

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Structural Studies of Polyketide Synthase Biosynthetic Machinery

Permalink

<https://escholarship.org/uc/item/9xv0t3f7>

Author

Valentic, Timothy Robert

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Structural Studies of Polyketide Synthase Biosynthetic Machinery

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Timothy Robert Valentic

Dissertation Committee:
Professor Shiou-Chuan Tsai, Chair
Associate Professor Naomi Morrisette
Chancellor's Professor Thomas Poulos

2016

DEDICATION

This dissertation is dedicated to my family, who always loved and encouraged me to follow my dreams. Without their love and support I would have never made it this far. To my mother, Patricia, for her never ending kindness, her strength, and her devotion to our family. To my father, John, for his constant faith in my success, the many inspiring speeches he has given and for instilling a love of science in me from an early age. To my brother Jonathan, who is one of the strongest people I know. Since I was born, Jon has always looked out for me and has been a caring older brother. To my sister Christine, who has always been so kind to me and has overcome so much. To Karlie Barnes, my long-lost niece now found; I am so happy to have her in my life. I know the family has a bright future ahead. To the many friends I've met throughout the years that I have had the opportunities to laugh, cry and share both the good and bad times of life. Finally, to my girlfriend Amanda Gilliam who, through thick and thin, always stood by my side.

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
LIST OF FIGURES	vi
LIST OF TABLES	x
ACKNOWLEDGMENTS	xi
CURRICULUM VITAE	xvi
ABSTRACT OF THE DISSERTATION	xix
CHAPTER 1: Introduction	1
1.1 Natural Products: A Brief History	1
1.2 Natural Products: Primary and Secondary Metabolism	4
1.3 Primary Metabolism: Fatty Acid Biosynthesis	7
1.4 Secondary Metabolism: Polyketide Synthases	10
1.5 Type II ARO/CYCs: Discovery and Overall Function	25
1.6 Type II Ketoreductases	31
1.7 Secondary Metabolism: Nonribosomal Peptide Synthetases	36
1.8 Secondary Metabolism: Hybrid Systems	40
1.9 Acyl-CoA Carboxylases	42
1.10 Crotonyl CoA Carboxylases/Reductases	48
1.11 Thesis Overview	51
1.12 References	51
CHAPTER 2: Structural Insights into Chain Length and Cyclization Specificities of Bacterial Type II Polyketide Intermediates in an Aromatase/Cyclase Domain	62
2.1 Abstract	62
2.2 Introduction	63

2.3	Results and Discussion	66
2.4	Materials and Methods	78
2.5	References	82
CHAPTER 3: Structural Basis for Acyl-CoA Carboxylase-Mediated Assembly of Unusual Polyketide Synthase Extender Units Incorporated into the Stambomycin Antibiotics		85
3.1	Abstract	85
3.2	Introduction	86
3.3	Results and Discussion	90
3.4	Conclusions	105
3.5	Experimental Procedures and Supplemental Information	106
3.6	References	147
CHAPTER 4: Structural Basis for Condensation Domain Mediated Chain Release During Enacyloxin Biosynthesis		150
4.1	Abstract	150
4.2	Introduction	150
4.3	Results and Discussion	154
4.4	Materials and Methods	169
4.5	References	175
CHAPTER 5: Structural and Functional Studies of Enzymes Involved in Pentangular Polyphenol Biosynthesis		178
5.1	Abstract	178
5.2	Introduction	179
5.3	Results and Discussion	182
5.4	Materials and Methods	193
5.5	References	202

CHAPTER 6: Conclusions and New Directions	205
6.1 Overview	205
6.2 Chapter 2: WhiE	206
6.3 Chapter 3: MccB	207
6.4 Chapter 4: Bamb_5915	208
6.5 Chapter 5: ARX Pathway	210
6.6 New Directions	211
6.7 References	217

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 1.1. Acetyl-CoA is central to both primary and secondary metabolic pathways.	5
Figure 1.2. Different classes of natural products.	7
Figure 1.3. Domain architecture of various polyketide and fatty acid synthases.	8
Figure 1.4. Chain elongation and processing steps of polyketide and fatty acid biosynthesis.	10
Figure 1.5. A collection of polyketide natural products produced by Modular type I PKSs, bacterial type II PKSs, or fungal type I iterative PKSs.	11
Figure 1.6. Biosynthesis of fungal aromatic polyketides by non-reducing PKSs.	16
Figure 1.7. Schematic of type I modular PKS biosynthesis.	17
Figure 1.8. Structures of type I modular PKS chain extension and processing domains.	18
Figure 1.9. A structural comparison of FAS and PKS megasynthase structures.	20
Figure 1.10. The crystal structures of the enzymes involved in type II PKS biosynthesis.	22
Figure 1.11. A comparison of non-reducing C9-C14 type II PKS pathways to C7-C12 reducing and non-reducing type II PKS pathways.	23
Figure 1.12. A schematic diagram of the different ARO/CYC types.	27
Figure 1.13. Depiction of the KR activity and substrate specificity in representative type II PKSs.	32
Figure 1.14. The crystal structures and proposed mechanisms of type II PKS KRs.	35
Figure 1.15. An introduction to NRPS structure and function.	37
Figure 1.16. Side by side comparison of several C domains from different NRPSs.	39
Figure 1.17. A general cartoon schematic of different PKS/NRPS and FAS/PKSs hybrid systems.	41
Figure 1.18. The crystal structural and mechanism of propionyl-CoA carboxylase (PCC).	44
Figure 1.19. The extender diversity of polyketide natural product.	47
Figure 1.20. Crystal structure and function of the crotonyl-CoA carboxylase/reductase CinF.	49
Figure 2.1. The different clades of type II PKS ARO/CYC and their natural products.	64
Figure 2.2. The chemical structures of the four polyketide mimics used.	66
Figure 2.3. The overall architecture of WhiE ARO/CYC.	69

Figure 2.4. An inspection of active site residues of WhiE and interactions with the corresponding pantetheine linked mimic.	72
Figure 2.5. An active site overlay of the heptaketide-WhiE and octaketide-WhiE structures highlighting the inverse binding modes of 3 and 4 .	79
Figure 2.6. Sequence alignment of WhiE to other ARO/CYC homologues.	81
Figure 2.7. SA-Omit maps.	82
Figure 3.1. Structures of the four main components of the stambomycin complex.	87
Figure 3.2. Organization of the stambomycin biosynthetic gene cluster in <i>S. ambofaciens</i> ATCC23877.	88
Figure 3.3. FAS precursor incorporation.	91
Figure 3.4. Structure of stambomycin analogue 22 .	92
Figure 3.5. Incorporation of [3- ² H ₂]heptanoic acid 23 into the stambomycin analogue.	93
Figure 3.6. Mass spectra of stambomycin analogue 22 with heptanoic acid 21 and [3- ² H ₂]heptanoic acid 23 .	95
Figure 3.7. Structures of the reveromycins produced by <i>Streptomyces</i> sp. SN-593.	96
Figure 3.8. Mass spectra of reveromycin D 32 .	97
Figure 3.9. Extracted ion chromatograms for reveromycin E 35 .	98
Figure 3.10. X-ray crystal structure of <i>apo</i> -MccB.	99
Figure 3.11. X-ray crystal structure of MccB with four molecules of hexanoyl-CoA bound.	101
Figure 3.12. Structure of the binding region.	102
Figure 3.13. Structure of azide-tagged stambomycin analogue 37 .	104
Figure 3.14. Acetylene-tagged stambomycin analogue 39 .	105
Figure 3.15: Mass spectra from UHPLC-ESI-Q-TOF-MS analysis of methanolic mycelial extracts from the [U- ² H]L-valine 11 incorporation experiments.	110
Figure 3.16: Mass spectra from UHPLC-ESI-Q-TOF-MS analysis of methanolic mycelial extracts from the [U- ² H]butyric acid 12 incorporation experiments.	111
Figure 3.17: Mass spectra from UHPLC-ESI-Q-TOF-MS analysis of methanolic mycelial extracts from the [U- ² H]L-isoleucine 9 incorporation experiments.	112
Figure 3.18: Mass spectra from UHPLC-ESI-Q-TOF-MS analysis of methanolic mycelial extracts from the [U- ² H]L-leucine 10 incorporation experiments.	113

Figure 3.19: Mass spectra of stambomycin analogue 22 feeding studies with heptanoic acid 21 .	114
Figure 3.20. Comparison of the ^1H NMR spectra of 22 and stambomycins C/D 3/4 .	117
Figure 3.21. ^{13}C -APT NMR spectrum of 22 .	118
Figure 3.22. Overlaid COSY NMR spectra of 22 and stambomycins C/D 3/4 .	118
Figure 3.23. Overlaid TOCSY NMR spectra of 22 and stambomycins C/D 3/4 .	119
Figure 3.24. Overlaid HMBC NMR spectra of 22 and stambomycins C/D 3/4 .	119
Figure 3.25. Comparison of the ^1H NMR spectra of $[3\text{-}^2\text{H}_2]$ heptanoic acid 23 and unlabeled heptanoic acid 21 .	124
Figure 3.26. Comparison of the ^{13}C NMR spectra of $[3\text{-}^2\text{H}_2]$ heptanoic acid 23 and unlabeled heptanoic acid 21 .	125
Figure 3.27. Mass spectra from stambomycin analogue 22 feeding studies with $[3\text{-}^2\text{H}_2]$ heptanoic acid 23 .	126
Figure 3.28. Mass spectra of reveromycin D 32 feeding studies with $[3\text{-}^2\text{H}_2]$ heptanoic acid 23 .	128
Figure 3.29. Mass spectra of reveromycin D 32 feeding studies with labeled heptanoic acid.	128
Figure 3.30. Sequence alignment of MccB from <i>S. ambofaciens</i> with other ACC β -subunits.	135
Figure 3.31. Superimposition of PccB from <i>S. coelicolor</i> with MccB from <i>S. ambofaciens</i> .	136
Figure 3.32. Structural comparison of MccB, PccB, and AccD5 active site residues.	136
Figure 3.33. Stereo view of hexanoyl-CoA bound to MccB.	137
Figure 3.34. Stereo view of the apo MccB hexamer.	137
Figure 3.35. Stereo views of MccB hexamer with quadruple binding.	138
Figure 3.36. Stereo views of MccB hexamer with single binding.	139
Figure 3.37. Mass spectra of stambomycin analogue 37 feeding studies.	142
Figure 3.38. Chromatograms of extracts grown in the presence or absence of 6-heptynoic acid.	143
Figure 3.39. Mass spectra of stambomycin analogue 39 feeding studies.	144
Figure 3.40. Comparison of ^1H -NMR spectra for stambomycin analogue 39 and stambomycins C/D 3/4 .	145

Figure 3.41. NMR spectrum of biotinylated stambomycin analogue derived from the click reaction of stambomycin analogue 39 with the azide-PEG3-biotin conjugate.	146
Figure 4.1. The overall architecture of enacyloxin PKS/NRPS hybrid complex (ena PKS).	152
Figure 4.2. The crystal structure of bamb_5915.	155
Figure 4.3. A structural comparison of bamb_5915 to other C domain structures.	156
Figure 4.4. Superimposition of bamb_5915 with SgcC5.	158
Figure 4.5. COM domains involved in tubulysin and enacyloxin biosynthesis.	160
Figure 4.6. Biosynthetic genes from characterized pathways utilizing COM _{2N} domains.	162
Figure 4.7. Docking of bamb_5915 with PCP_5917.	163
Figure 4.8. <i>In vitro</i> reconstitution of enacyloxin chain release machinery.	165
Figure 4.9. Structures of natural products utilizing COM junctions.	166
Figure 5.1. The proposed biosynthesis of arixanthromycin A.	180
Figure 5.2. A schematic of the differing KR clades and their substrate specifics.	181
Figure 5.3. The overall architecture of ARX 21.	183
Figure 5.4. The co-factor and proposed substrate binding site of ARX 21.	184
Figure 5.5. Electrostatic potential surfaces of ARX 21 and other KR homologues.	185
Figure 5.6. Residues proposed to be important for ARX 21 substrate binding and recognition.	188
Figure 5.7. Sequence alignment of ARX 21 to functionally characterized C-19, C-17, C-15, and C-9 reducing KRs.	189
Figure 5.8. HPLC analysis of ARX 19 and ARX 21 activities at 280 nM.	191
Figure 5.9. <i>In vitro</i> reconstitution and protein purification of ARX 19 and ARX 21.	199
Figure 5.10. Sequence alignment of ARX 21 to functionally characterized C-19 reducing KRs.	200
Figure 6.1. Examples of polyketide isosteres currently being developed.	212
Figure 6.2. The purification of the hexanoyl-PCP_5917/bamb_5915 complex, COM _{2N} and bamb_5915ΔCOM _{2N} .	215

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 2.1. Crystallographic statistics for data collection, modeling and refinement.	80
Table 3.1: Bacterial strains and plasmids.	107
Table 3.2: Oligonucleotide primers.	108
Table 3.3: Comparison of ^1H and ^{13}C NMR data for stambomycin D 4 and stambomycin analogue 22 .	115
Table 3.4. Crystallographic statistics.	133
Table 4.1. Primers for the generation of COM _{2N} -C and bamb_5915 Δ COM _{2N} .	172
Table 4.2. Data collection and refinement statistics	174
Table 5.1. Primers used to generate pET28-ARX 21 and pET28-ARX19.	193
Table 5.2. Data collection and refinement statistics.	201

ACKNOWLEDGMENTS

When I left my small home town of Ukiah California in 2011 to pursue a doctorate in biochemistry, I never could have predicted the long lasting connections I would make with mentors, friends, and mentees that I met during my graduate school experience. Without the support of friends both old and new, my family, and mentors none the work presented here would have been possible.

First, I would like to express my sincere gratitude to my PhD advisor Professor Sheryl Tsai for being such a kind, patient, and enthusiastic mentor over the past 5 years. I don't think Sheryl realized how much she really helped turn my life around (I was stuck in Ukiah after all!). When I interviewed at UCI in 2011, she was only the professor that I talked to that put me at ease and genuinely seemed excited about research. Sheryl, perhaps unknowingly, took a risk by allowing me into her lab. I think Sheryl may have realized this when she received an email from myself in 2011, drunk, in Barcelona requesting a rotation in her lab. Luckily for myself, Sheryl agreed, although this wouldn't be the last time she would encounter my drunk escapades. Nonetheless, Sheryl has always expressed her confidence and faith in my abilities as a person and a scientist. She has provided a warm, friendly, and inviting environment that truly fosters the love of science. Sheryl is incredibly intelligent and an excellent mentor; being in her presence these past 5 years has been happily humbling. Clearly, without her help and guidance I wouldn't be where I am today. Sheryl, I will do my best to not test the patience of my next advisor... well, at least not as much. I can't thank Sheryl enough for everything she has done for me over the past 5 years. I will always have good memories of the times I spent working with her.

Next, I would like to thank Professors Naomi Morrisette and Tom Poulos. Both have been instrumental for my success at UCI. I have taught for Professor Morrisette on multiple occasions and it has been a wonderful experience filled with interesting conversations about science. Professor Morrisette has also given me myriad advice in regards to finding a good postdoc, classes that I need to complete to graduate, and life in general. Professor Morrisette genuinely cares about graduate students and her love of science is contagious. She also understands the struggles left handed people face every day. Professor Naomi Morrisette, I wish you and your the family best; the future is bright.

I would like to express my sincere gratitude to Professor Poulos. I hope Professor Poulos is aware of the positive impact he has on the second floor labs of Natural Sciences 1. On several occasions I've stopped Professor Poulos to ask him many questions about crystallography and he has always been happy to answer them. He has also provided a very collegial environment within the neighboring labs at UCI. On many occasions I have come in on weekends and have seen Professor Poulos roaming the halls of Natural Sciences 1, which has encouraged me and my fellow Tsai lab graduate students to work harder.

Several other professors have been essential for my success at UCI, specifically Professor Celia Goulding and Professor Greg Challis. From the start of my graduate student career Professor Goulding has been supportive and has given me great advice. She helped guide me during the postdoc application process and wrote a letter of recommendation on my behalf for a postdoctoral position at Stanford. Professor Goulding is incredibly knowledgeable and has mentored a group of friendly and talented scientists with whom I had the great pleasure of working. My experience at UCI would not have been the same without her support, help, and advice.

I would like to acknowledge Professor Greg Challis at the University of Warwick for his help and contributions toward Chapter 3. Professor Challis is an excellent mentor and his obsession for natural products biosynthesis is irrepressible. In addition, Professor Challis not only offered me a postdoctoral position in his lab, but he even proceeded to write me a letter of recommendation for a postdoctoral position in Stanford even after I had to decline his offer. He is someone I look up to and admire. I was lucky to have had the opportunity to collaborate with his group as it was very fruitful.

I was fortunate to have many senior graduate student mentors and friends that greatly helped me along the way. Dr. Joel Bruegger, Dr. Jesus Barajas, Dr. Dave Jackson, Dr. Chi Nguyen, Dr. Grace Caldera and Dr. Gaurav Shakya have been paramount to my success at UCI. When I first started at UCI, I was extremely nervous and my bench skills were not the best. Joel turned all of this around in a single rotation and has always been a caring mentor and friend. I've had countless conversations over beers with Dave and Jesus about science and life. Many times after botched experiments they would take me out for a beer at the pub. Joel, Dave and Jesus have always looked out for me and will be long lasting friends. During the times that I thought of giving up they always encouraged me to persevere. I could write an entire thesis about the crazy adventures we've had but that shall be for another time. Chi Nguyen and Grace Caldera were always kind and patient with me and also were happy to answer any of my questions no matter how trivial they were. Gaurav Shakya taught me much about life and chemistry. All of these individuals trained me and positively influenced my graduate career. I look forward to seeing where the future takes these talented individuals as I believe the sky is truly the limit.

Gabe Moreno, Alex White, Jake Milligan, and Andrew Schaub have also played a large role in my success at UCI. Gabe and I joined the Tsai lab around the same time and he always

provided a calming presence. He is a very kind person, patient, a master of crystallography robots and type II PKS ketoreductases. I was lucky to have such a good labmate. Alex White is an awesome chemist with powerful natural product extraction powers. Alex has been a blast to work with (especially while TAing) and has pioneered the development of new oxetane polyketide probes. Jake and Andrew joined the lab during my third year and are incredibly talented scientists. I think had I joined the Tsai lab during the same time they did, I may have had to quit the business! That aside, I have learned much from Jake in regards to crystallography and data processing. Andrew is an MD guru and is also from Ohio. Like the other friends and mentors I've encountered at UCI, Jake and Andrew have been extremely supportive and fun to work with. I wish all these individuals best in the later years of their graduate careers and look forward to reading their future publications.

I would like to acknowledge Jessica Kim who was a “super undergraduate” that I trained. She purified many liters of protein and her work contributed to the completion of Chapters 2 and 5. In addition I would like to thank Alec Follmer who was rotation student that I had the privilege to train. I know he will continue to be successful at UCI.

Dr. Joey Carlson and Alex Chao are two important individuals that I met at UCI and who deserve acknowledgment. Joey was my roommate for 4 years and was my first friend at UCI. We had many of good times over drinking beers, talking about chemistry, shooting zombies and playing Bugs. Bugs is a video game where gigantic insects invade the earth and the player must defeat the invaders with the help of pretty ladies, jet packs and plasma rifles. Needless to say, the world is now safe thanks to Dr. Carlson and myself (although I would like to also note that Joel, Dave, and Jesus also played a role in defending earth from bugs as well) In all seriousness, Joey made graduate school very fun and enjoyable. He introduced me to many good people and really

helped make graduate school a better experience. Alex Chao is a good friend and lab neighbor from the Goulding lab. Alex and I have had many conversations about biochemistry and life at the pub. He is a very talented scientist and I know he will go very far. The Tsai lab considers him an honorary lab member because he is socially out-going and likes to contribute to the engaging scientific atmosphere in the halls of Natural Sciences 1.

Scott Morgan is a friend from Ukiah that deserves recognition. Scott and I have known each other for almost 20 years and he has always been a very loyal friend who has been there during tough times. The same applies to Erik Koski, David Silva and many others from my small town of Ukiah. Without these friends I never would have made it out of Ukiah, let alone the alley behind the Saturday Afternoon Club.

I would like to thank my long-time girlfriend Dr. Amanda Gilliam for her love and support especially during the last three years of graduate school. Amanda is incredibly intelligent, thoughtful, and caring. She has always been there for me during the hard times and makes every day better. Together we have had many delightful conversations about natural products biosynthesis, which has positively influenced my work. The thought of graduate school without Amanda is analogous to a nightmare. Navigating graduate school truly would have been impossible without her by my side; she always manages to find a way to calm me down and make see the bright side of life. I look forward to our future together in Palo Alto and the many adventures we will have. I'm a lucky guy to have such a sweet and amazing girl by his side.

Finally, I would like to thank my family for their love and for always being there to prop me back up when I have fallen.

CURRICULUM VITAE

Timothy R. Valentic

707-972-5067
tvalent1@uci.edu

EDUCATION

University of California, Irvine , Irvine, CA Ph.D., <i>Molecular Biology and Biochemistry</i> –	2011 – 2016
University of California, Santa Cruz , Santa Cruz, CA Bachelor of Science, <i>Biochemistry and Molecular Biology</i> , <i>Cum Laude</i>	2005 – 2009

RESEARCH EXPERIENCE

Graduate Student , University of California, Irvine, <i>Molecular Biology and Biochemistry</i> Primary Investigator – Dr. Sheryl Tsai	2011 – 2016
▪ Utilized post translational mechanism based crosslinkers to study protein-protein interactions in Type I iterative PKSs and FASs. ▪ Successfully solved co-crystal structures of an aromatase cyclase bound to several polyketide mimics. ▪ Solved the crystal structures of both apo and hexanoyl CoA bound forms of a novel long chain Acyl CoA carboxylase beta subunit. ▪ Determined the crystal structure of the first NRPS/PKS hybrid condensation domain and identified regions necessary for protein-protein recognition.	
Graduate Student , University of California, Irvine, <i>Molecular Biology and Biochemistry</i> Primary Investigator – Dr. Rachel Martin	Jan. 2012 – Mar. 2012
<i>Basic biophysical characterization of jellyfish γ-crystallins.</i>	
Undergraduate Researcher , University of California, Santa Cruz, <i>Molecular Biology and Biochemistry</i> Primary Investigator – Dr. Hinrich Boeger	2009
<i>Studied the molecular basis for chromatin packaging in <i>Saccharomyces cerevisiae</i>.</i>	

RELEVANT TECHNIQUES

-
- Protein Expression and Purification
 - X-Ray Crystallography Including MAD/SAD Phasing
 - Cloning and Site Directed Mutagenesis
 - Posttranslational Modification of Carrier Proteins
 - *In vitro* Reconstitution of PKS Pathway Components
 - Dynamic Light Scattering (DLS)
 - Circular Dichroism (CD)
 - HPLC Analysis of Natural Product Precursors
 - MALDI/TOF Mass Spec
 - LC/MS and Electrospray Mass Spec

TEACHING EXPERIENCE

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

Course:	Quarters Taught:
Microbiology Lecture	Spring/Winter 2015
Molecular Biology Lecture	Spring 2015/2014
Biochemistry Lecture	Winter 2015/2014/2013
Microbiology Lab	Spring/Winter/Fall 2014
Human Parasitology Lecture	Fall 2014
Biochemistry Lab	Winter 2013

INVITED PRESENTATIONS

“Structural studies of a crosslinked type I fatty acid synthase from <i>M. tuberculosis</i> ” University of California, Irvine, MB&B Department Research in Progress Talks	2013
“Catching cyclization in action using polyketide mimics.” University of California, Irvine, MB&B Department Research in Progress Talks	2014
“Structural basis for the generation of unusual extender units involved in stambomycin biosynthesis.” University of California, Irvine, MB&B Department Research in Progress Talks	2014
“Learning Drug Design from Nature: Acyl CoA Carboxylase Mediated Biosynthesis of Unusual Polyketide Synthase Extender Units.” Annual ASBMB Conference	2016

HONORS AND ACHIEVEMENTS

Golden Key International Honour Society University of California, Santa Cruz	2009
--	------

FELLOWSHIPS

Cellular and Molecular Biosciences Fellowship University of California, Irvine	2011 – 2012
--	-------------

PUBLICATIONS

Pre-publication (abstracts available upon request)

Valentic, T.R.; Barajas, J.F.; Shakya, G.; Kim, J.; Burkart, M.D.; Tsai, S.C. Probing the cyclization specificity and chain length control of the non-reducing type II polyketide aromatase/cyclase using atom replacement probes. *In preparation.*

Valentic, T.R.; Ray, L.; Miyazawa, T.; Song, L.; Takahashi, S.; Milligan, J.; Osada, H.; Tsai, S.C.; Challis, G.L. A crotonyl-CoA carboxylase/reductase independent pathway for unusual extender unit biosynthesis. *In preparation.*

Valentic, T.R.; Jenner, M.; Ray, L.; Barajas, J.F.; Tsai, S.C.; Challis, G.L. Non-elongating ketosynthase domains circumvent carrier protein incompatibility in trans-AT polyketide synthases. *In preparation.*

Valentic, T.R.; Follmer, A.; Milligan, J.; Schaub, A.J.; Brady, S.F.; Tsai, S.C. Structural and functional characterization of the enzymes involved in the biosynthesis of pentangular polyphenols. *In preparation.*

Barajas, J.F.; **Valentic, T.R.**; Finzel, K.; Shakya, G.; Gamara, N.; Martinez, D.; Beld, J.; Vagstad, A.; Townsend, C.A.; Burkart, M.D.; Tsai, S.C. Structural and biochemical analysis of protein-protein interactions between the acyl carrier protein and the product template domain. *In preparation.*

Jackson, D. R.; Mohammed, L.; Vasilakis, K.; Wattana-amorn, P.; **Valentic, T.R.**; Long, P. F.; Crump, M. P.; Crosby, J.; Tsai, S. C. Determining the molecular basis for starter unit selection during daunorubicin biosynthesis. *In preparation.*

In Press

Jackson, D. R.; Festin, G.; Barajas, J. F.; **Valentic, T. R.**; Aguilar, S.; Patel, A.; Nguyen, M.; Sasaki, E.; Vo, M.; Khanna, A.; Liu, H. W.; Tsai, S. C. Structure and function of two di-domain ARO/CYCs from type II polyketide synthases. *Accepted to PNAS*

REFERENCES

Prof. Shiou Chuan (Sheryl) Tsai, Ph.D.
University of California, Irvine
Department of Molecular Biology and Biochemistry
Irvine, CA 92697
(949) 824 4486
sctsai@uci.edu

Prof. Michael D. Burkart, Ph.D.
University of California, San Diego
Department of Chemistry and Biochemistry
La Jolla, CA 92093
(858) 534 5673
mburkart@ucsd.edu

Prof. Gregory Challis, Ph.D.
University of Warwick, Coventry
Department of Chemistry
Coventry, CV4 7AL
United Kingdom
+44 (0)2476 574024
g.l.challis@warwick.ac.uk

Prof. Celia Goulding, Ph.D.
University of California, Irvine
Department of Molecular Biology and Biochemistry
Irvine, CA 92697
(949) 824-0337
celia.goulding@uci.edu

ABSTRACT OF THE DISSERTATION

Structural Studies of Polyketide Synthase Biosynthetic Machinery

by

Timothy Robert Valentic

Doctor of Philosophy in Molecular Biology & Biochemistry

University of California, Irvine, 2016

Professor Shiou-Chuan Tsai, Chair

Bacterial, fungal, and plant secondary metabolic pathways afford a huge library of elaborate organic molecules known as natural products. Polyketides and nonribosomal peptides are natural products that are biosynthesized from multienzyme complexes known as polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs), respectively. PKSs and NRPs utilize simple acyl-CoA or amino acid building blocks for the assembly of structurally diverse natural products. Because both polyketides and nonribosomal peptides exhibit a wide range of biological activities, they have been utilized as therapeutics and are important for drug discovery. The objective of this dissertation is to improve the structural understanding of how the biosynthetic machinery associated with polyketide and nonribosomal peptide biosynthesis generates such a large array of complex natural products.

In type II PKS biosynthesis, aromatase/cyclases (ARO/CYCs) are responsible for the generation of set C7-C12 or C9-C14 ring patterns through the regiospecific cyclization of reactive poly-beta-ketone intermediates. In Chapter 2, the first co-crystal structures of WhIE bound to four distinct pantetheine linked polyketide chain mimics are discussed. This work provides the first snapshot of ARO/CYC binding modes which, until now, have remained hypothetical. Further, this is the first successful application of polyketide mimics to investigate the structures and functions of ARO/CYCs.

Methyl-, ethyl-, and malonyl-CoA extender units are the most common building blocks employed by type I modular PKSs for natural product biogenesis. However, certain type I modular PKSs incorporate non-canonical long chain extender units to augment the structural diversity of a polyketide scaffold. Typically, crotonyl-CoA carboxylase/reductases (CCRs) are responsible for the generation of extravagant PKS extender units, whereas acyl-CoA carboxylases (ACCases) are responsible for the production of canonical short chain extender units. In Chapter 3, we discuss the structure and function of a novel ACCase β - subunit, MccB from the stambomycin biosynthetic pathway that is capable of generating long chain extender units. The crystal structures of *apo* and hexanoyl-CoA bound MccB were solved, providing a structural basis for unusual ACCase-mediate extender formation during stambomycin biogenesis.

Typically, in type I modular PKS biosynthesis, thioesterases (TEs) are responsible for product hydrolysis from the enzymatic assembly line. In some PKS/NRPS hybrid systems, however, a TE domain is absent and instead a terminal condensation domain (C domain) is required for product liberation from the enzymatic assembly line. The enacyloxin biosynthetic pathway is an example of a PKS/NRPS hybrid that utilizes a standalone NRPS condensation

domain, bamb_5915, for product release through an intermolecular transesterification reaction between shikimate and a mature polyketide chain. The crystal structure and functional analysis of bamb_5915 is described in Chapter 4. In addition, bioinformatic analysis of PCP_5917 and other NRPS/PKS hybrid pathways is discussed. The information gleaned from these studies will aid in the generation of “mixed” PKS/NRPS natural products.

Genome mining of environmental DNA is an emerging method that has led to the identification of novel PKS natural product pathways. The arixanthomycins are a group of recently discovered natural products that were identified through the heterologous expression of environmental DNA derived gene clusters in *Streptomyces*. Chapter 5 reports the crystal structure of ARX 21, a novel C-17 and C-19 reducing ketoreductase, and *in vitro* reconstitution of an unusual C9-C14 reducing ARO/CYC from the arixanthomycin pathway.

CHAPTER 1: Introduction

1.1 Natural Products: A Brief History

Natural products are organic molecules biosynthesized from living organisms that have long been utilized by humans. It is thought that Neanderthals 60,000 years ago were aware of plant-based therapeutics because of the presence of pollen deposits found from excavated prehistoric grave sites.[1] Ancient Mesopotamian texts dating near 2500 BC contained recipes for making extracts from *Papaver somniferum* for the treatment of pain, and the use of resin from *Commiphora myrrha* for its antiseptic properties. The *Ebers Papyrus* of ancient Egypt and the Chinese *Wu Shi Er Bing Fang*, are two thousand year old texts which outline prescriptions for hundreds of natural remedies, one of which involves the use aloe vera extracts for cosmetic purposes.[1, 2] The use of natural products is deeply entwined with the history of mankind, and has undoubtedly helped shape the quality of human life for millennia.

Prior to the advances in analytical and structural chemistry, natural products were typically isolated in the form of crude extracts obtained from plants, fungi or bacteria. The most infamous ancient use of natural product extracts involved the death of Socrates, who met his demise after consuming a hemlock concentrate that contained the hybrid alkaloid-polyketide toxin, coniine.[3] It was not until the end of the 19th century that individual natural products were isolated, and it was not until the 20th century that their individual chemical structures were determined to high resolution.[4] Before the mid 20th century, natural products chemists were limited to primitive calorimetric and spectroscopic techniques to obtain the chemical formulas of a given bioactive compound.[4] In addition, the individual biological molecules that were obtained were often of low purity either due to the presence of similar chemical isoforms or

undetectable containments from the extraction processes.[5] Given the challenges associated with natural product isolation, crude extracts isolated from plants, fungi, and bacteria were still in common use to screen for novel compounds with potent biological activity until the late 1980's.[6] With the advent of improved HPLC, NMR and mass spec technologies along with improved high throughput screening techniques (HTS), the use of natural product extracts for drug discovery was slowly abandoned.[5, 6] Further, with the advent of improved organic chemistry, total or combinatorial syntheses of natural products became possible, and these methods were desirable for several reasons. First, overall yields of the extracted molecules could be extremely low and was dependent on the organism from which they were derived.[7] If the organisms of interest were not readily available in ample amounts, this would limit the supply of a given therapeutic. Further, many important molecules were only biosynthesized by a particular organism under specific environmental conditions, and they required stress or necessary developmental cues for their biosynthesis.[8-10] Second, many of these natural products themselves were difficult to purify to homogeneity; therefore, it was difficult to assess the concentration or efficacy of a single bioactive molecule in a given extract.[11, 12] Lastly, natural products could be easily modified through various synthetic transformation to modulate a given compounds biological activity.[13] Combinatorial or total syntheses of natural products therefore represented a means to bypass the barriers of ancient natural products use; further paving the way for the mass distribution of natural products based drugs and is still in use today.[7]

While technological developments in analytical and organic synthetic chemistry that aided in natural product discovery and isolation spanned between the 19th to late 20th centuries, a conceptual understanding of how these important molecules were made by living cells remained elusive until the 1940s.[14] Up until the mid-20th century, the conceptual framework behind

natural product biogenesis was incoherent and mostly speculative. There were some notable theoretical natural products chemists who, given the lack of technology at their disposal, had a strong intuition about the biosynthetic logic that encompasses anabolic processes. A historical overview of the founding fathers of natural products research that laid the foundation for the understanding of secondary metabolite biosynthesis is not within the context of this thesis; however, John Norman Collie is an important exception. Collie was a successful chemist and mountaineer; through his work studying naphthalenoid and benzenoid materials he isolated from plants, he coined the term polyketide in 1907 to describe natural products with multiple ketene groups (later known as ketide groups).[3, 15] Unfortunately, he proceeded to erroneously propose that the lactone of tetracetic acid was dehydroacetic acid, which was later proved to be incorrect [15] and perhaps as a consequence, most of his experimental findings went unnoticed for decades. Regardless of Collie's faults as a scientist, the term polyketide, is still used to this day to describe secondary metabolites that are derived from the decarboxylative condensation of malonyl-CoA.[3, 16]

The dawn of stable radiolabeled isotopes in the mid-1940s gave rise to a new era of natural products research, because chemists were equipped with a means to experimentally validate biosynthetic hypotheses.[17] Fueled by the discovery of penicillin two decades earlier, a strong basis for the investigation of the natural product biogenesis was established by 1965, through the implementation of feeding experiments that utilized simple radiolabeled organic molecules.[3, 14] The fact that most metabolites were derived from acetate or ammonia as precursors soon became common knowledge.[18, 19] In addition, microbial knock-out mutants were also useful in the understanding of the sequential transformative steps of a given substrate involved in a given metabolic pathway. It was near this time that the field of classical

enzymology blossomed, which provided additional methodologies such cell free extracts and enzyme assays that aided in the *in vitro* identification of individual enzymes involved in secondary metabolite formation.[20] In parallel, advances in molecular biology and microbiology, aided in the mass production of desired biologically active compounds *in vivo*. [14] By the mid-1980s, techniques for the heterologous expression of the genes responsible for the biosynthesis of a given secondary metabolite in model microbes, such as *Streptomyces*, were developed.[21] Improved culturing techniques and fermentation capabilities additionally aided in augmented *in vivo* production of microbial natural products.[22] Over the course of about 50 years, (1945-1995) a clear understanding of the rules and players in secondary metabolite biogenesis came to light.[18]

Contemporary molecular biology, biochemistry, and microbiology has since paved the road for the future engineering of “mixed-and-matched” natural product pathways, leading to the creation of “unnatural” natural products through a combination of different biosynthetic genes.[23] The development of “omics” that encompasses genomics, transcriptomics, and metabolomics has greatly facilitated the discovery of new natural products from uncharacterized transcriptionally “silent” biosynthetic pathways.[24] The future of natural products research is bright as there is a large expanse of uncharted biosynthetic territory left to be explored.

1.2 Natural Products: Primary and Secondary Metabolism

Primary metabolites encompass a broad class of molecules that are mandatory for life.[18, 25] Nucleic acids, amino acids, fatty acids, polysaccharides, and co-factor precursors are examples of primary metabolites that are required for the survival of a cell.[25] When food sources are ingested, the extracted primary metabolites are broken down into simple carbon and

nitrogen constituents that can be utilized to generate energy or, depending on the needs of the cell, recycled.[18, 19] Polysaccharides and fats, for example, are ultimately broken down into two carbon (acetyl-CoA) building blocks, which can be used to obtain energy via the tri-carboxylic acid (TCA) cycle or for anabolic processes such as the replenishment of TCA cycle intermediates.[19] Protein obtained from food sources is broken down to yield free amino acids and acetyl-CoA that can be rerouted to biosynthesize nucleotide and or other proteins required for life (Figure 1.1A).[18]

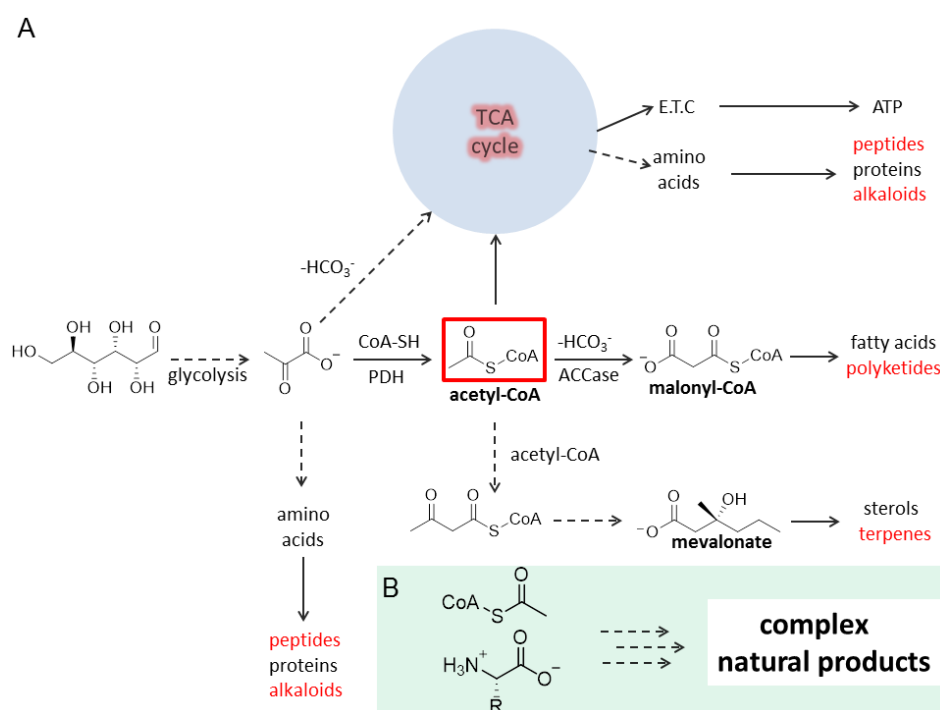


Figure 1.1. Acetyl-CoA is central to both primary and secondary metabolic pathways.

(A) Glucose is ultimately broken down to yield acetyl-CoA, which can be routed to both primary and secondary metabolic pathways. Depending on the needs of the cell, acetyl-CoA can be fed directly into the TCA cycle to yield energy (ATP and NADH) or enzymatically transformed to generate complex molecules (fatty acids, polyketides, terpenes, and alkaloids). Amino acids are biosynthesized from the recycled components of primary metabolic pathways. Like acetyl-CoA, free amino acids can be used to generate diverse secondary metabolites (nonribosomal peptides and alkaloids). (B) Acetyl-CoA and free amino acids are the precursors to many complex natural products.

Natural products are secondary metabolites derived from the recycled components of primary catabolic pathways.[26] The biosynthetic pathways of secondary metabolites utilize the same simple carbon and nitrogen building blocks derived from primary metabolic pathways such as acetyl-CoA and free amino acids, and are typically biosynthesized using elaborate multi enzyme systems.[3, 18, 19, 26, 27] Unlike primary metabolites, secondary metabolites are not required for cell survival, but they do provide environmental advantages to the organism. Antibiotics and toxins produced by bacteria or fungi, for example, are secondary metabolites that are illustrative of a chemical means of protecting an ecological niche.[28] There are many different types of secondary metabolites; however, most can be classified based on their chemical structures and their mode of biosynthesis into the following categories: polyketides, nonribosomal peptides, alkaloids, terpenes, isoprenoids, and complex lipids (Figure 1.2).[3, 26, 29] Secondary metabolites can be found in plants, bacteria, and fungi.[27, 28, 30, 31] In the case of bacterial and fungal secondary metabolism, the majority of the genes encoding the enzymatic machinery necessary for their biosynthesis are typically clustered together.[32] Biosynthetic genes in higher organisms such as plants are often not directly clustered together and, due to the presence of introns, can be arduous to decipher.[33] Nevertheless, a common biosynthetic theme for both primary and secondary metabolism is the utilization of free amino acids and/or acetyl-CoA as building blocks that ultimately are assembled into complex molecules (Figure 1.1B).[14, 26, 34]

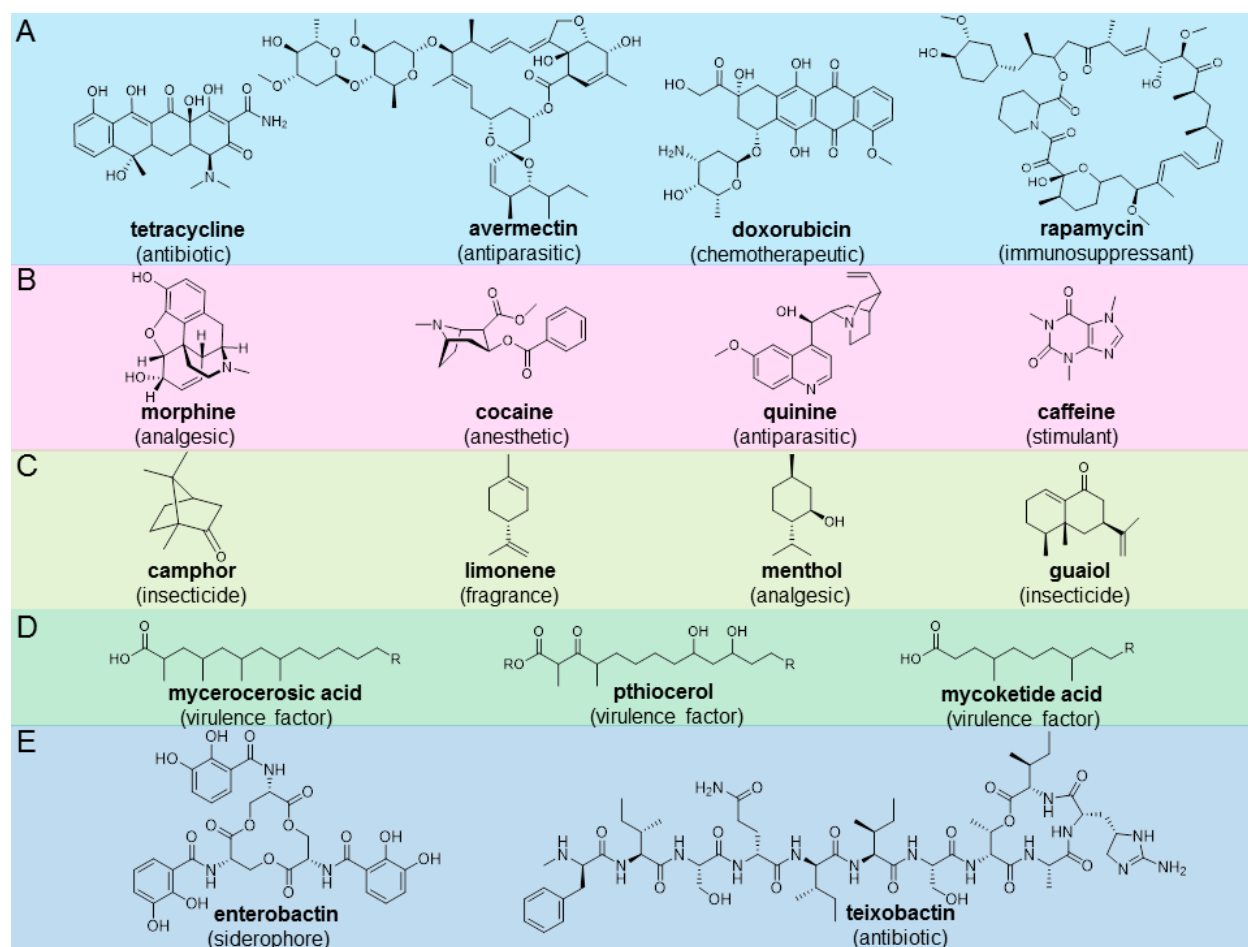


Figure 1.2. Different classes of natural products.

(A) Polyketides. (B) Alkaloids. (C) Terpenoids (D) Complex fatty acids. (E) Nonribosomal peptides.

