R$_{work}$ values of 21.7 % and 18.6 %, respectively. Data collection and refinement statistics can be found in Table 6-2.

References

1.   Chan, D. I., and Vogel, H. J. (2010) Current understanding of fatty acid biosynthesis and the acyl carrier protein, *Biochem J 430*, 1-19.
2.   Rock, C. O., and Cronan, J. E., Jr. (1981) Acyl carrier protein from Escherichia coli, *Methods Enzymol 71 Pt C*, 341-351.
3.   Crosby, J., and Crump, M. P. (2012) The structural role of the carrier protein--active controller or passive carrier, *Nat Prod Rep 29*, 1111-1137.
4.   Magnuson, K., Jackowski, S., Rock, C. O., and Cronan, J. E., Jr. (1993) Regulation of fatty acid biosynthesis in Escherichia coli, *Microbiol Rev 57*, 522-542.
5.   Joshi, A. K., Witkowski, A., Berman, H. A., Zhang, L., and Smith, S. (2005) Effect of modification of the length and flexibility of the acyl carrier protein-thioesterase interdomain linker on functionality of the animal fatty acid synthase, *Biochemistry 44*, 4100-4107.
6.   Issartel, J. P., Koronakis, V., and Hughes, C. (1991) Activation of Escherichia coli prohaemolysin to the mature toxin by acyl carrier protein-dependent fatty acylation, *Nature 351*, 759-761.
7.   Agarwal, V., Lin, S., Lukk, T., Nair, S. K., and Cronan, J. E. (2012) Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis, *Proc Natl Acad Sci U S A 109*, 17406-17411.
8.   Anderson, M. S., Bulawa, C. E., and Raetz, C. R. (1985) The biosynthesis of gram-negative endotoxin. Formation of lipid A precursors from UDP-GlcNAc in extracts of Escherichia coli, *J Biol Chem 260*, 15536-15541.
9.   Jordan, S. W., and Cronan, J. E., Jr. (1997) A new metabolic link. The acyl carrier protein of lipid synthesis donates lipoic acid to the pyruvate dehydrogenase complex in Escherichia coli and mitochondria, *J Biol Chem 272*, 17903-17906.
10.   Lu, Y. J., Zhang, Y. M., Grimes, K. D., Qi, J., Lee, R. E., and Rock, C. O. (2006) Acyl-phosphates initiate membrane phospholipid synthesis in Gram-positive pathogens, *Mol Cell 23*, 765-772.
11.   Roujeinikova, A., Baldock, C., Simon, W. J., Gilroy, J., Baker, P. J., Stuitje, A. R., Rice, D. W., Rafferty, J. B., and Slabas, A. R. (2002) Crystallization and preliminary X-ray crystallographic studies on acyl-(acyl carrier protein) from Escherichia coli, *Acta Crystallogr D Biol Crystallogr 58*, 330-332.
12.   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.
13.   Roujeinikova, A., Simon, W. J., Gilroy, J., Rice, D. W., Rafferty, J. B., and Slabas, A. R. (2007) Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates, *J Mol Biol 365*, 135-145.
14.   Nguyen, C., Haushalter, R. W., Lee, D. J., Markwick, P. R., Bruegger, J., Caldara-Festin, G., Finzel, K., Jackson, D. R., Ishikawa, F., O'Dowd, B., McCammon, J. A., Opella, S. J., Tsai, S. C., and Burkart, M. D. (2014) Trapping the dynamic acyl carrier protein in fatty acid biosynthesis, *Nature 505*, 427-431.
15.   Endo, K., Helmkamp, G. M., Jr., and Bloch, K. (1970) Mode of inhibition of beta-hydroxydecanoyl thioester dehydrase by 3-decynoyl-N-acetylcysteamine, *J Biol Chem 245*, 4293-4296.

16. Ishikawa, F., Haushalter, R. W., and Burkart, M. D. (2012) Dehydratase-specific probes for fatty acid and polyketide synthases, *J Am Chem Soc 134*, 769-772.