1.3 Primary Metabolism: Fatty Acid Biosynthesis

Fatty acids are primary metabolites and play many roles in the maintenance and survival of the cell.[19] Fatty acids serve as sources of storage for excess glycolytic intermediates, signaling platforms for intracellular signal transduction, mammalian immune response regulators, and are the primary components of phospholipids that constitute the plasma membrane of cells.[35-37] Fatty acids are biosynthesized by enzymes known as fatty acid

synthases (FAS), which utilize activated acyl-CoA thioesters as their primary substrates. FASs can generate both branched and linear fatty acids of various carbon lengths, with the 16 carbon palmitic acid being the favored product.[35] There are two distinct types of FAS: types I and II. Type I FASs are composed of multiple catalytic domains tethered together in a single polypeptide chain and are commonly found in mammals and fungi (Figure 1.3A).[38] Type II FASs are composed of multiple dissociative enzymes which act *in trans* to produce fatty acids; they are found in bacteria and plants (Figure 1.3B).[39] Both type I and type II FASs converge on their use of a conserved set of enzymatic activities to iteratively grow and process fatty acid chains.[29, 35, 39, 40]

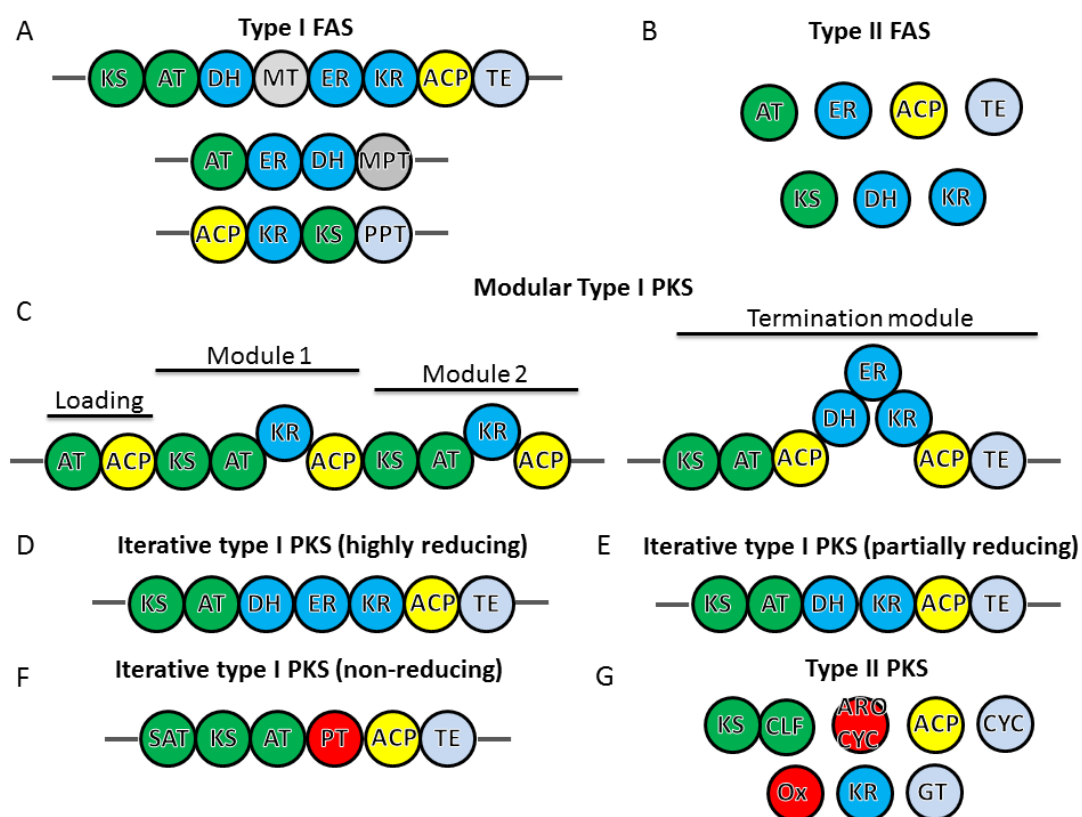


Figure 1.3. Domain architecture of various polyketide and fatty acid synthases.

(A) Type I mammalian FAS (top) and Yeast FAS (bottom). (B) Type II FAS (C) Modular type I PKS. (D) Iterative highly reducing type I PKS (HR-PKS). (E) Iterative partially reducing type I PKS (PR-PKS). (F) Iterative non-reducing PKS (NR-PKS). (G) Type II PKS.

The domains or enzymes that constitute an active FAS are highly conserved; they include: acyl carrier protein (ACP), ketosynthase (KS), malonyl-CoA ACP transacylase (MCAT or MAT), 3-ketoacyl-ACP reductase (KR), dehydratase (DH), enoyl reductase (ER) and thioesterase (TE).[40] The ACP is responsible for the transport of acyl intermediates, and it interacts with all downstream enzymes throughout the entire biosynthetic process. *In vivo*, ACPs are first translated in their inactive *apo* form, and must be post-translationally modified with a 4'phosphopantetheine (PPT) group, derived from CoA, to become active (*holo*-ACP). The enzymes that convert *apo*-ACP to *holo*-ACP are known as phosphopantetheinyl transferases (PPTase).[41] To initiate FAS biosynthesis, the *holo*-ACP is first loaded with an acyl starter unit by the priming KS (KAS III). The MCAT then catalyzes the transfer of a malonate group to the thiol group of the PPT moiety of ACP for subsequent chain extensions with an acylated KS. The KS next catalyzes the condensation of malonyl-ACP with the acyl group that is covalently linked to the KS's active site cysteine, leading to the extension of a fatty acyl intermediate by 2 carbon units. The ketoreductase (KR) is an NADPH dependent enzyme that proceeds to reduce β -ketoacyl-ACP to β -hydroxyacyl-ACP. The dehydratase (DH) is responsible for the generation of unsaturated fatty acids via the dehydration of the hydroxyl group of β -hydroxyacyl-ACP. The resulting alkene is reduced by the NADPH-dependent enoyl reductase (ER), thereby producing completely saturated acyl intermediates. Repetitive rounds of the above catalytic cycle occur until the acyl intermediate has reached a mature chain length (Figure 1.4).[35] The full-length acyl-ACP is hydrolyzed by the TE to generate a free fatty acid, or the acyl group is directly coupled to carrier molecules such as carnitine by auxiliary acyl transferases.[19, 39]

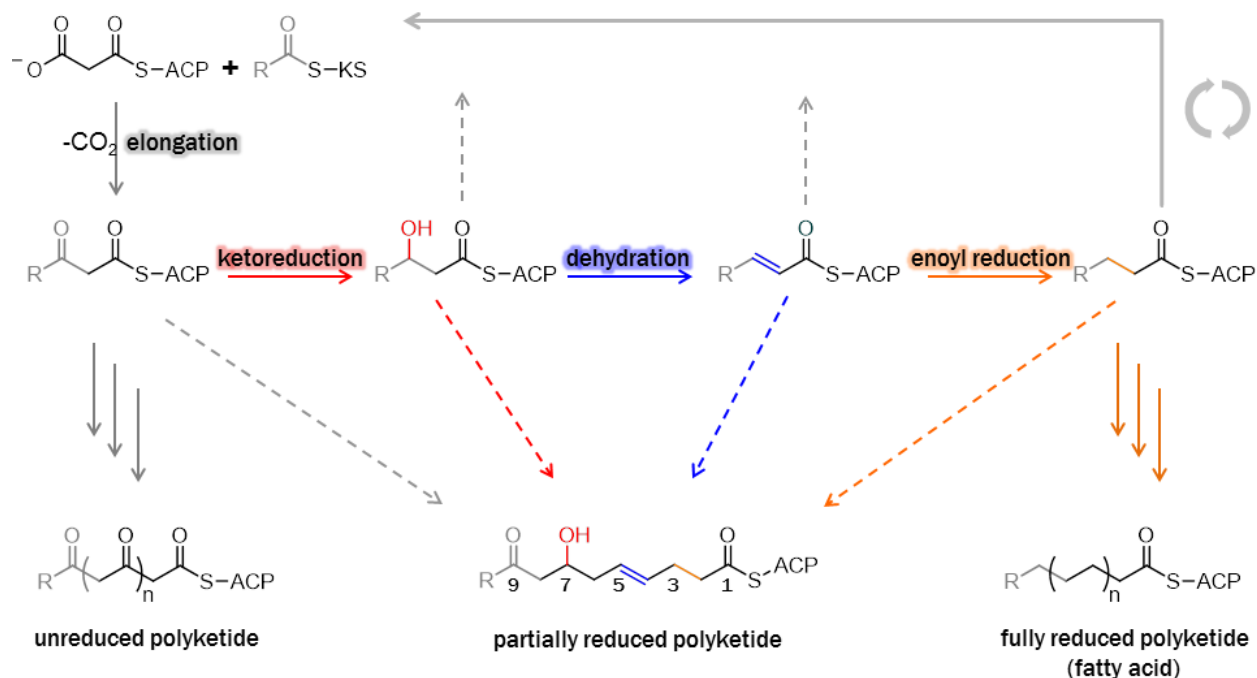


Figure 1.4. Chain elongation and processing steps of polyketide and fatty acid biosynthesis.

Fatty acids and polyketides are both assembled through decarboxylative condensations of malonyl-ACP with KS to afford an unacylated ACP-linked ketide intermediates. FASs utilize KR, DH and or ER domains to reduce the β -ketone group of a growing fatty acid intermediate to an unsaturated or saturated α - β bond prior to a subsequent round of chain extension resulting in the production of fatty acids. NR-PKSs do not contain DH, KR or ER domains, and produce long poly- β -ketone intermediates that are ultimately cyclized to afford aromatic natural products. Partially reducing PKSs contain KR and or DHs that are used to generate polyketides with α - β unsaturated bonds or β -hydroxyl groups. Highly reducing PKSs are very similar to FASs and produce polyketides that contain saturated α - β bonds.

1.4 Secondary Metabolism: Polyketide Synthases

Overall Function

Polyketide natural products are biosynthesized by polyketide synthases (PKS) which, similar to FASs, are complex macromolecular machines.[42] Polyketides are widely known for their broad medically relevant biological activities and serve as antibiotics (tetracycline and erythromycin) [43, 44], immunosuppressants (rapamycin) [45], chemotherapeutics (doxorubicin

and stambomycin) [10, 30], antiemetics [46] (tetrahydrocannabinol), and antiparasitics (ivermectin) [47]. The 2015 prize for medicine and physiology was awarded to research groups for the discovery of ivermectin, highlighting the significance of polyketide based therapeutics.[48] Humanitarians and animal lovers alike should deeply appreciate the discovery of ivermectin because it paved the way for the synthetic gold standard for the treatment of nematode-based parasitic diseases, ivermectin.[49] Given their chemical diversity and inherent therapeutic potential, PKS natural products have garnered much attention from the scientific community over the past three decades.

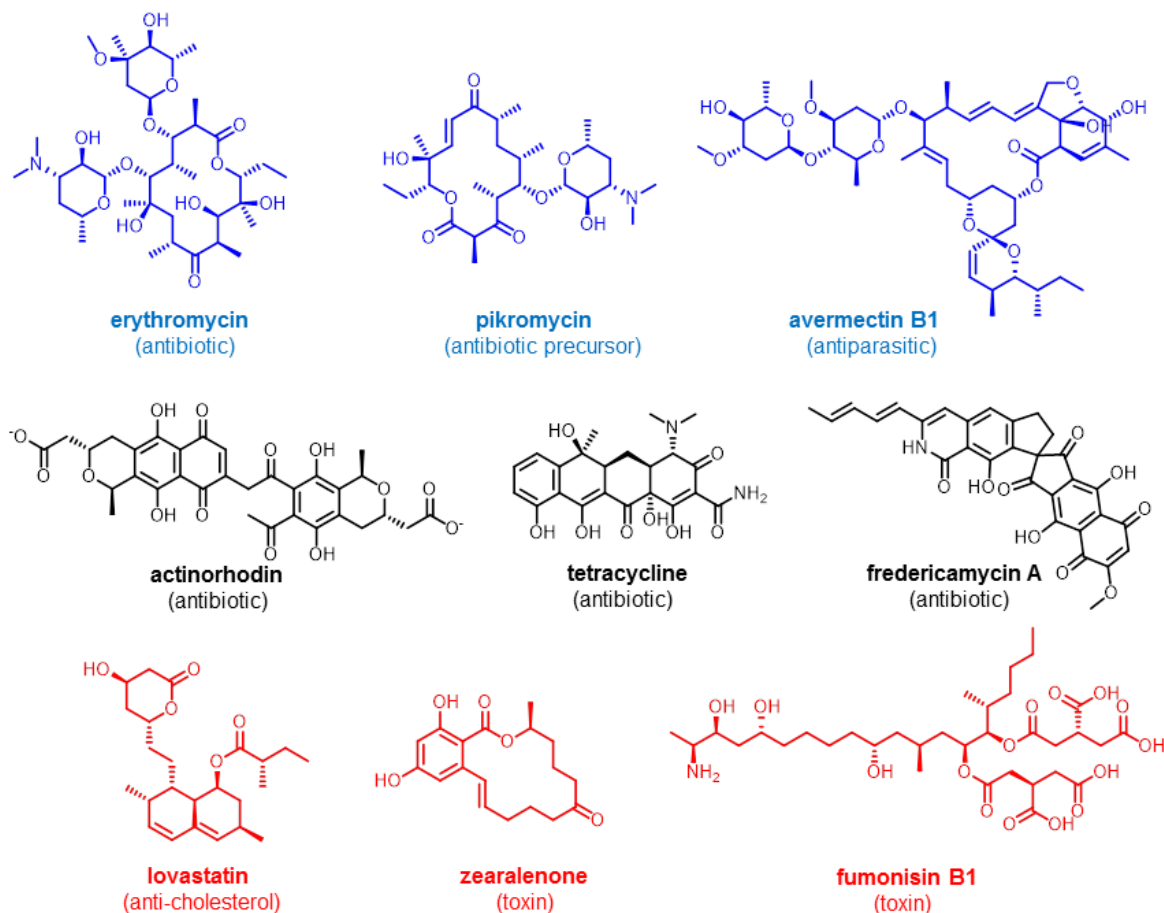


Figure 1.5. A collection of polyketide natural products produced by Modular type I PKSs, bacterial type II PKSs, or fungal type I iterative PKSs.
Modular type I PKSs (blue), bacterial type II PKSs (black), fungal type I iterative PKSs (red).

PKSs share a high degree of structural and sequence similarity to type I and type II FASs, and they utilize a very similar catalytic approach in the biosynthesis of polyketide natural products.[16] Most PKSs, for example, harbor homologous enzymes of FASs that include: ACP, KS, AT, DH, KR, ER and TE [30, 42, 43]. PKSs further resemble FASs in the use of malonyl-CoA extender units to build the polyketide backbone. However, unlike fatty acids, polyketide natural products are much more chemically diverse, because PKSs are not solely restricted to the production of linear or branched chained fatty acids.[3] PKSs can produce simple monocyclic aromatic compounds such 6-methylsalicylic acid, heterocyclic aromatic scaffolds such as the anthracycline skeleton of doxorubicin, or large macrolides such as the 51 membered aglycone core of stambomycin.[10, 16, 50] An assortment of important polyketides is shown in Figure 1.5.

PKSs can be divided into three main types: type I, type II, and type III.[42] Type I PKSs are often referred to as “megasyntases” in which a single protein encompasses all or most of the catalytic domains. Type I PKSs can be further categorized based upon the mode in which they biosynthesize polyketide natural products. Modular type I PKSs biosynthesize polyketides in an assembly line like fashion using multiple catalytic modules, whereas iterative type I PKS reuse the same set of domains of only one module during polyketide biosynthesis (Figure 1.3C).[30, 43] Type II PKSs are similar to type II FASs in that they are composed of discrete dissociative enzymes that act in *trans* to produce polyketides in an iterative fashion.[50] Lastly, type III PKSs are simple KS homodimers that act in an ACP-independent manner to biosynthesize polyketide natural products.[51] Regardless of the system, all PKSs utilize activated CoA thioesters derived from short chain fatty acids to generate a large array of complex biologically active molecules.

Fungal Type I Iterative PKSs

Fungal iterative type I PKSs are responsible for the biosynthesis of an assortment of therapeutically and agriculturally relevant molecules. Lovastatin, for example, is biosynthesized from a type I iterative PKS and has garnered much fame for its commercial success as a highly profitable cholesterol-lowering drug.[52] Aflatoxin B1, fumonisin B1 and zearalenone, on the other hand, are also biosynthesized from type I PKSs as agricultural toxins: exposure to these can cause hepatocellular carcinomas or other cancers.[53-55] Type iterative I PKSs generate natural products from repetitive catalytic cycles that involve the reuse of several domains in a single module.[30]

Type I iterative PKSs are subdivided by the degree in which the poly- β -ketone frame is reductively tailored. Partially reducing PKSs, for example, have a KR domain and can produce polyketides with hydroxyl groups via the reduction of various β -ketone groups (Figure 1.3E). Highly-reducing PKSs (HR-PKSs) can reduce poly- β ketone intermediates to a saturated carbon-carbon bond due to the presence of KR, DH and a *trans* or *cis* acting ER (Figure 1.3D).[56, 57] Examples include lovastatin.[52] Non-reducing PKSs (NR-PKSs) do not contain KRs or ERs and therefore produce long poly- β ketone intermediates that are cyclized to yield heterocyclic aromatic compounds.[30] Examples of NR-PKS derived natural products include bikaverin, aflatoxin B1, and cercosporin (Figure 1.3D).[30, 58, 59] Over the past decades, efforts towards the understanding of type I PKS biosynthesis has been vigorously pursued through the *in vitro* and *in vivo* reconstitution of the various enzymatic components associated with fungal PKSs.[60-62]

Fungal Type I NR-PKSs

Fungal NR-PKSs have a conserved order and domain architecture that includes the following domains: SAT, KS, AT, PT, ACP, and TE/CYC.[56] The Townsend lab at Johns Hopkins Univ. has pioneered the “deconstruction” of fungal NR-PKSs to decipher the roles of the individual domains involved in NR-PKS mediated biosynthesis.[61] Starting with the NR-PKS component involved in aflatoxin B1 biosynthesis, mono-, di-, and tri-domain constructs were generated and their individual activities were dissected.[63] The starter unit acyl transferase (SAT) domain was found to be essential for the priming of *holo*-ACP with an acyl starter unit to initiate biosynthesis.[64] *In vitro* transacylation assays have demonstrated that SATs have a certain degree of substrate tolerance and can act on a large panel of synthetic acyl-CoAs.[65, 66] SATs have also been shown to be important for efficient biosynthesis. Assays performed in the absence of the SAT domain lead to poor product yields due to poor *holo*-ACP priming, a phenomenon known as the “starter unit effect”.[67] The crystal structure of the SAT domain from the chaetomugilin A biosynthetic cluster has recently been solved (Figure 1.6).[31] The SAT was revealed to be a di domain enzyme that consists of α/β hydrolase core juxtaposed to a C-terminal ferredoxin-like subdomain. In addition, the SAT was found to share a high degree of structural similarity to previously solved FAS MCATs and type I PKS loading ATs.[31] Crystal structures of various SAT domains with differing substrate specificities will be useful for understanding the molecular basis for starter unit selection and incorporation in NR-PKS biosynthesis. The KS and AT domains of NR-PKSs retain their conserved roles in the extension of an unreduced polyketide chain in concert with the ACP domain.[30, 56, 62] The product template domain (PT) was a previously uncharacterized PKS domain and was shown to be important for maintaining proper chain length and the predisposition of a set cyclization pattern to a given natural product.

Different PT homologues can produce distinct cyclization patterns such as: C2-C7, C3-C8, C4-C9, and C6-C11.[68] The first crystal structure of the PT domain was solved by the Tsai group at the University of California, Irvine in 2009.[30] The structure revealed a dimeric “double hot dog in a bun” fold characteristics of DH domains, suggesting the two are evolutionarily related.[30] The TE/CYC domain was shown to be essential for efficient offloading of a mature polyketide intermediate through a final Claisen-type cyclization reaction.[69] The crystal structure of the TE/CYC involved in the biosynthesis of aflatoxin B1 was solved in 2010, revealing that TE/CYCs strongly resemble TEs from FAS and type I modular PKS pathways.[69] Little is known about how these domains recruit the ACP in the context of iterative biosynthesis. As of now, only the structures of the three domains associated with NR-PKSs have been solved (PT, TE/CYC, and SAT domains) leaving a bright future for those pursuing NR-PKS crystal structures.

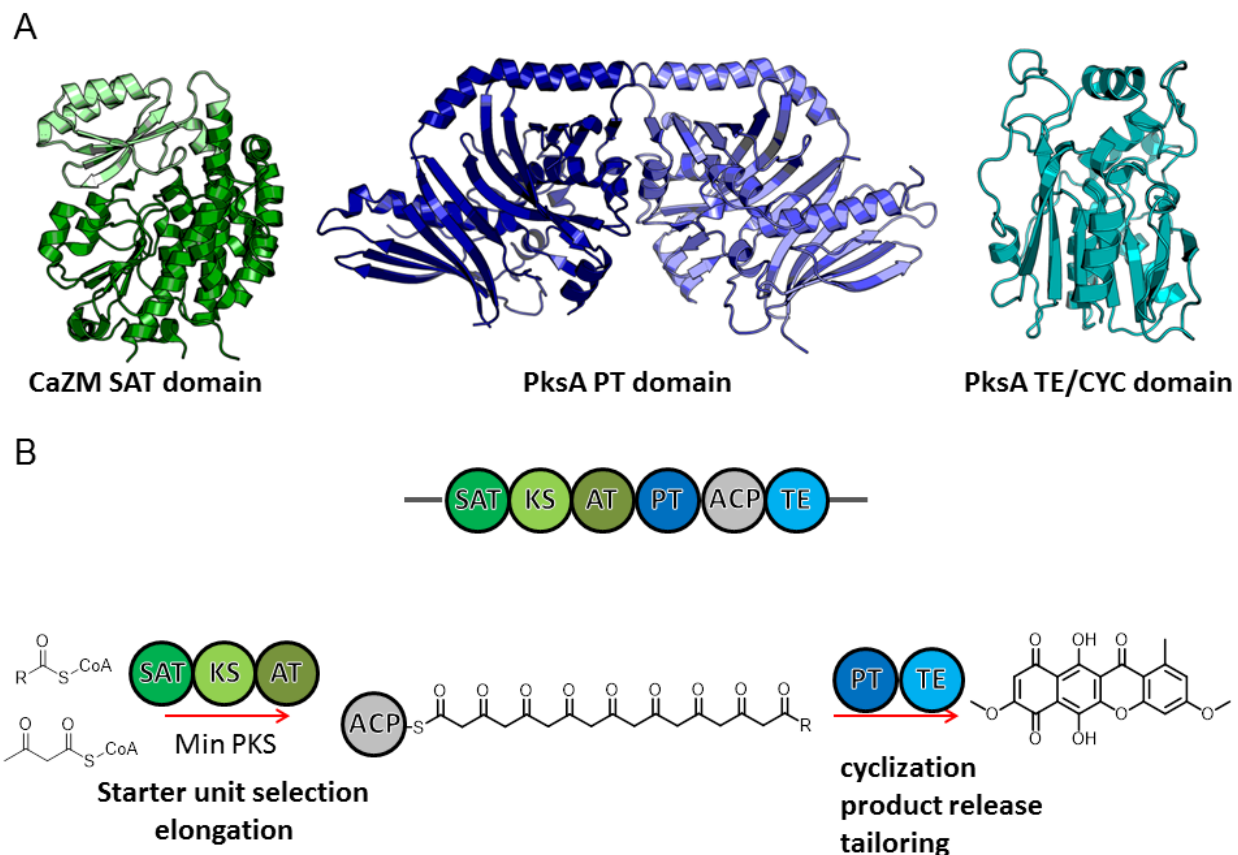


Figure 1.6. Biosynthesis of fungal aromatic polyketides by non-reducing PKSs.

(A) The crystal structures of three NR-PKS domains: CaZM SAT domain (left), PksA PT domain (middle), and PksA TE/CYC domain (right). (B) During NR-PKS biosynthesis, a minimal PKS consisting of the SAT, KS, and MAT domains generate a long poly- β -ketone chain, which is subsequently cyclized and released from the megasynthase through the concerted efforts of the PT and TE/CYC domains.

Type I Modular PKSs

Type I modular PKSs are responsible for the biosynthesis of myriad linear and macrocyclic compounds. The aglycone moieties of erythromycin, avermectin and enacyloxins are the byproducts of type I modular PKS biosynthesis.[9, 47, 70] Type I modular PKSs can be dissected into two groups: *cis* modular PKSs and *trans*-AT modular PKSs.[71] *Cis* modular PKSs can be described as megasynthases in which every catalytic domain is packaged in a single

polypeptide chain that contains multiple modules.[43] In contrast, the *trans*-AT modular often lack AT domains and instead have separate genes that encode functional AT(s) that act in *trans* to supply the building blocks.[71] Both *cis* and *trans*-AT type I modular PKSs biosynthesize polyketides in a collinear manner.[43, 71]

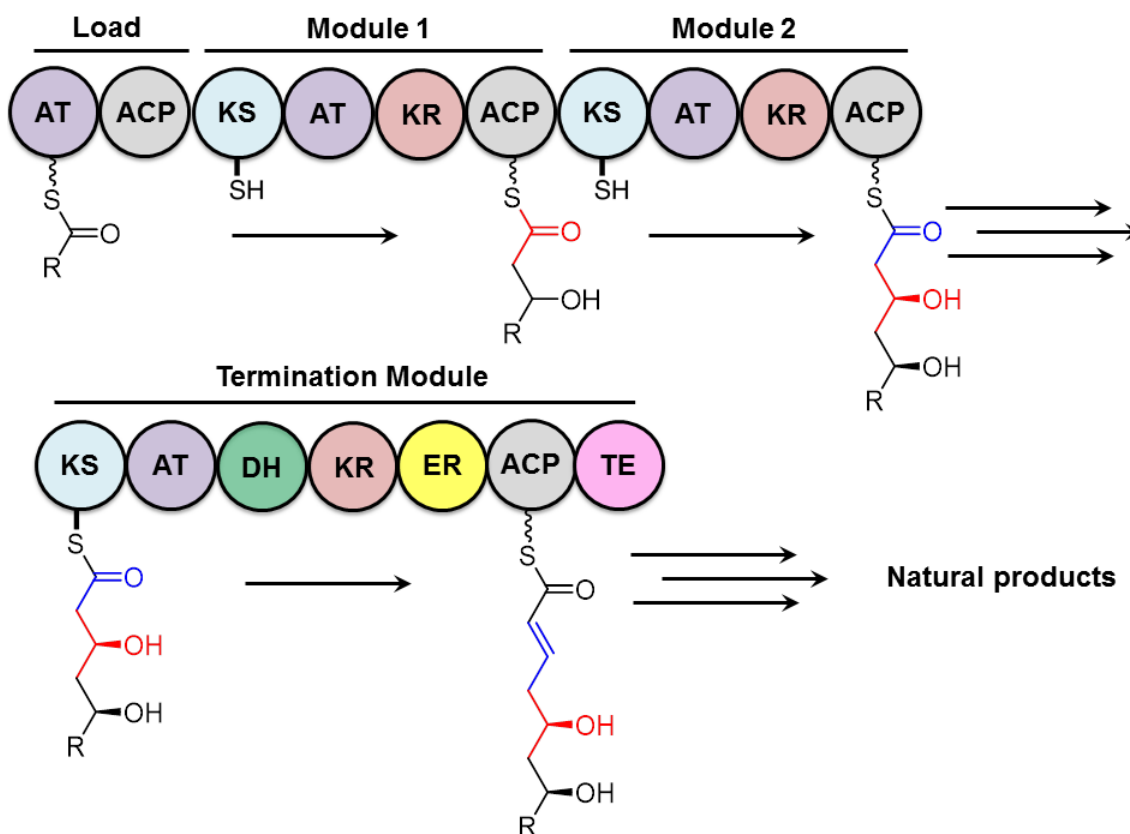


Figure 1.7. Schematic of type I modular PKS biosynthesis.

Modular polyketides are constructed in an assembly line fashion unlike the single module complexes responsible for the biosynthesis of fatty acids and aromatic polyketides.

In the modular PKS, each module contains a set of catalytic domains necessary to achieve one round of chain extension and processing (Figure 1.7).[16] To initiate biosynthesis, a loading module consisting of KS^Q (a decarboxylative, non-extending KS), AT and ACP domains is used to prime its ACP with an acyl starter unit. The newly formed acyl-ACP transfers the acyl

starter unit to the KS domain belonging to the next module, where it is extended by the concerted efforts of its ACP, KS, AT, and then processed by KR, DH, and ER if they are present in the module.[43] Once the acyl intermediate has undergone one round of chain extension and processing, it is transferred to the downstream module for an additional round of chain propagation and tailoring.[72] Eventually, the polyketide chain reaches its mature length and is transferred to a termination module containing the TE domain, where it is hydrolyzed from the assembly line.[16] Because each module effectively grows and tailors a specific portion of the natural product, modular type I PKS have been shown to biosynthesize their products in a linear order. In other words, the order and arrangement of both the extension and tailoring domains within a module dictates the chemical modifications that will take place throughout polyketide maturation.[71, 72]

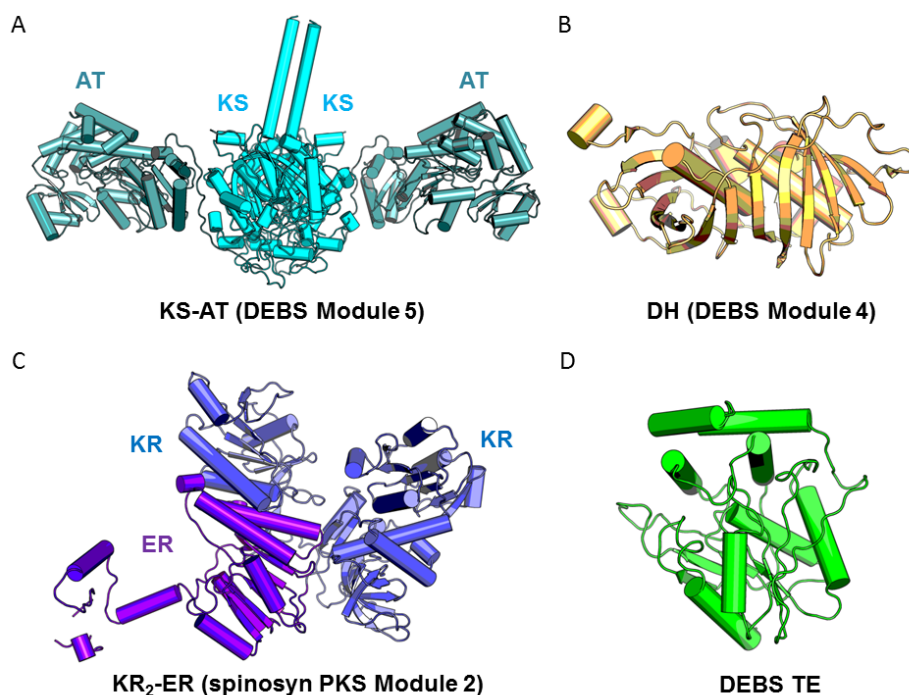


Figure 1.8. Structures of type I modular PKS chain extension and processing domains.

(A) The KS_AT di domain from DEBS module 5. (B) DH dimer from DEBS module 4. (C) KR₂-ER di-domain from the 2nd module of spinose PKS. (D) DEBS TE domain.

Modular PKSs have been studied extensively, and the protein structures of many enzymatic domains have been solved.[43] The first PKS structure was that of the TE of the termination module of DEBS (6-deoxyerythronolide B synthase).[73] Since then, many mono- and di-domain structures have been solved such as the KS_AT from DEBS module 5 and the KR₂_ER di-domain from the 2nd module of spinosyn PKS.[70, 74, 75] In most cases, the individual domains share a high degree of structural similarity to those associated with their type I and type II FASs relatives (Figure 1.8).[42, 43] More recently, the first cryo-EM structure of an intact module from the pikromycin PKS has been reported, which provided the first glimpse of complete type I modular PKS architecture.[76] Unlike mammalian and yeast type I FASs, the quaternary structure of the 6th module of pikromycin PKS does not resemble a gingerbread man or a barrel, but rather adopts the shape of a bat folding its wings.[39, 76, 77] The cryo-EM structure of the 5th module of pikromycin unveiled that these megasynthases are dimeric, where the KS homodimer forms the “head” of the bat, whereas the “wings” are composed of the AT and KR domains (Figure 1.9). The ACP “hands” are located at the bottoms of the wings (near the KR domain) and are positioned within the modules reaction chamber or localized near the AT domain. This domain arrangement allows for successful shuttling of ACP-tethered substrates to different enzymatic components that demarcate a single ACP chamber in the interior of module 5.

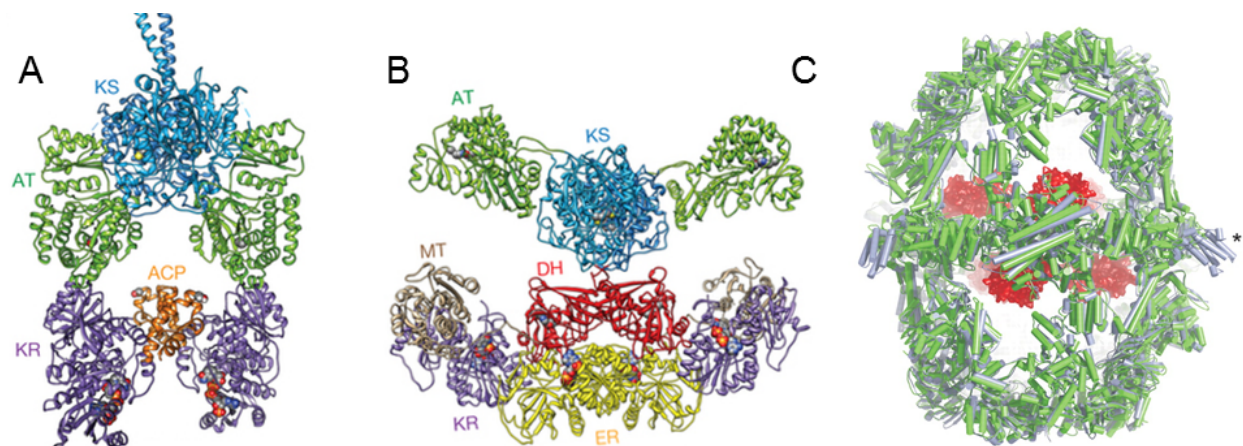


Figure 1.9. A structural comparison of FAS and PKS megasynthase structures.

(A) Cryo-EM structure of the 5th module of Pikromycin PKS (B) Crystal structure of mammalian FAS (C) The crystal structure of yeast FAS. Figure adapted from references 77 and 76.

Interestingly, the ACP was found to be primarily localized near either the KS, AT or KR domains in the megasynthase structure. Loading of the ACP with methylmalonyl-CoA (MM-CoA) yielded solutions that placed the MM-ACP near the KS-AT domains, suggesting that the various ACP positions obtained from the cryo-EM structure can be attributed to the substrate-enzyme interaction between a given enzyme domain with a specific acyl intermediate that is tethered to the ACP. The observation that ACP localization is dependent on the length of a tethered polyketide provides an explanation as to how the ACP recognizes the appropriate intra-module or inter-module domain at the correct time. In summary, the cryo-EM structure of the first intact type I PKS module demonstrates how the spatial arrangement of the different intra-modular PKS domains is necessary for high fidelity during biosynthesis.[76]

Type II PKSs

Type II PKSs, contrary to their type I relatives, are composed of several standalone enzymes which catalyze polyketide formation *in trans* (Figure 1.2G).[50] Type II PKSs are responsible for the production of a diverse set of heterocyclic aromatic compounds.[16] The anthracycline core of doxorubicin, the discoid skeleton of resistomycin, and the hexacyclic xanthone scaffold of the arixanthomycins are examples of chemical motifs that are generated by different type II PKSs.[78-80] The genes encoding individual enzymes involved in type II polyketide biosynthesis are localized in clusters and are highly conserved. The enzymes responsible for the formation of an unreduced polyketide chain are ACP and ketosynthase-chain length factor (KS-CLF also known as KS $_{\alpha}$ -KS $_{\beta}$). This set of enzymes is often referred to as a “minimal PKS” (Min PKS) due to their capacity to generate a full length, albeit untailored, poly- β -ketone intermediate in the presence of malonyl-CoA *in vivo* and *in vitro*. Whether or not an auxiliary MCAT is a mandatory requirement for Min PKS chain propagation is a contentious topic.[39] Because individual MCATs are absent in type II PKS gene clusters, it has been postulated they are not required for biosynthesis; however, others have argued that MCATs from FAS pathways could act as surrogates, which has been shown *in vitro*. [81] Synthetic type II *holo*-ACPs have also been shown to undergo self-malonation under high concentrations of malonyl-CoA, fueling additional controversy surrounding the components of a functional Min PKS.[82] That aside, KS-CLF is unique to the type II system. Both KS and CLF are structural and functional analogs of the KS domains found in FASs and type I PKSs. KS-CLF is a heterodimer, and both KS and CLF are composed of two structurally similar thiolase folds. The KS contains the conserved active site cysteine, histidine, and asparagine necessary for acylation and PKS chain propagation. On the contrary, the CLF, proposed to control polyketide chain-

length control, does not contain a conserved active site cysteine necessary for chain extension, but instead has an active site glutamine that is involved in the generation of acetyl-CoA starter units via decarboxylation of malonyl-ACP [83]. Additional tailoring enzymes such as KRs, aromatase/cyclases (ARO/CYC)s, cyclases (Cyc), and oxidases (Ox) are also found in type II PKS pathways (Figure 1.10). These enzymes are important for the proper folding and oxidative tailoring of a poly- β -ketone intermediate necessary to achieve a mature polyketide natural product.

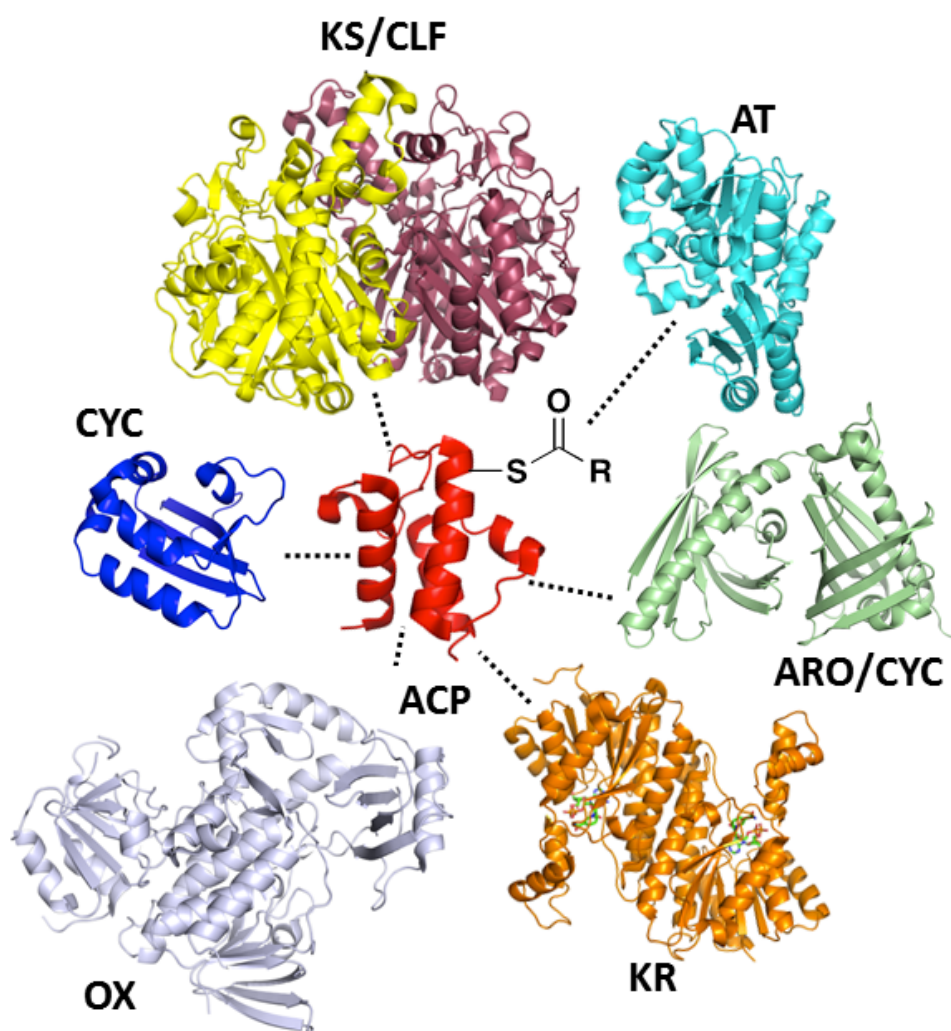


Figure 1.10. The crystal structures of the enzymes involved in type II PKS biosynthesis.

Similar to type I iterative fungal systems, type II systems can be categorized into reducing and non-reducing systems, depending on the presence or absence of a KR. In the reducing systems, the KR mediates the first ring cyclization of a mature poly- β -ketone chain between positions C7 and C12 prior to reduction of the carbonyl group at the C9 position.[84] The newly formed C9-hydroxyl group is then dehydrated by an ARO/CYC, leading to the aromatization of the first C7-C12 cyclized ring. Additional downstream secondary, tertiary, and quaternary ring cyclases further cyclize and aromatize the nascent polyketide chain for the establishment of the polyketide skeleton.[68, 85] Polyketides produced from type II reducing systems are characterized by the C7-C12 first ring cyclization motif, as exemplified by the actinorhodin and donurubicin biosynthetic pathways.[86-88]

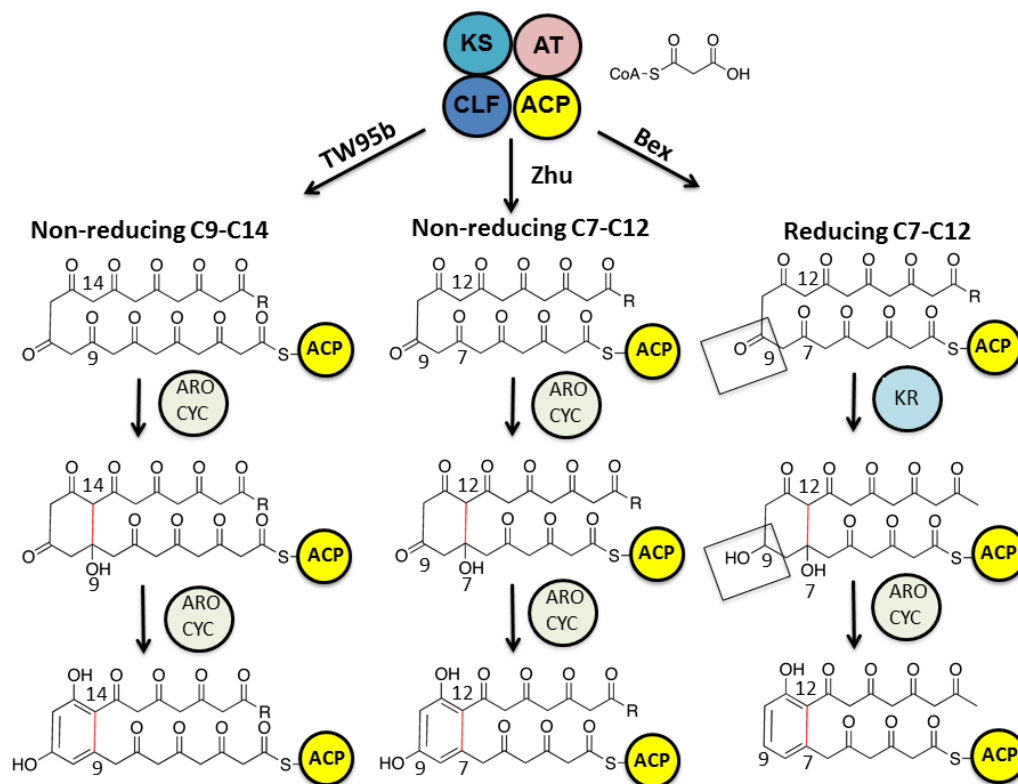


Figure 1.11. A comparison of non-reducing C9-C14 type II PKS pathways to C7-C12 reducing and non-reducing type II PKS pathways.

In non-reducing type II PKSs, a mature poly- β -ketone chain is first delivered to an ARO/CYC that catalyzes the first ring cyclization between positions C9 and C14, followed by second ring cyclization between positions C7 and C16 or, in the case of the R1128 biosynthetic pathway and certain di-domain ARO/CYCs, a single C7-C12 cyclization (Figure 1.11).[68, 89, 90] Similar to reducing type II PKSs, auxiliary CYCs catalyze the formation of the third and fourth rings of a given polyketide.[85] Tetracycline is an archetypal example of a type II polyketide produced from non-reducing PKSs that harbor a conserved C9-C14 cyclization pattern.[89, 91] It is worth noting that the ARO/CYCs associated with type II non-reducing pathways catalyze the formation of first and second rings, although the ZhuI ARO/CYC (only the first ring) and di-domain ARO/CYCs from non-reducing pathways (C7-C12 cyclization) are notable exceptions.[68, 90]

Type II PKS biosynthesis is an iterative process and can be dissected into three main steps: priming, extension and processing. During the priming step, a *holo*-ACP is first primed with an acyl starter unit. ACP priming can be achieved by a priming acyl transferase that transfers an acyl group from an acyl-CoA to the ACP or from the decarboxylation of malonyl-ACP by the KS-CLF. Next, an acyl-ACP transfers its acyl group to the active site cysteine of KS, setting the stage for polyketide chain extension. Once the KS-CLF is primed with an acyl starter unit, the concerted efforts of the ACP and KS-CLF lead to the formation of full length polyketone chain through repetitive decarboxylative condensations of malonyl-ACP with the acylated KS-CLF. Processing events post poly- β -keto chain propagation result in the formation of set ring systems, glycosylation and oxidation patterns associated with a fully furnished polyketide. How a fully processed polyketide is released from ACP remains an enigma. The canonical chain release enzyme for mega-synthases, the TE, is absent in type II PKS, suggesting

that thioester cleavage occurs through a non-enzymatic route. It has been speculated that the formation of an α -pyrone from a 3,5 diketo ester could be a chemically feasible means of chain release in type II PKSs.[16] This chemical mechanism for type II PKS product release; however, cannot explain how these systems avoid ill-timed spontaneous product hydrolysis from the ACP. Although type II PKSs have been extensively studied, there are still many unanswered questions with regard to the *modus operandi* of these *trans* acting enzymes.

1.5 Type II ARO/CYC: Discovery and Overall Function

The establishment of a set cyclization motif is an additional means of achieving polyketide chemical diversity. During type II PKS biosynthesis, ARO/CYCs facilitate the folding of linear poly- β -ketone intermediates to produce natural products with nondegenerate aromatic rings.[68, 88] ARO/CYCs were first functionally characterized in 1994 by Khosla and co-workers through heterologous expression of various actinorhodin biosynthetic genes in *S. coelicolor* CH999 (a strain that lacks the endogenous PKS gene clusters).[87] In the absence of the genes encoding ActKR, ActVII and ActVI, aberrant first dehydrations and second ring cyclization were observed. The addition of genes encoding ActKR, ActVII, and ActIV led to the accumulation of products with a dehydrated C9-hydroxyl group and a proper second ring register that are naturally present in actinorhodin. These findings suggested the ActVII and ActIV encode the aromatase and cyclase, respectively.[87] After more genetic studies, the enzymes that were involved in first ring cyclization and/or aromatization were renamed ARO/CYCs, whereas the enzymes involved in 2nd, 3rd, or 4th ring cyclization events were referred to as cyclases.[68, 89]

To our current knowledge, all ARO/CYCs cyclize and/or aromatize long polyketide chains that range from 16-26 carbons in length to produce C7-12 or C9-C14 first ring cyclization

patterns.[8, 62, 68, 89-91] Besides their inherent cyclase and aromatase activities, several ARO/CYCs from different biosynthetic pathways, such as HedE and WhiE ARO/CYCs, have been implicated in PKS chain-length control.[92, 93] *In vivo* assays containing minimal PKSs from the hedamycin and TW95b biosynthetic pathways led to the production of a wide variety of different shunt products with differing chain lengths and cyclization patterns.[92, 93] Inclusions of the cognate ARO/CYC (HedE and WhiE, respectively) *in vivo*, however, lead to the establishment of uniformly cyclized products with fixed chain lengths.[92, 93] These findings implied that certain ARO/CYCs such as WhiE and HedE may form a complex with their endogenous minimal PKSs (ACP and KS-CLF), thereby aiding in the stability of a nascent ACP-linked polyketide chain. ARO/CYCs have a certain degree of chain length tolerance and are capable of biasing the cyclization profile of PKSs with different chain lengths.[8, 62, 68, 90] Incubation of a C9-C14 or C7-C12 ARO/CYCs with Min PKSs that supply octaketide (actinorhodin) or nonaketide (Pks4) as ARO/CYC substrates led to the favoring of shunt products with C7-C12 or C9-C14 cyclization patterns.[62] ARO/CYCs can be stratified based upon the cyclization pattern that they promote, the requirement for a partially reduced poly- β -ketone chain, and the number of domains that constitute a fully functional enzyme. It is important to mention that ARO/CYCs that are affiliated with reducing pathways are presented pre-cyclized substrates and have “dehydratase like” activity that is required for first ring aromatization. In contrast, the ARO/CYCs of non-reducing PKSs do not require first ring dehydration and can be described as polyketide “chaperones” whose substrate binding pockets act like molds to guide a linear poly- β -ketone intermediate into the proper orientation to achieve a specific cyclization pattern as discussed below.

Mono-Domain ARO/CYCs: Polyketide Chaperones

C9-C14 ARO/CYCs are responsible for the establishment of a C9-C14 and C7-C16 di ring system through intramolecular aldol cyclization reactions. To date, all structurally and biochemically characterized C9-C14 ARO/CYCs are mono-domain enzymes that include WhiE and TcmN ARO/CYCs.[8, 89] The only exception is ZhuI, a mono-domain ARO/CYC from the R1128 biosynthetic pathway, which produces C7-C12 cyclized products and does not influence the regiospecificity of the second ring.[90]

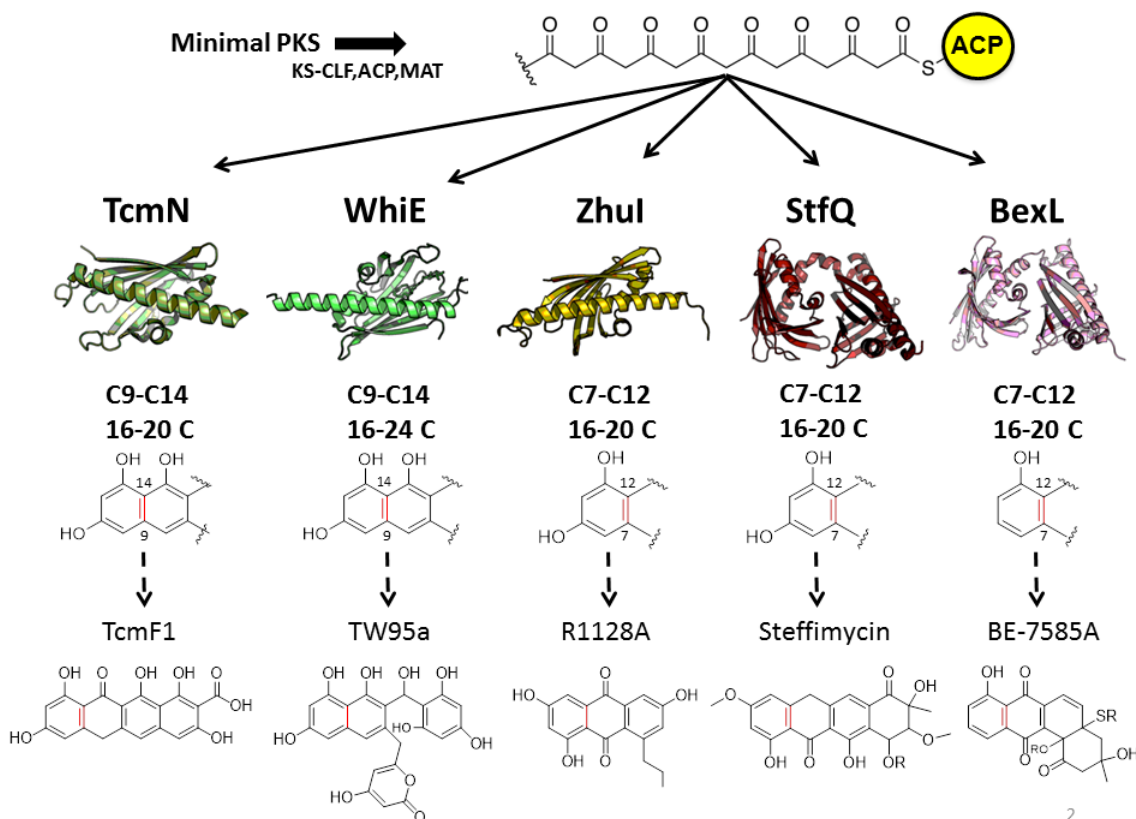


Figure 1.12. A schematic diagram of the different ARO/CYCs types.

The mono-domain ARO/CYCs TcmN and WhiE accept non-reduced polyketide intermediates to yield aromatic compounds that harbor C9-C14 cyclization patterns. The mono-domain ARO/CYC ZhuI and di-domain ARO/CYC StfQ cyclize non-reduced polyketides to produce aromatic products with C7-C12 cyclization patterns. The di-domain ARO/CYC BexL is responsible for the aromatization of a C9-reduced and C7-C12 cyclized substrate via the dehydration of the C9 hydroxyl group.

The structures of TcmN, WhiE and ZhuI ARO/CYCS have been solved, shedding insight for the molecular basis of polyketide cyclization and aromatization.[91] All three mono-domain ARO/CYCs belong to the Bet V1-like protein superfamily and exhibit a characteristic helix-grip-fold.[8, 89, 90] These ARO/CYCs are comprised of a long central α -helix adjacent to a seven stranded anti parallel β -sheet sandwiched between a shorter distal helix. The crystal structures of these enzymes revealed a large active site pocket located between the central α -helix and the interior β -sheet of the ARO/CYC (Figure 1.12). *In silico* docking simulations with TcmN, ZhuI, and WHiE with their putative decaketide or dodecaketide substrates revealed that the interior active site pocket can easily accommodate both linear and partially cyclized poly- β -ketone intermediates. Further, the size and dimensions of the substrate binding pockets were found to correlate with the natural length of their polyketide substrates of a given ARO/CYC. For example, WhiE ARO/CYC cyclizes a dodecaketide with C9-C14 and C7-C16 first and second ring cyclizations; it has the largest substrate binding pocket with the highest abundance of smaller amino acid residues in its active site.[8] In contrast, ZhuI, which acts on a much smaller nonaketide substrate and promotes the cyclization of a single ring, has the smallest pocket of the mono-domain ARO/CYCs due to the high abundance of bulky amino acid residues lining the interior of its active site.[90] The information gleaned from individual crystal structures and docking simulations aided in the identification of key active site residues essential for ARO/CYC catalysis. For the structurally similar C9-C14 ARO/CYCs such as TcmN and WhiE, Y35, R69 and R82 were hypothesized to be involved in substrate orientation and possible acid-base catalysis.[8, 89] Key residues identified for the C7-C12 cyclase activity of ZhuI were R66, H109, and D146, which were proposed to participate in acid-base chemistry or stabilization of the enolate intermediates for intramolecular aldol condensation of a mature polyketide chain.[90]

Superimposition of the C9-C14 ARO/CYC crystal structures with that of ZhuI revealed that C α atoms of R66 and R69 are structurally conserved and may serve as crucial anchoring points during cyclization. The guanidinium moiety of R66/69 however, was found to be shifted 2.8 Å toward the top of the active site pocket in the ZhuI ARO/CYC crystal structure with respect to TcmN and WhiE ARO/CYCs. Further, H109 and D146 of ZhuI ARO/CYC are localized in distinct positions to that of R82 and Y35 of WhiE and TcmN ARO/CYCs, suggesting that the arrangement of the residues within the active site pocket contributes to the different cyclization patterns produced by C9-C14 and C7-C12 mono-domain ARO/CYCs.[8, 89, 90] Chapter 2 will present the first polyketide mimetic bound crystal structures of WhiE ARO/CYC to further elucidate the chain length control and cyclization specificity of the ARO/CYCs.

Di-Domain ARO/CYCs: StfQ and BexL

The di-domain ARO/CYCs comprise the second major class of ARO/CYCs, and are characterized by their production of first ring C7-C12 cyclization patterns in type II polyketides. However, unlike their mono-domain relatives, di-domain ARO/CYCs can be found in both reducing and non-reducing type II PKS pathways. Di-domain ARO/CYCs from non-reducing pathways are functionally similar to mono-domain ARO/CYCs in that they act more like chaperones than dehydratases. Mithramycin and steffimycin are examples of natural products from non-reducing pathways and are characterized by a C7-C12 first cyclization pattern and an intact hydroxyl group in position C9.[91, 94] The di-domain ARO/CYCs, MtmQ and StfQ, are required for the establishment of the C7-C12 first ring cyclization motif of mithramycin and steffimycin, respectively.[88, 94] Reducing pathways, on the other hand, have a set requirement of dehydration at the C9 hydroxyl group to achieve first ring aromatization (Figure 1.11).[90, 91]

These enzymes harbor a catalytic His that is proposed to behave in a similar manner as the catalytic histidine of known dehydratases.[90, 91] Examples of natural products produced from reducing pathways include actinorhodin and BE-7585A.[91, 95]

Recently, the crystal structures of two di-domain ARO/CYCs, StfQ and BexL, from non-reducing and reducing type II PKS pathways have been determined (Figure 1.12).[91] The structures of these enzymes are highly homologous to the mono-domain ARO/CYCs described above with the exception of an additional domain. The structures of both StfQ and BexL resemble the unison of two individual mono-domain ARO/CYCs covalently attached to one another in the same orientation. Both the N-terminal and C-terminal domains of StfQ and BexL are representative of a helix-grip-fold as described above for TcmN, WhiE and ZhuI.[91] It had been previously hypothesized that di-domain ARO/CYCs were a product of microbial gene duplication events which, given the high degree of structural similarity of both the C-term and N-term domains of StfQ and BexL to another, appears to be valid.[88] The major structural deviations between StfQ, a non-reducing ARO/CYC, and BexL, a reducing ARO/CYC, occurs at the location of the substrate binding pocket. In the StfQ crystal structure, the N-terminal domain harbors a smaller pocket than its corresponding C-terminal domain. On the contrary, the N-terminal and C-terminal domains of BexL house large cavities that could accommodate a mature polyketide chain. Docking simulations with putative substrates were performed with StfQ and BexL that aided in the identification of active site residues important for cyclization. Mutagenesis of the predicted catalytic residues of StfQ revealed that its C-term domain was vital for catalysis. Complimentary mutagenesis experiments with BexL demonstrated that its ARO/CYC activity was attributed to the N-term pocket. Excision of the non-catalytic domains of BexL and StfQ lead to insoluble protein, suggesting that duplicate domains play a possible role

in protein stability. Of the many residues tested through mutagenesis, R218, H259, and N295 of StfQ, and R66, H109, and E135 of BexL were shown to be essential for di-domain ARO/CYC mediated cyclization. The structural conservation of an active site arginine for both mono-domain and di-domain ARO/CYCs alike provides further evidence that retained placement of positively charged Arg in the ARO/CYC active site may be crucial for the proper orientation of a full length poly- β -ketone chain to accomplish controlled regiospecific C7-C12 intramolecular cyclization. Additional experiments need to be implemented to determine how BexL, along with other reducing ARO/CYCs catalyze C9-OH dehydration, which still remains unclear.[91] The BexL-StfQ work was published by PNAS in 2015, in which I am a co-author (not included in this thesis).

1.6 Type II Ketoreductases

The ketoreductases (KR) are NADPH-dependent enzymes that are responsible for the reduction of a plethora of reactive poly- β -ketone intermediates during the biosynthesis of aromatic type II polyketides.[50, 95] PKS KRs can be classified based upon the site of ketone reduction, which has been used to construct four basic KR clades that encompass C-9, C-15, C-17, and C-19 reducing KRs (Figure 1.13).[96]

The C-9 reducing KRs are the most common and have been the most structurally and biochemically characterized.[97] C-9 KRs aid in the formation of anthracyclines such as doxorubicin and epirubicin.[78] Along with their capacity for C-9 reduction, the C-9 reducing KRs have been associated with chain length control and influencing the establishment of first-ring C7-C12 cyclization in reducing systems, which is crucial for productive biosynthesis.[84, 98]

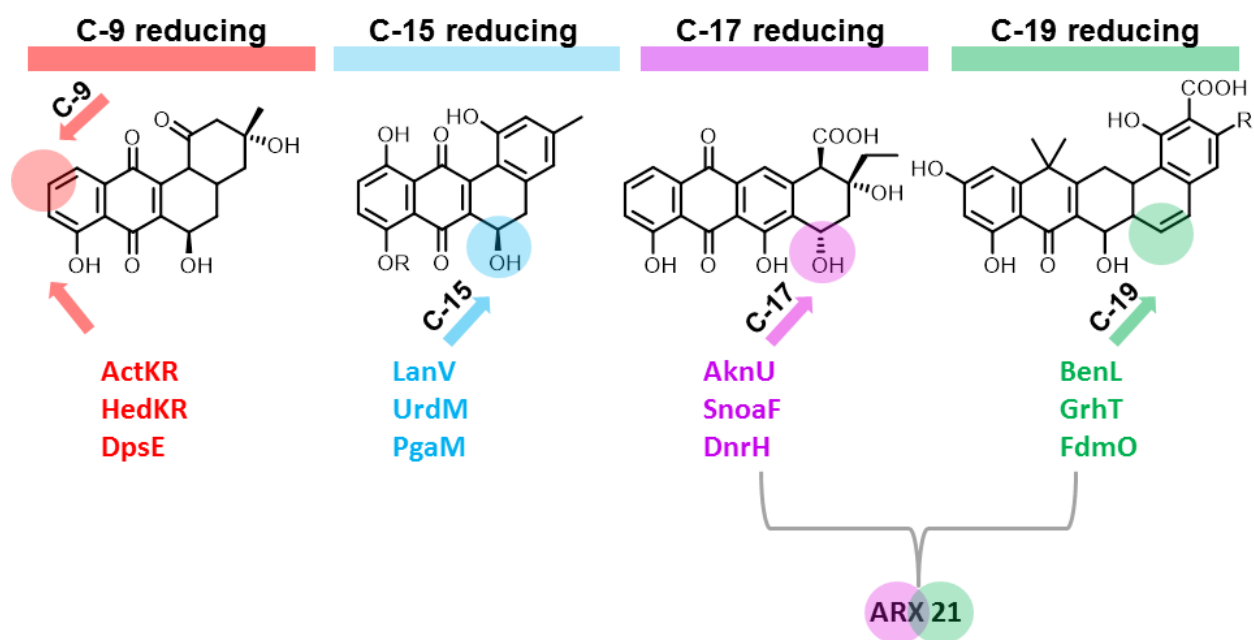


Figure 1.13. Depiction of the KR activity and substrate specificity in representative type II PKSs.

In vivo heterologous expression of the dodecaketide producing hedamycin min PKS in the absence of its endogenous C-9 reducing KR (HedKR) resulted in the production of additional triketide pyrone shunt product.[92] Similar experiments performed with HedKR led to the consistent biosynthesis of products with longer chain lengths, suggesting that HedKR plays a part in the stabilization of the growing polyketide chain.[99] There is strong evidence that C-9 KRs also influence first ring cyclization events.[100] *In vivo* expression of the Act Min PKS yields one major shunt product with C7-C12 (SEK4) cyclized ring and another minor C5-C10 cyclized product (SEK4b). *In-vivo* expression of Act Min PKS with ActKR results in one dominant C7-C12 cyclized product, mutactin, implicating that C-9 reducing KRs bias PKS biosynthesis toward C7-C12 cyclized products.[84]

The C-15 reducing KRs (also known as C-6 KRs) are involved in the reductive tailoring of angucycline-based natural products such as landomycin D and urdaymycin A.[101, 102] Several C-15 reducing KRs have been structurally characterized which include LanV and UrdMred.[101] *In vitro* assays demonstrated that LanV and UrdMred are only capable of reducing the carbonyl group at the C-15 position after the angucycline core is enzymatically oxidized by two separate NADPH- and O₂-dependent flavoprotein monooxygenases.[103] A C-6 hydroxylation and/or C-17 dehydration catalyzed by two flavoprotein monooxygenases is required to predispose the KR toward C-15 reduction.[102]

C-17 reducing KRs are associated with the biosynthetic pathways of aclacinomycin, steffimycin, and chartreusin and are not well characterized.[104-106] Multiple groups have reported their presence in the biosynthesis of glycosylated anthracycline-based natural products, yet their activities *in vivo* and *in vitro* using recombinant *Streptomyces* mycelial extracts have only been examined once.[107] There are currently no crystal structures of C-17 reducing KRs to provide a structural basis for substrate selection and C-17 reduction.

The C-19 reducing KRs are involved in the biosynthesis of pentangular polyphenolic polyketides. Benastatin A, fredericamycin A, and pradimicin are examples of natural products that are thought to require a C19 KR for their complete biosynthesis.[96] In 2008, the Hertwick group reported the first *in vivo* characterization of the C-19 KR BenL through the heterologous expression in *S. albus* of various gene components involved in benastatin A biosynthesis.[96] Novel benastatins, mainly benastatin K, were produced with C-19-hydroxyl group from transformants lacking the BenL gene. Although this work demonstrated that C-19 ketoreduction was not mandatory for the complete formation of the benastatin scaffold, a C-19 KR knockout provides an attractive strategy for the producing C-19 hydroxylated pentangular

polyphenols.[96] Currently the enzymology of C-19 KRs have not been investigated. The first crystal structure of a C-19 reducing KR has been solved by Tsai lab research group, which will be discussed in Chapter 5.

Structures of Type II KRs

The crystal structures of several C-9 KRs (Act, Hed, and Dps) and C-15 KRs (Lan V and UrdMred) from type II PKS biosynthetic pathways have been solved, providing a structural framework for region- and stereo-specific ketoreduction.[95, 99, 101] All type II PKS aromatic KRs belong to the short chain dehydrogenase (SDR) superfamily and consequently share a high degree of structural similarity to the bacterial type II FAS KR, FabG.[97] Proteins that belong to the SDR superfamily are characterized by a conserved Rossmann fold necessary for NADPH co-factor binding, which consists of a central 7-stranded parallel β -sheet flanked on two sides by three parallel α -helices. The KR active site is located between the α 4, α 5 and α 6 and α 7 helices, where the NADPH binding site resides.[95] In addition to a conserved tertiary structure, type II KRs house either a conserved active site triad or tetrad (Ser, Tyr, Lys or Asn, Ser, Tyr, Lys for C-9 reducing KRs) that includes a catalytic tyrosine proposed to foster a proton relay network, a common mechanism that is associated with other known SDR enzymes. The active site Tyr, Ser and Lys are located on the β 5- α 5 loop, while the active site Asn is located on the α -4 helix of C9-reducing KRs.[95, 101] Superimposition of the individual KR crystal structures reveals that the majority of the secondary structural features and active site residues are conserved (Figure 1.14). However, discrepancies in the α 7 flap lid region of the KR are present amongst different KR homologues that had been shown to contribute to substrate specificity and the site of ketoreduction (Tsai lab, unpublished work). Because all known KR structures have conserved

Rossmann folds and active site triads/tetrads, a catalytic mechanism has been proposed based on the crystal structures of FabG and ActKR as summarized in Figure 1.14.[95] While much is known about the SDR mechanism, how ketoreduction occurs in the context of an ACP-tethered polyketide substrate remains nebulous. Currently, there are no published PKS KR-ACP structures that would further strengthen our understanding of the protein-protein and protein-substrate interactions that govern C9 ketone reduction in type II PKS reducing pathways.

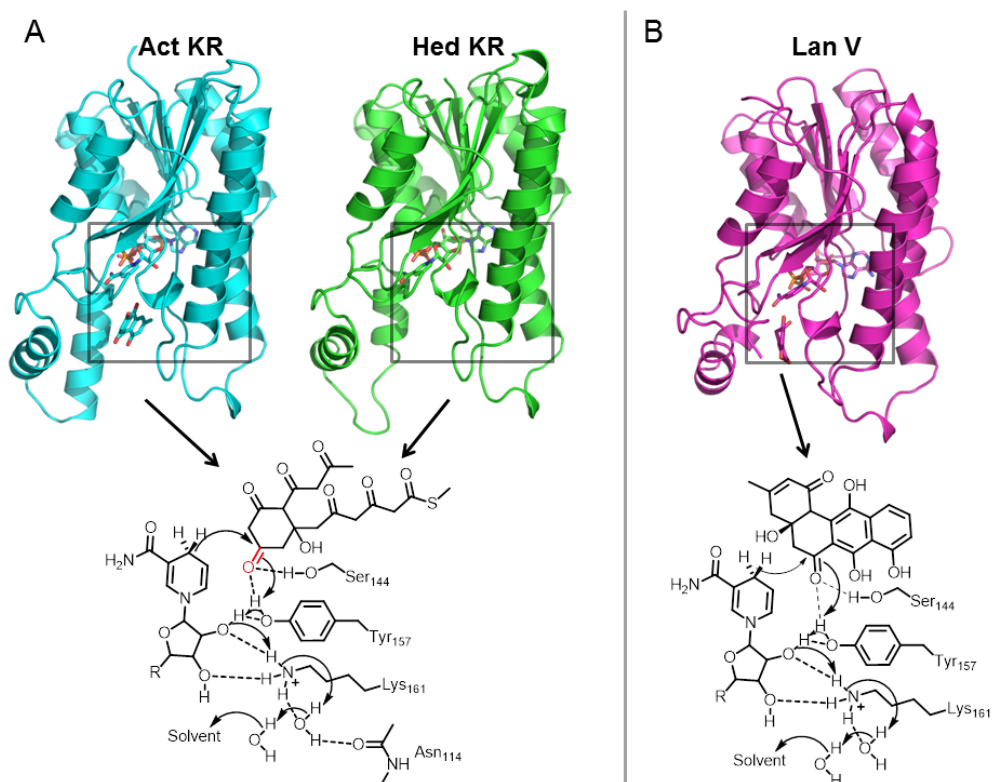


Figure 1.14. The crystal structures and proposed mechanisms of type II PKS KRs.

(A) The crystal structures of two C-9 reducing KRs: ActKR and HedKR. C-9 ketone reduction is thought to follow a conserved mechanism. The active site serine and tyrosine form an oxyanion hole via hydrogen bond contacts with the ketone group targeted for reduction. After a hydride from NADPH is transferred to the ketone substrate, the alkoxide is stabilized by the oxyanion hole and abstracts a proton from the active site tyrosine. In order to replenish the tyrosine proton, extensive proton relay occurs involving consecutive proton abstraction from the tyrosine oxyanion, the 2-OH Ribose of NADP⁺, the active site Lys, and finally four water molecules. (B) The crystal structure of the C15-reducing KR LanV bound to 11-deoxyandomycinone and NADP⁺. The mechanism for LanV mediated reduction C15 ketoreduction is similar to (B) however, LanV does not contain an active site Asn or act upon an ACP-linked substrate.

1.7 Secondary Metabolism: Nonribosomal Peptide Synthetases

Similar to PKS, nonribosomal peptide synthetases (NRPS) are known for the biosynthesis of a large repertoire of therapeutically valuable natural products through the condensation of activated amino acids.[108] Nonribosomal peptides are biosynthesized in a non-RNA templated manner by large multi-domain enzyme complexes known as non-Ribosomal synthetases (NRPS).[109] Vancomycin, surfactin and teixobactin are examples of clinically salient NRPS derived antibiotics.[110, 111] Nonribosomal peptides can be formed from L, D or unnatural amino acids. Like PKSs, NRPS can be divided into three separate types based on their mode of biosynthesis: type A, B and C.[112] Type A NRPS resemble type I modular PKSs in their architecture that contains multiple modules, as well as the use of each module in the extension and processing of a single peptide bond. Type B NRPSs mirror iterative PKSs with their reuse of core set of catalytic modules during the formation of their natural products. Finally, type C NRPSs are nonlinear synthetases that can be described as a hybrid of type A and B systems, where the majority of the catalytic events proceed in a linear order with the addition of a few iterative extension steps.[112] Nevertheless, all NRPSs contain a set of conserved catalytic domains necessary for the formation of their peptide products that include: the adenylation domain (A), condensation domain (C), thiolation or peptidyl carrier protein domain (T or PCP), and thioesterase domain (TE) (Figure 1.15).[108, 109, 112] The A domain is required for the activation of amino acid precursors via acyl adenylation using ATP and free amino acids as substrates. The C-domain mediates peptide bond formation from a free or tethered amino acid to a peptide chain attached to the terminal PPT arm of PCP. The PCP is analogous to the ACPs of PKSs and FASs, and it is responsible for the transport of peptidyl intermediates from one enzymatic domain to another throughout the biosynthesis. The TE domain is similar to the TEs

of PKS and FAS and acts to liberate mature or derailed NRPS products yielding free or macrocyclized peptides.[108, 109, 112, 113].

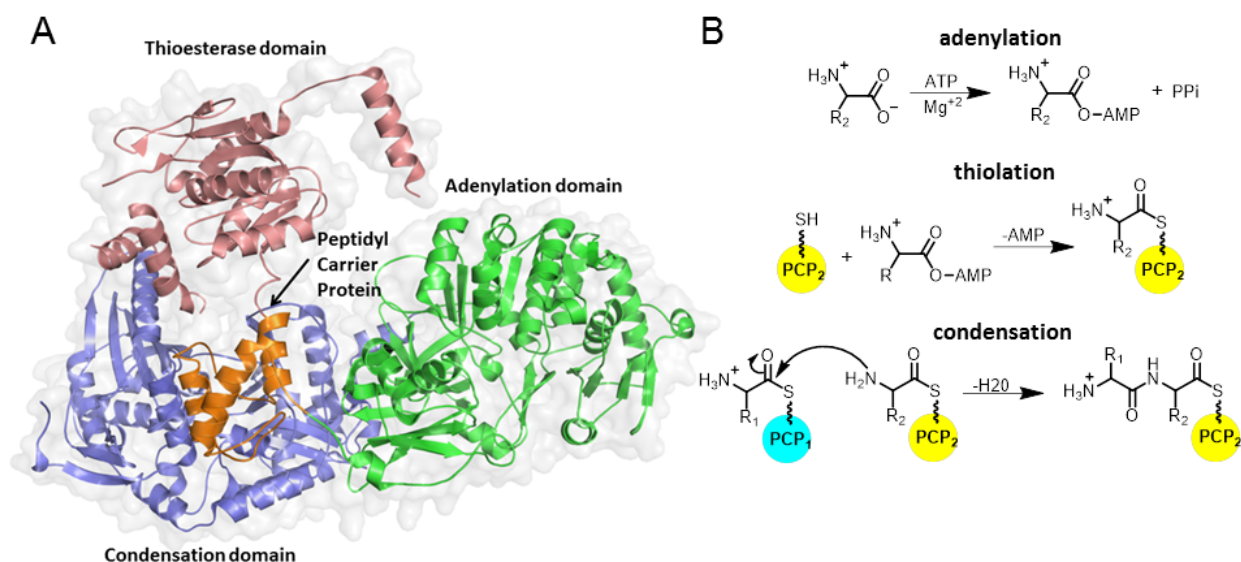


Figure 1.15. An introduction to NRPS structure and function.

(A) The crystal structure of an intact termination module from surfactin NRPS. (B) A schematic of NRPS biosynthesis. The ATP dependent A domain generates an activated amino acid through adenylation. The adenylated amino acid is next transferred to the phosphopantetheine prosthetic group of the PCP via an acyl addition elimination reaction with the terminal thiol group of phosphopantetheine. The C domain then catalyzes peptide bond formation between PCP tethered donor (blue PCP) and acceptor amino acids (yellow PCP).

Condensation Domains

The condensation domain (C domain) catalyzes peptide bond formation during NRP biosynthesis.[114] The C domain directs the amino group of a free or PCP-linked amino acid into the proper orientation necessary to achieve productive nucleophilic attack on the thioester of an amino acid chain attached to an upstream PCP, leading to the extension of the peptide chain by a single amino acid. C domains are not solely limited to the coupling of amino acids to a growing NRPS peptide chain.[115] The free standing C domains from vibriobactin and enacyloxin biosynthetic pathways, VibH and bamb_5915, are atypical in that they utilize

norspermidine and shikimate as acceptor nucleophiles that simultaneously extend and release a mature peptide from a terminal PCP domain.[9, 113] Several crystal structures of C domains have been determined which include: VibH (vibrobactin synthetase), TycC (PCP-C di-domain from tyrocidine synthetase), CDA-C1 (calcium-dependent antibiotic synthetase), and the C domain from the intact termination module of surfactin synthetase.[110, 113, 116, 117] All four C domain monomers consist of two separate chloramphenicol acyl transferase (CAT) like domains that are arranged in a V shape to form a pseudodimer. Each domain can be further described as an α - β -sandwich with a β strand from each monomer to form the central β -sheet of an adjoining monomer.[110, 113, 116, 117] All known active C domains harbor a HHXXDG motif that is thought to be involved in catalysis.[113, 116] The active site of the C domain lies at the interface of its two CAT domains and is characterized by a solvent channel that allows access to the conserved HHXXDG active site consensus residues from two different faces (Figure 1.16). This domain arrangement allows for the positioning of a donor PCP-linked polypeptide chain with a free or downstream PCP-tethered acceptor amino acid near the active site.[113, 116] Because of the lack of substrate-bound crystal structures, the exact role the two consensus histidine residues play during peptide bond formation remains unclear.[118] Given the position of the second histidine in the C domain active site, it has been postulated that the second histidine is an active site base that deprotonates the amino group of an incoming acceptor amino acid.[109] However, active site mutagenesis experiments demonstrated in several cases that the second active site His is not essential for condensation activity.[113] Recent work published by the Schmeing group utilized an active site Cys mutant that could react with bromoalkyl-alanine analogues to resemble the natural PPT-conjugated substrate of C domains.[119] This result suggested that the second histidine does not act like an active site base, but rather is important for

proper steering of the acceptor amino acid nucleophile toward the donor polypeptide chain during peptide bond formation.[119]

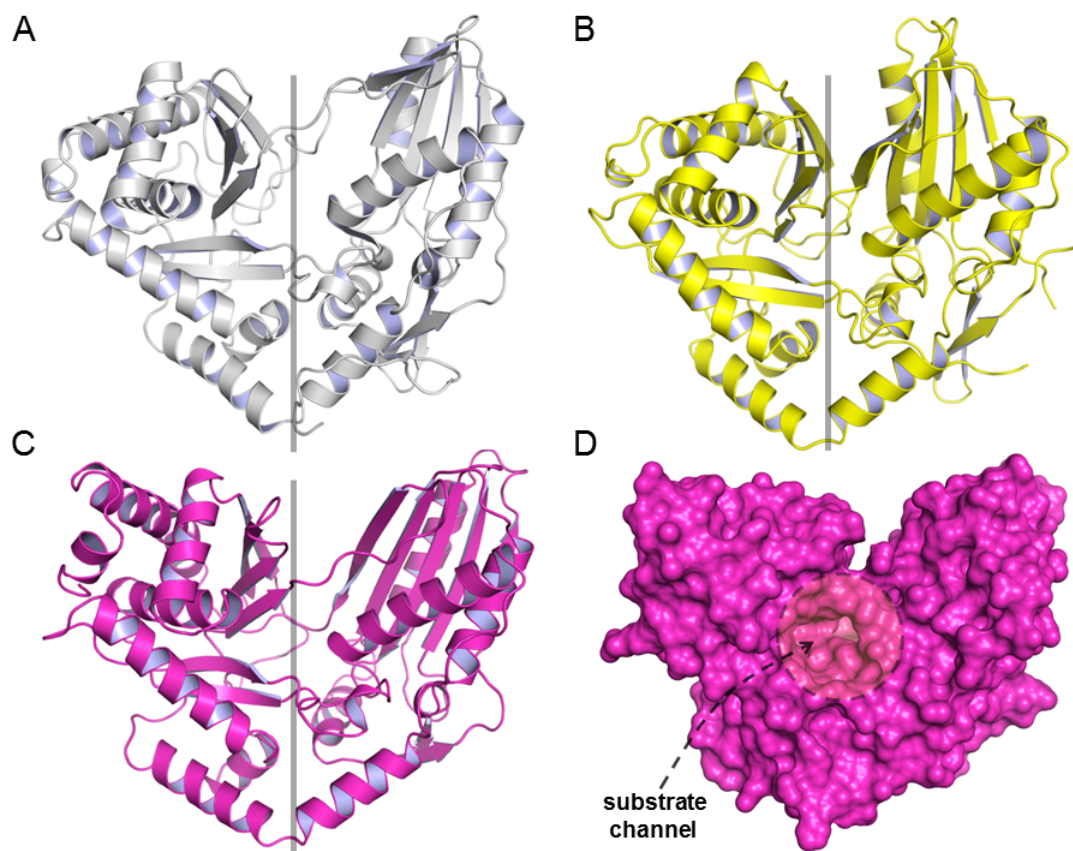


Figure 1.16. Side by side comparison of several C domains from different NRPSs. (A) TycC with PCP removed for clarity. (B) CDA-C1. (C) VibH. (D) A surface representation of VibH that identifies the C domain substrate channel.

Recently, the crystal structure of the X domain from the teicoplanin biosynthetic pathway has been reported. The X domain is structurally similar to known C domains, yet it is catalytically inactive and harbors a HRXXDG motif.[111, 119] The X domain has been shown to serve as a protein-protein interaction scaffold that is required for the recruitment of several trans acting oxygenases in the teicoplanin biosynthetic pathway.[111] Future structural and biochemical experiments are required to elucidate the active site residues that are important for

catalysis. Nonetheless, due to the relatively forward assembly line logic associated with NRPSs, C domain engineering is desired for the creation of new non-ribosomal peptides.[109]

1.8 Secondary Metabolism: Hybrid Systems

The biosynthesis of secondary metabolites is not a black and white process where the biogenesis of a set of natural products is solely restricted to one particular enzymatic pathway. Indeed, there are many shades of grey associated with secondary metabolite biosyntheses where natural products produced from one particular pathway serve as the substrates for other enzymatic assemblies [120]. In Nature, it is common to find secondary metabolites that are biosynthesized by hybrids of PKS, NRPS and FAS (Figure 1.17).[121] The biosynthesis of aflatoxin B₁, for example, involves both type I FAS and NR-PKS for the formation of norsolorinic acid anthrone, an aflatoxin B₁ precursor.[30] In this hybrid system, a dedicated yeast-like FAS, composed of two individual subunits (HexA and HexB), provides a hexanoyl-starter unit to an NR-PKS, PksA, which is then extended and cyclized to afford norsolorinic acid anthrone.[122] An additional example of hybrid secondary metabolite systems stems from the biosynthesis of enacyloxins found in *Burkholderia cepacia*. [9] Here, a type I *cis/trans*-AT modular PKS first generates the enacyloxin polyene core, which is later passed to a downstream NRPS PCP and then released via a C domain mediated transamidation using shikimate as an acceptor nucleophile.[9] Individuals infatuated with P450s may enjoy the fact that P450s are often required for the trans-hydroxylation of many natural product precursors (such as the stambomycins) for their maturation and/or macrocyclization.[123] The list of hybrid systems, including their enzyme partners and their corresponding molecules that they biosynthesize, could quite easily take up several hundred pages due to their prevalence in nature and in natural

product literature. Hybrid complexes of PKS, NRPS and FAS serve as a great source of inspiration for natural product researchers and enzymologists alike. After all, these mixed systems have succeeded in accomplishing combinatorial biosynthesis of different multi-domain enzymatic assemblies, a goal we humans have yet come close to achieving.

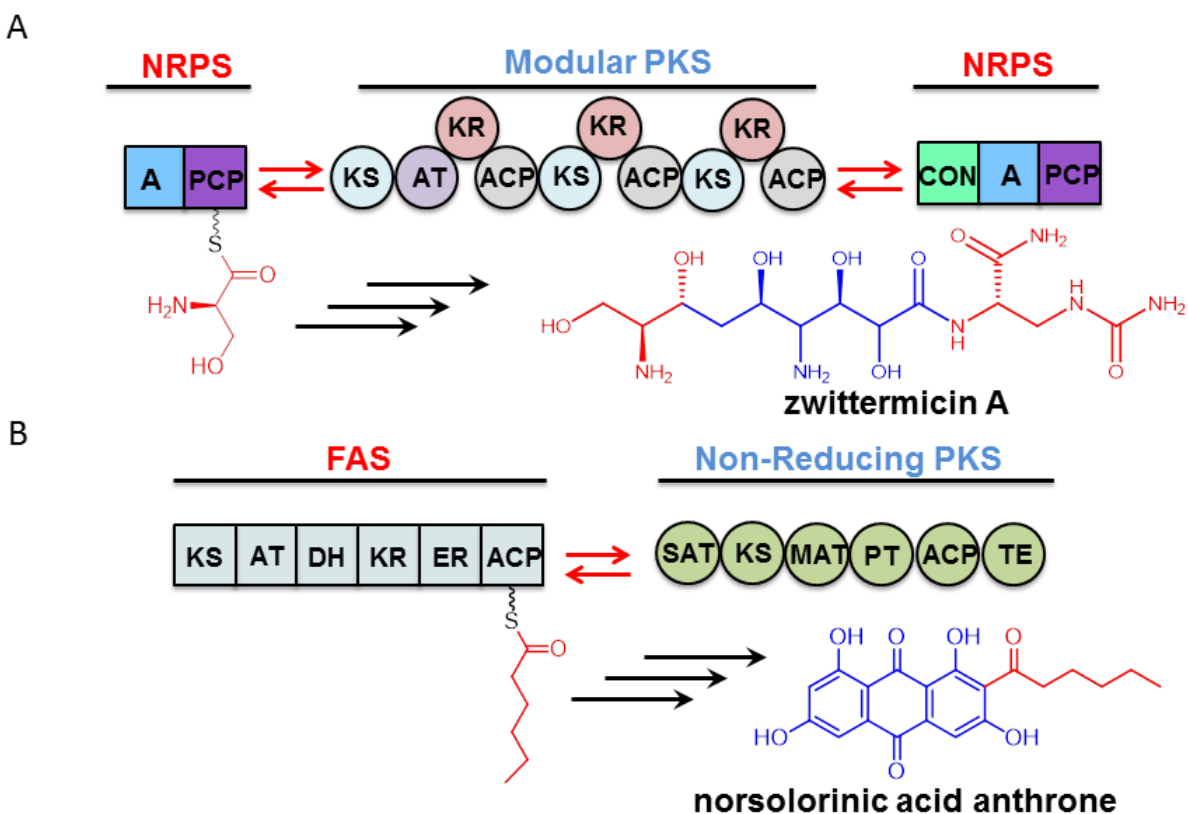


Figure 1.17. A general cartoon schematic of different PKS/NRPS and FAS/PKSs hybrid systems.

(A) The biosynthesis of zwittermycin A involves the cooperation of NRPS and PKS assemblies. The biosynthesis of aflatoxin B1 involves the use of a FAS/PKS hybrid for the generation of norsolorinic acid anthrone. The six-domain FAS supplies the hexanoyl group as the starter unit for the non-reducing PKS.

The rules that dictate permissive intersystem cross talk are still unknown, but they mostly likely involve the synergistic coordination of protein-protein and protein-substrate interactions between

various enzymatic components and the ACP domain.[121, 122] Crystal structures of hybrid complexes will provide a wealth of structural information that will illuminate the structural rationale behind intersystem communication in mix PKS/NRPS and FAS pathways. The above discussion may be the hubris of a natural products researcher speaking, because it is very easy to take for granted the time, selective pressure, and protein interface sampling that was required for efficient non-derailed combinatorial biosynthesis (personal experience). However, Nature is unafraid of bending the biosynthetic rules in order to salvage the useful chemical components from one pathway and combine them with another for the generation of natural products with greater functional and chemical complexities, and crystal structures of the hybrid enzymes are essential for us humans to learn from Nature. In Chapter 4, we will discuss the detailed characterization of the C domain from a hybrid NRPS/PKS system.

1.9 Acyl-CoA Carboxylases

The building blocks necessary for the genesis of fatty acids and polyketides are made by enzymes known as acyl-CoA carboxylases (hereafter referred to as ACCases).[124] ACCases encompass a wide spectrum of biotin dependent enzymes that are responsible for the incorporation of CO₂ into a variety of acyl-CoA substrates.[125] Examples of ACCases include acetyl-CoA carboxylases (ACC), propionyl CoA carboxylases (PCC), 3-methyl-crotonyl-CoA carboxylases (MCC), and geranyl-CoA carboxylases (GCC).[126] ACCases can be grouped based upon substrate preferences and the site of substrate carboxylation.[126] ACCs and PCCs, for instance, will be referred to as saturated ACCases, given their capacity to carboxylate the saturated α -carbons of acetyl-CoA and propionyl-CoA to form malonyl-CoA and methyl-malonyl-CoA, respectively.[127] MCC and GCC, on the contrary, are unsaturated ACCases that

act upon α - β unsaturated acyl-CoAs to generate 3-methylglutaconyl-CoA and carboxyl-geranyl-CoA through the addition of a CO₂ group to the γ -position of 3-methyl-crotonyl-CoA and geranyl-CoA, respectively.[128, 129] ACCases can be further subcategorized base on the number of subunits, subunit domain composition, and subunit stoichiometry that comprise an active ACCase. Multisubunit ACCases, for example, are typically found in yeast, archaea bacteria, and plants; whereas mammals encode the complete ACCase biosynthetic machinery on a single polypeptide chain.[126] All ACCases are mechanistically comparable and afford carboxylated products necessary for the biosynthesis of important primary and secondary metabolites.[125, 126, 130, 131]

The extender unit formation mediated by ACCase is a two-step process that requires three components and two distinct enzymatic activities. The enzymatic carboxylation process can be divided into: 1) the generation of carboxy-biotin, and 2) the coupling of a carboxyl group derived from carboxy-biotin to a saturated or unsaturated acyl-CoA. Biotin carboxylase (BC) catalyzes the transfer a carboxyl group from bicarbonate to the N1 atom of the ureido ring of biotin in an ATP dependent process (Figure 1.18B). The formation of carboxy-biotin proceeds through an acyl carboxyphosphate intermediate that is generated after an active site base of BC deprotonates a bicarbonate molecule, which in turn acts a nucleophile and attacks the γ -phosphate of ATP. Ultimately, the resulting carboxy-phosphate intermediate decomposes yielding CO₂ and orthophosphate. The newly liberated phosphate molecule then acts as a base, deprotonating the N1 nitrogen of biotin, thereby promoting ureido enolate formation that is followed by nucleophilic attack of the N1 atom of biotin onto CO₂ to generate carboxy-biotin. Mg⁺² divalent cations are required to stabilize the negatively charged phosphate in the intermediates that are formed during the generation of carboxy-biotin.[126, 130]

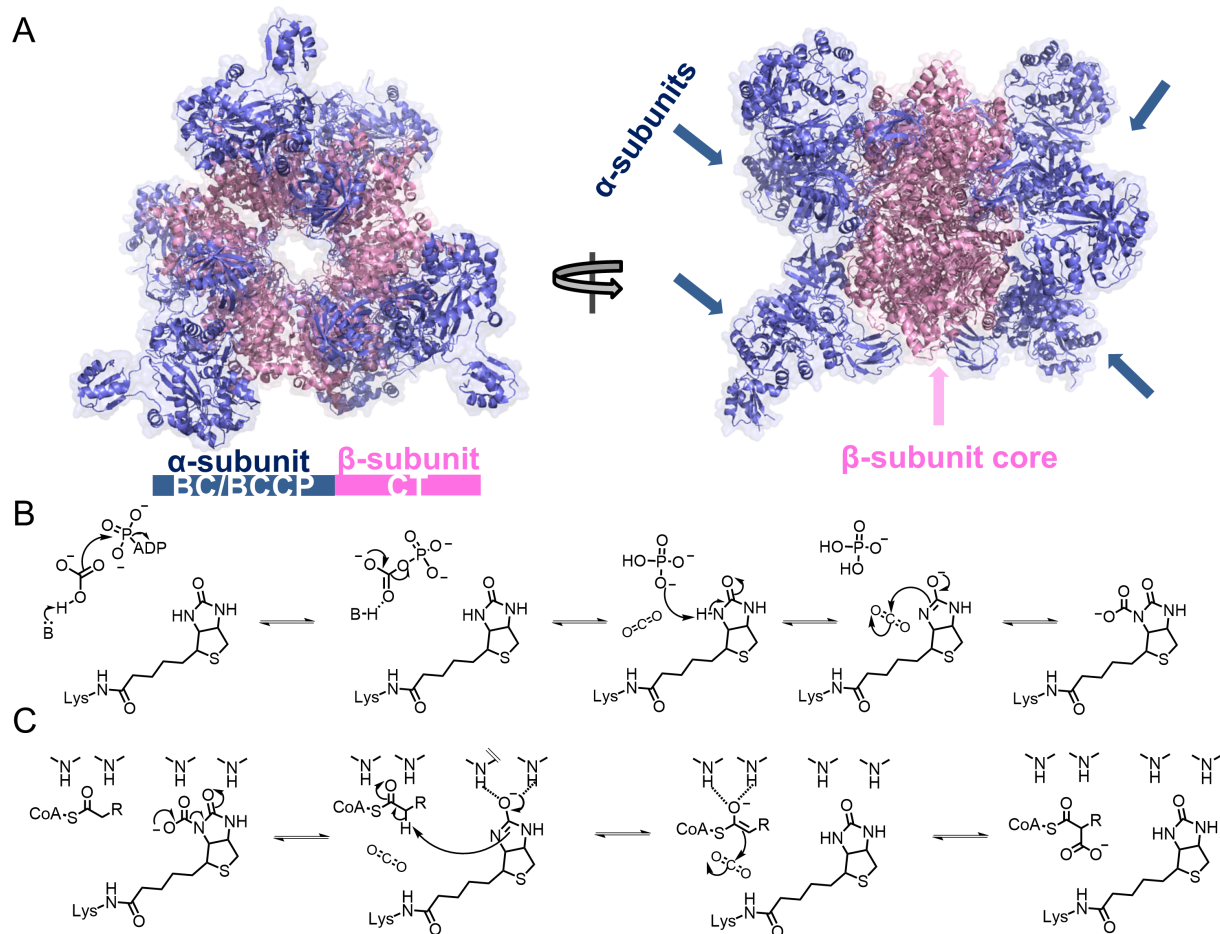


Figure 1.18. The crystal structural and mechanism of propionyl-CoA carboxylase (PCC).