17. Meier, J. L., and Burkart, M. D. (2009) The chemical biology of modular biosynthetic enzymes, *Chem Soc Rev 38*, 2012-2045.

18. Moynie, L., Leckie, S. M., McMahon, S. A., Duthie, F. G., Koehnke, A., Taylor, J. W., Alphey, M. S., Brenk, R., Smith, A. D., and Naismith, J. H. (2013) Structural insights into the mechanism and inhibition of the beta-hydroxydecanoyl-acyl carrier protein dehydratase from Pseudomonas aeruginosa, *J Mol Biol 425*, 365-377.

19. Zhuang, Z., Song, F., Zhao, H., Li, L., Cao, J., Eisenstein, E., Herzberg, O., and Dunaway-Mariano, D. (2008) Divergence of function in the hot dog fold enzyme superfamily: the bacterial thioesterase YciA, *Biochemistry 47*, 2789-2796.

20. Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., and Smith, J. L. (1996) Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site, *Structure 4*, 253-264.

21. Babu, M., Greenblatt, J. F., Emili, A., Strynadka, N. C., Reithmeier, R. A., and Moraes, T. F. (2010) Structure of a SLC26 anion transporter STAS domain in complex with acyl carrier protein: implications for E. coli YchM in fatty acid metabolism, *Structure 18*, 1450-1462.

22. Cryle, M. J., and Schlichting, I. (2008) Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(BioI) ACP complex, *Proc Natl Acad Sci U S A 105*, 15696-15701.

23. Parris, K. D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C. C., Seehra, J., and Somers, W. S. (2000) Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites, *Structure 8*, 883-895.

24. Guy, J. E., Whittle, E., Moche, M., Lengqvist, J., Lindqvist, Y., and Shanklin, J. (2011) Remote control of regioselectivity in acyl-acyl carrier protein-desaturases, *Proc Natl Acad Sci U S A 108*, 16594-16599.

25. Qiu, X., and Janson, C. A. (2004) Structure of apo acyl carrier protein and a proposal to engineer protein crystallization through metal ions, *Acta Crystallogr D Biol Crystallogr 60*, 1545-1554.

26. Kapur, S., Worthington, A., Tang, Y., Cane, D. E., Burkart, M. D., and Khosla, C. (2008) Mechanism based protein crosslinking of domains from the 6-deoxyerythronolide B synthase, *Bioorg Med Chem Lett 18*, 3034-3038.

27. Worthington, A. S., and Burkart, M. D. (2006) One-pot chemo-enzymatic synthesis of reporter-modified proteins, *Org Biomol Chem 4*, 44-46.

28. Worthington, A. S., Rivera, H., Torpey, J. W., Alexander, M. D., and Burkart, M. D. (2006) Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis, *ACS Chem Biol 1*, 687-691.

29. Worthington, A. S., Hur, G. H., Meier, J. L., Cheng, Q., Moore, B. S., and Burkart, M. D. (2008) Probing the compatibility of type II ketosynthase-carrier protein partners, *Chembiochem 9*, 2096-2103.

30. Worthington, A. S., Porter, D. F., and Burkart, M. D. (2010) Mechanism-based crosslinking as a gauge for functional interaction of modular synthases, *Org Biomol Chem 8*, 1769-1772.

31.   Meier, J. L., Mercer, A. C., Rivera, H., Jr., and Burkart, M. D. (2006) Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification, *J Am Chem Soc 128*, 12174-12184.

32.   Meier, J. L., Haushalter, R. W., and Burkart, M. D. (2010) A mechanism based protein crosslinker for acyl carrier protein dehydratases, *Bioorg Med Chem Lett 20*, 4936-4939.

33.   Haushalter, R. W., Worthington, A. S., Hur, G. H., and Burkart, M. D. (2008) An orthogonal purification strategy for isolating crosslinked domains of modular synthases, *Bioorg Med Chem Lett 18*, 3039-3042.