(A) The crystal structure of the $\alpha_6\beta_6$ PCC *holo* complex with α - and β -subunits colored as blue and pink, respectively. The α -subunit contains both the BCCP and BC domains, whereas the β subunit contains the CT domain. Note the arrangement of the α -subunit around a central β -subunit core. (B) Proposed mechanism for the formation of carboxy-biotin by the PCC α -subunit. Biotin is first conjugated to the BCCP domain of the PCC α -subunit through a conserved Lys residue by a biotin ligase. An active site base in the BC domain deprotonates bicarbonate, which subsequently undergoes nucleophilic attack on the γ -phosphate of ATP, leading to the formation of a carboxy-phosphate intermediate and ADP. The carboxy-phosphate intermediate decomposes, and the resulting free phosphate molecule deprotonates the N1 of biotin, leading to the formation of a ureido enolate. The reactive ureido enolate undergoes nucleophilic attack on carbon dioxide thereby generating carboxy-biotin. (C) The proposed CT mechanism. Carboxy-biotin is delivered to the active site of the CT bound to an acyl- CoA via the BCCP domain. A protein conformational change induces the dissociation of CO_2 from carboxy-biotin. The resulting ureido enolate undergoes deprotonates the α -carbon of an acyl-CoA they by generating an acyl-enolate. The acyl enolate proceeds to attack the residual CO_2 molecule thereby generating a carboxylated malonyl-CoA based extender unit.

The biotin carboxyl carrier protein (BCCP) serves as a carrying platform for biotin and carboxy-biotin, which must be delivered to both the BC and CT components, respectively. In order to be active, the BCCP must be covalently modified with a biotin prosthetic group, the so called “swinging arm”, which is appended to the epsilon amino group of a conserved lysine residue of BCCP by a biotin ligase. After leaving the BC, the BCCP forwards carboxy-biotin to the carboxyl transferase (CT), leading to the direct carboxylation of the α - or γ -carbon of acyl-CoA, thereby generating a carboxylated acyl product.[128, 132] Unlike the mechanism of BC that has been extensively investigated, the CT mechanism has proven to be difficult to study and remains partially hypothetical.[133] Nonetheless, a mechanism for CT-mediated carboxyl transfer for saturated ACCases has been proposed, based on the biotin and propionyl-CoA bound PCC β -subunit (PccB) cocrystal structures (Figure 1.18C).[131] The binding of carboxy-biotin to the active site of the CT is presumed to promote the liberation of CO₂ from biotin via a CT induced bending of the CO₂-N1-CO π orbital of carboxy-biotin. The resulting deprotonated N1 of biotin can then serve as a general base to deprotonate the α -carbon of an acyl-CoA, affording a reactive enolate species. The newly acquired enolate continues to attack free CO₂, leading to the addition of a carboxyl group in the α position of a saturated acyl-CoA. Both acyl and urideo enolate intermediates are stabilized by two separate oxyanion holes in the active site of the CT, which are formed by the peptide backbone amides and carbonyls of the residues lining the entrance to the CT substrate binding pocket.[131] A notable exception is *E.coli* ACC CT, which contains an active site cysteine that is proposed to facilitate the deprotonation of the C α of a saturated acyl-CoA to produce an enolate required for acyl-CoA carboxylation.[134]

In *Actinomyces* and *Rhodobacteraceae*, the ACCases are comprised of two separate α and β -subunits that form an active $\alpha_6\beta_6$ heterododecameric complex.[126, 130] In these biological

systems, the α -subunit contains the BC and BCCP domains, whereas the β -subunit contains the CT domain. Crystal structures of the individual subunits and the intact $\alpha_6\beta_6$ *holo* ACCase complexes have been solved, providing a structural basis of bacterial ACCase mediated extender unit formation (Figure 1.18A).[126, 130] In the PCC $\alpha_6\beta_6$ *holo*-complex crystal structure, the β -subunits form two stacks of trimers that pack together to form a toroid like hexamer, which serves as the central core of the ACCase *holo* complex. The monomeric α -subunits bind to individual β -subunit in an orthogonal manner with respect to the hexameric β -subunit core, which provides access to the CT active site pocket that lies at the dimer interface of two opposing β -subunit monomers. This subunit arrangement places the BCCP domain approximately 55 Å away from the CT active site pocket.[135] It was previously hypothesized that the BCCP domain of the α -subunit would bind to the β -subunit in a manner that would position its biotin sweeping arm in close proximity (16 Å) to both BC and CT active sites. However, based on the *holo* PCC complex crystal structure, this was shown to be incorrect, suggesting that the BCCP domain must translocate between the BC and CT active sites for productive catalysis. The quaternary structure of the *holo* PCC provides an excellent model that explains how the spatial organization of its α - and β -subunits facilitates the efficient production of malonyl-CoA based extender units.

The selection and incorporation of various dicarboxylic acid building blocks into a final natural product is yet another source for polyketide functional diversification. Polyketides such as salinosporamide A, reveromycin E, sanglifehrin A, and soce-thuggacin A are natural products whose polyketide scaffolds can be derived from the use of malonyl- methylmalonyl-, ethylmalonyl-, hexylmalonyl- and chloroethylmalonyl-CoA as their extender units (Figure 1.19).[136, 137] Short chain linear extender units such as malonyl and methylmalonyl-CoA, are

most commonly used for PKS chain propagation and are primarily generated by the ACCases. Long chain linear and branched alkyl malonyl units such as, hexyl- and octyl-malonyl-CoA are not found in high abundance in PKS natural products and are typically biosynthesized from NADPH dependent enzymes known as crotonyl-CoA carboxylase reductases (CCRs). In Chapter 3, we will present the studies on novel ACCases that are exceptions to this rule, as certain ACCase β -subunits are capable of generating long chain PKS extender units.

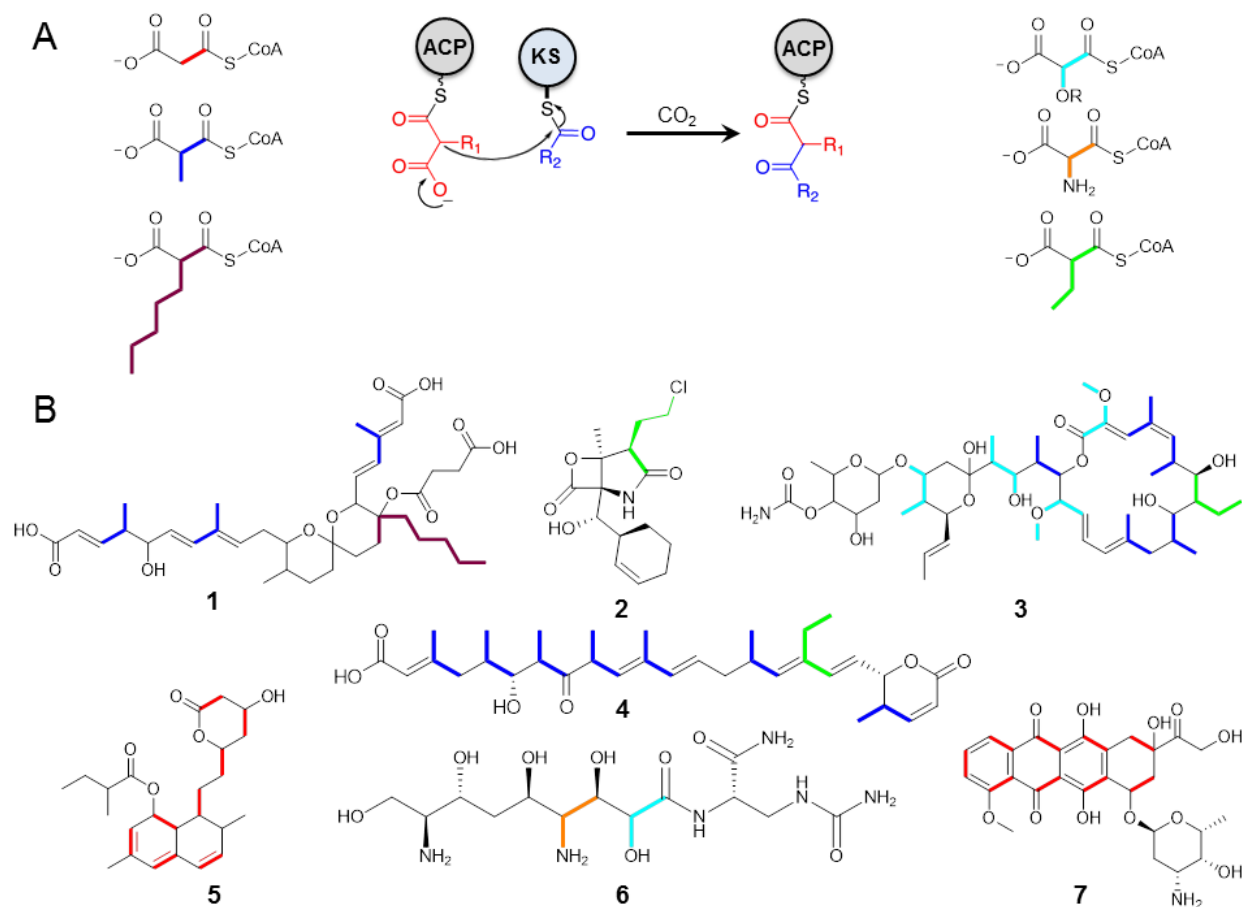


Figure 1.19. The extender diversity of polyketide natural product.

(A) A panel of natural and synthetic PKS extender units such as malonyl- (red), methyl- (blue), ethyl- (green), methylaminoyl- (orange), methoxyl- (cyan) and hexanoyl-CoA (purple). The mechanism of PKS chain extension is also shown. (B) A collection of polyketides that exhibit extender unit diversity (bonds colored in bold). (1) Reveromycin D, (2) salinosporamide A, (3) concanamycin A, (4) leptomycin B, (5) lovastatin, (6) zwittermycin A, and (7) doxorubicin.

The significance of extender unit selection lies in the observation that the biological activities of a given polyketide can be “fine-tuned” or modulated by a given extender unit’s alkyl side chain. Stambomycins A-D, for example, share a common 51-membered macrolide aglycone; however, the presence or absence of distinct alkyl side chains in position C26 leads to an apparent difference in antibiotic and anticancer activities of these polyketides.[10] The engineering of different PKS systems for the controlled incorporation of a specific extender unit is still a work in progress. The rate limiting step of designed extender unit incorporation lies within the AT, which is highly specific in terms of protein-protein and protein-substrate interactions.[138] ATs have been manipulated through active site mutagenesis or domain swapping from different type I modular systems to accept new substrates.[139] These experiments resulted in in poor yield or abolished biosynthesis.[138] Nonetheless, controlled incorporation of an extender unit of choice is an attractive means for achieving novel unnatural products with augmented biological activities.

1.10 Crotonyl CoA Carboxylases/Reductases

Crotonyl-CoA reductase/carboxylases (CCR) are enzymes involved in the biosynthesis of extravagant malonate based PKS extender units.[136, 140] CCR biosynthetic machinery lies at the intersection of biogenesis and catabolism. Their stringent substrate specificity towards α - β unsaturated acyl-CoAs, which are products of the β -oxidation of fatty acids.[19, 136] CCRs catalyze the reductive carboxylation of unsaturated α - β acyl-CoAs, thereby generating malonyl-based extender units in a biotin independent manner (Figure 1.20C).[141] CCRs are NADPH dependent enzymes that belong to the medium-chain dehydrogenase/reductase (MDR) super families, and have been implicated in the biosynthesis of hexyl-, octanoyl-, chloroethyl-, and

malonyl-CoA based extender units.[136] These non-canonical building blocks are rarely utilized more than once throughout the biosynthesis of PKS natural products, but still impart structural diversity to a given polyketide.

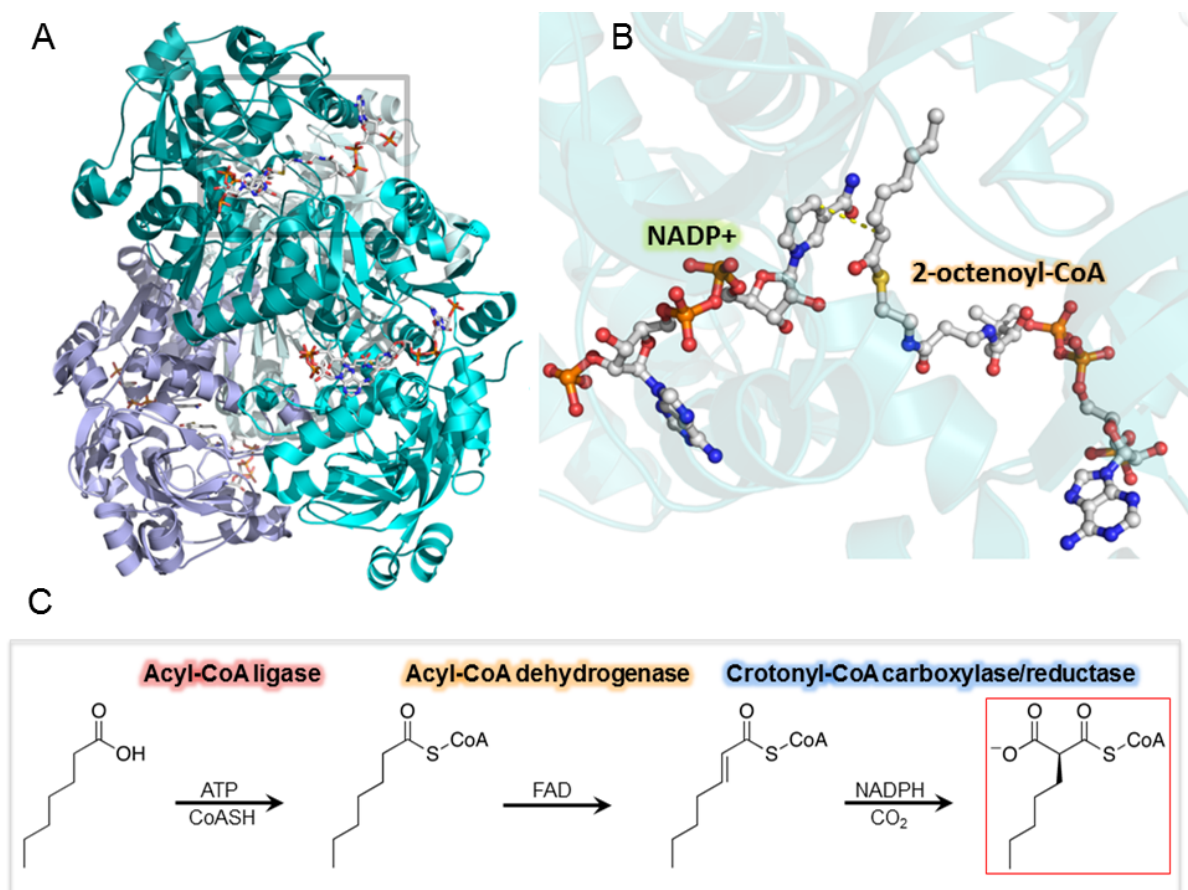


Figure 1.20. Crystal structure and function of the crotonyl-CoA carboxylase/reductase CinF.

(A) The crystal structure of the CinF tetramer with individual monomers colored differently. (B) CinF positions 2-octenoyl-CoA in close proximity to the nicotinamide ring of NADP+ for simultaneous regio-specific hydride transfer and carboxylation. (C) The biosynthetic pathway for CCR mediated extender unit formation. A fatty acid is converted to a fatty-acyl-CoA by an acyl-CoA ligase. The resulting acyl-CoA is oxidized by an acyl-CoA dehydrogenase to produce an α - β unsaturated acyl-CoA, which is subsequently reduced and carboxylated by a CCR to afford highly unusual building blocks.

Several crystal structures of CCRs have been solved, revealing their structural similarity to other known oxydoreductases such as human PIG3.[34, 140] The structure of the CinF, a CCR involved in cinnabaramide A production, was the first CCR crystal structure to be published.[34] Like many other oxydoreductases such as *E.coli* semialdehyde dehydrogenase (SsADH), CinF crystallized as tetramer with four CinF monomers per asymmetric unit that pack together to form a dimer of dimers (Figure 1.20A). The individual CinF monomers can be dissected into two domains: an NADPH binding domain, and a catalytic domain consisting of the first 200 residues of the N terminus and the last 100 amino acid residues of the C-terminus. The NADPH binding domain of CinF contains a conserved Rossman fold associated with NADPH binding enzymes, as discussed in the type II KR section above. The CinF catalytic domain is composed of eleven α -helices, a three-stranded antiparallel β -sheet, and a separate seven-stranded antiparallel β -sheet. The cocrystal structure of CinF bound to NADP⁺ and 2-octenoyl-CoA unveiled the CCR Co-factor and substrate binding sites.[34] The acyl binding pocket of CinF is defined by helices α 6, α 8, and α 17 of CinF, which favorably positions the negatively charged 3-phospho-ADP moiety of 2-octenoyl-CoA near the surface of one monomer allowing for van der Waals contacts with an additional dimeric partner. The eight-carbon hydrophobic tail of 2-octenoyl-CoA is buried within the substrate binding region of the catalytic domain, and it forms hydrophobic contacts with the residues lining the catalytic domain's pocket. The resulting substrate binding mode beautifully places the 2-alkene C-C double bond of 2-octenoyl-CoA in close proximity to the nicotinamide ring of NADP⁺ necessary to achieve a regiospecific hydride transfer (Figure 1.20B). Docking of CO₂ into the CinF ternary complex led to the identification of several candidate residues that through mutagenesis were confirmed to be catalytic.[34] To date, the exact role these residues play in the carboxylation of unsaturated CoAs remains mostly hypothetical.

1.11 Thesis Overview

Chapter 2 will focus on the structural analysis of WhiE ARO/CYC bound with atom-replaced polyketide mimics. Chapter 3 will focus on the structural and functional analyses of a medium-chain acyl-CoA carboxylase beta subunit (MccB). Chapter 4 will describe the structural basis for C domain-catalyzed product release from a mixed NRPS-PKS hybrid assembly line. In Chapter 5, the crystal structure of a novel PKS KR and successful *in vitro* reconstitution of an unusual ARO/CYC will be presented. The overall conclusion and future directions will be discussed in Chapter 6.

1.12 References

1. Ji, H.F., X.J. Li, and H.Y. Zhang, *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, 2009. **10**(3): p. 194-200.
2. Aboelsoud, N.H., *Herbal medicine in ancient Egypt*. Journal of Medicinal Plants Research, 2009. **04**(02): p. 082-086.
3. Bentley, R. and J.W. Bennett, *Constructing polyketides: from collie to combinatorial biosynthesis*. Annu Rev Microbiol, 1999. **53**: p. 411-46.
4. Hoffmann, R.W., *Natural product synthesis: changes over time*. Angew Chem Int Ed Engl, 2013. **52**(1): p. 123-30.
5. Koehn, F.E. and G.T. Carter, *The evolving role of natural products in drug discovery*. Nat Rev Drug Discov, 2005. **4**(3): p. 206-20.
6. Dias, D.A., S. Urban, and U. Roessner, *A historical overview of natural products in drug discovery*. Metabolites, 2012. **2**(2): p. 303-36.
7. Paterson, I. and I. Lyothier, *Total synthesis of (+)-discodermolide: an improved endgame exploiting a Still-Gennari-type olefination with a C1-C8 beta-ketophosphonate fragment*. Org Lett, 2004. **6**(26): p. 4933-6.

8. Lee, M.Y., B.D. Ames, and S.C. Tsai, *Insight into the molecular basis of aromatic polyketide cyclization: crystal structure and in vitro characterization of WhiE-ORFVI*. Biochemistry, 2012. **51**(14): p. 3079-91.
9. Mahenthiralingam, E., et al., *Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic Burkholderia ambifaria Genomic Island*. Chem Biol, 2011. **18**(5): p. 665-77.
10. Laureti, L., et al., *Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens*. Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6258-63.
11. He, H., et al., *Mannopeptimycins, novel antibacterial glycopeptides from Streptomyces hygroscopicus, LL-AC98*. J Am Chem Soc, 2002. **124**(33): p. 9729-36.
12. Garcia, A.E., et al., *Isolation, synthesis, and antimicrobial activities of naturally occurring theta-defensin isoforms from baboon leukocytes*. Infect Immun, 2008. **76**(12): p. 5883-91.
13. Ortholand, J.Y. and A. Ganesan, *Natural products and combinatorial chemistry: back to the future*. Curr Opin Chem Biol, 2004. **8**(3): p. 271-80.
14. Bentley, R., *Secondary metabolite biosynthesis: the first century*. Crit Rev Biotechnol, 1999. **19**(1): p. 1-40.
15. Bentley, R., *John Norman Collie: Chemist and mountaineer*. Journal of Chemical Education, 1999. **76**(1): p. 41-47.
16. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review*. Nat Prod Rep, 2001. **18**(4): p. 380-416.
17. Buchanan, J.M. and B. Hastings, *The use of isotopically marked carbon in the study of intermediary metabolism*. Physiol Rev, 1946. **26**: p. 120-55.
18. Voet, D. and J.G. Voet, *Biochemistry*. 4th ed. 2011, Hoboken, NJ: John Wiley & Sons. xxv, 1428, 53 p.
19. Nelson, D.L., A.L. Lehninger, and M.M. Cox, *Lehninger principles of biochemistry*. 6th ed. 2013, New York: W.H. Freeman. 1 vol. (various pagings).
20. Whitaker, J.R., *Principles of enzymology for the food sciences*. 2nd ed. Food science and technology. 1994, New York: Marcel Dekker. xv, 625 p.
21. Malpartida, F. and D.A. Hopwood, *Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host*. 1984. Biotechnology, 1992. **24**: p. 342-3.

22. Chen, Y., M.J. Smanski, and B. Shen, *Improvement of secondary metabolite production in Streptomyces by manipulating pathway regulation*. Appl Microbiol Biotechnol, 2010. **86**(1): p. 19-25.
23. Sun, H., et al., *Recent advances in combinatorial biosynthesis for drug discovery*. Drug Des Devel Ther, 2015. **9**: p. 823-33.
24. Patti, G.J., O. Yanes, and G. Siuzdak, *Innovation: Metabolomics: the apogee of the omics trilogy*. Nat Rev Mol Cell Biol, 2012. **13**(4): p. 263-9.
25. Drew, S.W. and A.L. Demain, *Effect of primary metabolites on secondary metabolism*. Annu Rev Microbiol, 1977. **31**: p. 343-56.
26. Hanson, J.R., *Natural products : the secondary metabolites*. Tutorial chemistry texts. 2003, Cambridge: Royal Society of Chemistry. vi, 147 p.
27. Franck, B., *Key building blocks of natural product biosynthesis and their significance in chemistry and medicine*. Angew Chem Int Ed Engl, 1979. **18**(6): p. 429-39.
28. Maplestone, R.A., M.J. Stone, and D.H. Williams, *The evolutionary role of secondary metabolites--a review*. Gene, 1992. **115**(1-2): p. 151-7.
29. Gokhale, R.S., R. Sankaranarayanan, and D. Mohanty, *Versatility of polyketide synthases in generating metabolic diversity*. Curr Opin Struct Biol, 2007. **17**(6): p. 736-43.
30. Crawford, J.M. and C.A. Townsend, *New insights into the formation of fungal aromatic polyketides*. Nat Rev Microbiol, 2010. **8**(12): p. 879-89.
31. Winter, J.M., et al., *Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis*. J Am Chem Soc, 2015. **137**(31): p. 9885-93.
32. Andersen, M.R., et al., *Accurate prediction of secondary metabolite gene clusters in filamentous fungi*. Proc Natl Acad Sci U S A, 2013. **110**(1): p. E99-107.
33. Osbourn, A., *Gene clusters for secondary metabolic pathways: an emerging theme in plant biology*. Plant Physiol, 2010. **154**(2): p. 531-5.
34. Quade, N., et al., *Unusual carbon fixation gives rise to diverse polyketide extender units*. Nat Chem Biol, 2012. **8**(1): p. 117-24.
35. Cronan, J.E. and J. Thomas, *Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways*. Methods Enzymol, 2009. **459**: p. 395-433.
36. Pike, L.J., *The challenge of lipid rafts*. J Lipid Res, 2009. **50 Suppl**: p. S323-8.

37. Rouzer, C.A. and L.J. Marnett, *Cyclooxygenases: structural and functional insights*. J Lipid Res, 2009. **50 Suppl**: p. S29-34.
38. Maier, T., M. Leibundgut, and N. Ban, *The crystal structure of a mammalian fatty acid synthase*. Science, 2008. **321**(5894): p. 1315-22.
39. Beld, J., D.J. Lee, and M.D. Burkart, *Fatty acid biosynthesis revisited: structure elucidation and metabolic engineering*. Mol Biosyst, 2015. **11**(1): p. 38-59.
40. Campbell, J.W. and J.E. Cronan, Jr., *Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery*. Annu Rev Microbiol, 2001. **55**: p. 305-32.
41. Beld, J., et al., *The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life*. Nat Prod Rep, 2014. **31**(1): p. 61-108.
42. Tsai, S.C. and B.D. Ames, *Structural enzymology of polyketide synthases*. Methods Enzymol, 2009. **459**: p. 17-47.
43. Keatinge-Clay, A.T., *The structures of type I polyketide synthases*. Nat Prod Rep, 2012. **29**(10): p. 1050-73.
44. Griffin, M.O., et al., *Tetracyclines: a pleiotropic family of compounds with promising therapeutic properties. Review of the literature*. Am J Physiol Cell Physiol, 2010. **299**(3): p. C539-48.
45. Benjamin, D., et al., *Rapamycin passes the torch: a new generation of mTOR inhibitors*. Nat Rev Drug Discov, 2011. **10**(11): p. 868-80.
46. Vickery, C.R., et al., *Harvesting the biosynthetic machineries that cultivate a variety of indispensable plant natural products*. Curr Opin Chem Biol, 2016. **31**: p. 66-73.
47. Wang, F., et al., *Structural and functional analysis of the loading acyltransferase from avermectin modular polyketide synthase*. ACS Chem Biol, 2015. **10**(4): p. 1017-25.
48. Davey, S., *2015 Nobel Prize in Physiology or Medicine: Punishing parasites*. Nat Chem, 2015. **7**(12): p. 949.
49. Barragry, T.B., *A review of the pharmacology and clinical uses of ivermectin*. Can Vet J, 1987. **28**(8): p. 512-7.
50. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork*. Nat Prod Rep, 2007. **24**(1): p. 162-90.
51. Weng, J.K. and J.P. Noel, *Structure-function analyses of plant type III polyketide synthases*. Methods Enzymol, 2012. **515**: p. 317-35.

52. Tobert, J.A., *Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors*. Nat Rev Drug Discov, 2003. **2**(7): p. 517-26.
53. Zinedine, A., et al., *Review on the toxicity, occurrence, metabolism, detoxification, regulations and intake of zearalenone: an oestrogenic mycotoxin*. Food Chem Toxicol, 2007. **45**(1): p. 1-18.
54. Stockmann-Juvala, H. and K. Savolainen, *A review of the toxic effects and mechanisms of action of fumonisin B1*. Hum Exp Toxicol, 2008. **27**(11): p. 799-809.
55. Hamid, A.S., et al., *Aflatoxin B1-induced hepatocellular carcinoma in developing countries: Geographical distribution, mechanism of action and prevention*. Oncol Lett, 2013. **5**(4): p. 1087-1092.
56. Cox, R.J., *Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets*. Org Biomol Chem, 2007. **5**(13): p. 2010-26.
57. Chiang, Y.M., et al., *Unraveling polyketide synthesis in members of the genus Aspergillus*. Appl Microbiol Biotechnol, 2010. **86**(6): p. 1719-36.
58. Linnemannstons, P., et al., *The polyketide synthase gene pks4 from Gibberella fujikuroi encodes a key enzyme in the biosynthesis of the red pigment bikaverin*. Fungal Genet Biol, 2002. **37**(2): p. 134-48.
59. Newman, A.G. and C.A. Townsend, *Molecular Characterization of the Cercosporin Biosynthetic Pathway in the Fungal Plant Pathogen Cercospora nicotianae*. J Am Chem Soc, 2016.
60. Parascandolo, J.S., et al., *Insights into 6-Methylsalicylic Acid Bio-assembly by Using Chemical Probes*. Angew Chem Int Ed Engl, 2016. **55**(10): p. 3463-7.
61. Townsend, C.A., *Aflatoxin and deconstruction of type I, iterative polyketide synthase function*. Nat Prod Rep, 2014. **31**(10): p. 1260-5.
62. Zhang, W., Y. Li, and Y. Tang, *Engineered biosynthesis of bacterial aromatic polyketides in Escherichia coli*. Proc Natl Acad Sci U S A, 2008. **105**(52): p. 20683-8.
63. Crawford, J.M., et al., *Deconstruction of iterative multidomain polyketide synthase function*. Science, 2008. **320**(5873): p. 243-6.
64. Crawford, J.M., et al., *Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase*. Proc Natl Acad Sci U S A, 2006. **103**(45): p. 16728-33.
65. Crawford, J.M., et al., *Starter unit specificity directs genome mining of polyketide synthase pathways in fungi*. Bioorg Chem, 2008. **36**(1): p. 16-22.

66. Huitt-Roehl, C.R., et al., *Starter unit flexibility for engineered product synthesis by the nonreducing polyketide synthase PksA*. ACS Chem Biol, 2015. **10**(6): p. 1443-9.
67. Newman, A.G., et al., *Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function*. Chem Commun (Camb), 2012. **48**(96): p. 11772-4.
68. Zhou, H., Y. Li, and Y. Tang, *Cyclization of aromatic polyketides from bacteria and fungi*. Nat Prod Rep, 2010. **27**(6): p. 839-68.
69. Korman, T.P., et al., *Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis*. Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6246-51.
70. Tang, Y., et al., *The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase*. Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11124-9.
71. Piel, J., *Biosynthesis of polyketides by trans-AT polyketide synthases*. Nat Prod Rep, 2010. **27**(7): p. 996-1047.
72. Hertweck, C., *Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology*. Trends Biochem Sci, 2015. **40**(4): p. 189-99.
73. Tsai, S.C., et al., *Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel*. Proc Natl Acad Sci U S A, 2001. **98**(26): p. 14808-13.
74. Bretschneider, T., et al., *Vinylogous chain branching catalysed by a dedicated polyketide synthase module*. Nature, 2013. **502**(7469): p. 124-8.
75. Zheng, J., et al., *Divergence of multimodular polyketide synthases revealed by a didomain structure*. Nat Chem Biol, 2012. **8**(7): p. 615-21.
76. Dutta, S., et al., *Structure of a modular polyketide synthase*. Nature, 2014. **510**(7506): p. 512-7.
77. Lomakin, I.B., Y. Xiong, and T.A. Steitz, *The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together*. Cell, 2007. **129**(2): p. 319-32.
78. Rajgarhia, V.B. and W.R. Strohl, *Minimal Streptomyces sp. strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis*. J Bacteriol, 1997. **179**(8): p. 2690-6.
79. Jakobi, K. and C. Hertweck, *A gene cluster encoding resistomycin biosynthesis in Streptomyces resistomycificus; exploring polyketide cyclization beyond linear and angucyclic patterns*. J Am Chem Soc, 2004. **126**(8): p. 2298-9.

80. Kang, H.S. and S.F. Brady, *Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols*. ACS Chem Biol, 2014. **9**(6): p. 1267-72.
81. Tang, Y., et al., *Ketosynthases in the initiation and elongation modules of aromatic polyketide synthases have orthogonal acyl carrier protein specificity*. Biochemistry, 2003. **42**(21): p. 6588-95.
82. Arthur, C.J., et al., *Self-malonylation is an intrinsic property of a chemically synthesized type II polyketide synthase acyl carrier protein*. Biochemistry, 2005. **44**(46): p. 15414-21.
83. Bisang, C., et al., *A chain initiation factor common to both modular and aromatic polyketide synthases*. Nature, 1999. **401**(6752): p. 502-5.
84. Javidpour, P., et al., *The determinants of activity and specificity in actinorhodin type II polyketide ketoreductase*. Chem Biol, 2013. **20**(10): p. 1225-34.
85. Pickens, L.B. and Y. Tang, *Oxytetracycline biosynthesis*. J Biol Chem, 2010. **285**(36): p. 27509-15.
86. Otten, S.L., K.J. Stutzman-Engwall, and C.R. Hutchinson, *Cloning and expression of daunorubicin biosynthesis genes from Streptomyces peucetius and S. peucetius subsp. caesius*. J Bacteriol, 1990. **172**(6): p. 3427-34.
87. Mcdaniel, R., et al., *Engineered Biosynthesis of Novel Polyketides - Act(Vii) and Act(Iv) Genes Encode Aromatase and Cyclase Enzymes, Respectively*. Journal of the American Chemical Society, 1994. **116**(24): p. 10855-10859.
88. Osbourn, A.E. and R.J. Goss, *Natural products : discourse, diversity and design*. 1 online resource.
89. Ames, B.D., et al., *Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides*. Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5349-54.
90. Ames, B.D., et al., *Structural and biochemical characterization of Zhul aromatase/cyclase from the R1128 polyketide pathway*. Biochemistry, 2011. **50**(39): p. 8392-406.
91. Caldara-Festin, G., et al., *Structural and functional analysis of two di-domain aromatase/cyclases from type II polyketide synthases*. Proc Natl Acad Sci U S A, 2015. **112**(50): p. E6844-51.
92. Das, A. and C. Khosla, *In vivo and in vitro analysis of the hedamycin polyketide synthase*. Chem Biol, 2009. **16**(11): p. 1197-207.

93. Shen, Y., et al., *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, 1999. **96**(7): p. 3622-7.
94. Florez, A.B., et al., *Transcriptional regulation of mithramycin biosynthesis in Streptomyces argillaceus: dual role as activator and repressor of the PadR-like regulator MtrY*. Microbiology, 2015. **161**(Pt 2): p. 272-84.
95. Korman, T.P., et al., *Structural analysis of actinorhodin polyketide ketoreductase: cofactor binding and substrate specificity*. Biochemistry, 2004. **43**(46): p. 14529-38.
96. Lackner, G., et al., *Biosynthesis of pentangular polyphenols: deductions from the benastatin and griseorhodin pathways*. J Am Chem Soc, 2007. **129**(30): p. 9306-12.
97. Paananen, P., et al., *Structural and functional analysis of angucycline C-6 ketoreductase LanV involved in landomycin biosynthesis*. Biochemistry, 2013. **52**(31): p. 5304-14.
98. Javidpour, P., et al., *Structural and biochemical analyses of regio- and stereospecificities observed in a type II polyketide ketoreductase*. Biochemistry, 2011. **50**(21): p. 4638-49.
99. Javidpour, P., et al., *Structural and biochemical studies of the hedamycin type II polyketide ketoreductase (HedKR): molecular basis of stereo- and regiospecificities*. Biochemistry, 2011. **50**(34): p. 7426-39.
100. Keatinge-Clay, A.T., et al., *An antibiotic factory caught in action*. Nat Struct Mol Biol, 2004. **11**(9): p. 888-93.
101. Patrikainen, P., et al., *Structure-based engineering of angucyclinone 6-ketoreductases*. Chem Biol, 2014. **21**(10): p. 1381-91.
102. Patrikainen, P., et al., *Tailoring enzymes involved in the biosynthesis of angucyclines contain latent context-dependent catalytic activities*. Chem Biol, 2012. **19**(5): p. 647-55.
103. Kallio, P., et al., *Sequential action of two flavoenzymes, PgaE and PgaM, in angucycline biosynthesis: chemoenzymatic synthesis of gaudimycin C*. Chem Biol, 2008. **15**(2): p. 157-66.
104. Gullon, S., et al., *Isolation, characterization, and heterologous expression of the biosynthesis gene cluster for the antitumor anthracycline steffimycin*. Appl Environ Microbiol, 2006. **72**(6): p. 4172-83.
105. Tsukamoto, N., et al., *Nucleotide sequence of the aknA region of the aklavinone biosynthetic gene cluster of Streptomyces galilaeus*. J Bacteriol, 1994. **176**(8): p. 2473-5.
106. Xu, Z., et al., *Biosynthesis of the antitumor agent chartreusin involves the oxidative rearrangement of an anthracyclic polyketide*. Chem Biol, 2005. **12**(5): p. 579-88.

107. Dickens, M.L., J. Ye, and W.R. Strohl, *Cloning, sequencing, and analysis of aklaviketone reductase from Streptomyces sp. strain C5*. J Bacteriol, 1996. **178**(11): p. 3384-8.
108. Strieker, M., A. Tanovic, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics*. Curr Opin Struct Biol, 2010. **20**(2): p. 234-40.
109. Challis, G.L. and J.H. Naismith, *Structural aspects of non-ribosomal peptide biosynthesis*. Curr Opin Struct Biol, 2004. **14**(6): p. 748-56.
110. Tanovic, A., et al., *Crystal structure of the termination module of a nonribosomal peptide synthetase*. Science, 2008. **321**(5889): p. 659-63.
111. Haslinger, K., et al., *X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis*. Nature, 2015. **521**(7550): p. 105-9.
112. Felnagle, E.A., et al., *Nonribosomal peptide synthetases involved in the production of medically relevant natural products*. Mol Pharm, 2008. **5**(2): p. 191-211.
113. Keating, T.A., et al., *The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains*. Nat Struct Biol, 2002. **9**(7): p. 522-6.
114. Rausch, C., et al., *Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution*. BMC Evol Biol, 2007. **7**: p. 78.
115. Finking, R. and M.A. Marahiel, *Biosynthesis of nonribosomal peptides I*. Annu Rev Microbiol, 2004. **58**: p. 453-88.
116. Bloudoff, K., D. Rodionov, and T.M. Schmeing, *Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases*. J Mol Biol, 2013. **425**(17): p. 3137-50.
117. Samel, S.A., et al., *Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase*. Structure, 2007. **15**(7): p. 781-92.
118. Kittila, T. and M.J. Cryle, *Capturing the Structure of the Substrate Bound Condensation Domain*. Cell Chem Biol, 2016. **23**(3): p. 315-6.
119. Bloudoff, K., D.A. Alonzo, and T.M. Schmeing, *Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases*. Cell Chem Biol, 2016. **23**(3): p. 331-9.
120. Boettger, D. and C. Hertweck, *Molecular diversity sculpted by fungal PKS-NRPS hybrids*. Chembiochem, 2013. **14**(1): p. 28-42.

121. Du, L., C. Sanchez, and B. Shen, *Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules*. Metab Eng, 2001. **3**(1): p. 78-95.
122. Foulke-Abel, J. and C.A. Townsend, *Demonstration of starter unit interprotein transfer from a fatty acid synthase to a multidomain, nonreducing polyketide synthase*. Chembiochem, 2012. **13**(13): p. 1880-4.
123. Song, L., et al., *Cytochrome P450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis*. J Antibiot (Tokyo), 2014. **67**(1): p. 71-6.
124. Choi, J.W. and N.A. Da Silva, *Improving polyketide and fatty acid synthesis by engineering of the yeast acetyl-CoA carboxylase*. J Biotechnol, 2014. **187**: p. 56-9.
125. Arabolaza, A., et al., *Crystal structures and mutational analyses of acyl-CoA carboxylase beta subunit of Streptomyces coelicolor*. Biochemistry, 2010. **49**(34): p. 7367-76.
126. Tong, L., *Structure and function of biotin-dependent carboxylases*. Cell Mol Life Sci, 2013. **70**(5): p. 863-91.
127. Rodriguez, E. and H. Gramajo, *Genetic and biochemical characterization of the alpha and beta components of a propionyl-CoA carboxylase complex of Streptomyces coelicolor A3(2)*. Microbiology, 1999. **145** (Pt 11): p. 3109-19.
128. Huang, C.S., et al., *An unanticipated architecture of the 750-kDa alpha6beta6 holoenzyme of 3-methylcrotonyl-CoA carboxylase*. Nature, 2012. **481**(7380): p. 219-23.
129. Jurado, A.R., et al., *Structure and substrate selectivity of the 750-kDa alpha6beta6 holoenzyme of geranyl-CoA carboxylase*. Nat Commun, 2015. **6**: p. 8986.
130. Tong, L., *Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery*. Cell Mol Life Sci, 2005. **62**(16): p. 1784-803.
131. Diacovich, L., et al., *Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity*. Biochemistry, 2004. **43**(44): p. 14027-36.
132. Cronan, J.E., Jr., *The biotinyl domain of Escherichia coli acetyl-CoA carboxylase. Evidence that the "thumb" structure is essential and that the domain functions as a dimer*. J Biol Chem, 2001. **276**(40): p. 37355-64.
133. Cronan, J.E., Jr. and G.L. Waldrop, *Multi-subunit acetyl-CoA carboxylases*. Prog Lipid Res, 2002. **41**(5): p. 407-35.

134. Blanchard, C.Z. and G.L. Waldrop, *Overexpression and kinetic characterization of the carboxyltransferase component of acetyl-CoA carboxylase*. J Biol Chem, 1998. **273**(30): p. 19140-5.
135. Huang, C.S., et al., *Crystal structure of the alpha(6)beta(6) holoenzyme of propionyl-coenzyme A carboxylase*. Nature, 2010. **466**(7309): p. 1001-5.
136. Wilson, M.C. and B.S. Moore, *Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity*. Nat Prod Rep, 2012. **29**(1): p. 72-86.
137. Chan, Y.A., et al., *Biosynthesis of polyketide synthase extender units*. Nat Prod Rep, 2009. **26**(1): p. 90-114.
138. Williams, G.J., *Engineering polyketide synthases and nonribosomal peptide synthetases*. Curr Opin Struct Biol, 2013. **23**(4): p. 603-12.
139. McDaniel, R., et al., *Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products*. Proc Natl Acad Sci U S A, 1999. **96**(5): p. 1846-51.
140. Zhang, L., et al., *Rational Control of Polyketide Extender Units by Structure-Based Engineering of a Crotonyl-CoA Carboxylase/Reductase in Antimycin Biosynthesis*. Angew Chem Int Ed Engl, 2015. **54**(45): p. 13462-5.
141. Erb, T.J., et al., *Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase*. Proc Natl Acad Sci U S A, 2009. **106**(22): p. 8871-6.

CHAPTER 2: Structural Insights into Chain Length and Cyclization Specificities of Bacterial Type II Polyketide Intermediates in an Aromatase/Cyclase Domain

2.1 Abstract

Ring formation is a crucial biosynthetic operation for diversifying natural products such as polyketides, which are known for their abundance of diverse biological activity with a wide variety of therapeutic value. The aromatase/cyclase (ARO/CYC) domain from type II PKS promotes cyclization through the highly-specific folding and alignment of a nascent linear polyketone intermediate. Although previous efforts have provided crystal structure and mutagenesis data from nearly all classes of ARO/CYC, these structures collected to date lack a bound substrate. These studies have identified residues essential to ARO/CYC function; however, how the ARO/CYC interact with polyketide substrates and cyclized intermediates remains unknown. The major technical hurdle lies in the availability of the putative polyketide substrates, which remain highly reactive. Herein we present the first report of co-crystal structures of the WhiE ARO/CYC bound to series of atom-replaced mimics. In particular, we have established structures the WhiE ARO/CYC bound to tetraketide, hexaketide, heptaketide and octaketide mimics, which contain both the linear polyketide isosterers and pantetheine groups. The four cocrystal structures reveal how an ARO/CYC traps an immature linear polyketide outside of its substrate-binding pocket or misorients the immature chain near the active site, and how the mature linear substrate is properly oriented for polyketide cyclization. These findings provide an important structural platform to tailor ARO/CYCs with the goal of engineering the biosynthesis of new polyketides with diversified cyclization patterns.

2.2 Introduction

Polyketides are an ever-expanding class of natural products with a high degree of commercial and therapeutic value.[1] This underscores the potential marketability and importance of polyketides in healthcare. The rise in rates of cancer, heart disease, and antibiotic resistant pathogens is an increasing problem, leading to an augmented demand for new therapeutics. Both natural and synthetically-derived polyketide leads are discovered on a continual basis, strengthening the promise that this extensive and still-growing class of compounds may contain viable solutions to these problems.[2, 3] The number of publications on polyketides produced in the past decade alone scores over 5000, reinforcing the notion that there is a high degree of untapped potential in the area of polyketide design.

Polyketide synthases (PKS) are multi-enzyme complexes responsible for the biosynthesis of polyketide natural products. PKSs fall into three distinct categories: Types I, II, and III.[1] This study focuses on the aromatase/cyclase (ARO/CYC) domain of type II PKS (Figure 2.1). Type II PKSs produce an array of conjugated aromatic natural products proven to possess important pharmaceutical value. These include: tetracycline (an antibiotic) and daunorubicin (a chemotherapeutic agent). Type II PKSs incorporate unique starter units (such as propionyl-CoA) and building blocks (most commonly in the form of malonyl-CoA) in combination with a wide variety of reduction and cyclization reactions to achieve a unique structure and its associated biological activity. Both domain-domain and protein-substrate interactions are important to guide polyketide production.[4, 5] Type II PKSs are composed of several stand-alone enzymes that act together in a highly coordinated fashion to iteratively elongate and cyclize a linear polyketide chain.[6] The priming of the ketosynthase with an acyl starter unit is the first step in type II PKS biosynthesis. The polyketide chain is sequentially extended and modified with malonyl units

through concerted interactions with the malonyl-CoA:ACP transacylase (MAT), acyl carrier protein (ACP), ketosynthase/chainlength factor (KS/CLF), ketoreductase (KR), and aromatase/cyclase (ARO/CYC) (Figure. 2.1). The growing polyketide chain intermediate is tethered to the phosphopantetheine (PPant) prosthetic group of ACP during each step of polyketide chain elongation and modification.[7-9] The precise timing and coordination of ACP with the downstream processing domains is essential for the successful production of type II polyketides. However, the interactions between ACP and its partners, in particular the ARO/CYC, are not well understood.