34.   Bruegger, J., Haushalter, B., Vagstad, A., Shakya, G., Mih, N., Townsend, C. A., Burkart, M. D., and Tsai, S. C. (2013) Probing the selectivity and protein.protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers, *Chem Biol 20*, 1135-1146.

35.   Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., Lindquist, Y., and Larsen, S. (1999) The X-ray crystal structure of beta-ketoacyl [acyl carrier protein] synthase I, *FEBS Lett 460*, 46-52.

36.   Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., and Larsen, S. (2001) Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery, *Structure 9*, 233-243.

37.   Price, A. C., Choi, K. H., Heath, R. J., Li, Z., White, S. W., and Rock, C. O. (2001) Inhibition of beta-ketoacyl-acyl carrier protein synthases by thiolactomycin and cerulenin. Structure and mechanism, *J Biol Chem 276*, 6551-6559.

38.   Zhang, Y. M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O., and White, S. W. (2001) Identification and analysis of the acyl carrier protein (ACP) docking site on beta-ketoacyl-ACP synthase III, *J Biol Chem 276*, 8231-8238.

39.   von Wettstein-Knowles, P., Olsen, J. G., McGuire, K. A., and Henriksen, A. (2006) Fatty acid synthesis. Role of active site histidines and lysine in Cys-His-His-type beta-ketoacyl-acyl carrier protein synthases, *FEBS J 273*, 695-710.

40.   Pappenberger, G., Schulz-Gasch, T., Kusznir, E., Muller, F., and Hennig, M. (2007) Structure-assisted discovery of an aminothiazole derivative as a lead molecule for inhibition of bacterial fatty-acid synthesis, *Acta Crystallogr D Biol Crystallogr 63*, 1208-1216.

41.   Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol 276*, 307-326.

42.   Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography 40*, 658-674.

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

44.   Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D 64*, 61-69.

45.   Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D 68*, 352-367.

46.   Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Crystallogr D 65*, 1074-1080.

Chapter 7

Conclusions and Future Directions

7.1 Conclusions

Polyketides and fatty acids play important roles in many cellular processes. Polyketides are ecologically relevant secondary metabolites that have been used as potent pharmaceutical compounds and leads in drug discovery(1). Fatty acids are essential for life, and recently fatty acids have been used to develop advanced biofuels, which are necessary to strengthen our national energy security, as well as address the growing concerns about fossil fuel dependence(2).

These two classes of molecules are linked evolutionarily and share many common enzymatic components during their biosynthesis. The work presented in this dissertation provides insights into the structures and functions of enzymes involved in fatty acid and polyketide biosynthesis.

7.1.1 ARO/CYCs (Chapter 2)

To understand cyclization and aromatization of reduced and non-reduced poly-β-ketone intermediates during type II polyketide biosynthesis, the structures and functions of the di-domain ARO/CYCs BexL and StfQ were solved. BexL aromatizes a reduced substrate and StfQ cyclizes/aromatizes a non-reduced substrate, despite the same domain arrangement and fold present in both structures. BexL and StfQ both consist of two mono-domains, which look nearly identical with each other, as well as with previously characterized mono-domain ARO/CYCs(3-5). Type II minimal PKS functional assays and mutagenesis studies revealed that the N-terminal

domain for BexL is responsible for its aromatase activity, whereas the C-terminal domain of StfQ is responsible for cyclization and aromatization. The C-terminal domain of BexL and the N-terminal domain of StfQ were found to have no catalytic activity; however, they may be important for overall protein stability. This work is important because we present the first molecular view and functional analysis of two di-domain ARO/CYCs with different substrate specificities and we are currently finalizing the manuscript for submission to PNAS.

7.1.2 DpsC (Chapter 3)

DpsC is a unique bi-functional acyltransferase/ketosynthase that installs a three-carbon starter unit during daunorubicin biosynthesis(6, 7). Several interesting features of DpsC activity have remained unknown since its initial characterization, such as the molecular basis for the bi-functionality of DpsC, as well as the substrate specificity of DpsC for propionyl-CoA. In order to unravel these mysteries, we solved the crystal structures of five forms of DpsC, including the *apo*-DpsC structure, multiple acyl-DpsC structures, and a DpsC co-crystal structure with a phosphopantetheine analogue. Crystallographic studies of DpsC revealed a nucleophilic active site serine juxtaposed to an oxyanion hole and hydrophobic acyl-binding region. Furthermore, we used a pantetheine analogue to induce a complex formation between DpsG (the Dps ACP) and DpsC. Combined with functional assays done by our collaborator Prof. John Crosby at University of Bristol, this work sheds light, for the first time, on a unique ketosynthase that is also an acyltransferase. This work is significant because the combined structural and functional studies of DpsC will allow for the engineering of starter unit specificity in type II PKSs. We are currently finalizing the manuscript for submission to PNAS.

7.1.3 AuaEII (Chapter 4)

AuaEII initiates a novel priming strategy for an unusual starter unit during aurachin biosynthesis through the conversion of anthranilic acid into anthranoyl-CoA. Subsequently, anthranoyl-CoA is then integrated into type II polyketide biosynthesis(8). The crystal structure of AuaEII was solved bound to an anthranoyl-AMP intermediate, which allowed for identification of the aryl-binding region. A homology model of the anthranoyl-CoA:ACP acyltransferase (AuaE) was generated to gain insight into how AuaE has evolved to possess acyltransferase activity, despite having a CoA ligase fold. By structure and sequence comparisons of AuaE to other know CoA-ligases, we hypothesize that AuaE may catalyze acyl-transfer the reaction using a nucleophilic active site residue such as serine or cysteine. This work is significant because it improves our understanding of starter unit selection in type II PKSs, which will allow for the generation of new bioactive natural product analogues. We are currently finalizing the manuscript for submission to Chemistry and Biology.

7.1.4 BexE (Chapter 5)

BE-7585A is a unique type II PKS angucyclic natural product with multiple oxygenases in its gene cluster. Initial studies on BE-7585A biosynthesis suggested that a novel oxidative rearrangement occurs, and the oxygenase BexE is hypothesized to play an important role in this process(9). To elucidate oxidation steps during BE-7585A biosynthesis, we solved the crystal structure of BexE and conducted preliminary structural analysis, supplemented with molecular docking studies. Our collaborators in Jurgen Rohr's lab at the University of Kentucky have recently developed an *in vitro* assay for BexE, and we are in the process of structurally characterizing the true substrate and product of BexE. This work is significant because a

thorough understanding of oxidative tailoring in type II PKSs will allow for the engineering of new bioactive angucyclines. We are currently finalizing the manuscript for submission to Chemistry and Biology.

7.1.5 AcpP=FabA and AcpP=FabB (Chapter 6)

A long-term goal of our lab is to understand the role of protein-protein interactions in biosynthesis. We used the model type II fatty acid synthase from *E. coli* to gain insight into how the acyl carrier protein (AcpP) interacts with the dehydratase (DH) FabA and ketosynthase (KS) FabB(10). We applied mechanism-based crosslinking, NMR, and molecular dynamics to better understand how AcpP delivers substrates into the FabA active site for catalysis(11). In this work, we captured a snapshot of dehydratase activity by solving the crystal structure of the AcpP=FabA crosslinked protein complex. Dr. Chi Nguyen primarily did this work and I assisted in protein crystallization and structural determination of the AcpP=FabA complex. As a logical extension of this project, we applied mechanism-based crosslinking to obtain an AcpP=FabB crystal structure. This allowed for the direct comparison of AcpP interactions with different enzymes in the *E. coli* type II FAS. This work has a high impact because the AcpP=FabB crystal structure represents the first ever structure of an ACP in complex with a KS and provides a rational basis for engineering protein-protein interactions in the versatile *E. coli* type II FAS, which is used for the production of biofuels. We will submit the final manuscript to Nature.