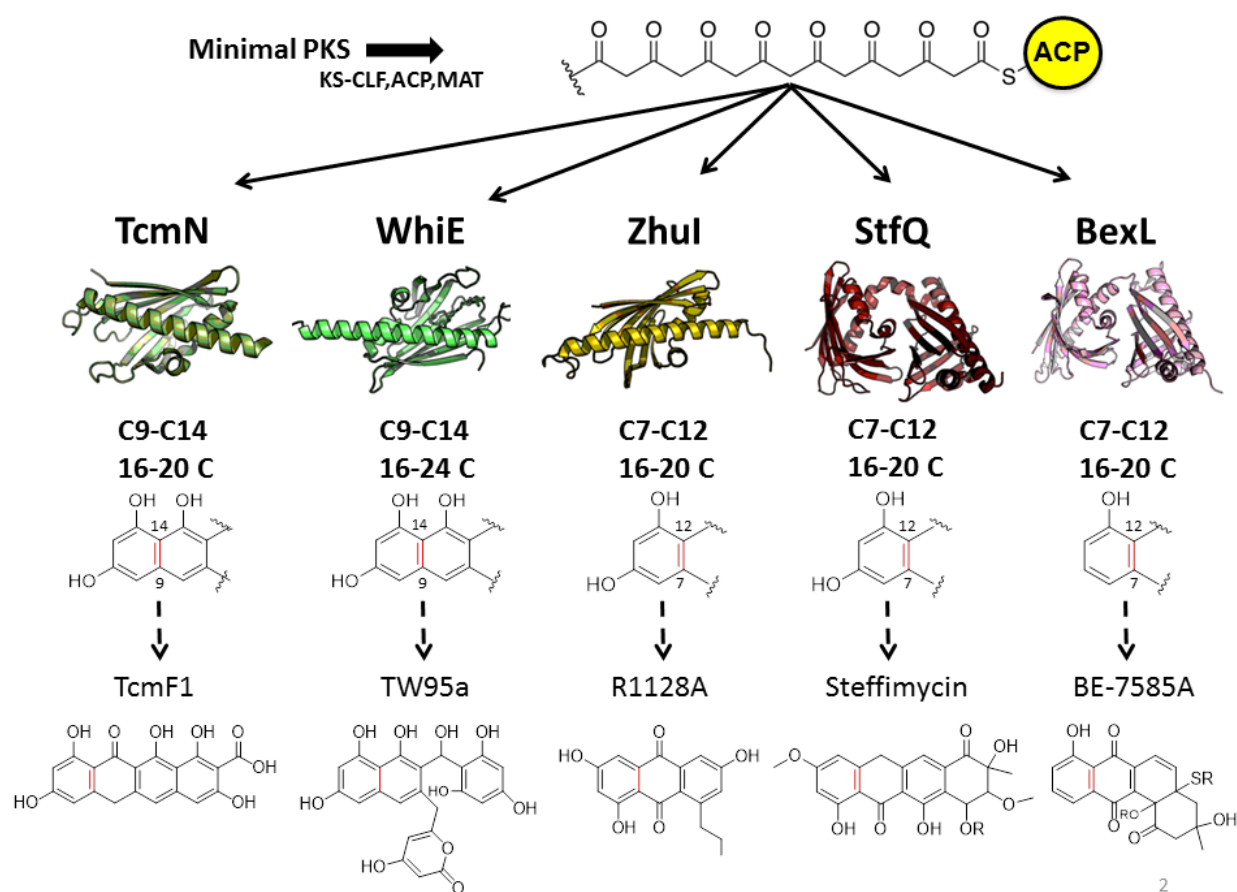


Figure 2.1. The different clades of type II PKS ARO/CYC and their natural products.

From left to right, WhiE and TcmN are representatives of non-reducing C9-C14 ARO/CYCs. ZhuI, BexL, and StfQ are members of the C7-C12 clades of ARO/CYC. All *apo* structures of ARO/CYCs were reported by this group.[11-13, 21, 22]

Structural efforts in recent years have resulted in the solution of crystal structures from virtually every class of ARO/CYCs that vary in their substrate specificity and cyclization patterns (Figure 2.1). However, these structures are *apo*; in lacking a bound substrate, the mechanism in which the ARO/CYCs maintain a highly regio-specific cyclization motif while simultaneously influencing the product chain length remains unknown.[9-13] While docking studies of ARO/CYCs have predicted essential active site catalytic residues, the role of auxiliary residues necessary for substrate binding and chain length determination remain hypothetical and partially undetermined.[7, 12-14] Furthermore, the exact role of residues that define the interior of each active site pocket in determining the final cyclization pattern has yet to be fully elucidated. Therefore, understanding the protein-substrate interactions involved in regio-specific cyclization in the ARO/CYC active site, coupled with knowledge of the protein-protein interactions between the ACP and ARO/CYC domains, would have a high impact for future efforts in engineering type II polyketide natural products.[15, 16] However, the transient nature of the cyclization reaction intermediates and the lack of stable linear polyketone intermediates bound to the phosphopantetheine (PPant) arm of ACP has proven to be a great hurdle.

Recently, we reported a method that designs and synthesizes chemically stable pantetheine-linked polyketide isosteric mimics (Figure 2.2).[17] As part of the program to explore the use of these materials probe for the selectivity in substrate binding pockets of PKS domains, herein we present the cocrystal structures of the WhiE ARO/CYC bound with four different chain-length mimics, with the goal of understanding the role of chain length and cyclization specificity in the associated ARO/CYC domains. To the best of our knowledge, this is the first comprehensive structural study of substrate-ARO/CYC interactions through the implementation of a robust panel of pantetheine-linked polyketide mimics. These substrate

mimic-bound cocrystal structures provide an important first step in developing a structural understanding as to how a PKS domain can set cyclization pattern into non-cognate biosynthetic pathways. Understanding this information is essential for expanding both the structural diversity and the biological activity of type II PKS natural products.

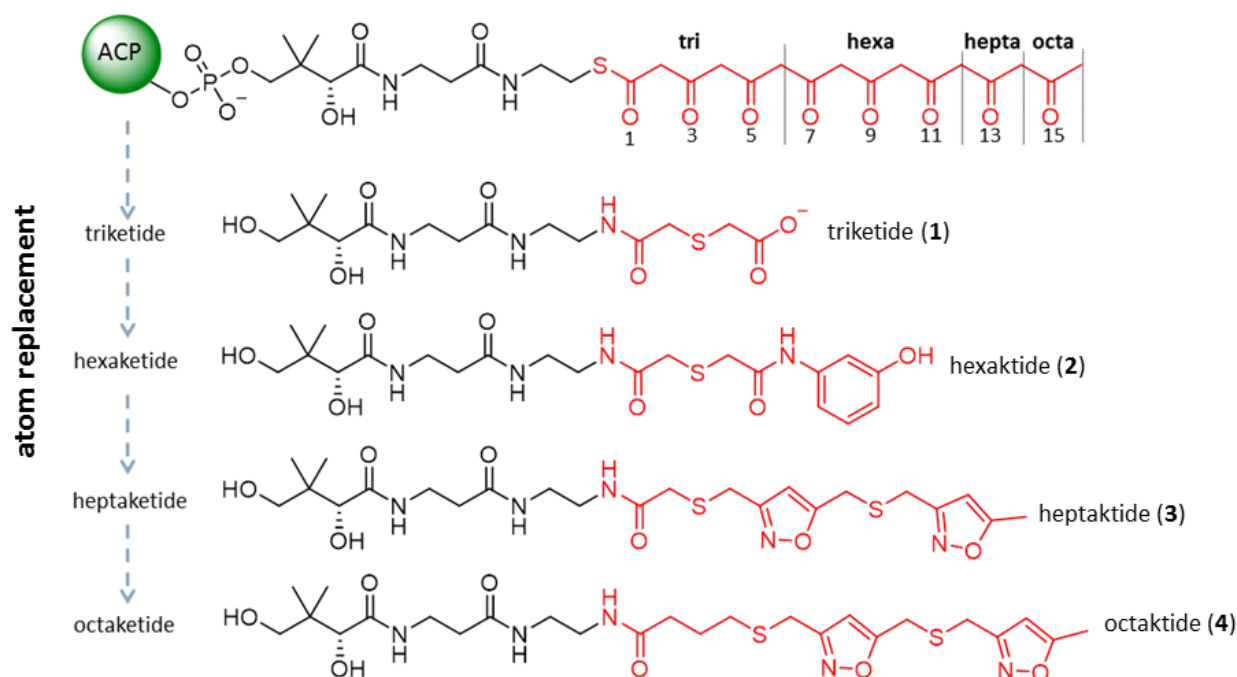


Figure 2.2. The chemical structures of the four polyketide mimics used.

These mimics used to obtain WhiE ARO/CYC cocrystal structures consist of triketide **1**, hexaketide **2**, heptaketide **3** and octaketide **4**.

2.3 Results and Discussion

Rationale for designing polyketide isosteric mimics 1-4 and cocrystallization with WhiE ARO/CYC.

Knowledge regarding PKS substrate-enzyme interactions and processivity has been impeded by the lack of stable aromatic polyketide intermediates. Due to the inherent tendency to undergo intra- or inter-molecular Aldol/Claisen condensations, long polyketone chains cannot be

feasibly used for cocrystal or other PKS structural studies. Furthermore, the direct chemical syntheses of aromatic polyketide intermediates have also been difficult to achieve, thereby limiting PKS mechanistic studies. Here we apply four pantetheine-linked mimics as innovative tools to probe ARO/CYC function, because they represent the closest substrate intermediates for ARO/CYCs to date (Figure 2.2). Compounds **1-4** resemble the polyketone backbone of a natural polyketide; however, the carbonyls are strategically substituted with N and S atoms to prevent spontaneous inter- or intra-molecular reactions. Additional modifications, such as isoxazole and phenol rings, have been installed to mirror diketides and aromatic rings that are formed during different steps of iterative PKS biosynthesis. These atom-replaced mimics are sufficiently stable for use in cocrystallization and *in vitro* assays, which can be used to dissect the substrate preferences of a specific PKS domain.

An array of pantetheine-linked analogues were synthesized by Dr. Gaurav Shakya, ranging from 6 to 16 carbons in the polyketide moiety.[17] and they were used to assess ARO/CYC substrate-interactions (Figure 2.2). After extensive crystals trials, optimizations, and multiple rounds of data collection, we have successfully applied these atom-replaced mimics to obtain the first ligand-bound cocrystal structures of ARO/CYC. The pantetheine (Pant)-linked analogues have different carbon chain lengths: triketide (6 carbons, **1**), hexaketide (12 carbons, **2**), heptaketide (14 carbons, **3**), and octaketide (16 carbons, **4**). The heptaketide **3** and octaketide **4** cocrystal structures posed a challenge to obtain because of increased protein precipitation upon incubation with these compounds. In total, over 100 crystals underwent diffraction analysis and 30 data sets for crystals containing compounds **3** and **4** were collected before these cocrystal structures could be solved to high resolution ($< 2.8 \text{ \AA}$, Table 2.1).

The cocrystal structure of WhiE bound with the six-carbon triketide mimic demonstrates that chain length is a stringent criterion for entrance into WhiE's active site pocket.

The cocrystal structure of WhiE bound to a 6-carbon triketide **1** was solved to 1.85 Å (Figure 2.3B). Similar to the *apo*-structure [11], the triketide-WhiE crystals have a P2₁ space group, with one monomer per asymmetric unit. The triketide-WhiE structure maintains the same tertiary structure as that of the apo form, with an RMSD of 0.1 Å. Try63 and Trp65 of the active site, however, exhibit minor perturbations and are shifted ~0.5 Å toward the triketide analogue (Figure 2.4A, 2.5A). These concerted movements compress the interior pocket. Strikingly, instead of entering the interior pocket, the triketide **1** threads along the exterior of the WhiE pocket and is sandwiched between Trp63 and Trp65 (Figure 2.3A). The carbonyl group of **1** forms a single hydrogen bond contact with Arg82, locking the compound in place (Figure 2.4A). Several aromatic compounds have been reported to form π - π stacking interactions with the surrounding tryptophans in ARO/CYCs [11, 12]; however, this is the first observation that a polyketide analog with an immature chain length (only six carbons) indeed does not enter the ARO/CYC pocket.

Two solvent exposed residues, Asp59 and Asp57, are located near the surface of the substrate pocket entrance. These two acidic residues are in close proximity with **1**. The presence of two negatively charged residues in close proximity to one another near the entrance of substrate pocket provides an unfavorable environment for the negatively charged **1** to enter the interior pocket of WhiE. Consequently, the triketide moiety of **1** is oriented toward the solvent to minimize contacts with Asp59 and Asp57.

The information gleaned from the cocrystal structure of WhiE bound to the triketide mimic raised the question about role of these two negatively-charged residues as the “gate

keeper” residues that prevent immature polyketide substrates (such as the triketide) from entering the ARO/CYC pocket. Sequence and structural alignments revealed that Asp 59 is highly conserved across mono-domain ARO/CYCs, whereas Asp57 is Asn in TcmN, ZhuI and ARX 19 (Figure 2.6). The conservation of these residues suggests that Asp59 and Asp/Asn57 are additional auxiliary residues that may serve as substrate gatekeepers.

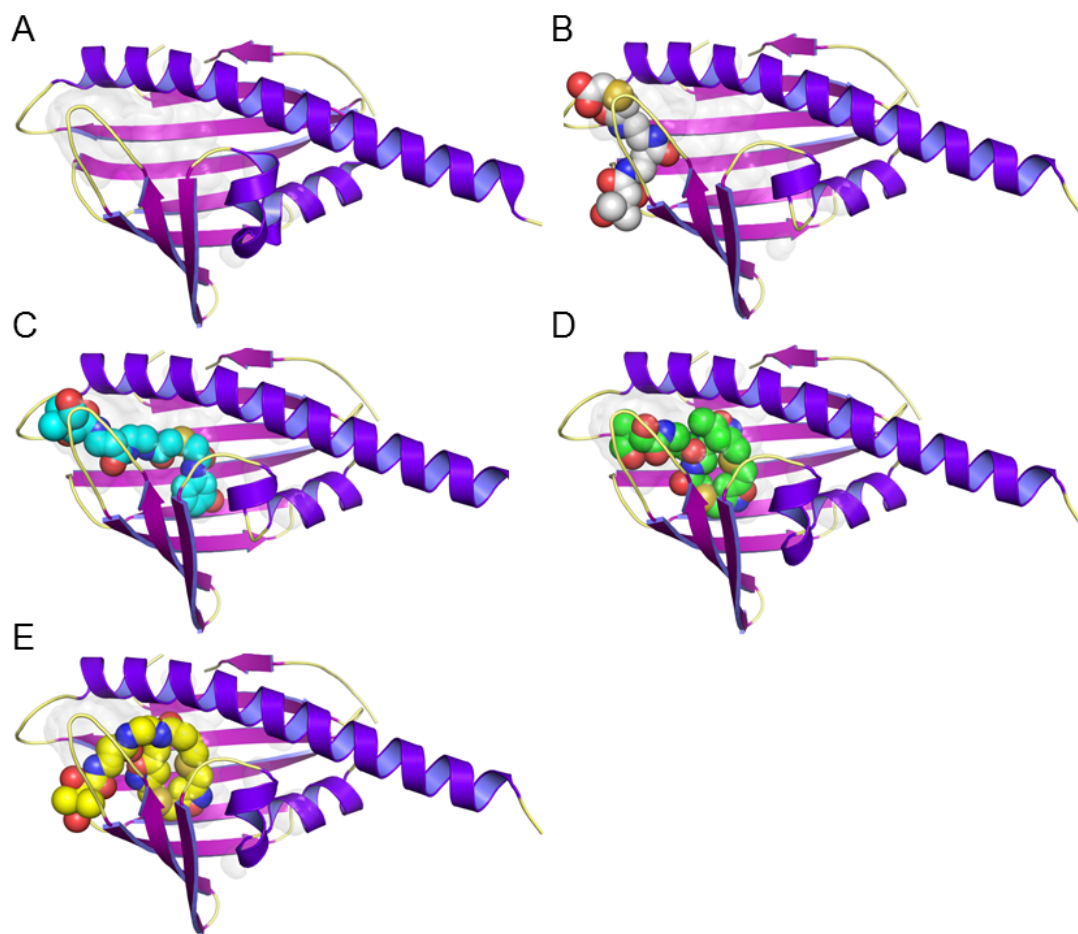


Figure 2.3. The overall architecture of WhiE ARO/CYC.

Shown here are (A) *apo*-WhiE, (B) triketide **2** bound WhiE, (C) hexaketide **3** bound WhiE, (D) heptaketide **4** bound WhiE, and (E) heptaketide **5** bound WhiE. All five structures are shown in the same orientation. The ketide mimics are shown in spheres. WhiE binds the triketide **1** on its surface, while hexaketide **2**, heptaketide **3** and octaketide **4** occupy the substrate pocket.

Because type II PKS involves separate enzymes, finely choreographed interactions with a full-length substrate and highly specific protein-protein interactions are essential to control the product outcome.[7, 9, 16] Therefore, it is well known that *in vivo* ARO/CYCs only cyclize mature polyketide chains that reach their full-lengths, ranging from 16 to 30-carbons. It is not surprising, therefore, that WhiE would occlude a short (immature) precursor mimic with only 6 carbons (a triketide) from its active site pocket. Based on the triketide-WhiE structure, we conclude that WhiE is able to trap the immature chain outside of the pocket, so that only substrates with a mature polyketide chain length can enter the cyclization pocket.

Three cocrystal structures of WhiE bound with hexa-, hepta- and octaketide mimics show that the binding modes of different polyketide chain lengths are determinants of a given cyclization pattern.

Next, we solved the cocrystal structures of WhiE bound to the mimics of hexaketide **2** (Figures 2.3B and 2.4B), heptaketide **3** (Figure 2.3C and 2.4C), and octaketide **4** (Figures 2.3D and 2.4D), which were solved to 1.94, 1.85 and 1.90 Å, respectively. These structures help visualize the binding modes of longer polyketide intermediates. Similar to the triketide **1**-WhiE structure, these three structures were solved in P2₁ space group with one monomer per asymmetric unit. All three structures have little conformational change in comparison to the *apo*-structure with an averaged RMSD of 0.1 Å.

The cocrystal structure of WhiE with hexaketide mimic 2 showed that the hexaketide chain was misoriented.

The cocrystal structure of hexaketide **2**-WhiE was solved to 1.9 Å (Figures 2.3B and 2.4B). Mimic **2** contains both the pantetheine moiety and a monocyclic hexaketide (Figure 2.2). Well-defined electron density of **2** was found in the interior pocket of WhiE. Mimic **2** was observed bent near its terminal sulfur group, and its phenolic hydroxyl group forms a hydrogen bond with Arg69. The binding of **2** in the WhiE pocket is stabilized by both polar and nonpolar interactions (Figure 2.4B). Specifically, Arg82 interacts with the carbonyl group of pantetheine in **2** *via* both hydrogen bonds and coordination of a crystalline water molecule that may serve as a hydrogen bond acceptor. The thioether sulfur in the position in the C3 position forms direct hydrogen bond contacts with Gln110. The carbonyl oxygen in the C5 position and the amide nitrogen 3 of **2** were also coordinated by two crystalline water molecules that form hydrogen bonds with Trp65 and Ser67, respectively (Figure 2.4B). Tyr35 further forms a hydrogen bond with the amide N6 of **2**. Finally, the N1 of Trp95 forms a hydrogen bond with the C5 carbonyl oxygen of **2**, further anchoring the hexaketide into the interior pocket of WhiE. These interactions stabilize **2** in the interior pocket of WhiE.

Although the electron density of **2** is well-defined in the interior pocket of WhiE, it does not display a productive binding mode. The probe adopts a rigid, non-productive conformation in the interior pocket of WhiE. Specifically, Gln110 contacts with the thioether sulfur of **2**, which significantly decreases the flexibility of **2** inside the pocket. Probe **2** also has extensive interactions with Trp95 and R69, two highly conserved residues in mono-domain ARO/CYCs that have been shown to be crucial for catalysis. The strong hydrogen bonds between **2** and Trp65-R69 restrict the probe from free movements in the interior pocket. These findings imply

that unproductive substrate binding of premature polyketide chains plays an important role in preventing aberrant cyclization events.

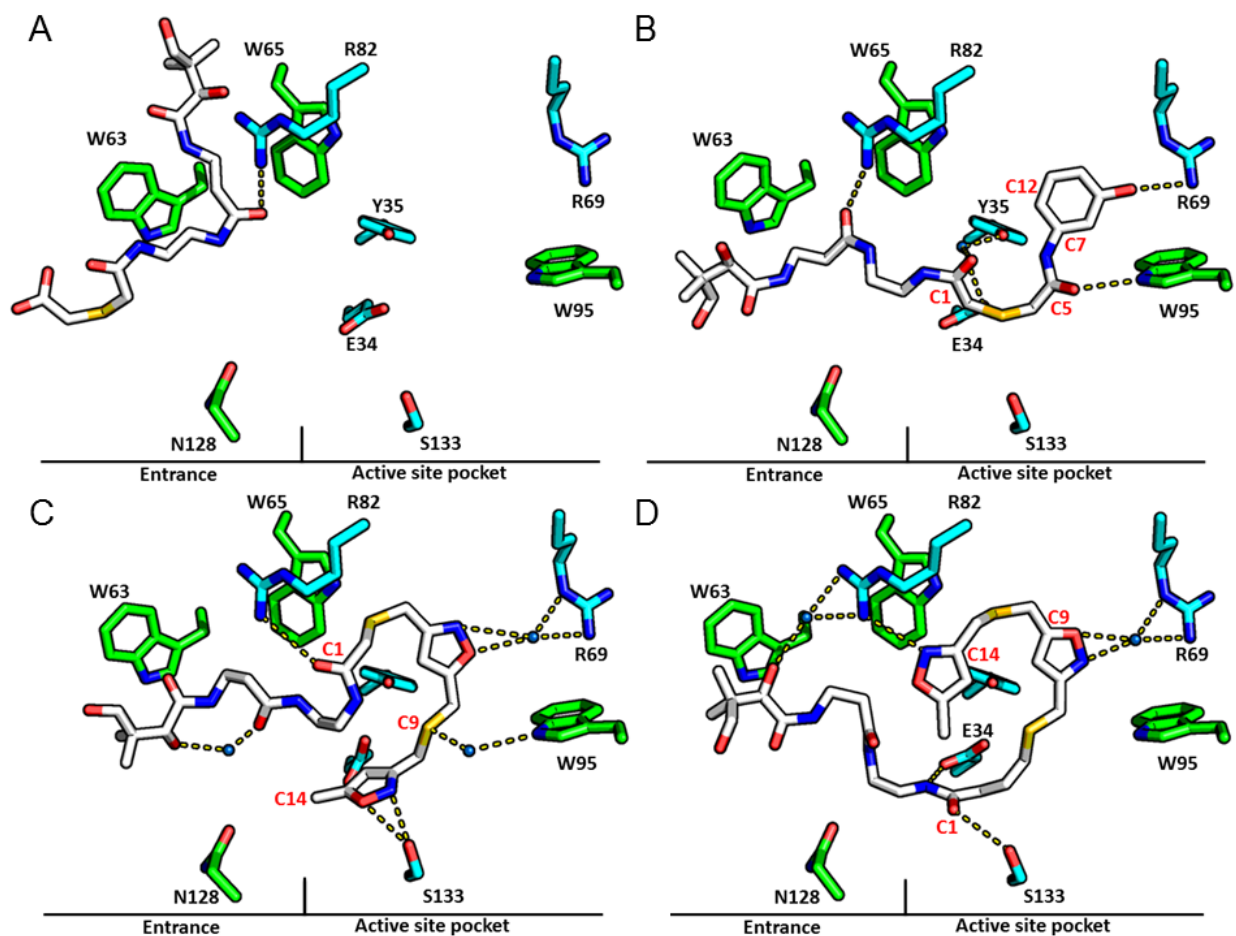


Figure 2.4. An inspection of active site residues of WhiE and interactions with the corresponding pantetheine linked mimic.

Known catalytic residues are highlighted in blue with structural residues in green. The mimics are shown in white. (A) The triketide **1** is trapped on the protein surface. (B) The hexaketide **2** is oriented in the substrate pocket to prevent access to the active site residues. (C) The heptaketide **3** is oriented in the substrate pocket to prevent cyclizations. (D) Only the octaketide **4**, a known substrate of WhiE, is oriented correctly, with all active site residues aligned, to conduct a C9-C14 first ring cyclization.

The cocrystal structure of WhiE with heptaketide mimic 3 was also misoriented

Mimic **3** is a linear heptaketide mimic linked to the pantetheine moiety. Well-defined electron density was detected in the WhiE interior pocket (Figure 2.4C). Unlike hexaketide **2**, the heptaketide **3** adopts a kinked U-shape in the interior pocket, thus placing the first and second isoxazole rings within close proximity of Arg69 and Ser132, respectively. However, in the cocrystal structure of **3**-WhiE, the polyketide C9 and C14 are 6.1 Å apart, which is too long a distance to promote intra-molecular aldol cyclization between C9 and C14. Further, the bending of **3** is achieved through several key interactions in the active site. Arg69 mediates the formation of hydrogen bonds with the first isoxazole ring through a tightly coordinated water. Arg82 then helps orient **3** via a hydrogen bond with the C1 carbonyl oxygen of **3**. A network of hydrogen bonds to the second isoxazole ring further stabilizes the bent conformation of **3** with Ser132 and a water molecule that interacts with Gln110. Consequently, the C9 and C14 of polyketide in **3** is oriented away from both Arg69 and Arg82, two key catalytic residues that promote cyclization. We found that the orientation of the heptaketide **3** was very different to that of the hexaketide **2** inside the binding pocket (Figure 2.4B-C). However, both the hexaketide **2** and heptaketide **3** are ‘misoriented’ in the substrate pocket, and no cyclization can occur due to the misorientation. This is consistent with the observation that WhiE ARO/CYC cannot cyclize a hexaketide or heptaketide substrate *in vivo*.^[11]

The misorientation of **3** can be important for the observed chain length specificity of WhiE. It is well known that long, linear poly-beta-ketone chains are highly reactive and are unstable in solution. Instead of trapping the heptaketide mimic **3** on the exterior surface of the substrate binding pocket, WhiE may protect the immature, 14-carbon polyketide chain in the interior pocket with an unproductive binding mode, until the polyketide undergoes one more

round of chain elongation to 16 carbons, so that C9-C14 and C7-C16 cyclizations can occur. This sequestration mechanism serves to prevent aberrant cyclization events that are energetically costly.

In summary, we found that the ARO/CYC pocket is capable of differentiating substrates with chain length and chemical moieties (such as the aryl ring of **2** and isoxazole ring of **3**), which can radically change the orientation of an incoming polyketide substrate, leading to different ways of folding the linear chain and thus different cyclization patterns.

The cocrystal structure of WhiE with octaketide mimic 4 identified the correct orientation for cyclization

We observed well-defined electron density of the octaketide mimic **4** in the active site of WhiE, where it has a completely different conformation from those of **2** and **3** discussed above (Figure 2.4D). In the cocrystal structure of heptaketide **3**-WhiE, the C1 carbonyl oxygen forms a hydrogen bond with Arg82, causing the second isoxazole ring to adopt an inverse confirmation that places the oxygen of the second isoxazole ring in close proximity to Ser132. For the octaketide **4**, the C1 carbonyl oxygen forms a hydrogen bond with Ser132 that allows the second isoxazole ring oxygen and its nitrogens to form hydrogen bonds with Arg82, resulting in the complete opposite orientation of the long polyketide chain, as compared with that of **3** (Figure 2.4C *versus* Figure 2.4D). Much like what was seen with **2** and **3**, Arg69 maintains a conserved role of anchoring the first isoxazole ring of the octaketide mimic into the active site pocket through coordination of a crystalline water molecule that mediates hydrogen bonds between the first isoxazole ring and Arg69. Other predicted catalytic residues Glu34 and Tyr35 lie within 3.2 Å of terminal amide nitrogen and 3.8 Å away from the second isoxazole ring of the octaketide,

respectively. Collectively, these interactions act to lock and stabilize the octaketide mimic in the active site of WhiE in an orientation distinct from that of **3**.

The cocrystal structure of **4**-WhiE showed that C9 and C14 are in close proximity, thus it represents the first snapshot of a polyketide ready to be cyclized by an ARO/CYC. The set of cocrystals structures from WhiE bound with **1-4** emphasizes the importance that polyketide chain length plays in guiding the specificity of cyclization events in the type II PKSs. In the case of C9-C14 ARO/CYCS such as WhiE, entrance to the active site can only be permitted if the nascent chain is fully mature and has grown to the appropriate length (namely greater than the heptaketide **3**). Further, the “correct” orientation for C9-C14 cyclization (as shown in Figure 2.4D with the octaketide **4**) can only be achieved with a longer chain length.

In type II PKS, how ACP, KS, CLF, MAT, and ARO/CYC interact is poorly understood [1, 6, 16]. HSQC NMR experiments have shown that the ACP surface is disturbed when bound to polyketide analogues of differing chain length from 6 carbons (triketide) to 16 carbons (octaketide), and that these perturbations may play a role in domain synchronization during natural product biosynthesis [18]. Presumably, during elongation steps, the *holo*-acyl ACP has a higher affinity toward the KS-CLF than downstream processing enzymes. Once proper chain length has been reached, ARO/CYC-ACP interactions are favored and polyketide processing is the next committed step.

The structure is also consistent with past reports that WhiE accept polyketide substrates with a chain length equal or longer than 16 carbons (an octaketide) [11]. It has been previously shown that WhiE ARO/CYC has a permissive *in vitro* substrate tolerance. Combinatorial experiments where WhiE was incubated with different minimal PKSs that produce polyketide chains of varying length (16 and 18 carbons) lead to a favored C9-C14 product ratio [11]. WhiE

therefore must act on these 16- and 18-carbon polyketides in similar fashion as it does with its longer, 24 carbon native substrate. There is currently no evidence that WhiE ARO/CYC can act on a polyketide chain shorter than 16 carbons. Based on the cocrystal structures of **1-4**, chain length appears to be the first criteria necessary for successful processing by WhiE ARO/CYC. The second mandatory rule for the proper installation of controlled cyclization motifs is a proper substrate-binding mode in the interior of the ARO/CYC active site pocket. A productive ARO/CYC binding mode is achieved through tight coordination of substrate-enzyme interactions between conserved active site residues and the mature polyketide chain tethered to the ACP (Figure 2.4D). The current rationale is that the positions of the catalytic residues in respect to the active site pocket and ligand are determinants of a given cyclization mode. This is supported by the above structural comparison of **1-4** (Figure 2.4), as well as comparing WhiE ARO/CYC with the C7-C12 ARO/CYC ZhuI, where the essential catalytic Asp 146 and H109 are orthogonal to WhiE's active site residues Tyr35 and Glu34. [12]. Therefore, it would appear that based on the above cocrystal structures of **1-4**, a productive binding event could only occur if the polyketide chain has grown to a sufficient length as to adopt the proper geometry necessary for cyclization catalysis within the active site of the ARO/CYC.

Biological Significance

The generation of new pharmaceuticals through the manipulation of biosynthetic enzymes is a major focal point in natural products research. However, the engineering of mega-synthases for the installation of specific functional groups and/or cyclization patterns requires comprehensive knowledge of enzyme-substrate and enzyme-enzyme interactions, which are often too transient or chemically unstable to capture.

To address this issue, we have determined the cocrystal structures of WhiE ARO/CYC, a C9-C14 mono-domain ARO/CYC, bound to four synthetic substrate mimics for triketide, hexaketide, heptaketide, and octaketide mimics (**1-4**). Each cocrystal structure exhibits unique probe binding modes with WhiE ARO/CYC that has helped to further elucidate how ARO/CYCs recognize and orient the substrate. The shorter 6C mimic (**1**) was found to be localized on the exterior surface of the ARO/CYC substrate-binding, where **1** makes hydrogen bonds with R82 and may have electronic clashes with the negatively-charged Asp57 and Asp59. In contrast, the larger 12C, 14C and 16C mimics **2-4** were found in the substrate-binding pocket. Of all four structures, only the octaketide mimic adopts the proper orientation necessary to achieve C9-C14 first-ring cyclization. It can be concluded that substrate chain length is a mandatory criterion for successful orientation of ACP- tethered polyketide intermediate in the substrate-binding pocket interior of ARO/CYC. In addition, the mis-orientation of polyketide mimics in the interior pocket of WhiE supports a sequestration mechanism for polyketide chain length control. Specifically, Asp57, Asp59, Arg82 and Gln110 are identified as key residues for substrate orientation and chain-length recognition. To the best of our our knowledge, this is the first report of comprehensive dissection of chain length control for any type of PKSs. The structural information obtained from this study has shed light on ARO/CYC mediated regiospecific cyclization. We showed that the atom replaced probe strategy described in this study is a powerful chemical biology tool for the investigation of transient substrate-protein and protein-substrate interactions. This knowledge will facilitate efforts toward the generation of new “unnatural” natural products” through the implementation of programmable cyclization patterns during polyketide biosynthesis.

2.4 Materials and Methods

Protein expression and purification

The expression plasmid (pWhiE) has been described previously [11]. Briefly, *E. coli* BL21 (DE3) cells containing pWhiE were grown in 2 X 1L of LB media supplemented with kanamycin (50 µg/mL) to $Ab_{860nm} = 0.4 - 0.6$ at 37 °C. Cultures were induced with the addition of 0.5 mM IPTG and expressed overnight (16 hr) at 18 °C. Cells were harvested by centrifugation at 5500 rcf for 10 minutes, resuspended in lysis buffer (50 mM Tris·HCL pH 7.5, 300 mM NaCl, 20 mM imidazole, and 10 % glycerol), sonicated, and the cell debris was removed by centrifugation (21000 rcf for 70 min). The clarified cell lysate was batch bound to 5 mL of Ni-IMAC resin (Bio-Rad Laboratories) for 1 hr at 4 °C. The nickel slurry was transferred to a gravity-flow column, washed with 50 ml of lysis buffer, and again with another 50 ml of lysis buffer. Protein was eluted with 7 X 5 mL fractions of increasing imidazole concentration in lysis buffer. Protein purity was assessed by SDS-PAGE. Fractions containing the highest purity (≥ 90 %) were collected and dialyzed overnight at 4 °C against 50 mM Tris·HCl pH 7.5, 50 mM NaCl, 10 % glycerol and 1 mM DTT. The dialyzed protein was concentrated to 5-10 mg/mL using a centrifugal filter (Millipore) with a 3500 Da molecular weight cutoff, flash frozen, and stored at -80 °C.

Protein crystallization

Frozen WhiE aliquots were thawed on ice and desalted into minimal buffer (20 mM HEPES pH 7.0) using a PD-10 column (GE Healthcare). Protein concentration was determined using Bradford assay (Bio-Rad Laboratories). WhiE at 5 mg/mL was incubated with 5 mM compound (100 mM stock resuspended in DMSO) for 1 hr and filtered using a 0.2 µm

centrifugal filter device (Pall Life Sciences) prior to crystal trials. Cocrystals were obtained by combining 2 μL of well solution (0.1 M Bis-Tris, pH 5.0-6.0, 0.2 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, and 28 - 30 % PEG 3350) with 2 μL of the WhiE mixture equilibrated over a well volume of 500 μL at 4 $^\circ\text{C}$ (sitting drop vapor diffusion method). Cocrystals grew after 6 days and were flash frozen in liquid nitrogen with a cryoprotectant of well solution containing 20 % glycerol.

Data Collection

X-ray diffraction data corresponding to WhiE Aro/Cyc bound to TALS and TAPS were collected to 1.8-1.9 $\text{\AA}$ on beamline 8.2.1 of Advanced Light Source (ALS). Diffraction Data for WhiE bound to PGS22a were obtained at 1.8 $\text{\AA}$ from Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 12-2. The statistics are listed in Table 2.1.

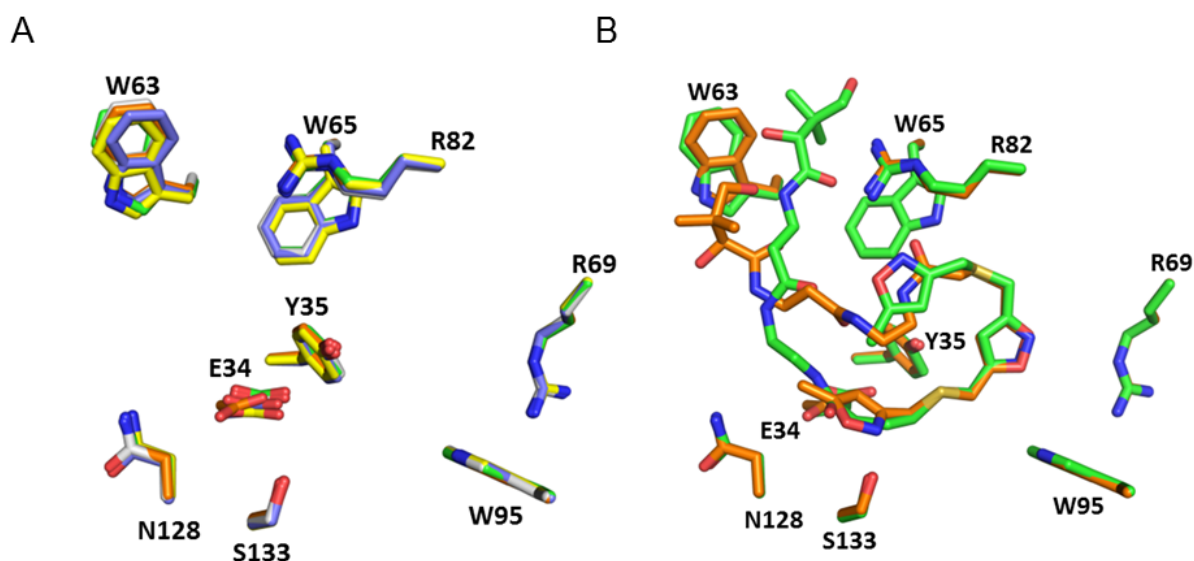


Figure 2.5. An active site overlay of the heptaketide-WhiE and octaketide-WhiE structures highlighting the inverse binding modes of 3 and 4.

(A) An overlap of the active site between the apo-WhiE (yellow) and triketide-WhiE (grey), hexaketide-WhiE, (dark blue) heptaketide-WhiE (orange), and octaketide-WhiE (green). (B) Note that triketide-WhiE does not enter the substrate pocket.

Table 2.1. Crystallographic statistics for data collection, modeling and refinement.

Structure	Triketide (1)	Hexaketide (2)	Heptaketide (3)	Octaketide (4)
Wavelength (Å)	0.91998	0.91998	0.99999	0.99995
Resolution range (Å)	48.09 - 1.769 (1.832 - 1.769)	47.98 - 1.87 (1.937 - 1.87)	48.07 - 1.783 (1.847 - 1.783)	47.96 - 1.852 (1.918 - 1.852)
Space group	P 1 2 1 1	P 1 2 1 1	P 1 2 1 1	P 1 2 1 1
Unit cell	47.179 41.994 48.106 90 91.37 90	47.193 41.991 47.992 90 91.52 90	47.783 42.046 48.012 90 91.36 90	47.614 42.101 47.984 90 91.67 90
Total reflections	136325	70216	104003	70524
Unique reflections	18538 (1777)	18591 (1324)	17924 (1723)	16356 (1607)
Completeness (%)	99.60 (96.00)	97.02 (85.64)	97.55 (94.72)	99.76 (98.11)
Mean I/sigma(I)	142 (10.8)	83.5 (5.4)	254.2 (17.4)	145.2 (10)
Wilson B-factor	19.77	19.66	26.04	19.81
R-work	0.2206 (0.2253)	0.1847 (0.2201)	0.1967 (0.2408)	0.1796 (0.2173)
R-free	0.2567 (0.2829)	0.2286 (0.2778)	0.2259 (0.2791)	0.2232 (0.2713)
RMS(bonds)	0.015	0.007	0.007	0.007
RMS(angles)	1.53	1.11	1.12	1.1
Ramachandran favored (%)	98	99	99	98
Ramachandran outliers (%)	0.06	0	0	0
Average B-factor	24.6	28.4	26.9	26.1
Macromolecules	24.2	27.1	25.8	24.6
Ligands	44.2	47	46.8	44.9
Solvent	32.4	35.4	33.4	34.5

Phasing, Model Building, and Refinement.

All data sets used for structure solution were obtained from individual crystals and processed using HKL2000. The initial phases for WhiE ARO/CYC cocrystals were determined by molecular replacement using Phaser (Phenix) [19] with the refined WhiE ARO/CYC structure (PDB:3TVR) as a search model [13]. Similar to the native structure, the cocrystal structures contained one WhiE molecule in the asymmetric unit. The coordinate and topology files for each compound were prepared using Elbow, and fitting was performed using LigandFit [19]. The structures were refined using Phenix.refine in the Phenix suites along with COOT [20]. Table 2.1 lists the data collection, refinement and model statistics.

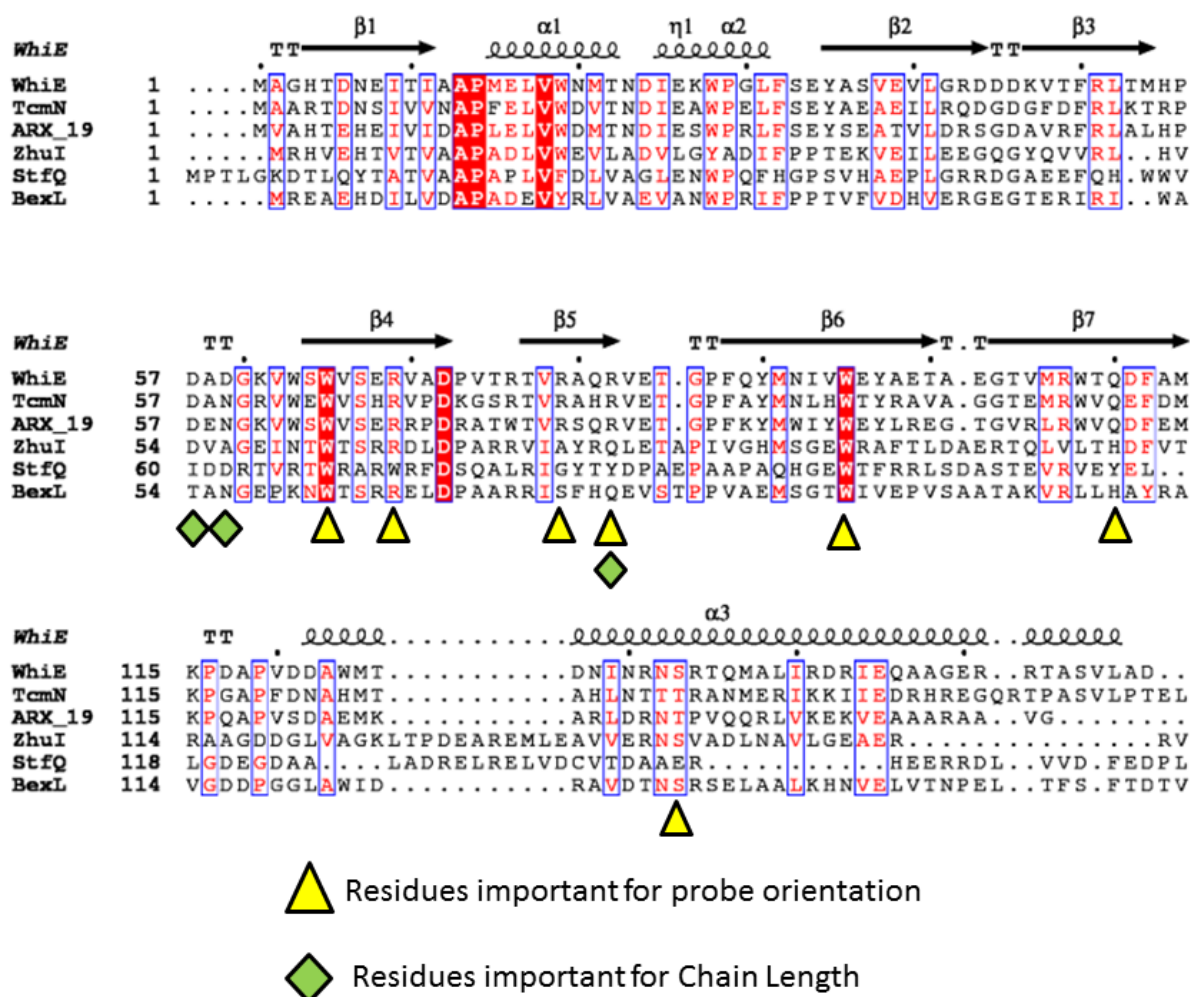


Figure 2.6. Sequence alignment of WhiE to other ARO/CYC homologues.

Green diamonds demarcate residues that were found to interact with the triketide mimic whereas yellow triangles are indicative of residues that interacted with the longer chain mimics (12C, 14C, and 16C).

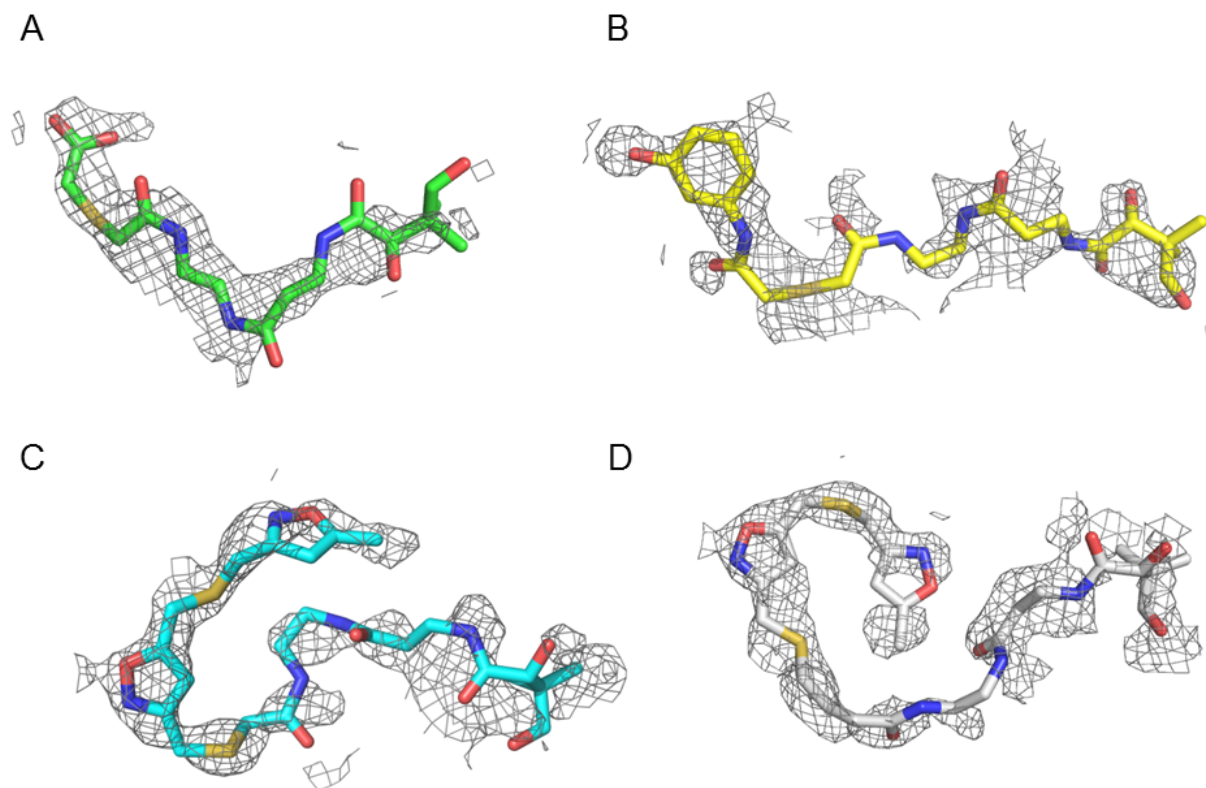


Figure 2.7. SA-Omit maps.

2FO-FC SA-omit maps for the triketide-WhiE (A), hexaketide-WhiE (B), hepta-ketide (C) and octaketide-WhiE (D) structures contoured at 0.7 sigma.

2.5 References

1. Weissman, K.J. and P.F. Leadlay, *Combinatorial biosynthesis of reduced polyketides*. Nature reviews. Microbiology, 2005. **3**(12): p. 925-36.
2. Xu, Y., et al., *Thioesterase domains of fungal nonreducing polyketide synthases act as decision gates during combinatorial biosynthesis*. Journal of the American Chemical Society, 2013. **135**(29): p. 10783-91.
3. Huang, W., et al., *FW-04-806 inhibits proliferation and induces apoptosis in human breast cancer cells by binding to N-terminus of Hsp90 and disrupting Hsp90-Cdc37 complex formation*. Molecular cancer, 2014. **13**: p. 150.

4. Crawford, J.M. and C.A. Townsend, *New insights into the formation of fungal aromatic polyketides*. Nat Rev Microbiol, 2010. **8**(12): p. 879-89.
5. Bruegger, J., et al., *Probing the selectivity and protein-protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers*. Chemistry & biology, 2013. **20**(9): p. 1135-46.
6. Hertweck, C., *The biosynthetic logic of polyketide diversity*. Angewandte Chemie, 2009. **48**(26): p. 4688-716.
7. Zhang, W., Y. Li, and Y. Tang, *Engineered biosynthesis of bacterial aromatic polyketides in Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(52): p. 20683-8.
8. Zawada, R.J. and C. Khosla, *Heterologous expression, purification, reconstitution and kinetic analysis of an extended type II polyketide synthase*. Chemistry & biology, 1999. **6**(9): p. 607-15.
9. Shen, Y., et al., *Ectopic expression of the minimal whiE polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(7): p. 3622-7.
10. Petkovic, H., et al., *Disruption of an aromatase/cyclase from the oxytetracycline gene cluster of Streptomyces rimosus results in production of novel polyketides with shorter chain lengths*. The Journal of biological chemistry, 1999. **274**(46): p. 32829-34.
11. Lee, M.Y., B.D. Ames, and S.C. Tsai, *Insight into the molecular basis of aromatic polyketide cyclization: crystal structure and in vitro characterization of WhiE-ORFVI*. Biochemistry, 2012. **51**(14): p. 3079-91.
12. Ames, B.D., et al., *Structural and biochemical characterization of Zhul aromatase/cyclase from the R1128 polyketide pathway*. Biochemistry, 2011. **50**(39): p. 8392-406.
13. Ames, B.D., et al., *Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(14): p. 5349-54.
14. Zhou, H., Y. Li, and Y. Tang, *Cyclization of aromatic polyketides from bacteria and fungi*. Nat Prod Rep, 2010. **27**(6): p. 839-68.
15. Pickens, L.B. and Y. Tang, *Decoding and engineering tetracycline biosynthesis*. Metabolic engineering, 2009. **11**(2): p. 69-75.
16. Crosby, J. and M.P. Crump, *The structural role of the carrier protein--active controller or passive carrier*. Natural product reports, 2012. **29**(10): p. 1111-37.

17. Shakya, G., et al., *Modeling linear and cyclic PKS intermediates through atom replacement*. J Am Chem Soc, 2014. **136**(48): p. 16792-9.
18. Haushalter, R.W., et al., *Binding and "pKa" modulation of a polycyclic substrate analogue in a type II polyketide acyl carrier protein*. ACS chemical biology, 2011. **6**(5): p. 413-8.
19. McCoy, A.J., et al., *Phaser crystallographic software*. Journal of Applied Crystallography, 2007. **40**: p. 658-674.
20. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallographica Section D-Biological Crystallography, 2004. **60**: p. 2126-2132.
21. Caldara-Festin, G., et al., *Structural and functional analysis of two di-domain aromatase/cyclases from type II polyketide synthases*. Proc Natl Acad Sci U S A, 2015. **112**(50): p. E6844-51.
22. Tsai, S.C. and B.D. Ames, *Structural enzymology of polyketide synthases*. Methods in enzymology, 2009. **459**: p. 17-47.

CHAPTER 3: Structural Basis for Acyl-CoA Carboxylase-Mediated Assembly of Unusual Polyketide Synthase Extender Units Incorporated into the Stambomycin Antibiotics

3.1 Abstract

Type I modular polyketide synthases assemble diverse bioactive natural products. Such multienzymes typically use malonyl and methylmalonyl-CoA thioesters as building blocks for polyketide chain assembly. However, in several cases more exotic alkylmalonyl-CoA extender units are known to be incorporated alongside these conventional building blocks. In all examples studied to date, these unusual extender units are biosynthesized via reductive carboxylation of an α , β -unsaturated thioester catalyzed by a crotonoyl-CoA reductase/carboxylase (CCRC). Here we show that a group of unusual pentyl- and hexylmalonyl-CoA-based extender units incorporated into the stambomycins complex of macrolides is assembled via a CCRC-independent mechanism. Incorporation of stable isotope-labeled precursors indicated that these unusual extender units originate from fatty acid metabolism. Retention of both deuterium atoms in the stambomycin analogue derived from incorporation of $[3\text{-}^2\text{H}_2]$ heptanoic acid ruled out the involvement of a CCRC in the assembly of the unusual extender units. Instead an acyl-CoA carboxylase (ACC) β -subunit was shown to be involved in this process via expression in a *revT* mutant of *Streptomyces* sp. SN-593, in which the CCRC-dependent pathway for assembly of *n*-pentylmalonyl-CoA (and related) extender units incorporated into the reveromycins is inactivated. X-ray crystal structures of the ACC β -subunit alone and in complex with hexanoyl-CoA revealed the molecular basis for substrate recognition. Taken together, our results establish a unique ACC-dependent pathway for unusual polyketide extender unit biosynthesis. The potential to exploit this pathway for selective polyketide derivatization was demonstrated by

precursor directed biosynthesis of azide- and alkyne-tagged stambomycin analogues via the incorporation of 6-azidohexanoic acid and 8-nonynoic acid.

3.2 Introduction

Modular polyketide synthases (PKSs) are bacterial molecular machines responsible for the assembly of numerous structurally-complex bioactive natural products with diverse applications in medicine and agriculture.[1] Important examples of modular PKS products include: erythromycin, rifamycin and fidaxomicin (antibacterials); amphotericin and nystatin (antifungals); rapamycin and FK-506 (immunosuppressants); spinosyns (insecticides) and avermectins (insecticides and anthelmintics). These remarkable multienzymes employ various tactics to generate structural diversity, such as variation in the number of modules within the assembly line, and of the α - and β -carbon-modifying domains within each module; *in trans* recruitment of a range of enzymes that modify the growing polyketide during chain assembly; utilization of several distinct mechanisms for termination of chain assembly, and a wide variety of starter and extender units.[1]

The stambomycins are a novel class of macrolide antibiotics recently identified as products of a modular PKS via a rational activation of a silent biosynthetic gene cluster in *Streptomyces ambofaciens* ATCC23877 (Figure 3.1).[2, 3] The majority of the stambomycin PKS chain elongation modules are predicted to employ widely-used malonyl- and (2*S*)-methylmalonyl-CoA extender units (Figure 3.2).[2] The exception is module 12, which is postulated to incorporate several atypical pentyl- and hexylmalonyl-CoA-based extender units, giving rise to the structural variants of the stambomycin complex (Figure 3.2).[2]

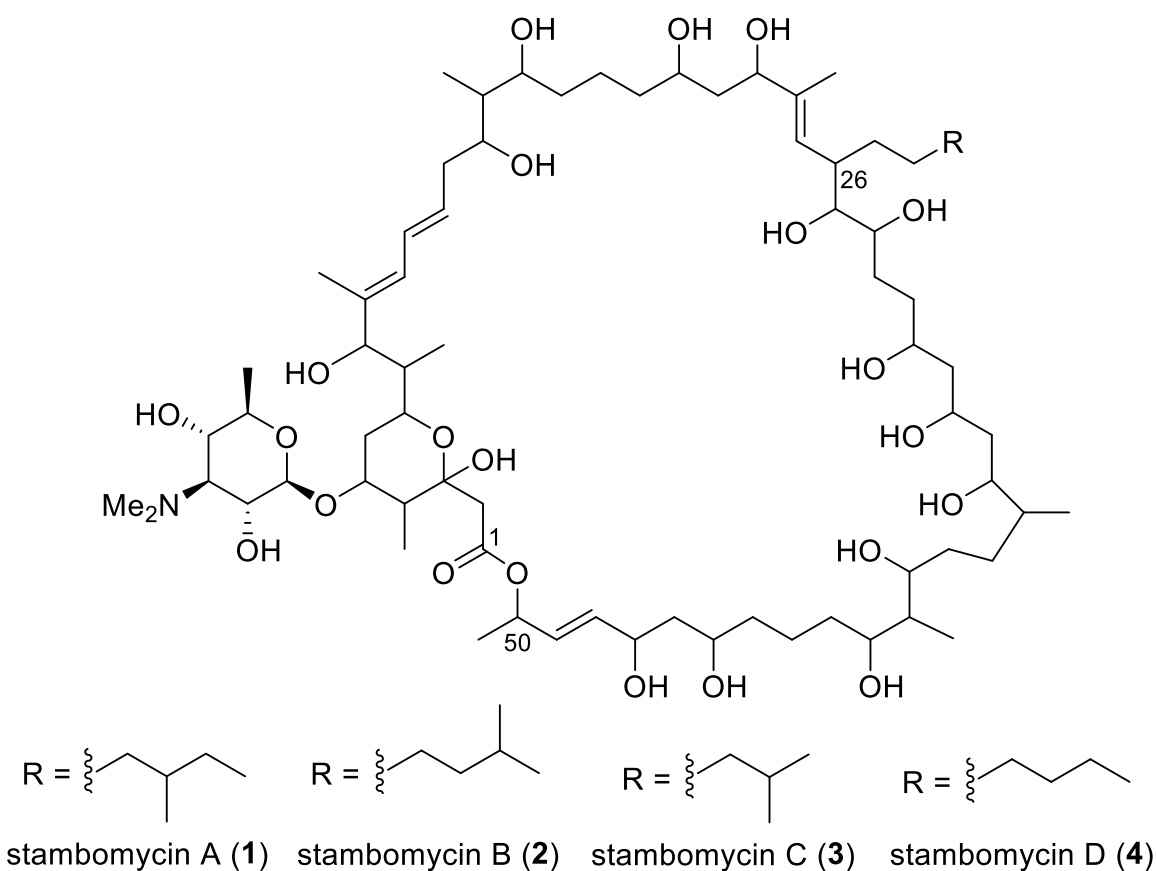


Figure 3.1. Structures of the four main components of the stambomycin complex.

The source of the differences among these antibiotics arise from the differing natures of the C-26 side chain.

Several unusual extender units are known to be utilized by other modular PKSs, including those responsible for the assembly of salinosporamide A (chloroethylmalonyl-CoA) [4], cinnabaramide A (*n*-hexylmalonyl-CoA) [5], filipin III (*n*-hexylmalonyl-CoA) [6], antimycins (*n*-butylmalonyl-CoA, *n*-hexylmalonyl-CoA) [7], splenocins (benzylmalonyl-CoA) [8], FK-506 (allylmalonyl-CoA) [9], chlorizidines (dichloropyrrolylpropylmalonyl-ACP) [10], and the reveromycins (*n*-butylmalonyl-CoA, *iso*-butylmalonyl-CoA, *n*-pentylmalonyl-CoA and *n*-hexylmalonyl-CoA)[11].

reaction with carbon dioxide.[12-14] The genes encoding such CCRC homologs are invariably clustered with the genes encoding the PKSs they supply.[4-11] Remarkably, no genes encoding CCRC homologs can be found in the stambomycin biosynthetic gene cluster (BGC) [2], suggesting that the unusual extender units incorporated into the stambomycins may be assembled by a CCRC-independent mechanism. The *samR0483* gene within the stambomycin cluster encodes an acyl-CoA carboxylase (ACC) β -subunit that, in partnership with the primary metabolic ACC α -subunit, could fulfill this role (Figure 3.2).[2] However, while ACCs are well known for their ability to generate the malonyl- and methylmalonyl-CoA extender units commonly utilized by modular PKSs via carboxylation of acetyl- and propionyl-CoA, respectively,[15, 16] the biosynthesis of unusual PKS extender units by such enzymes is without precedent. Moreover, the *S. ambofaciens* genome contains several genes encoding homologs of CCRCs, albeit not clustered with the stambomycin biosynthetic genes. Thus, the biosynthetic mechanism for the unusual extender units incorporated into the stambomycins is unclear.

Here we report a series of experiments that elucidate the biosynthetic origin and mechanism of assembly of the unusual stambomycin PKS extender units. Incorporation of stable isotope-labeled precursors indicated that they originate from fatty acid metabolism. Retention of both deuterium atoms in the stambomycin analogue derived from feeding of [3-²H₂]heptanoic acid to *S. ambofaciens* ruled out the involvement of a CCRC in unusual extender unit biosynthesis. Direct evidence for the involvement of *samR0483* in this process was provided by expression of this gene in a *revT* mutant of *Streptomyces* sp. SN-593, in which the CCRC-dependent pathway for the biosynthesis of *n*-pentylmalonyl-CoA (and related) extender units incorporated into the reveromycins has been abolished. X-ray crystal structures of SamR0483 alone and in complex with hexanoyl-CoA revealed the molecular basis for medium chain acyl-

CoA recognition by this unusual ACC β -subunit, prompting us to rename it medium chain acyl-CoA carboxylase beta subunit (MccB). Taken together, these results establish the existence of a unique ACC-dependent pathway for unusual polyketide extender unit biosynthesis. Finally, the potential of this pathway to be harnessed for selective polyketide derivatization was demonstrated by feeding 6-azidohexanoic acid and 8-nonynoic acid to *S. ambofaciens*, resulting in production of the corresponding azide- and alkyne-tagged stambomycin analogues.

3.3 Results and Discussion

Metabolic origin of the C-26 side chains of the stambomycins

Stambomycins A-D (**1-4**) differ in the length and methylation pattern of their C-26 side chains (Figure 3.1) [2], which are of unknown biosynthetic origin. In *Streptomyces* species, the FabH component of the core metabolic fatty acid synthase (FAS) catalyzes elongation of 2-methylbutyryl-, 3-methylbutyryl-, *iso*-butyryl-, and *n*-butyryl-CoA (**5-8**) starter units with malonyl-ACP.[17] Subsequent rounds of chain elongation are catalyzed by FabF.[18] 2-Methylbutyryl-CoA (**5**), 3-methylbutyryl-CoA (**6**) and *iso*-butyryl-CoA (**7**) derive from transamination and oxidative decarboxylation of L-isoleucine, L-leucine and L-valine respectively (Figure 3.3).[18] We hypothesized that the C-26 side chains of the stambomycins are derived from **5-8**. To test this hypothesis, we fed universally deuterium labeled [U-²H]isoleucine (**9**), [U-²H]leucine (**10**), [U-²H]valine (**11**) and [U-²H]butyric acid (**12**) separately to cultures of *S. ambofaciens* W130 (a mutant of the ATCC23877 strain with an extra copy of *samR0484*, which encodes a pathway-specific transcriptional regulator of the stambomycin BGC, integrated into its chromosome under the control of the constitutive *ermE** promoter). LC-ESI-TOF-MS analyses of mycelial extracts from these experiments showed that L-isoleucine and L-leucine are specifically incorporated into

1 and **2**, whereas L-valine and *n*-butyric acid are specifically incorporated into **3** and **4**, respectively (Figure 3.3 and section 3.5).

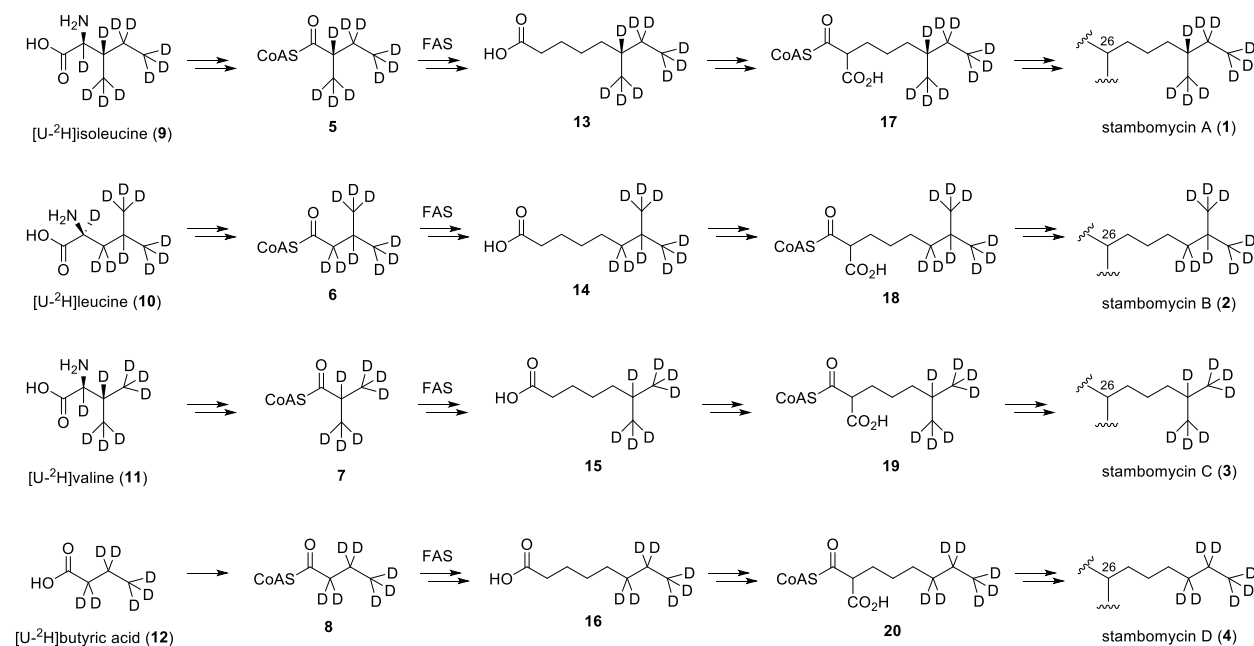


Figure 3.3. FAS precursor incorporation.

Pattern of incorporation observed for labeled precursors of FAS starter units (**5-8**) into the C-26 side chains of the stambomycins.

These data are consistent with biosynthesis of the (4'-methylhexyl)-, (5'-methylhexyl)-, (4'-methylpentyl)-, and *n*-hexylmalonyl-CoA extender units (**17-20**) incorporated into stambomycins A-D **1-4**, respectively, from the corresponding fatty acids, i.e. 6'-methyloctanoic acid (**13**), 7'-methyloctanoic acid (**14**), 6'-methylheptanoic acid (**15**) and *n*-octanoic acid (**16**) (Figure 3.3).

These fatty acids are probably converted to their CoA thioesters by SamR0482, a homologue of the medium chain fatty acyl-CoA ligase RevS, which was recently shown to participate in the biosynthesis of unusual extender units for the reveromycin PKS.[11] Feeding of heptanoic acid (**21**) to *S. ambofaciens* W130 resulted in the production of a novel stambomycin derivative (**22**) with an *n*-pentyl side chain at C-26 (Figure 3.4 and section 3.5), providing further support for the

proposed fatty acid origin of the unusual extender units and highlighting the ability of the acyltransferase (AT) domain within module 12 of the stambomycin PKS to tolerate unnatural substrates.

*Synthesis of [3-²H₂]heptanoic acid **23** as a mechanistic probe for unusual extender unit biosynthesis*

The finding that feeding of heptanoic acid **21** to *S. ambofaciens* W130 results in production of a stambomycin analogue **22** (Figure 3.4) inspired us to synthesize [3-²H₂]heptanoic acid (**23**) as a mechanistic probe to determine whether the unusual extender units incorporated into the stambomycins are assembled by a CCRC- or ACC-dependent pathway.

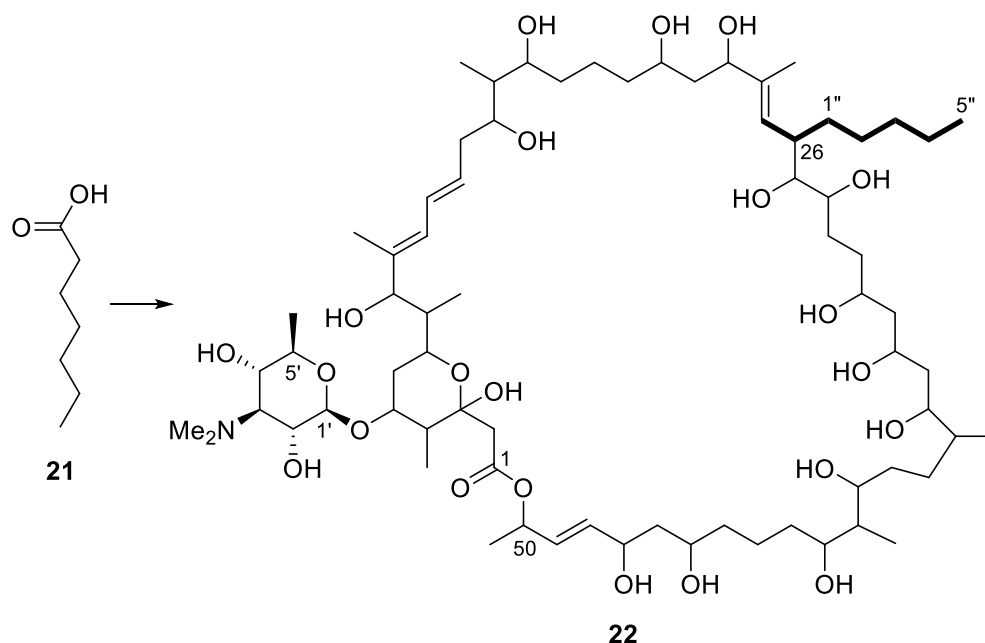


Figure 3.4. Structure of stambomycin analogue 22.

The novel stambomycin analogue **22** is produced when heptanoic acid **21** is fed to *S. ambofaciens* W130. The position of incorporation of **21** into the analogue is highlighted with bold bonds.

Incorporation of [3-²H₂]heptanoic acid **23** into the stambomycin analogue **22** via a CCRC-dependent pathway would be expected to proceed via initial SamR0482-catalyzed conversion to the CoA thioester (**24**), followed by β -oxidation to yield the corresponding α , β -unsaturated thioester (**25**) and reductive carboxylation to yield *n*-pentylmalonyl-CoA (**26**) containing only a single deuterium label (Figure 3.5). In contrast, the ACC-dependent pathway would proceed via direct carboxylation of [3-²H₂]heptanoyl-CoA **24**, resulting in *n*-pentylmalonyl-CoA **26** containing two deuterium labels (Figure 3.5). Thus, both of the deuterium atoms would be expected to be retained in the stambomycin analogue if *n*-pentylmalonyl-CoA is biosynthesized via the ACC-dependent pathway, whereas one of the deuterium labels would be lost if this extender unit is assembled by the CCRC-dependent pathway.

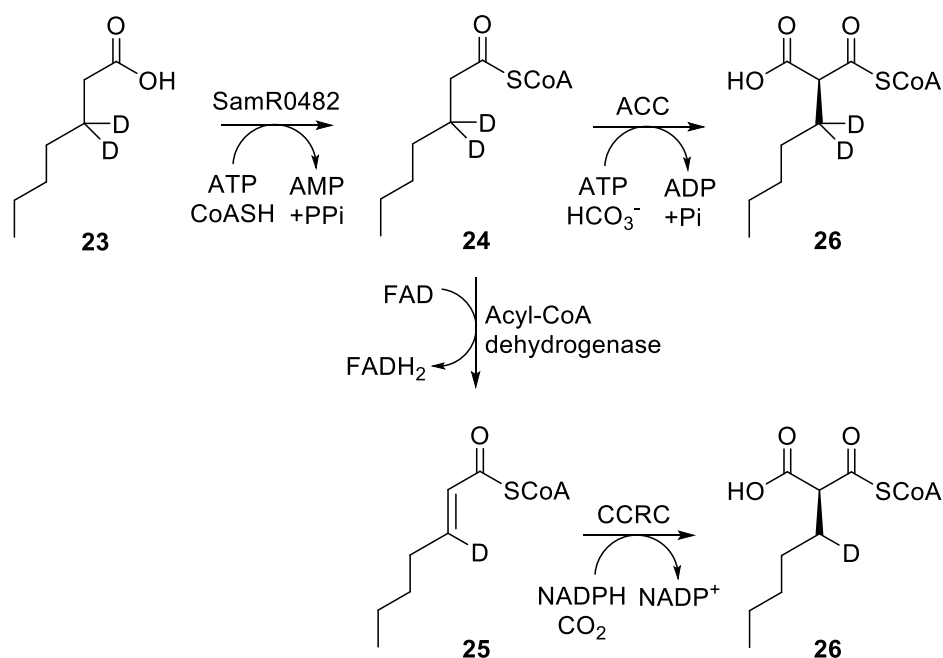
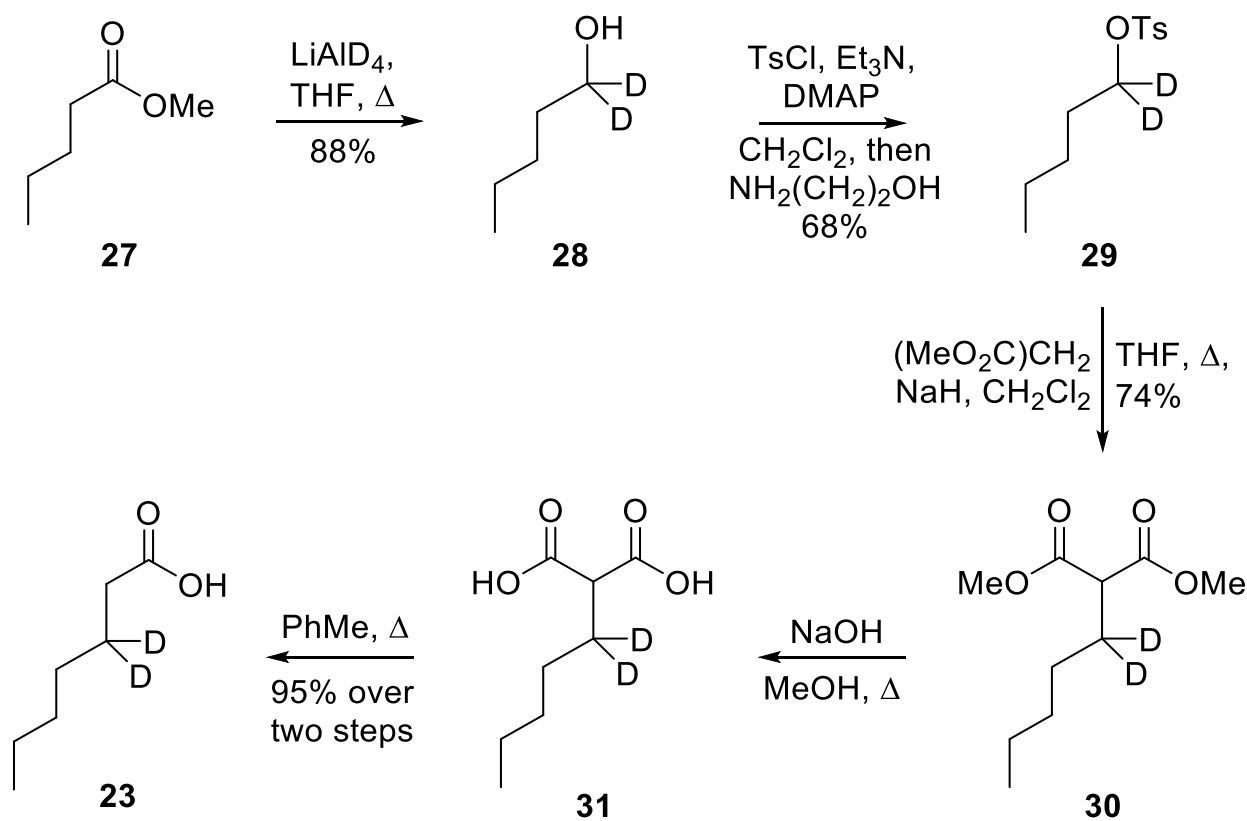


Figure 3.5. Incorporation of [3-²H₂]heptanoic acid **23 into the stambomycin analogue.**

Conversion of [3-²H₂]heptanoic acid **23**, via its CoA thioester **24**, to pentylmalonyl-CoA **26** by an ACC-dependent pathway would result in retention of both deuterium labels, whereas one of the deuterium labels would be lost in pentylmalonyl-CoA **26** formed by a CCRC-dependent pathway, because CoA thioester **24** must undergo desaturation to **25** in order to be reductively carboxylated.

[3-²H₂]heptanoic acid **23** was synthesized in five steps from methyl-*n*-pentanoate **27** (Scheme 3.1). Reduction of methyl-*n*-pentanoate (**27**) with LiAlD₄ yielded [1-²H₂]pentanol (**28**), which was converted to its tosylate derivative (**29**) by reaction with TsCl, DMAP and Et₃N. Deprotonation of dimethyl malonate with sodium hydride followed by alkylation with **29** gave (**30**). Hydrolysis of **30** with sodium hydroxide in methanol afforded the corresponding diacid (**31**), which was decarboxylated by refluxing in toluene to give yield the mechanistic probe **23**.



Scheme 3.1. Synthesis of [3-²H₂]heptanoic acid **23.**

Comparison of the mass spectra for stambomycin analogue **22** from LC-ESI-TOF-MS analyses of mycelial extracts of *S. ambofaciens* W130 cultures that had been fed with labeled and unlabeled heptanoic acid showed that the [M+2H]²⁺ parent ions differ by one *m/z* unit (Figure

3.6). Thus both atoms of deuterium are retained when $[3-^2\text{H}_2]$ heptanoic acid **23** is incorporated into stambomycin analogue **22**, because the parent ion is doubly charged (Figure 3.6). These results confirmed that the pentylmalonyl-CoA extender unit incorporated into stambomycin analogue **22** is biosynthesized via an ACC-dependent pathway.

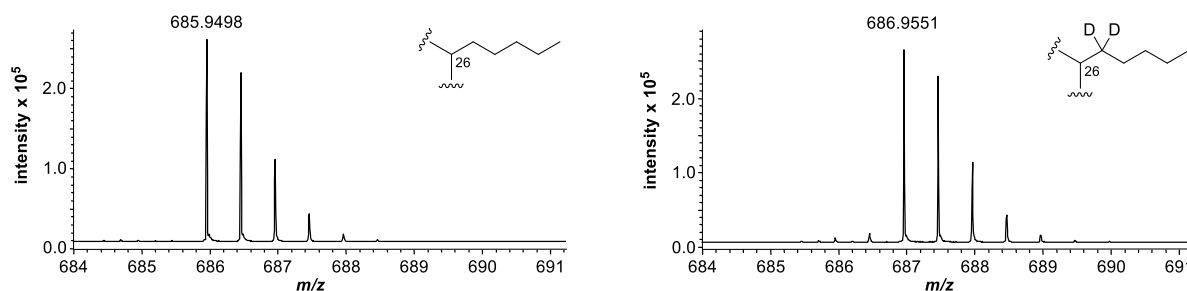
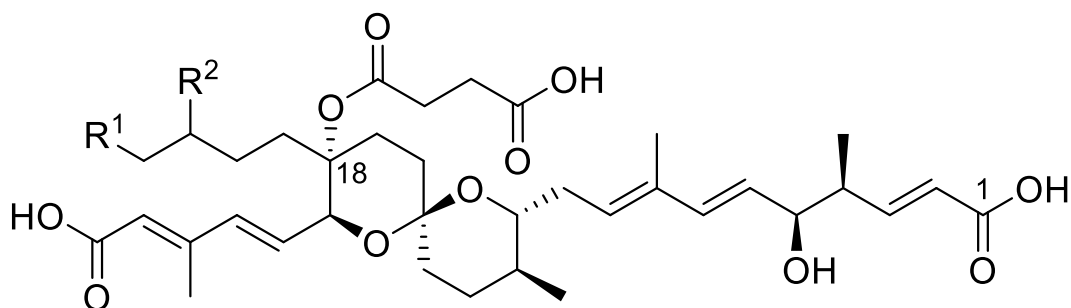


Figure 3.6. Mass spectra of stambomycin analogue **22 with heptanoic acid **21** and $[3-^2\text{H}_2]$ heptanoic acid **23**.**

LC-MS analyses of mycelial extracts of *S. ambifaciens* W130 cultures supplemented with unlabeled heptanoic acid **21** (left spectrum) and $[3-^2\text{H}_2]$ heptanoic acid **23**, (right spectrum). The shift of one m/z unit for the doubly-charged parent ion in the right spectrum indicates that both deuterium labels are retained when $[3-^2\text{H}_2]$ heptanoic acid **23** is incorporated into stambomycin analogue **22**.

To further demonstrate the ability of our mechanistic probe to distinguish between CCRC-dependent and CCRC-independent pathways for unusual polyketide extender unit biosynthesis, we investigated the incorporation of $[3-^2\text{H}_2]$ heptanoic acid **23** into revermomycin D (**32**). *Streptomyces* sp. SN-593 produces reveromycins A, C, D and E (**33**, **34**, **32** and **35**), a family of structurally-related polyketides that differ in the nature of their C-18 side chains (Figure 3.7).[19] The structural differences between the reveromycins result from the utilization of various unusual extender units by module 4 of the reveromycin PKS.[19] The *revR*, *revS* and *revT* genes, which encode a FabH homolog, a medium chain fatty acyl-CoA ligase, and a CCRC, respectively, have recently been implicated in the biosynthesis of these extender units.[11]



- $R^1 = R^2 = H$; reveromycin A (**33**)
 $R^1 = H, R^2 = Me$; reveromycin C (**34**)
 $R^1 = Me, R^2 = H$; reveromycin D (**32**)
 $R^1 = Et, R^2 = H$; reveromycin E (**35**)

Figure 3.7. Structures of the reveromycins produced by *Streptomyces* sp. SN-593.

Reveromycin D **32** incorporates *n*-pentylmalonyl-CoA, but its abundance is low relative to reveromycin A **33**, which is derived from an *n*-butylmalonyl-CoA extender unit. Deletion of *revR* suppresses the production of reveromycin A **33** relative to the other congeners.[11] Thus, we fed $[3-^2H_2]$ heptanoic acid **23** to a *revR* mutant of *Streptomyces* sp. SN-593. LC-MS analysis of an ethyl acetate extract of this culture showed that $[3-^2H_2]$ heptanoic acid **23** was incorporated into reveromycin D **32** with retention of only a single deuterium label, consistent with the assembly of *n*-pentylmalonyl-CoA in *Streptomyces* sp. SN-593 via a CCRC-dependent pathway (Figure 3.8).[11] Moreover, high resolution ESI-MS analysis of partially purified reveromycins A, C and D confirmed that $[3-^2H_2]$ heptanoic acid **23** is incorporated exclusively into reveromycin D **32** (see section 3.5).

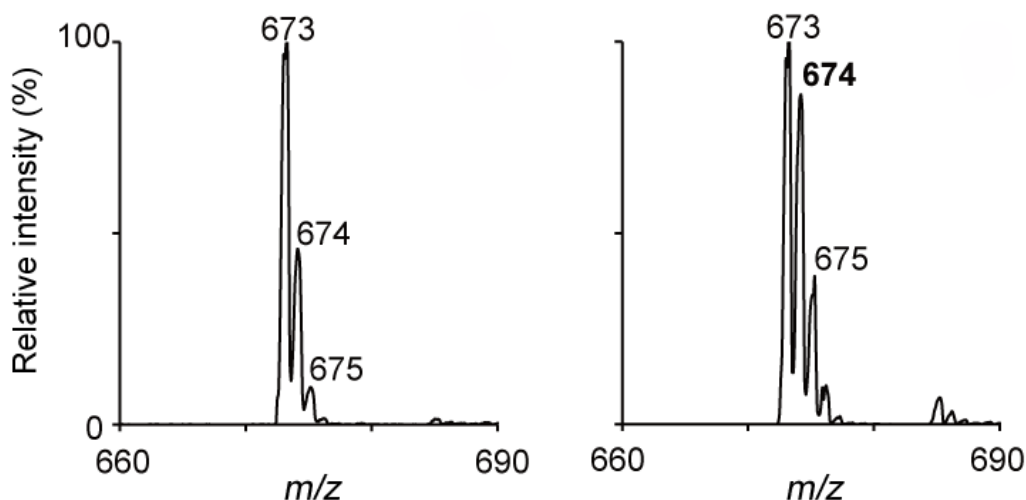


Figure 3.8. Mass spectra of reveromycin D 32.

Mass spectra of reveromycin D **32** from UHPLC-ESI-MS analyses of ethyl acetate extracts of *Streptomyces* sp. SN-593 *revR* mutant cultures grown in the absence (left spectrum) and presence (right spectrum) of [3-²H₂]heptanoic acid **23**. The increase in intensity of the *m/z* = 674 peak in the right spectrum is consistent with incorporation of [3-²H₂]heptanoic acid **23** into reveromycin D **32** with loss of one of the two deuterium labels.

Expression of samR0483 in a revT mutant of Streptomyces sp. SN-593

In-frame deletion of *revT* in *Streptomyces* sp. SN-593 has been recently reported to abolish reveromycin production.[11] Expression of *samR0483* in the *revT* mutant restored the production of reveromycin E **35** (Figure 3.9), demonstrating the ability of the ACC β-subunit encoded by this gene to biosynthesize *n*-hexylmalonyl-CoA. Only trace amounts of reveromycins A and D were observed in the *revT* mutant expressing *samR0483*. However, feeding of [3-²H₂]heptanoic acid **23** to this strain resulted in production of a modest amount of reveromycin D **32** in which both of the deuterium labels were retained (see section 3.5), consistent with assembly of *n*-pentylmalonyl-CoA via an ACC-dependent pathway.

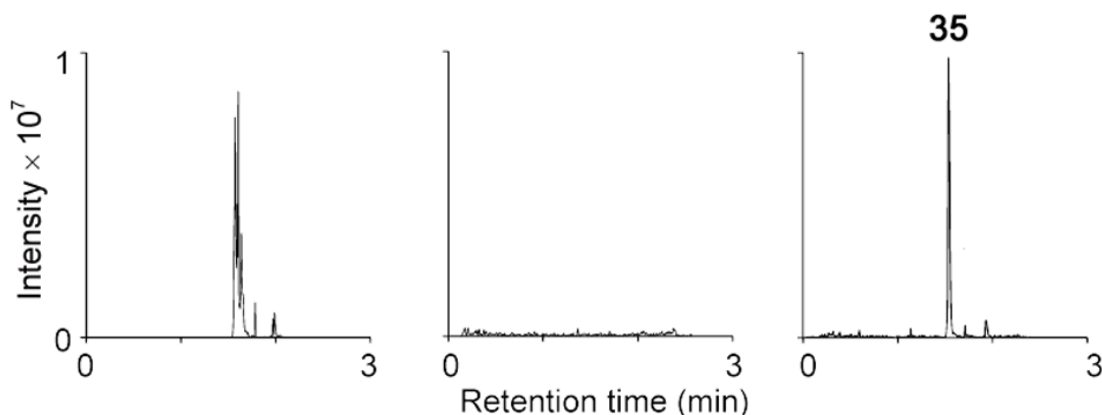


Figure 3.9. Extracted ion chromatograms for reveromycin E 35.

Extracted ion chromatograms at $m/z = 687-688$ corresponding to $[M-H]^-$ for reveromycin E from UHPLC-ESI-MS analyses of ethyl acetate extracts of wild type *Streptomyces* sp. SN-593 (left chromatogram), the *revT* mutant (middle chromatogram) and the *revT* mutant in which *samR0483* has been expressed (right chromatogram).

X-ray crystal structure of the acyl-CoA carboxylase β -subunit encoded by samR0483

To develop a better understanding of the molecular basis for the unique substrate specificity of the acyl-CoA carboxylase β -subunit encoded by *samR0483*, hereafter referred to as medium chain acyl-CoA carboxylase β subunit (MccB), we solved its crystal structure to 2.45 Å. MccB was crystallized as a hexamer in the orthorhombic $P2_12_12_1$ space group, and was solved via molecular replacement using the hexameric PccB D422A mutant structure (PDB Code:3IBB) as a search model. Similar to other bacterial ACC β -subunits, MccB has 3-2 symmetry and is composed of three individual dimers to form a ring-like hexamer (Figure 3.10). Each monomer can be divided into two domains harboring two copies of the crotonase fold that pack together in an antiparallel manner to form a dimer (Figure 3.10). The active site of MccB is located at the dimerization interface for two opposing monomers where residues 137-237 from the N-terminus of monomer B are packed against residues 392-492 from the C-terminus of the next proximal

monomer F (Figure 3.10). The individual monomers themselves superimpose well with a backbone C α RMSD of 0.167 Å.

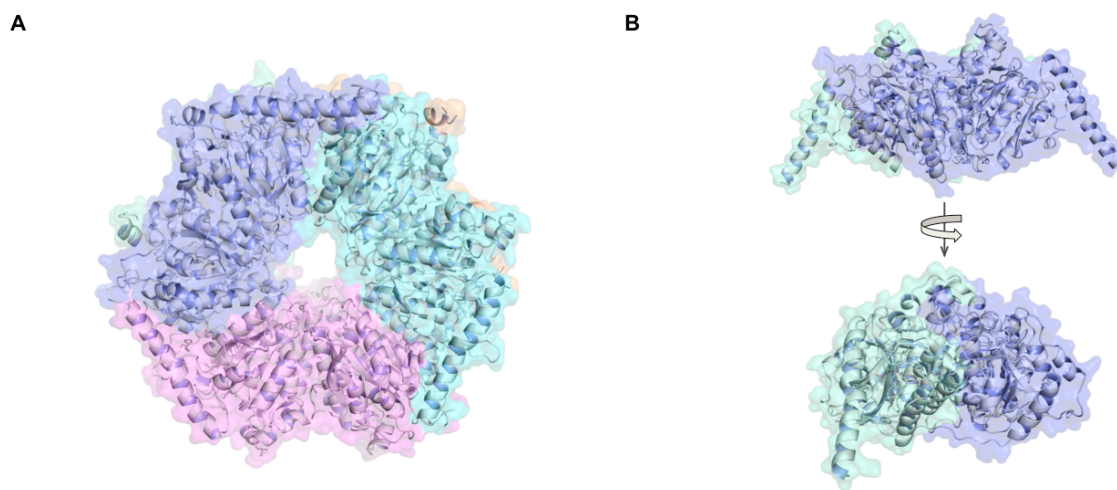


Figure 3.10. X-ray crystal structure of *apo*-MccB.

(A) The hexameric overall structure, and (B) the substrate binding site at the subunit interface

Overall, MccB has a high degree of structural similarity to the previously solved β -subunit of propionyl-CoA carboxylase PccB from *Streptomyces coelicolor* [20], with a C α backbone RMSD of 0.738 Å (see section 3.5). Consequently, the vast majority of the known substrate and cofactor binding residues of ACC β -subunits are both structurally and sequentially conserved for MccB, PccB, and AccD5 from *M. tuberculosis* (see section 3.5).[20, 21] The greatest structural deviations of MccB from previously solved structures occur near residues 33-46 and 452-477. Residues 33-46 of MccB form a long α -helix resulting in a larger cavity near the substrate binding zone. Residues 452-477 of MccB could not be modeled in the *apo* structure, highlighting this region's high degree of flexibility. In contrast, the region corresponding to residues 452-477 for PccB adopts a partially ordered helix-loop-helix motif. The high flexibility

of MccB in the vicinity of the substrate binding pocket is consistent with its ability to bind larger acyl-CoA substrates than PccB.

Molecular basis for hexanoyl-CoA recognition by MccB

Given the high degree of structural and sequence similarity between MccB and PccB, it was not immediately clear how the former is able to accommodate significantly longer acyl-CoA substrates than the latter. To address this question structures of MccB bound to either one or four molecules of hexanoyl-CoA were solved to 2.75 Å and 2.85 Å, respectively. These cocrystal structures were obtained via incubation of MccB with 5 mM or 10 mM hexanoyl-CoA prior to crystallization. Interestingly, the electron density corresponding to residues 452-477 of monomer F was only present in the crystal form bound to four molecules of hexanoyl-CoA, suggesting that the helix-loop-helix lid region of MccB is stabilized at higher substrate concentrations.

The hexanoyl-CoA bound structures reveal that some of the substrate-binding features of MccB are similar to those employed by other ACC β -subunits. For example, the residues that interact with the adenine ring of the substrate are conserved between MccB and PccB. However, MccB and PccB recognize the phosphopantetheine (PPant) and acyl moieties of hexanoyl-CoA quite differently (see below).

The residues that interact with hexanoyl-CoA are localized in three regions at the MccB dimer interface. The first region is located at the entrance of the substrate binding pocket, which is characterized by a network of hydrogen bonds between the adenosine moiety of hexanoyl-CoA and the protein. The exocyclic amine and N-1 of adenine form hydrogen bond contacts with the peptide backbone of Ile147, Ala145 and Gly143 monomer B (Figures 3.11 and 3.12). The adenosine moiety is further stabilized by a hydrogen bond from the backbone carbonyl group of

Pro453 to the 2' hydroxyl group of ribose (Figure 3.11). The second region is the PPant binding site, in which the backbone N-H groups of Ala145 and Gly184 in monomer B form hydrogen bonds to the carbonyl oxygen atom of the β -alanine component of PPant. In addition, the sulfur atom of the thioester is hydrogen bonded to the backbone N-H of Gly420 (Figure 3.11). The third region is the binding pocket for the aliphatic portion of the substrate, which is formed primarily by the hydrophobic side chains of Ile396, Phe399 and Ala423 in Monomer F, and Leu191, Tyr157, and Tyr187 in monomer B (Figure 3.11). Collectively, these three sets of interactions serve to anchor hexanoyl-CoA strongly into the active site of MccB.

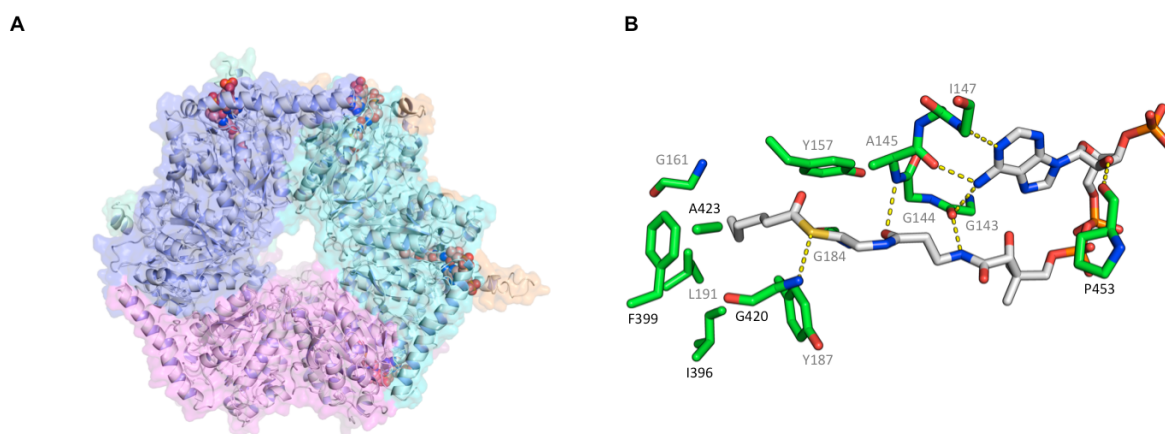


Figure 3.11. X-ray crystal structure of MccB with four molecules of hexanoyl-CoA bound. Figure shows (A) the overall architecture of the complex with hexanoyl-CoA in space filling representation and (B) the interactions between the substrate and protein. Residues labeled in black correspond to monomer F and residues labeled in grey correspond to monomer B of the MccB homo-dimer.