7.2 Future Directions

7.2.1 Gaining a Picture of Cyclization in Action

Thus far, the crystal structures of three mono-domain ARO/CYCs (TcmN, WhiE and ZhuI) and two di-domain ARO/CYCs (BexL and StfQ) have been solved(3-5). These structures have revolutionized people's understanding of cyclization and aromatization in type II PKSs. Nevertheless, we do not have substrate-analogue bound co-crystal structures of ARO/CYCs with different cyclization patterns. Recently, Tim Valentic solved multiple crystal structures of the C9-C14 mono-domain ARO/CYC WhiE bound to substrate analogues of varying length to address this concern. His work revealed that only substrates of the appropriate length can adopt the proper conformation for cyclization. However, our current set of substrate analogues contain conformationally restricted isoxazole rings, and could only be co-crystallized with the WhiE ARO/CYC. Accordingly, we are developing more flexible polyketide mimics that can elucidate cyclization in not only WhiE but also other ARO/CYCs with differing cyclization patterns.

7.2.2 Structural Studies on a Putative C9-C14 ARO/CYC From a Reducing PKS

Towards the end of my graduate career I noticed that a recently discovered type II PKS natural product class named the arixanthomycins may contain a C9-C14 mono-domain ARO/CYC that acts on a C11-reduced substrate (Figure 7-1)(12). This pathway is novel for two reasons: a C9-C14 ARO/CYC that acts on a C11-reduced substrate has never been characterized and the system contains an atypical C11-ketoreductase. I drafted a letter and initiated a collaboration with Dr. Sean Brady, the head of the lab at Rockefeller University who discovered the arixanthomycins. Consequently, we acquired the arixanthomycin gene cluster in its entirety, and initial cloning of the putative ARO/CYC and ketoreductases genes has been carried out by
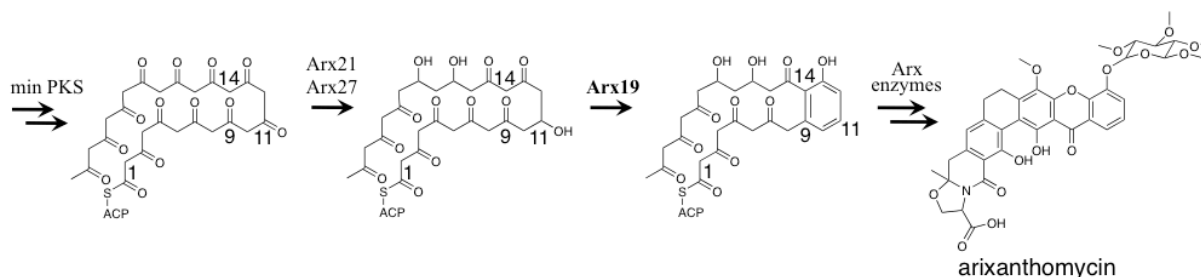
Figure 7-1. Proposed biosynthesis of arixanthomycin. The minimal PKS synthesizes a 26-carbon poly-β-ketone intermediate, which undergoes several regiospecific reductions. The reduced intermediate is cyclized between carbons C9 and C14 by Arx19, the putative ARO/CYC. Arx19 is also proposed to aromatize the terminal ring C9-C14 ring. The cyclized intermediate is then processed by additional Arx enzymes to yield arixanthomycin.

Tim Valentic. We would like to crystallize the novel ARO/CYC and do a structural comparison to other know ARO/CYCs to understand aromatization of C11-reduced, C9-C14 cyclized intermediates. Additionally, the arixanthomycin minimal PKS enzyme can be expressed and the assayed with the putative C11-ketoreductase and ARO/CYC to verify enzyme activity *in vitro*. This work is in accordance with our previous ARO/CYC studies and will shed light into how reducing ARO/CYCs are able to recognize their substrates.