The structures of MccB with hexanoyl-CoA bound were compared with the previously-reported structure of PccB bound to propionyl-CoA (Figure 3.12).[20] In the PccB cocrystal structure, propionyl-CoA binds close to the entrance of the active site, leaving the 3'phosphate of CoA more exposed to the solvent. In contrast, MccB sequesters its substrate deep into the active

site via hydrophobic interactions between the side chains of the residues lining the substrate binding pocket and the aliphatic portion of hexanoyl-CoA. Further, the adenine ring of hexanoyl-CoA is tilted approximately 90° with respect to that of the adenine-ring of propionyl-CoA in the PccB cocrystal structure. Residue 421 of PccB is Asp, whereas in MccB the corresponding residue is Ala. Mutation of Asp421 to Ala allows PccB to bind substrates with a longer aliphatic chain, such as butyryl-CoA.[22] This suggests that the negatively-charged Asp421 side chain of PccB diminishes the hydrophobic character of the binding pocket, thus biasing its substrate specificity towards acyl-CoAs with shorter aliphatic chains. In comparison, the MccB substrate binding pocket consists of more hydrophobic residues, with higher flexibility at the substrate entrance thus allowing it to accommodate longer aliphatic chains.

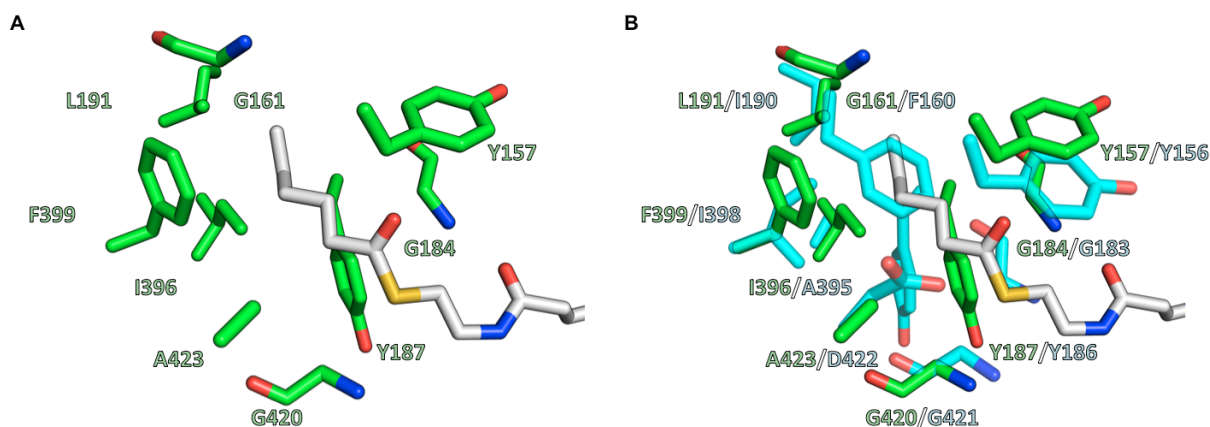


Figure 3.12. Structure of the binding region.

(A) A snapshot of the acyl binding region of MccB bound to hexanoyl-CoA (left panel). (B) Overlay of the structures of MccB (green) and PccB (cyan), highlighting residues lining the binding pockets for the aliphatic chains of the substrates.

Further insight into the molecular basis for the unique substrate specificity of MccB was gleaned from sequence alignments and detailed structural comparisons with other ACC β -subunits (Figure 3.12 and section 3.5). Structural alignment of MccB with PccB demonstrated

that Gly161 in the former corresponds to a bulky Phe residue in the latter. Superimposition of the substrate bound MccB and PccB structures strongly suggested that replacing Gly161 with Phe in MccB would cause a steric clash with the aliphatic chain of hexanoyl-CoA (Figure 3.12). Indeed, sequence alignments reveal that Phe is universally conserved in position 161 of all ACC β -subunits known to carboxylate acyl-CoAs with four carbons or less (see section 3.5).

Precursor-directed biosynthesis of alkyne- and azide-tagged stambomycin analogues

Inspired by the insights into the molecular basis for substrate specificity generated by our structural analysis of MccB, we sought to further probe its substrate tolerance by investigating the precursor-directed biosynthesis of additional stambomycin analogues. In particular, we targeted the production of analogues bearing an alkyne or azide group at the terminus of the C-26 substituent, with the aim of carrying out click chemistry on these functional groups to produce biotinylated or fluorescently-labeled stambomycin derivatives for future mechanism of action studies.

Thus, we fed commercially available 6-azidohexanoic acid (**36**) to cultures of *S. ambofaciens* W130. UHPLC-ESI-TOF-MS analyses of mycelial extracts identified a novel metabolite with the molecular formula $C_{70}H_{128}N_4O_{22}$ (calculated for $[C_{70}H_{129}N_4NaO_{22}]^{2+}$: 699.4414; found: 699.4419), corresponding to the stambomycin analogue (**37**) (Figure 3.13 and section 3.5). This analogue presumably arises from the conversion of 6-azidohexanoic acid **36** to its CoA thioester followed by MccB-mediated carboxylation to produce 4-azidobutylmalonyl-CoA, which is then utilized as an extender unit by module 12 of the stambomycin PKS. However, the production level of stambomycin analogue **37** is relatively modest compared with

that of stambomycins A-D **1-4** (see section 3.5), suggesting that azide-containing substrates are not well tolerated by SamR0482, MccB, or stambomycin PKS module 12.

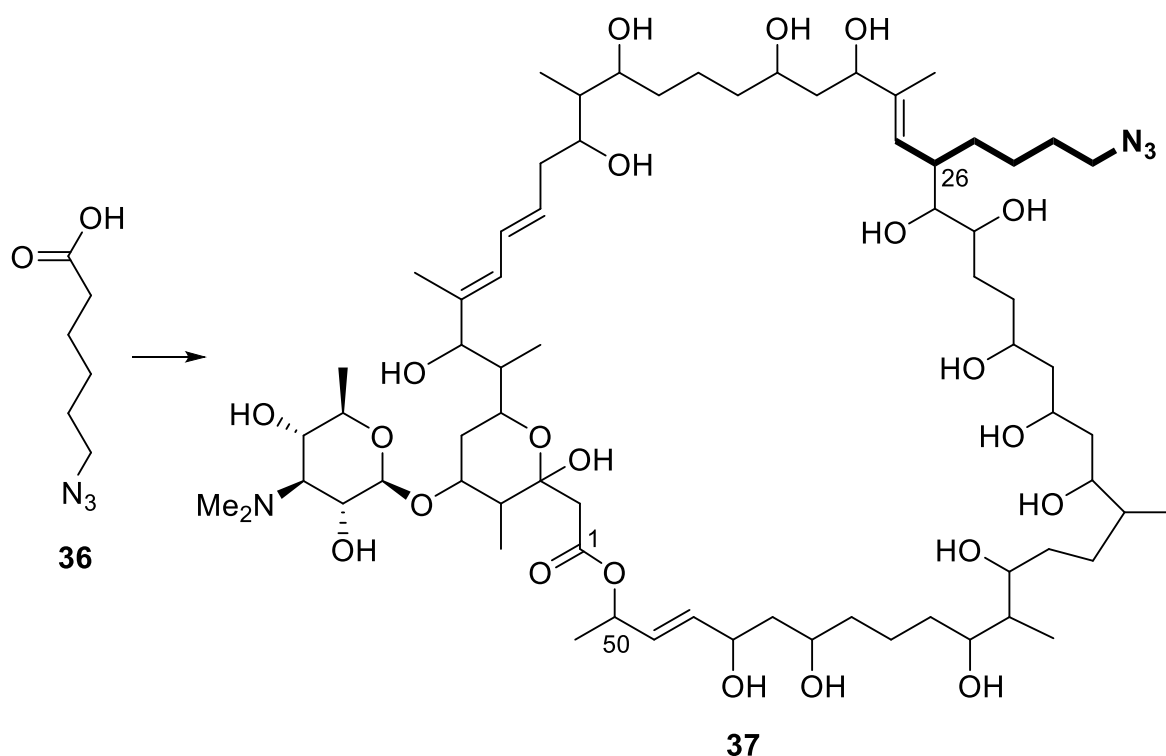


Figure 3.13. Structure of azide-tagged stambomycin analogue 37.

Azide-tagged stambomycin analogue **37** is produced when 6-azidoheptanoic acid **36** is fed to *S. ambofaciens* W130.

Surprisingly, feeding of commercially available 6-heptynoic acid (**38**) to *S. ambofaciens* W130 resulted in the production of a new metabolite with the molecular formula C₇₃H₁₂₉NO₂₂ (calculated for [C₇₃H₁₃₀NNaO₂₂]²⁺: 697.9486; found: 697.9548), corresponding to the stambomycin analogue (**39**) derived from 8-octynoic acid (Figure 3.14 and section 3.5). This analogue presumably arises from 2-carbon elongation of 6-heptynoic acid **38** by the primary metabolic FAS of *S. ambofaciens*, yielding 8-nonynoic acid (**40**), which is converted to its CoA thioester, carboxylated by MccB, and utilized as an extender unit by module 12 of the

stambomycin PKS. To test this hypothesis, Dr. Lauren Ray synthesized 8-nonynoic acid according to a literature procedure, and fed it to *S. ambofaciens* W130. This resulted in the production of stambomycin analogue **39** at a similar level to that of stambomycins A and B combined, allowing sufficient quantities of the analogue to be purified for ^1H NMR spectroscopic analysis (see section 3.5). The reaction of **39** with commercially available azide-PEG3-biotin resulted in conversion to the corresponding biotinylated compound (see section 3.5), providing proof of principle for the production of bio-orthogonally-tagged stambomycin derivatives.

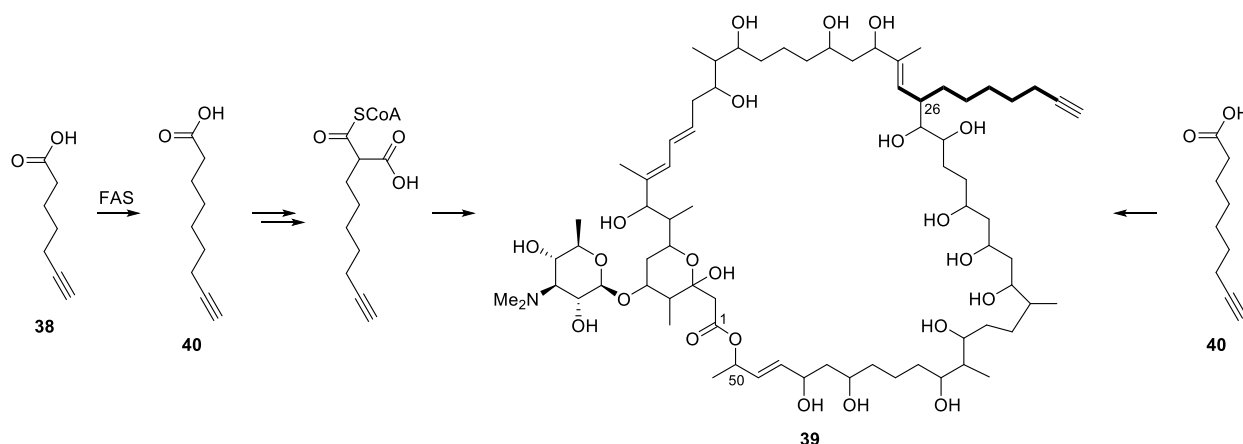


Figure 3.14. Acetylene-tagged stambomycin analogue **39.**

This analogue is produced when either 6-heptynoic acid **38** or 8-nonynoic acid **40** is fed to *S. ambofaciens* W130. 6-heptynoic acid **38** presumably undergoes conversion to its CoA thioester, 2-carbon elongation by the primary metabolic FAS and hydrolysis to form 8-nonynoic acid **40** prior to incorporation into **39**.

3.4 Conclusions

Several examples of unusual type I modular PKS extender units have been reported over the past half-decade. In all cases, these are assembled via a common final biosynthetic step, involving reductive carboxylation of an α , β -unsaturated CoA thioester by a CCRC homolog.

The data reported here reveal an alternative mechanism for assembly of unusual type I modular PKS extender units. This employs a unique ACC β -subunit (MccB) that, in partnership with the primary metabolic ACC α -subunit, is able to directly carboxylate medium chain CoA thioesters. X-ray crystallographic analysis revealed the structural basis for substrate recognition by MccB, inspiring the precursor-directed biosynthesis of novel azide- and alkyne-containing stambomycin analogues. Several biosynthetic engineering approaches have recently been utilized to introduce alkyne-terminated extender units containing 5-8 carbons into erythromycin and antimycin analogues.[23-26] Feeding of nonynoic acid to *S. ambofaciens* resulted in the production of a stambomycin analogue containing a 9-carbon alkyne-terminated extender unit in comparable quantities to the natural products. Thus, MccB, the AT domain from module 12 of the stambomycin PKS and the medium chain acyl-CoA synthetase encoded by *samR0482* are useful additions to the synthetic biology toolkit for bio-orthogonal tagging of polyketides.

3.5 Experimental Procedures and Supplemental Information

Bacterial strains and plasmids

The bacterial strains and plasmids used in the study are described in Table 3.1 and the oligonucleotide primers used are described in Table 3.2.

Table 3.1: Bacterial strains and plasmids.

Strain/plasmid	Relevant characteristics	Ref.
Strain		
<i>S. ambofaciens</i> ATCC 23877	Wild-type strain, in which stambomycin cluster is poorly expressed	[26]
<i>S. ambofaciens</i> W130	Stambomycin producing strain in which <i>samR0484</i> is integrated into the chromosome and is under control of the <i>ermE</i> * promoter	This work
<i>E. coli</i> ET12567/pUZ8002	Strain used for conjugal transfer of pOSV556t derivatives into <i>S. ambofaciens</i> ATCC 23877	[27]
<i>Streptomyces</i> sp. SN-593	Wild-type reveromycin-producing strain	[28]
<i>E. coli</i> GM2929 <i>hsdS</i> :: Tn10 / pUB307- <i>aph</i> ::Tn7	Strain used for conjugal transfer of pTYM19 derivatives into <i>Streptomyces</i> sp. SN-593	[29]
$\Delta revT$	<i>revT</i> disruptant of <i>Streptomyces</i> sp. SN-593	[30]
$\Delta revR$	<i>revR</i> disruptant of <i>Streptomyces</i> sp. SN-593	[30]
$\Delta revT::samR0483$	$\Delta revT$ mutant complemented with pTYM19- <i>P_{aph}-samR0483</i>	This work
Plasmid		
pOSV556t	Integrative <i>E. coli-Streptomyces</i> shuttle vector	[31]
pOSV556t- <i>samR0484</i>	<i>samR0484</i> inserted between the <i>HindIII</i> and <i>ClaI</i> sites of pOSV556t	This work
pOSV556t- <i>samR0483</i>	<i>samR0483</i> inserted between the <i>HindIII</i> and <i>ClaI</i> sites of pOSV556t	This work
pTYM19	Integrative <i>Streptomyces-E. coli</i> shuttle vector	[32]
pTYM19- <i>P_{aph}</i>	Derivative of pTYM19 in which the <i>aphII</i> promoter fragment is inserted into the <i>EcoRI</i> and <i>BamHI</i> sites.	[33]
pTYM19- <i>P_{aph}-samR0483</i>	<i>samR0483</i> inserted into <i>BamHI</i> and <i>HindIII</i> sites of pTYM19- <i>P_{aph}</i>	This work
pET151- <i>samR0483</i>	MccB overproduction construct	This work
pET28a- <i>samR0483</i>	MccB overproduction construct	This work

Table 3.2: Oligonucleotide primers.

Target gene (template)	Primers	Ref.
<i>samR0484</i>	pOSV_samR0484_FW: 5'-AAAGGGAAGCTTAGGAGGCCAGTCATTGCTGGTCC-3'	This work
(for cloning into pOSV556t between <i>Hind</i> III and <i>Cla</i> I)	pOSV_samR0484_RV: 5'-CCCTTTATCGATCGTGGGCAGGCTCTGCTC-3'	
<i>samR0483</i>	pOSV_samR0483_FW: 5'-AAGGGAAGCTTGCTGGTATGTCGTCCCA-3'	This work
(for cloning into pOSV556t between <i>Hind</i> III and <i>Cla</i> I)	pOSV_samR0483_RV: 5'-CCCTTTATCGATACGACGTCACA-3'	
<i>samR0483</i>	0483_BamHI: 5'-CGCGGATCCATGTCGCTCCAGGA-3'	This work
(pOSV556t- <i>samR0483</i>)	0483_HindIII: 5'-GGGAAGCTTTCACAGGGGAATGT-3'	
<i>samR0483</i>	TOPO_483_FW: 5'-CACCATGTCGCTCCAGGAGCCTGTCTC-3'	This work
(For TOPO cloning into pET151)	TOPO_483_RV: 5'-TCACAGGGGAATGTTCCCGTGTT-3'	
<i>samR0483</i> in <i>pET151</i>	pET28a-samR0483_FW: 5'-ATGCGCGCTAGCGATTCTACGGAAAAC-3'	This work
(For cloning into pET28a between <i>Nhe</i> I and <i>Not</i> I)	pET28a-samR0483_RV: 5'-TATATAGCGGCCGCTCACAGGGGAATGTT-3'	

Construction of the stambomycin overproducing strain S. ambofaciens W130

The LAL regulatory gene, *samR0484*, was PCR amplified from genomic DNA of *S. ambofaciens* ATCC 23877 using the primers pOSV_samR0484_FW and pOSV_samR0484_RV and cloned into the *Cla*I/*Hind*III restriction sites of pOSV556t, placing *samR0484* under the control of the constitutive *ermE** promoter. pOSV556t-*samR0484* was transferred into *S. ambofaciens* ATCC 23877 via conjugation from *E. coli* ET12567/pUZ8002, as previously described.[26] One of the resulting hygromycin-resistant exconjugants was picked and designated

as *S. ambofaciens* W130. Subsequent analysis, as previously described [26], showed that *S. ambofaciens* W130 overproduces the stambomycins.

Incorporation of labeled precursors into the C-26 side chain of the stambomycins

MP5 agar plates containing 5 mM [U-²H]L-valine **11**, [U-²H]butyric acid **12**, [U-²H]L-isoleucine **9**, or [U-²H]L-leucine **10** were overlaid with sterile permeable membranes (12-14000 Da molecular weight cut off, size 20). A 4 µL aliquot of *S. ambofaciens* W130 spore stock was spread onto each plate and they were incubated at 30 °C for 4 days. The mycelia were scraped off the membranes and extracted with methanol. The extracts were analyzed by UHPLC-ESI-TOF-MS on a Dionex 3000 RS UHPLC attached to a Bruker MaXis mass spectrometer (Figures 3.15-3.18). An Agilent Zorbax Eclipse plus column (C18, 100x2.1mm, 1.8µm) was used. Mobile phases consisted of A (water with 0.1% formic acid) and B (acetonitrile with 0.1% formic acid), the flow rate was 0.2 ml/min and absorbance at 240 nm was monitored. The elution profile was as follows: 0 min, 20% B; 5 min, 20% B; 20 min, 100% B. The mass spectrometer was operated in positive mode, with a scan range of 50-3,000 *m/z*, and calibrated with sodium formate (10mM) through a loop injection of 20 µL at the beginning of each run. Source conditions were as follows. End plate offset: -500V; capillary: -4500 V; nebulizer gas (N₂): 1.4 bar; dry gas (N₂): 8L/min; dry temperature: 180 °C. Ion transfer conditions were as follows. Ion funnel 1 RF: 200 Vpp; ion funnel 2 RF: 200 vpp; hexapole RF: 200 Vpp; quadrupole ion energy: 5 ev; quadrupole low mass: 55 *m/z*; collision energy: 5.0 ev; collision RF: ramped from 800 to 1500 Vpp; transfer time: from 100 to 155 µs; pre-pulse storage time: 5 µs.

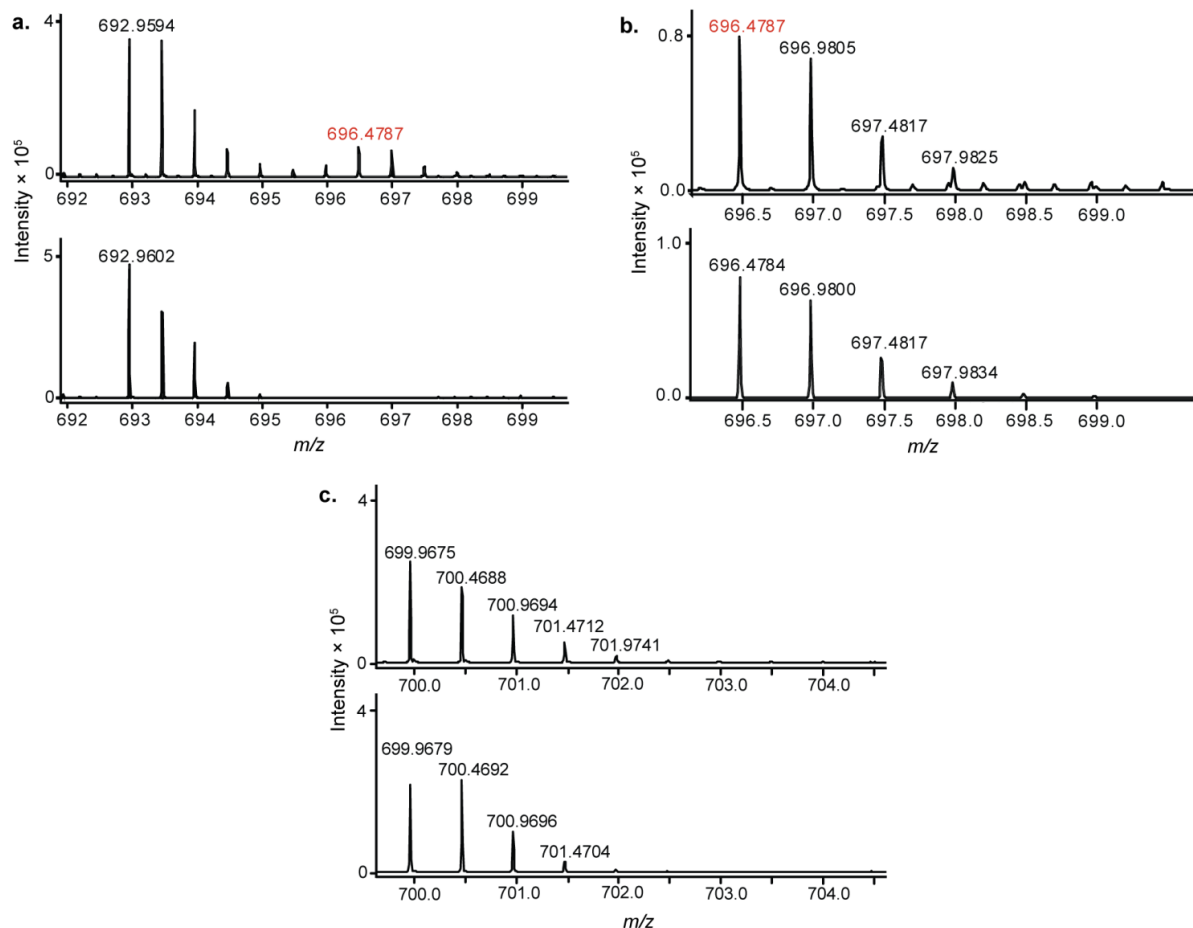


Figure 3.15: Mass spectra from UHPLC-ESI-Q-TOF-MS analysis of methanolic mycelial extracts from the [U-²H]L-valine **11 incorporation experiments.**

(a) Spectra resulting from addition of 5 mM [U-²H]L-valine **11** (top panel) and no [U-²H]L-valine (bottom panel) to the growth medium. The signal at $m/z = 692.9594$ corresponds to $[M+Na+H]^{2+}$ for stambomycins C/D **3/4** and the signal at $m/z = 696.4787$ corresponds to $[M+Na+H]^+$ for stambomycin C **3** derived from incorporation of [U-²H]L-valine **11**. (b) Comparison of the spectrum measured for labeled stambomycin C **3** derived from incorporation of [U-²H]L-valine **11** (top panel) with the spectrum calculated for the $[C_{72}H_{125}D_7NNaO_{22}]^{2+}$ ion (bottom panel). (c) Spectra for stambomycins A/B **1/2** resulting from addition of 5 mM [U-²H]L-valine **11** (top panel) and no [U-²H]L-valine (bottom panel) to the growth medium. No specific incorporation of the deuterium-labeled precursor is observed.

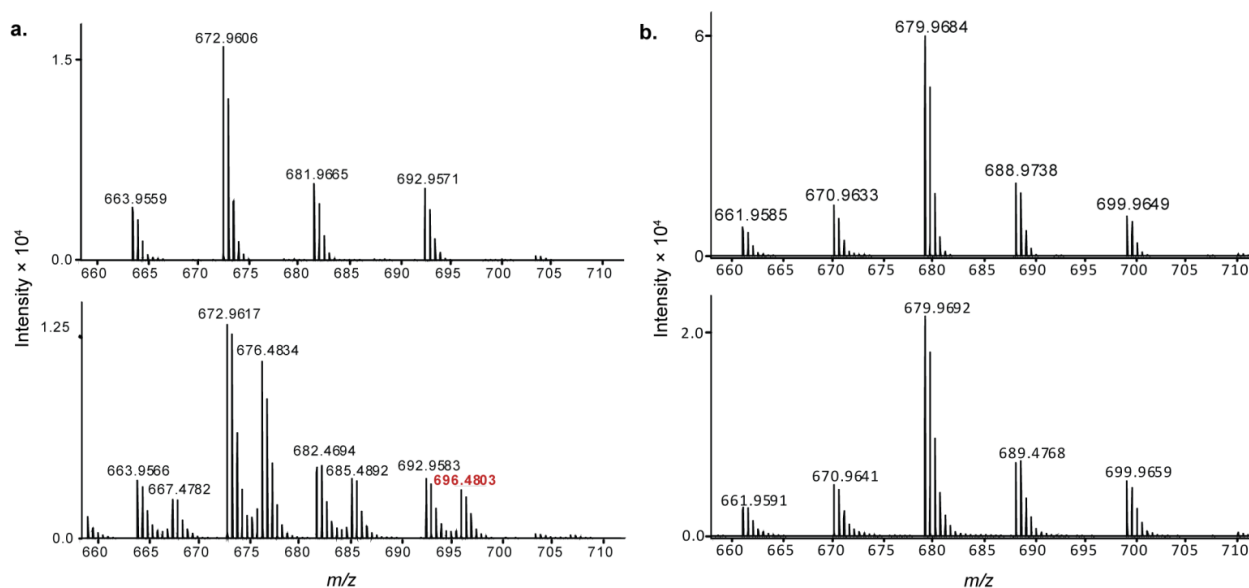


Figure 3.16: Mass spectra from UHPLC-ESI-Q-TOF-MS analysis of methanolic mycelial extracts from the [U-²H]butyric acid **12 incorporation experiments.**

(a) Spectra resulting from addition of 5 mM [U-²H]butyric acid **12** (bottom panel) and no [U-²H]butyric acid (top panel) to the growth medium. The signal at $m/z = 692.9583$ corresponds to $[M+Na+H]^{2+}$ for stambomycins C/D **3/4** and the signal at $m/z = 696.4803$ corresponds to $[M+Na+H]^{2+}$ for stambomycin D **4** derived from incorporation of [U-²H]butyric acid **12** (calculated for $[C_{72}H_{125}D_7NNaO_{22}]^{2+}$:696.4784). (b) Spectra for stambomycins A/B **1/2** resulting from addition of 5 mM [U-²H]butyric acid **12** (top panel) and no [U-²H]butyric acid (bottom panel) to the growth medium. No specific incorporation of the deuterium-labeled precursor is observed.

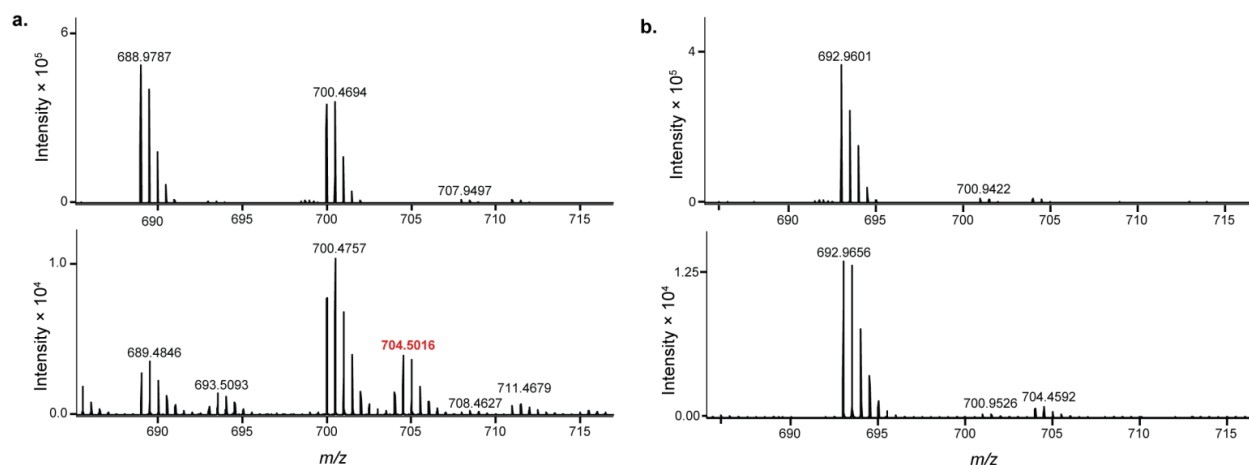


Figure 3.17: Mass spectra from UHPLC-ESI-Q-TOF-MS analysis of methanolic mycelial extracts from the [U-²H]L-isoleucine **9 incorporation experiments.**

(a) Spectra resulting from addition of 5 mM [U-²H]L-isoleucine **9** (bottom panel) and no [U-²H]L-isoleucine (top panel) to the growth medium. The signal at $m/z = 700.4757$ corresponds to $[M+Na+H]^{2+}$ for stambomycins A/B **1/2** and the signal at $m/z = 704.5016$ corresponds to $[M+Na+H]^{2+}$ for stambomycin A **1** derived from incorporation of [U-²H]L-isoleucine **9** (calculated for $[C_{73}H_{125}D_9NNaO_{22}]^{2+}$: 704.4925;). (b) Spectra for stambomycins C/D **3/4** resulting from addition of 5 mM [U-²H]L-isoleucine **9** (bottom panel) and no [U-²H]L-isoleucine (top panel) to the growth medium. No specific incorporation of the deuterium-labeled precursor is observed.

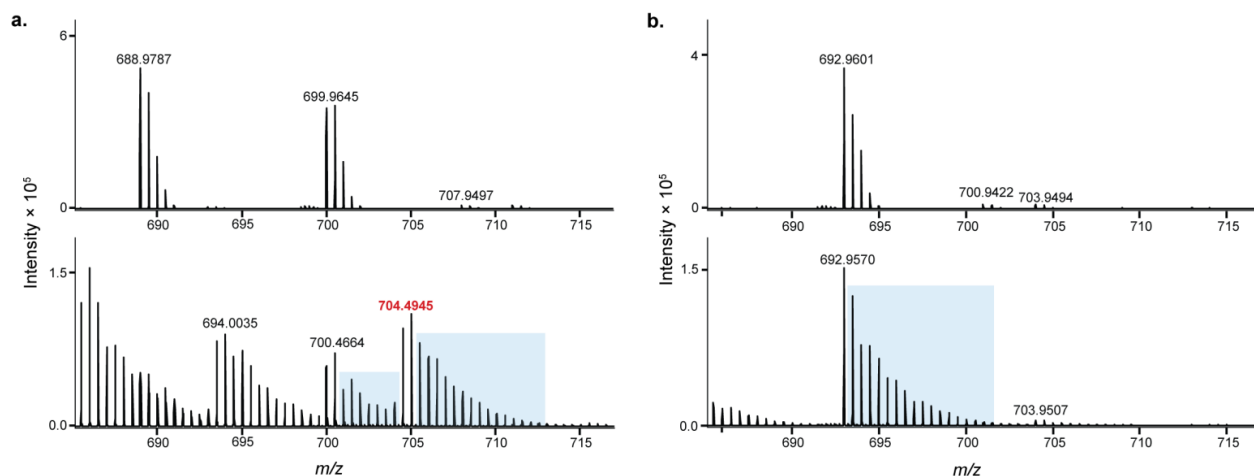


Figure 3.18: Mass spectra from UHPLC-ESI-Q-TOF-MS analysis of methanolic mycelial extracts from the [U-²H]L-leucine **10 incorporation experiments.**

(a) Spectra resulting from addition of 5 mM [U-²H]L-leucine **10** (bottom panel) and no [U-²H]L-leucine (top panel) to the growth medium. The signal at m/z = 699.9645 corresponds to $[M+Na+H]^{2+}$ for stambomycins A/B **1/2** and the signal at m/z = 704.4945 corresponds to $[M+Na+H]^{2+}$ for stambomycin B **2** derived from incorporation of [U-²H]L-isoleucine **10** (calculated for $[C_{73}H_{125}D_9NNaO_{22}]^{2+}$:704.4925). (b) Spectra for stambomycins C/D **3/4** resulting from addition of 5 mM [U-²H]L-leucine **10** (bottom panel) and no [U-²H]L-leucine (top panel) to the growth medium. No specific incorporation of the deuterium-labeled precursor was observed. The regions of the spectra highlighted by the blue boxes result from the incorporation of deuterium labeled acetyl-CoA derived from the catabolism of [U-²H]L-leucine **10**.

Precursor directed biosynthesis of stambomycin analogue 22

Heptanoic acid **21** was fed to cultures of *S. amboaciens* W130 as described in section titled “Incorporation of labeled precursors into the C-26 side chain of the stambomycins”. UHPLC-ESI-Q-TOF-MS analysis of the methanolic mycelial extracts, as described in the indicated section, showed that a new metabolite with a molecular formula corresponding to **22** was produced (Figure 3.19).

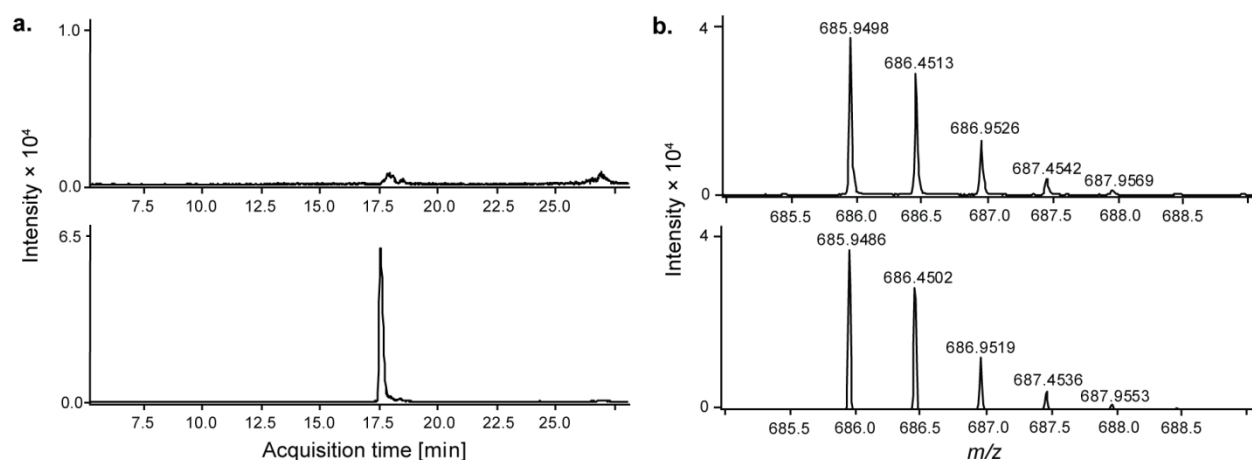


Figure 3.19: Mass spectra of stambomycin analogue 22 feeding studies with heptanoic acid 21.

(a) EIC at $m/z = 685.9$, corresponding to $[M+Na+H]^{2+}$ for stambomycin analogue **22**, from UHPLC-ESI-Q-TOF-MS analyses of methanolic mycelial extracts *S. ambofaciens* W130 grown in the absence (top) and presence (bottom) of 5 mM heptanoic acid **21**. (b) Top panel: Mass spectrum of **22** from UHPLC-ESI-Q-TOF-MS analyses of *S. ambofaciens* W130 fed with **21**. Bottom panel; Simulated spectrum for $C_{71}H_{130}NNaO_{22}^{2+}$.

The stambomycin analogue was purified from the mycelial extract by semi-preparative HPLC on an Agilent C18 column (100 x 21 mm, fitted with a C18 pre-column 10 x 21 mm) connected to an Agilent 1100 instrument. The mobile phases used were water (A) and methanol (B), both of which contained 0.1% formic acid. The total run time was 55 minutes with a flow rate of 5 mL/min and absorbance was monitored at 240nm. The elution conditions employed during the run are as follows: 0 minutes, 60% A / 40% B; 5 minutes, 60% A / 40% B; 30

minutes, 100% B; 35 minutes, 100% B; 40 minutes, 60% A / 40% B. Fractions containing the stambomycin analogue were identified by ESI-MS analysis, pooled and concentrated to dryness by rotary evaporation and lyophilisation. The residue was analyzed ^1H , ^{13}C -APT, COSY, TOCSY, and HMBC NMR spectroscopy (Table 3.3 and Figure 3.20-3.24).

Table 3.3: Comparison of ^1H and ^{13}C NMR data for stambomycin D 4 [26] and stambomycin analogue 22.

The atom numbering is shown in Figure 3.4 of the main manuscript.

Position	δ_{C} (ppm) in 4	δ_{C} (ppm) in 22	δ_{H} (ppm) in 4	δ_{H} (ppm) in 22	$J_{\text{H-H}}$ (Hz)
1	173.3	173.1	-	-	
2	44.3	44.3	2.52, 2.83	2.53, 2.76	14.5
3	100.0	100.1	-	-	
4	46.2	46.2	1.56	1.57	
4-Me	12.4	12.5	1.04	1.11	
5	81.0	81.0	3.66	3.66	
1'	104.6	105.8	4.44	4.44	7.0
2'	70.0	69.8	3.54	3.54	7.5, 10.5
3'	72.0	72.0	3.23	3.23	10.5
4'	70.8	70.5	3.38	3.39	10.0
5'	74.0	74.0	3.42	3.42	6.0, 9.0
6'	18.1	17.8	1.29	1.29	
3'-NMe ₂	42.2	42.3	3.03	3.00	
6	39.1	38.9	1.61, 1.90	1.91	
7	70.0	69.9	3.83	3.83	
8	40.8	41.0	1.56	1.58	
8-Me	9.3	9.4	0.93	0.89	
9	79.9	79.9	3.90	3.90	7.5
10	137.1	136.9	-	-	
10-Me	12.4	12.4	1.65	1.65	
11	128.0	127.8	6.06	6.04	11.0
12	129.9	129.5	6.35	6.36	11.0, 14.5
13	131.7	131.6	5.81	5.80	
14	39.8	39.8	2.34	2.34	6.5

15	75.8	75.7	3.77	3.78	
16	42.6	42.9	1.52	1.52	
16-Me	7.4	7.4	0.95	0.96	
17	75.5	75.5	3.72	3.72	
18	36.1	36.1	1.48	1.47	
19	23.5	23.3	1.32, 1.34	1.32	
20	38.5	38.5	1.47	1.45	
21	71.1	70.9	3.55	3.55	
22	43.1	43.0	1.61, 1.67	1.61, 1.68	
23	77.5	77.4	4.23	4.23	7.0
24	139.4	138.5	-	-	
24-Me	12.2	12.4	1.64	1.64	
25	129.3	129.9	5.23	5.21	10.5
26	41.9	41.8	2.45	2.44	
1''	32.4	32.4	1.18, 1.79	1.19, 1.81	7.0
2''	26.6	30.6	1.37	1.38	
3''	35.8	34.9	1.31, 1.44	1.29	
4''	34.0	33.0	1.43, 1.49	1.28	
5''	29.2	14.4	1.29	0.88	
6''	11.5	-	0.87	-	
27	79.7	79.6	3.36	3.36	
28	74.2	74.0	3.47	3.47	
29	27.8	27.8	1.43, 1.68	1.42, 1.66	
30	36.1	36.0	1.43, 1.70	1.69	
31	70.1	69.9	3.83	3.83	
32	46.4	46.2	1.50, 1.57	1.51, 1.57	
33	66.8	66.7	4.06	4.07	
34	42.0	41.8	1.48, 1.50	1.49, 1.50	
35	72.9	72.8	3.74	3.74	
36	40.8	40.6	1.48	1.47	
36-Me	15.6	15.6	0.92	0.93	
37	29.8	29.6	1.28, 1.12	1.12, 1.56	
38	36.0	36.0	1.50	1.49	
39	76.3	76.3	3.72	3.72	
40	42.3	42.3	1.51	1.51	
40-Me	7.0	6.9	0.91	0.92	

41	75.9	75.9	3.71	3.71	
42	33.8	33.6	1.43, 1.62	1.43, 1.62	
43	22.9	22.9	1.43	1.45	
44	38.8	38.9	1.45, 1.46	1.45, 1.46	
45	70.6	70.5	3.67	3.67	
46	44.8	44.9	1.60, 1.65	1.60	
47	71.1	71.0	4.28	4.27	6.5, 13.0
48	136.7	136.8	5.80	5.77	5.0, 10.5
49	130.8	130.9	5.72	5.73	6.0, 15.5
50	72.8	72.8	5.43	5.42	6.0
51	20.8	20.8	1.33	1.35	

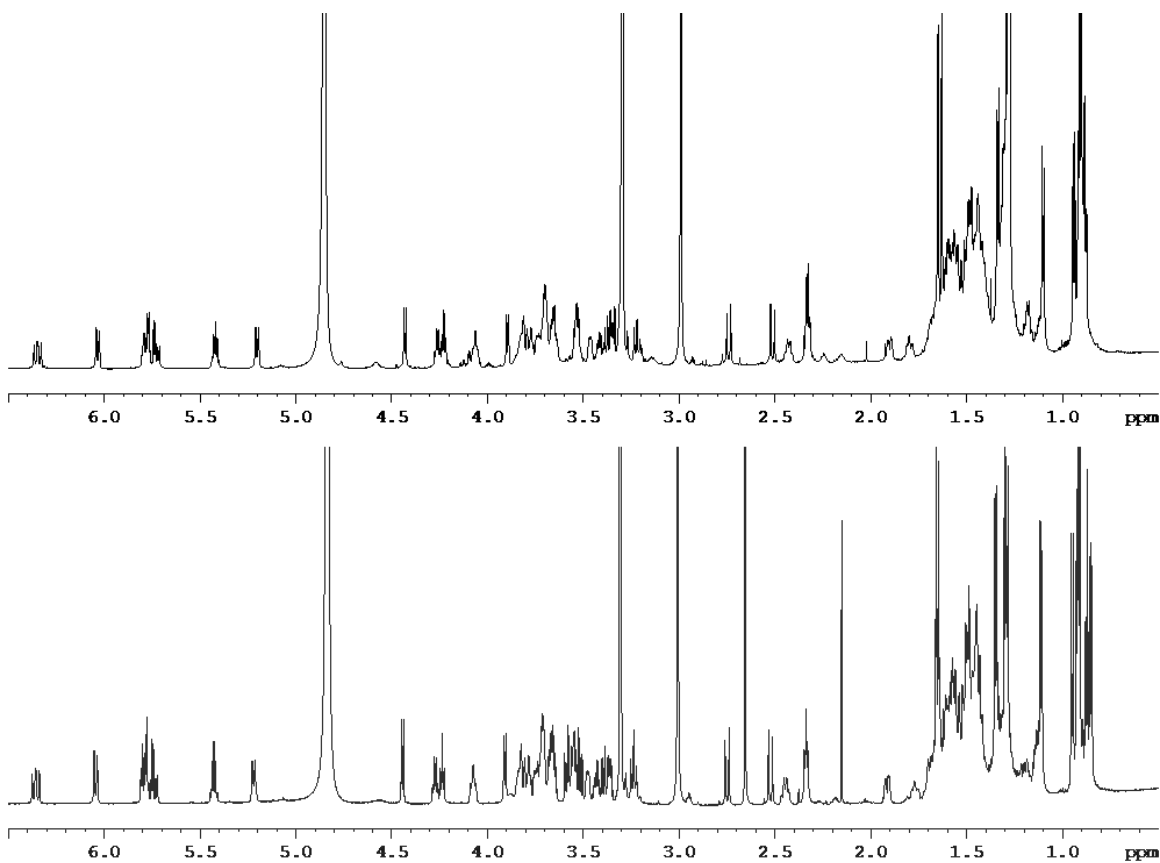


Figure 3.20. Comparison of the ¹H NMR spectra of 22 and stambomycins C/D 3/4.
¹H NMR spectra (700 MHz, d₄-MeOH) of 22 (top) and stambomycins C/D 3/4 (bottom)

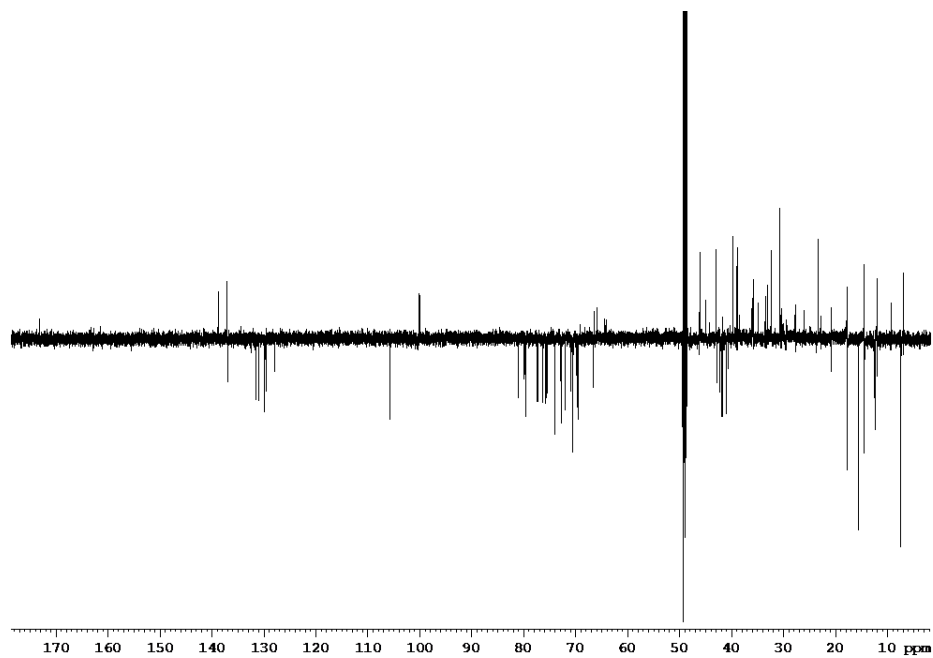


Figure 3.21. ^{13}C -APT NMR spectrum of **22**.
 ^{13}C -APT NMR spectrum (175 MHz, d_4 -MeOH)