7.2.3 Engineering DpsC to Accept New Substrates

The DpsC structural studies are part of an ongoing collaboration with Dr. John Crosby's laboratory at the University of Bristol in the United Kingdom. Dr. Crosby's laboratory has been conducting functional studies on DpsC substrate specificity and kinetics. In addition to the DpsC structural determination, I generated a series of single, double, and triple mutations aimed at rationally engineering the substrate specificity of DpsC based on comparative sequence analysis. Ultimately, the goal of this project is to gain a complete understanding of the KS and AT activity of DpsC in addition to directing substrate specificity.
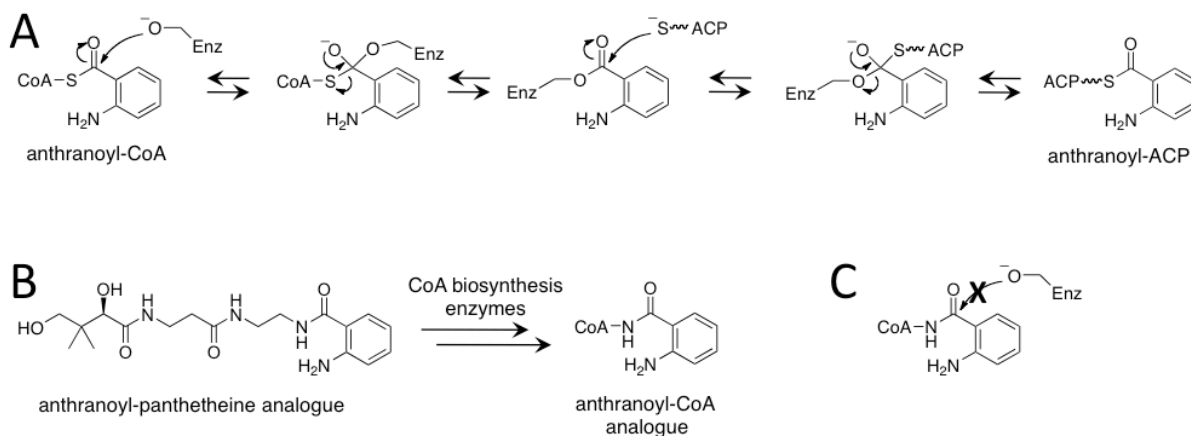
Figure 7-2. The design of an anthranoyl-CoA analogue for AuaE co-crystallization. (A) The proposed AuaE acyltransferase mechanism employing an active site serine for transacylation. (B) The synthesized anthranoyl-pantetheine analogue and conversion into an anthranoyl-CoA analogue using CoA biosynthesis enzymes. (C) The anthranoyl-CoA will interact tightly with AuaE but will not be able to undergo transacylation.

7.2.4 Using an Anthranoyl-CoA Analogue to Induce AuaE Crystallization

Despite extensive efforts, the anthranoyl-CoA:ACP acyltransferase AuaE could not be crystallized. This is most likely because of the high conformational flexibility of the C-terminal domain. I hypothesize that a substrate analogue is needed to stabilize the C-terminal domain and aid in crystallization. Alex White synthesized an anthranoyl-pantetheine analogue that can be enzymatically converted to a non-hydrolyzable anthranoyl-CoA analogue (Figure 7-2). As a last attempt to crystallize AuaE and determine the structural basis for its AT activity, I propose to co-crystallize AuaE with the anthranoyl-CoA analogue.

7.2.5 Identifying the True BexE Substrate

Our collaborators in Jurgen Rohr's laboratory developed an *in vitro* assay for BexE. At the present time, this work is still in progress and we are waiting for structural characterization of the substrate and products from this assy. We plan to co-crystallize BexE with its substrate and products to gain insight into its oxidation reaction and engineer BexE to accept other substrates.

7.2.6 Modulating Fatty Acid Production by Altering Protein-Protein Interactions

Ultimately, the goal of the AcpP=FabA and AcpP=FabB complex study is to use the knowledge gained in our structural studies to modulate the fatty acid profile *in vivo* in a predictable manner. We plan to mutate surface residues of AcpP and target enzymes to change the affinity of protein-protein interactions while AcpP carries the growing substrate through rounds of fatty acid biosynthesis. To our knowledge, this strategy has never been tested. Jake "The Snake" Milligan has already expressed and purified all the type II FAS enzymes and will follow a previously described protocol for reconstituting the entire FAS *in vitro* (13-15). Once the *in vitro* assay is mastered in our lab, we will begin to evaluate AcpP and target enzyme surface mutants for their ability to generate altered fatty acid profiles relative to the wild type enzymes. Depending on the results of the *in vitro* experiments, we plan to take interesting mutants into an *in vivo* setting in *E. coli* and evaluate the fatty acid profiles in cells. We hypothesize that altering protein-protein interactions *in vivo* will be an effective strategy to produce shorter fatty acids, which can be diverted into pathways for short-chain alkane biofuel production.

References

1.   Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms, *Curr Opin Chem Biol 7*, 285-295.
2.   Janssen, H. J., and Steinbuchel, A. (2014) Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels, *Biotechnol Biofuels 7*, 7.
3.   Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y., and Tsai, S. C. (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides, *Proc Natl Acad Sci U S A 105*, 5349-5354.
4.   Ames, B. D., Lee, M. Y., Moody, C., Zhang, W., Tang, Y., and Tsai, S. C. (2011) Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway, *Biochemistry 50*, 8392-8406.
5.   Lee, M. Y., Ames, B. D., and Tsai, S. C. (2012) Insight into the molecular basis of aromatic polyketide cyclization: crystal structure and in vitro characterization of WhiE-ORFVI, *Biochemistry 51*, 3079-3091.
6.   Bao, W., Sheldon, P. J., and Hutchinson, C. R. (1999) Purification and properties of the Streptomyces peucetius DpsC beta-ketoacyl:acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis, *Biochemistry 38*, 9752-9757.
7.   Bao, W., Sheldon, P. J., Wendt-Pienkowski, E., and Hutchinson, C. R. (1999) The Streptomyces peucetius dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide, *J Bacteriol 181*, 4690-4695.
8.   Pistorius, D., Li, Y., Mann, S., and Muller, R. (2011) Unprecedented anthranilate priming involving two enzymes of the acyl adenylating superfamily in aurachin biosynthesis, *J Am Chem Soc 133*, 12362-12365.
9.   Sasaki, E., Ogasawara, Y., and Liu, H. W. (2010) A biosynthetic pathway for BE-7585A, a 2-thiosugar-containing angucycline-type natural product, *J Am Chem Soc 132*, 7405-7417.
10.   White, S. W., Zheng, J., Zhang, Y. M., and Rock, C. O. (2005) The structural biology of type II fatty acid biosynthesis, *Annu Rev Biochem 74*, 791-831.
11.   Nguyen, C., Haushalter, R. W., Lee, D. J., Markwick, P. R., Bruegger, J., Caldara-Festin, G., Finzel, K., Jackson, D. R., Ishikawa, F., O'Dowd, B., McCammon, J. A., Opella, S. J., Tsai, S. C., and Burkart, M. D. (2014) Trapping the dynamic acyl carrier protein in fatty acid biosynthesis, *Nature 505*, 427-431.
12.   Kang, H. S., and Brady, S. F. (2014) Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols, *ACS Chem Biol 9*, 1267-1272.
13.   Kuo, J., and Khosla, C. (2014) The initiation ketosynthase (FabH) is the sole rate-limiting enzyme of the fatty acid synthase of Synechococcus sp. PCC 7002, *Metab Eng 22*, 53-59.
14.   Xiao, X., Yu, X., and Khosla, C. (2013) Metabolic flux between unsaturated and saturated fatty acids is controlled by the FabA:FabB ratio in the fully reconstituted fatty acid biosynthetic pathway of Escherichia coli, *Biochemistry 52*, 8304-8312.
15.   Yu, X., Liu, T., Zhu, F., and Khosla, C. (2011) In vitro reconstitution and steady-state analysis of the fatty acid synthase from Escherichia coli, *Proc Natl Acad Sci U S A 108*, 18643-18648.