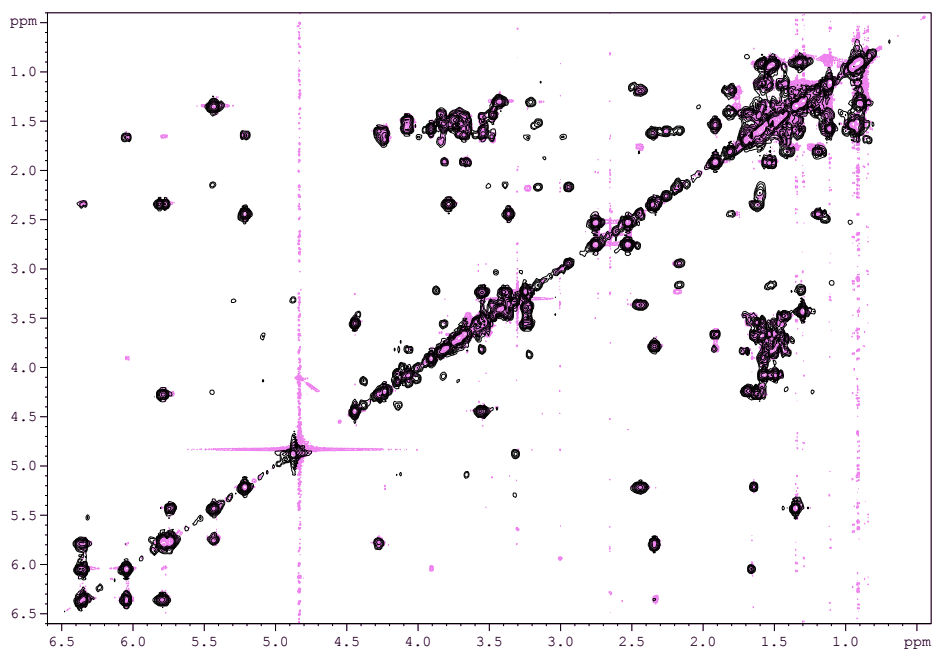


Figure 3.22. Overlaid COSY NMR spectra of **22** and stambomycins C/D **3/4**.
 COSY NMR spectra (700 MHz, d_4 -MeOH) of **22** (black) and stambomycins C/D **3/4** (pink)

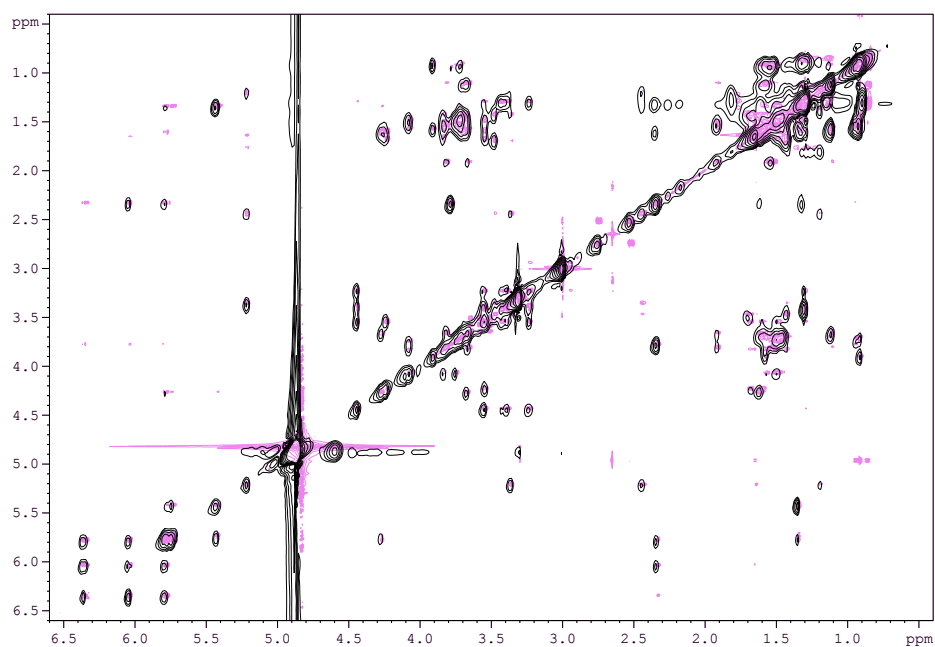


Figure 3.23. Overlaid TOCSY NMR spectra of 22 and stambomycins C/D 3/4.
TOCSY NMR spectra (700 MHz, d₄-MeOH) of 22 (black) and stambomycins C/D 3/4 (pink)

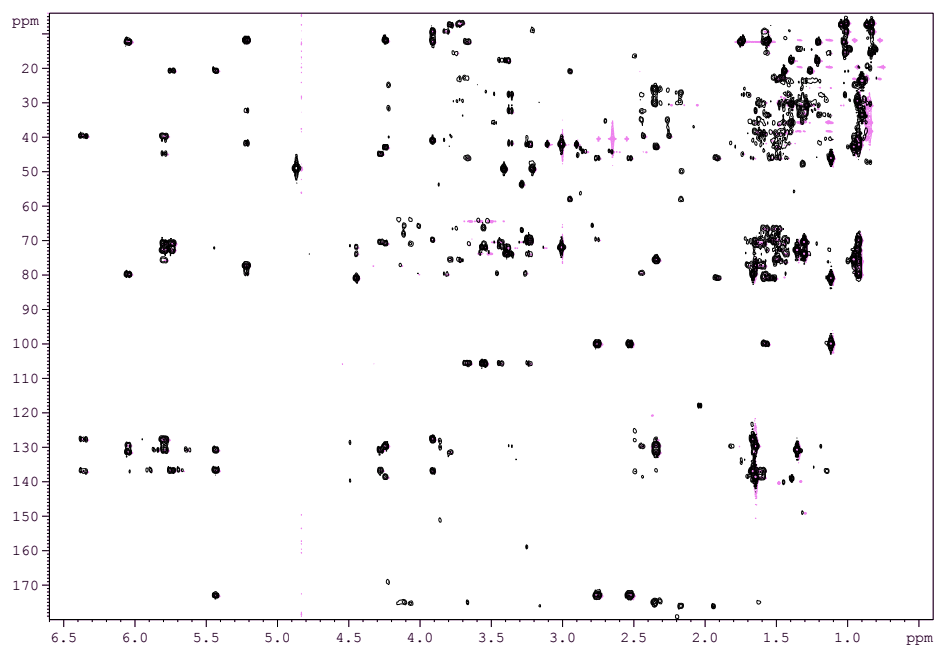


Figure 3.24. Overlaid HMBC NMR spectra of 22 and stambomycins C/D 3/4.
HMBC NMR spectra (700 MHz / 175 MHz, d₄-MeOH) of 22 (black) and stambomycins C/D 3/4 (pink)

Syntheses

All reagents (purchased from Sigma Aldrich, Alfa Aesar, Acros Organics and Jena Bioscience) and solvents were used as supplied, unless otherwise stated. THF and DCM were dried over calcium hydride, distilled and stored over 4Å molecular sieves under argon. Dry triethylamine was obtained by distillation over calcium hydride. Dry dimethyl malonate was obtained by vacuum distillation at 37 mbar pressure at 90 °C under argon and stored over 4 Å molecular sieves under argon. Commercially available *p*-toluenesulphonyl chloride was recrystallized before use. To a solution of *p*-toluenesulphonyl chloride (25 g) in chloroform (100 ml) was added hexane (500 ml). The resulting mixture was stirred for 30 min, the white precipitate formed was removed *via* filtration and the solvent removed under vacuum. Solvents were evaporated using a Buchi Rotavapor R-200 equipped with a Buchi Vacuubrand pump. Residual solvent was removed on a high vacuum line using an Edwards E2M5 vacuum pump. Flash chromatography was conducted on Fluka Silica Gel (40 - 63 µm, 60 Å). TLC was performed on aluminum backed plates pre-coated with Merck silica gel (60 F₂₅₄) and visualized by UV radiation and potassium permanganate. All glassware was dried by heating in a 140 °C oven for a minimum of two hours.

Low resolution ESI mass spectra were recorded using a Bruker Esquire 2000 spectrometer and high resolution mass spectra were measured by the University of Warwick Mass Spectrometry service on a Bruker MaXis ESI-Q-TOF mass spectrometer. NMR spectra were obtained using a Bruker DPX400 (¹H 400 MHz and ¹³C 100 MHz) equipped with a proton-carbon dual probe. High field spectra were measured by the University of Warwick NMR Service using Bruker AV III-600 (¹H 600 MHz and ¹³C 150 MHz) or AV II-700 (¹H 700 MHz

and ^{13}C 175 MHz) instruments. Chemical shifts are quoted in ppm with respect to the residual solvent peak. Deuterated solvents (Sigma Aldrich) were used as supplied.

*Synthesis of [1- $^2\text{H}_2$]n-pentanol **28***

To a stirred suspension of lithium aluminum deuteride (1.96 g, 51.7 mmol, 3 equiv) in 25 mL of dry THF under argon was added methyl valerate (2.25 mL, 17.2 mmol, 1 equiv) in 5 mL of dry THF drop wise. The resulting mixture was heated to 70 °C and the reaction was monitored by TLC (25% ethyl acetate in 40 – 60 °C pet. ether). Upon completion, the reaction was cooled to 0 °C and quenched with 2 M HCl, resulting in a white precipitate which was removed by vacuum filtration. The filtrate was concentrated *in vacuo* and diluted with 250 mL of water. The resulting solution was extracted with 3 x 100 mL ethyl acetate. The organics were combined, washed with brine and dried over magnesium sulfate. The solvent was removed *in vacuo* to yield the crude product, which was purified by flash column chromatography (silica, 25% diethyl ether in pentane) to yield the product as a colorless oil (1.37 g, 88%). ^1H -NMR (CDCl_3 , 400 MHz) δ_{H} : 0.86 (3H, $\text{CH}_3(\text{CH}_2)_3\text{CD}_2\text{OH}$, t, $J_{\text{H-H}} = 6.5$ Hz), 1.28 – 1.32 (4H, $\text{CH}_3(\text{CH}_2)_2\text{CH}_2\text{CD}_2\text{OH}$, m), 1.54 (2H, $\text{CH}_3(\text{CH}_2)_2\text{CH}_2\text{CD}_2\text{OH}$, t, $J_{\text{H-H}} = 6.5$). ^{13}C -NMR (CDCl_3 , 100 MHz) δ_{C} : 14.03 ($\underline{\text{C}}\text{H}_3(\text{CH}_2)_3\text{CD}_2\text{OH}$), 22.54 ($\text{CH}_3\text{-}\underline{\text{C}}\text{H}_2\text{-(CH}_2)_2\text{CD}_2\text{OH}$), 27.96 ($\text{C}_2\text{H}_5\text{-}\underline{\text{C}}\text{H}_2\text{CH}_2\text{CD}_2\text{OH}$), 32.35 ($\text{C}_3\text{H}_7\text{-}\underline{\text{C}}\text{H}_2\text{CD}_2\text{OH}$), 62.34 ($\text{C}_4\text{H}_9\text{-}\underline{\text{C}}\text{D}_2\text{OH}$, sep, $J_{\text{C-D}} = 19.7$ Hz).

*Synthesis of [1- $^2\text{H}_2$]n-pentane 4-methylbenzenesulfonate **29***

To a stirred solution of [1- $^2\text{H}_2$]n-pentanol **28** (1.17g, 12.9 mmol, 1 equiv) in 23 mL of dry DCM, under argon, was added 4-dimethylamino-pyridine (473 mg, 3.88 mmol, 0.3 equiv) and triethylamine (3.27g, 32.3 mmol, 2.5 equiv) followed by *p*-toluene sulfonyl chloride (2.711g,

14.4 mmol, 1.11 equiv). The reaction was stirred at room temperature under argon for 17 hours, after which TLC indicated that it had gone to completion. Ethanolamine (1.58g, 26.6 mmol, 2 equiv) was added and the resulting mixture was stirred for a further 2 hours. 125 mL of water was added, the layers were separated and the aqueous layer was extracted with 4 x 50 mL of DCM. The organics were combined, washed with brine and dried over magnesium sulfate. The solvent was removed *in vacuo* and the residue was purified via flash column chromatography (silica, 10% ethyl acetate in 40-60 °C pet ether) to give the product as a colorless oil (2.13 g, 68%). ¹H-NMR (CDCl₃, 400 MHz) δ: 0.84 (3H, CH₃-, t, *J*_{H-H} = 7.0 Hz), 1.23-1.28 (4H, CH₃-CH₂-CH₂-, m), 1.62 (2H, CH₃-(CH₂)₂-CH₂-, t, *J*_{H-H} = 7.0 Hz), 2.44 (3H, -O-S(O)₂-Ar-CH₃, s), 7.34 (2H, -O-S(O)₂-**meta** ArH-CH₃, d, *J*_{H-H} = 8.0 Hz), 7.79 (2H, -O-S(O)₂-**ortho** ArH-CH₃, d, *J*_{H-H} = 8.0 Hz). ¹³C-NMR (CDCl₃, 100 MHz) δ: 13.93 (CH₃(CH₂)₃-), 21.76 (-O-S(O)₂-Ar-CH₃), 22.16 (CH₃-CH₂-CH₂-), 27.53 (CH₃-CH₂-CH₂-), 28.44 (CH₃-(CH₂)₂-CH₂-), 70.18 (CH₃(CH₂)₃CD₂-, sep, *J*_{C-D} = 22.7 Hz), 127.99 (-O-S(O)₂-orthoArC-CH₃). HRMS (ESI⁺): *m/z* calculated for C₁₂H₁₆D₂NaO₃S⁺: 267.0994; found: 267.0990.

*Synthesis of dimethyl ([1-²H₂]n-pentyl)malonate **30***

To a stirred suspension of sodium hydride (471 mg, 19.65 mmol, 14.8 equiv) in 10 mL of dry DCM, under argon, was added distilled dimethyl malonate (2.70 g, 20.46 mmol, 5 equiv) in 15 mL dry of THF drop wise. The resulting mixture was stirred at 60 °C for 1 hour and [1-²H₂]n-pentyl-4-methylbenzenesulfonate **29** (1.00 g, 4.09 mmol, 1 equiv) in 10 mL of dry THF was added. Stirring was continued at 60 °C under argon for a further 36 hours, after which TLC indicated that the reaction had gone to completion. 125 mL of water was added cautiously and the aqueous layer was extracted with 4 x 100 mL of ethyl acetate. The organics were combined,

washed with brine and dried over magnesium sulfate. The solvent was removed *in vacuo* and the residue was purified by flash column chromatography (silica, 25% ethyl acetate in 40-60 °C pet ether), yielding the product as a colorless oil (616 mg, 74%). ¹H-NMR (CDCl₃, 400 MHz) δ: 0.84 (3H, CH₃-, t, *J*_{H-H} = 7.5 Hz), 1.22 – 1.29 (6H, CH₃-CH₂-CH₂-CH₂-, m), 3.30 (1H, -CD₂-CH-(COO-CH₃)₂, s), 3.69 (6H, -CD₂-CH-(COO-CH₃)₂, s). ¹³C-NMR (CDCl₃, 100 MHz) δ: 13.9 (CH₃-), 27.38 and 31.32 (CH₃(CH₂)₂), 26.84 (CH₃-CH₂-CH₂-CH₂-), 28.16 (C₄H₉-CD₂-CH-(COO-CH₃)₂, sep, *J*_{C-D} = 20.0 Hz), 51.60 (-CD₂-CH-(COO-CH₃)₂), 52.42 (-CD₂-CH-(COO-CH₃)₂), 170.00 (-CD₂-CH-(COO-CH₃)₂). HRMS (ESI⁺): *m/z* calculated for C₁₀H₁₆D₂NaO₄⁺: 227.1223 [M+Na]⁺. Found: 227.1221

Synthesis of [3-²H₂]heptanoic acid 23

To a stirred solution of dimethyl ([1-²H₂]*n*-pentyl)malonate **30** (500 mg, 2.45 mmol, 1 equiv) in 10 mL of methanol, was added 5 M sodium hydroxide (2.9 mL, 14.69 mmol, 6 equiv) drop wise. The resulting mixture was stirred under reflux for 18 hours, after which time TLC indicated that no starting material remained. The mixture was cooled to room temperature, acidified with 2 M HCl and the methanol was removed *in vacuo*. 100 mL of 20% brine was added to the residue and the resulting solution was extracted with 3 x 75 mL of ethyl acetate. The organics were combined, dried over magnesium sulfate and the solvent was removed *in vacuo* to yield diacid **31** as a crystalline solid. 15 mL of toluene was added and the resulting mixture was heated under reflux, at 125 °C for 96 hours. The mixture was cooled to room temperature and the toluene was removed *in vacuo* to give the product (306 mg, 95%). ¹H-NMR (CDCl₃, 600 MHz) δ: 0.80 (3H, CH₃-, t, *J*_{H-H} = 7.0 Hz), 1.24-1.29 (6H, CH₃-CH₂-CH₂-CH₂-, m), 2.25 (2H, -CD₂-CH₂-COOH, s). ¹³C-NMR (CDCl₃, 150 MHz) δ: 14.15 (CH₃-), 22.60 and 31.51 (CH₃(CH₂)₂),

24.05 ($\text{CH}_3(\text{CH}_2)_3\text{CD}_2^-$, sep, $J_{\text{C-D}} = 19.7 \text{ Hz}$), 28.66 ($\text{CH}_3(\text{CH}_2)_2\text{CH}_2^-$), 34.06 ($\text{CH}_3(\text{CH}_2)_3\text{CD}_2\text{CH}_2^-$), 180.50 ($-\text{CH}_2\text{COOH}$). HRMS (ESI): m/z calculated for $\text{C}_7\text{H}_{11}\text{D}_2\text{O}_2^-$: 131.1047; Found: 131.1046. The ^1H and ^{13}C NMR spectra for labeled $[3\text{-}^2\text{H}_2]\text{heptanoic acid } \mathbf{23}$ and unlabeled heptanoic acid $\mathbf{21}$ are compared in Figures 3.25 and 3.26.

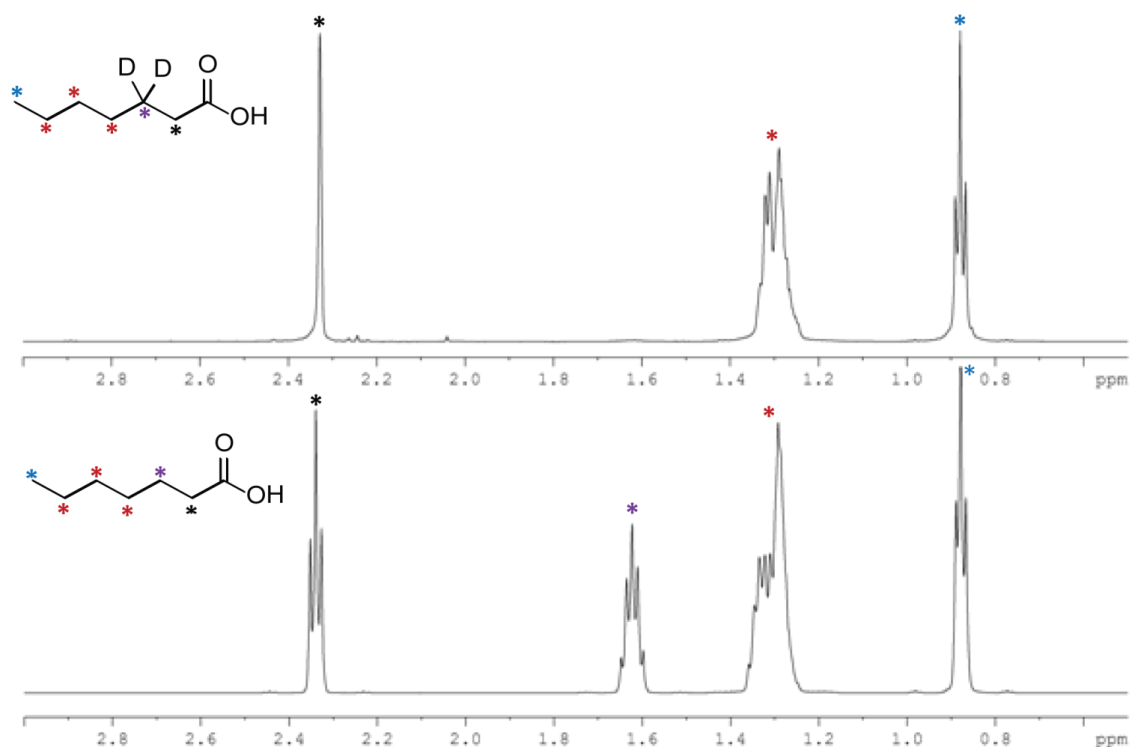


Figure 3.25. Comparison of the ^1H NMR spectra of $[3\text{-}^2\text{H}_2]\text{heptanoic acid } \mathbf{23}$ and unlabeled heptanoic acid $\mathbf{21}$.

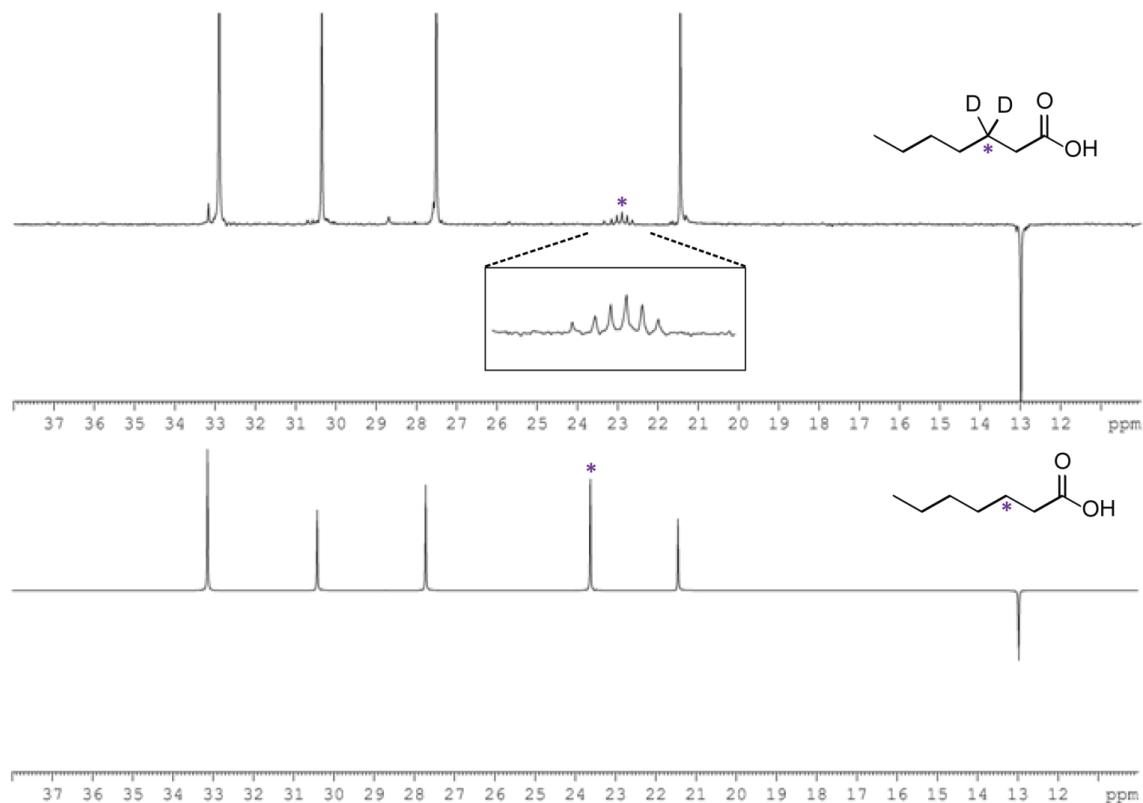


Figure 3.26. Comparison of the ^{13}C NMR spectra of $[3\text{-}^2\text{H}_2]$ heptanoic acid **23 and unlabeled heptanoic acid **21**.**

The splitting of the C-3 signal as a result of the attached deuterium atoms is shown in the inset.

*Incorporation of $[3\text{-}^2\text{H}_2]$ heptanoic acid **23** into stambomycin analogue **22** and reveromycin D **32***

$[3\text{-}^2\text{H}_2]$ heptanoic acid **23** was fed to cultures of *S. ambofaciens* W130 and the mycelial extract was prepared as described in section titled “Precursor directed biosynthesis of stambomycin analogue **22**”. UHPLC-ESI-Q-TOF-MS analysis of the extract, as described in section titled “Incorporation of labeled precursors into the C-26 side chain of the stambomycins”, showed that both of the deuterium atoms had been retained in the resulting **22** (Figure 3.27).

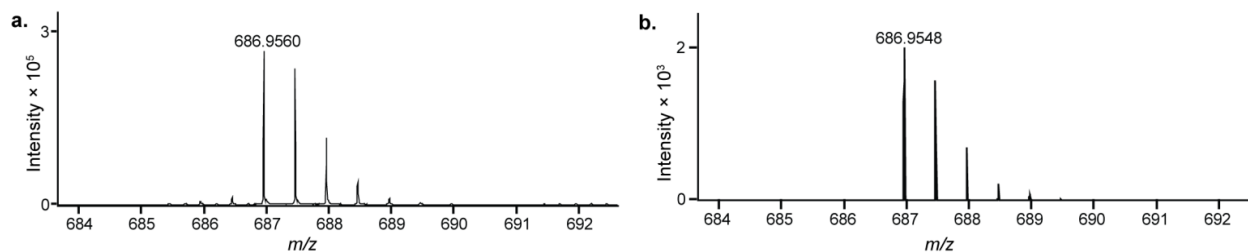


Figure 3.27. Mass spectra from stambomycin analogue 22 feeding studies with [3-²H₂]heptanoic acid 23.

UHPLC-ESI-Q-TOF-MS (a) Measured spectrum; (b) Simulated spectrum for C₇₁H₁₂₈D₂NNaO₂₂²⁺ corresponding to [M+Na+H]⁺ for doubly labeled **22**.

Incorporation of [3-²H₂]heptanoic acid 23 into reveromycin D 32

The $\Delta revR$ and $\Delta revT::samR0483$ mutants (Table 3.1) were grown at 28 °C on MS agar (2% (w/v) soy flour, 2 % (w/v) D-mannitol, 2% (w/v) agar). Spores were used to inoculate a 500-mL cylindrical flask containing 70 mL of SY medium (0.1% (w/v) yeast extract (Difco), 0.1% (w/v) NZ-amine (Wako) and 1% (w/v) starch; pH 7.0) and the resulting pre-culture was grown for 2 days at 28 °C at 150 rpm. 1 mL of the pre-culture was used to inoculate a 500 mL cylindrical flask containing 70 mL of production medium (RM-PM; 2% (w/v) potato dextrose (Difco), 1% (w/v) malt extract (Difco), 1% (w/v) dried yeast (Asahi beer), (w/v) tomato juice (Table land, Maruzen Food), 0.1%(w/v) K₂HPO₄, 0.1% (w/v) NaCl, 0.03% (w/v) MgSO₄ · 7H₂O, 0.01% (w/v) NaNO₃, 0.005% (w/v) ZnSO₄ · 7H₂O and 0.005% (w/v) CuSO₄ · 5H₂O; pH 6.5 before autoclaving). After 2 days' growth at 28 °C and 150 rpm, 0.3 mM of [3-²H₂]heptanoic acid **23** was added and the culture was grown for three further days. 4 mL of culture broth was extracted with an equal volume of acetone and residual acetone was removed by evaporation. The pH was adjusted to 4 by adding acetic acid and the culture broth was extracted twice with an equal volume of ethyl acetate. The organic layer was concentrated *in vacuo* and dissolved in 1.2 mL of methanol. The resulting sample from the $\Delta revR$ mutant culture was analyzed by UPLC-

ESI-MS on an Applied Biosystems API 3200 connected to a Waters ACQUITY UPLC and a Waters ACQUITY UPLC BEH C18 column (1.7 μm , 2.1 mm internal diameter $\times$ 50 mm). The conditions were as follows: flow rate 0.7 mL min⁻¹; solvent A, water containing 0.05% formic acid; and solvent B, acetonitrile. After injection of 1 μl of the sample into a column that was equilibrated with 30% solvent B, the column was developed with a linear gradient from 30% to 100% solvent B in 1.9 min and kept at 100% solvent B for 0.96 min. Mass spectra were recorded in negative ion mode.

To separate reveromycin D **32** from other reveromycin derivatives, the ethyl acetate extracts from the $\Delta revR$ and $\Delta revT::samR0483$ mutants were applied to a PEGASIL ODS column (5 μm , 4.6 $\times$ 250 mm). The HPLC conditions were as follows: flow rate, 1 mL min⁻¹; solvent A, water containing 0.05% formic acid; and solvent B, MeOH, isocratic system of 78% MeOH. The fraction containing reveromycin D **32** from the $\Delta revT::samR0483$ mutant cultures with and without supplementation of [3-²H₂]heptanoic acid **23** was concentrated and analyzed by UPLC-ESI-MS on an Applied Biosystems API 3200 connected to a Waters ACQUITY UPLC and a Waters ACQUITY UPLC BEH C18 column (1.7 μm , 2.1 $\times$ 50 mm) (Figure 3.28). The flow rate was 0.2 mL min⁻¹ and the column was eluted with isocratic 50% water containing 0.05% formic acid and 50% acetonitrile. Mass spectra were recorded in negative ion mode.

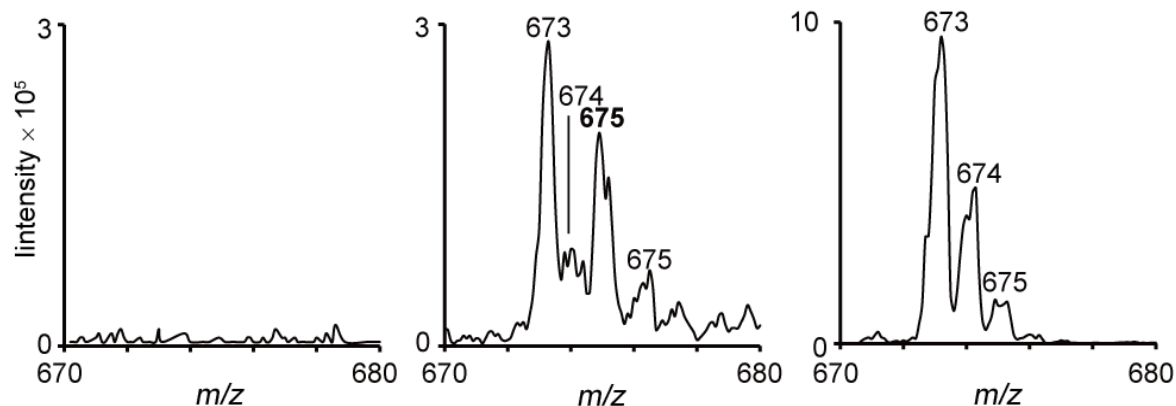


Figure 3.28. Mass spectra of reveromycin D **32 feeding studies with [3-²H₂]heptanoic acid **23**.**

Mass spectra of reveromycin D **32** from LC-MS analyses of partially purified extracts from cultures of the $\Delta revT$ mutant (left spectrum), the $\Delta revT::samR0483$ mutant to which [3-²H₂]heptanoic acid **23** has been fed (middle spectrum) and wild type *Streptomyces* sp. SN-593.

Partially purified reveromycin D **32** from the $\Delta revR$ mutant cultures with and without supplementation of [3-²H₂]heptanoic acid **23** was further analyzed by negative ion ESI-TOF-MS on a Waters Synapt G2 mass spectrometer (Figure 3.29).

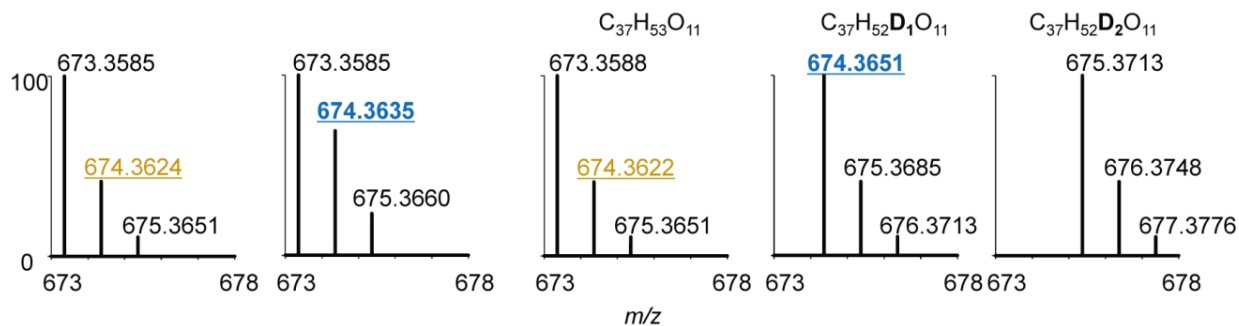


Figure 3.29. Mass spectra of reveromycin D **32 feeding studies with labeled heptanoic acid.**

Comparison of ESI-TOF mass spectra of partially purified reveromycin D **32** from cultures of *Streptomyces* sp. SN-593 $\Delta revR$ mutant grown in the absence (left) and presence (second from left) of [3,3-²H₂]heptanoic acid. The rightmost three spectra show simulated isotope distributions for unlabeled, singly-labeled and doubly-labeled reveromycin D. The data are consistent with 20% incorporation of singly-labeled heptanoic acid into **32**.

Expression of samR0483 in the revT mutant of Streptomyces sp. SN-593

The *samR0483* gene was amplified by PCR from genomic DNA of *S. ambofaciens* ATCC 23877 using the primers pOSV_samR0483_FW and pOSV_samR0483_RV and cloned into the *Cl*I/*H*indIII restriction sites of pOSV556t. The resulting plasmid, pOSV556t-*samR0483*, was used as the template for PCR amplification of *samR0483* using the primers 0483_BamHI and 0483_HindIII. The amplicon was cloned into the *B*amHI and *H*indIII sites of pTYM19-P_{aph} yielding pTYM19-P_{aph}-*samR0483*, which was used to transform *E. coli* GM2929 *hsdS*::Tn10 (pUB307::Tn7). pTYM19-P_{aph}-*samR0483* was transferred by intergenic conjugation from *E. coli* GM2929 *hsdS*::Tn10 (pUB307::Tn7) to *Streptomyces* sp. SN-593 as described previously.[33]

Cloning of samR0483 into E. coli expression vectors

The MccB-encoding gene, *samR0483*, was amplified by PCR from genomic DNA of *S. ambofaciens* ATCC 23877 using the primers TOPO_483_FW and TOPO_483_RV, and cloned into pET151 (Invitrogen) following the manufacturer's instructions to give pET151-*samR0483*. It was then amplified from pET151-*samR0483* using the primers pET28a-samR0483_FW and pET28a-samR0483_RV and the resulting amplicon was cloned into the *N*heI and *N*otI sites of pET28a (Novagen) to produce pET28a-*samR0483* encoding an N-terminal hexahistidine MccB fusion protein.

Overproduction and purification of MccB

E. coli BL21(DE3) cells containing pET28a-*samR0483* were grown in 4 X 1L of LB medium supplemented with kanamycin (50 µg/mL) to A₆₀₀ = 0.6-0.8 at 37 °C and 180 rpm.

Cultures were induced by the addition of 0.7 mM IPTG and incubated overnight (16 hours) at 19 °C and 180 rpm. Cells were harvested by centrifugation at 5500 RCF for 10 minutes, re-suspended in lysis buffer (50 mM Tris·HCl pH 8.0, 300 mM NaCl, 20 mM imidazole and 10% glycerol), sonicated, and the cell debris was removed by centrifugation (21000 RCF for 70 minutes). The clarified cell lysate was batch bound to 5 mL of Ni-IMAC resin (Bio-Rad Laboratories) for 1 hour at 4 °C. The resulting slurry was transferred to a gravity-flow column and washed twice with 50 mL of lysis buffer. His₆-MccB was eluted with 5 ml fractions containing 40, 100, 150, 250, and 450 mM imidazole in lysis buffer. Protein purity was assessed by SDS-PAGE. Fractions containing the highest purity (≥90%) were collected and dialyzed overnight at 4°C against 50 mM Tris·HCl pH 8.0, 50 mM NaCl, 10% glycerol and 1 mM DTT. Dialyzed protein was concentrated to 5-10 mg/mL using a 30,000 Da molecular weight cutoff centrifugal filter (Millipore) and injected onto a Superdex 200 column (GE Healthcare) pre-equilibrated with 50 mM Tris pH 8.0 and 1 mM DTT. Fractions were collected, analyzed by SDS-PAGE for purity, concentrated to 9 mg/ml and flash frozen for storage at -80°C.

For cocrystal studies, native His₆-MccB was treated with TEV protease (10 U/mg of MCC) post nickel affinity chromatography and dialyzed into 50 mM Tris·HCl pH 8.0, 50 mM NaCl, 10% glycerol, 1 mM DTT, and 1 mM EDTA at 18 °C for 18 hours. Complete digestion was verified using MALDI-TOF mass spectrometry. The cleaved protein was dialyzed into crystallization buffer (50 mM Tris pH 8.0 and 1 mM DTT) and subjected to anion exchange chromatography (High Trap Q, GE Healthcare) to remove excess TEV protease. Cleaved MccB was eluted using a linear gradient of 0-1 M NaCl and was dialyzed into 50 mM Tris pH 8.0 and 1 mM DTT for cocrystallization screens.

Selenomethionine *derivatized* (*SeMet*) His₆-MccB protein was produced in BL21(DE3) cells harboring pET28a-*samR0483*. Cells were grown in 4 X 1L of LB medium supplemented with kanamycin (50 µg/mL) at 37 °C and 180 rpm until A₆₀₀ reached 0.6. Prior to induction, the cells were harvested by centrifugation at 5500 RCF for 10 minutes, re-suspended and washed three times with 100 ml of M9 medium then transferred to 4 x 1L flasks containing M9 medium with the following additives: kanamycin (50 µg/mL); lysine, phenylalanine, threonine (100 mg/L); isoleucine, leucine, valine 50 (mg/L); and L-selenomethionine (80 mg/L)(Sigma). The temperature was then lowered to 19°C and protein expression was induced upon the addition of 0.7 mM IPTG. The SeMet substituted protein was purified following the same procedure as native His₆-MccB. Selenomethionine incorporation was confirmed using MALDI-TOF mass spectrometry.

Crystallization of apo and hexanoyl-CoA-bound forms of MccB

Frozen SeMET aliquots of His₆-MccB were thawed on ice and incubated with 5 mM biotin and 5 mM heptanoic acid **21** (50 mM stocks in 0.1 M Hepes pH 7.2 and 100% DMSO, respectively) (Sigma-Aldrich) at 4 °C for 1 hour prior to crystallization screens. Small crystals of *apo*-MccB formed overnight in 30% methyl propanediol (MPD) (Qiagen), 0.2 M magnesium acetate and 0.1 M MES pH 6.5. Diffraction-quality crystals were obtained by mixing 2 µL of protein solution with 2 µL of well solution (5-11% MPD, 0.2 M magnesium acetate and 0.1 M MES pH 6.5-6.8) using the hanging drop vapor diffusion method. *Apo*-MccB crystals were further optimized via four consecutive rounds of seeding using a Seed Bead (Hampton Research). The crystals were harvested, cryo-protected with 25% ethylene glycol and flash frozen in liquid nitrogen for subsequent X-ray diffraction data collection.

Prior to crystallization, cleaved native MccB at 7 mg/ml was incubated with 5 mM biotin and either 5 mM or 10 mM hexanoyl-CoA (50 mM stock in 0.1 M Hepes pH 7.2) (Sigma-Aldrich). Crystals of the MccB-hexanoyl-CoA complex were obtained by mixing 2 μ L of protein solution with 2 μ L of mother liquor (26-32 % PEG 5000 MME, 0.2 M ammonium sulfate and 0.1 M MES pH 6.5 – 7.0) using the hanging drop vapor diffusion method. MccB cocrystals grew at 25 °C over a period of 5 days. MccB cocrystals were cryo-protected in well solution plus 25% glycerol and flash frozen in liquid nitrogen for subsequent X-ray diffraction data collection.

Data Collection, Structure Determination, and Refinement

Data were collected at the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory, using beamlines 8.2.1 or 8.2.2. Data sets corresponding to the *apo* and mono-hexanoyl-CoA cocrystal forms of MccB were indexed, integrated, and scaled using Mosflm or HKL2000, respectively.[34, 35] Two separate data sets from a single crystal were collected for the tetra-hexanoyl-CoA MccB cocrystal form with 1s and 2s exposure times. The individual data sets were processed using Mosflm and scaled together using Aimless.[36] Phasing of *apo*-MccB was accomplished via molecular replacement (Phaser [37]) using the hexameric *apo*-PccB D422A structure (PDB Code: 3IBB) as a search model. An initial model was built using Autobuild, and this model was further improved through iterative rounds of model building (COOT [38]) and refinement (PHENIX.REFINE [39]). For the cocrystal structures, electron density corresponding to hexanoyl-CoA was identified using PHENIX LigandFit and ligand restraints were generated using PHENIX eLBOW. Both the *apo* and cocrystal structures were validated using PHENIX.VALIDATE and PDB_REDO.[39, 40] Resolution cutoffs were made

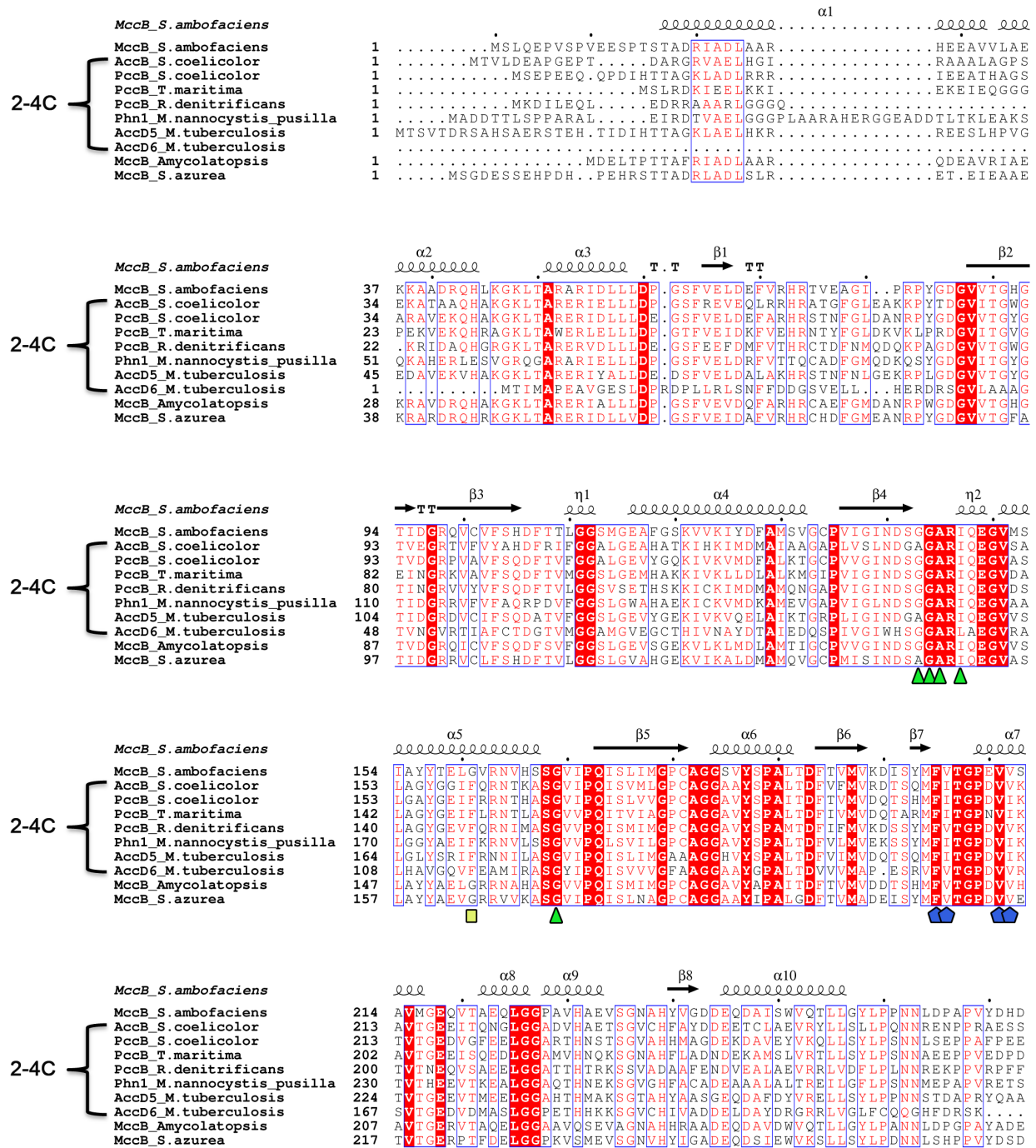
based on a combination I/σ , $CC(1/2)$, and percent completeness values.[41] Crystallographic statistics for the three MccB crystal forms are listed in Table 3.4.

Table 3.4. Crystallographic statistics.

All structures were determined from a single crystal. Highest resolution shell is shown in parenthesis.

	<i>apo</i> -MccB	MccB + 1 Hex-CoA	MccB + 4 Hex-CoA
Data collection	ALS BL821	ALS BL822	ALS BL821
Wavelength (Å)	0.9999	0.9998	1.000
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	120.56, 163.44, 186.55	110.072 165.453 190.261	110.978, 161.526, 187.294
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	58.12 - 2.45	42.22 - 2.75	48.44 - 2.85
<i>R</i> _{merge}	0.112 (0.966)	0.208 (0.837)	0.206 (0.731)
<i>I</i> / <i>I</i>	11.8 (1.6)	14.4 (3.5)	10 (3.6)
Completeness (%)	99 (97.3)	100 (99.9)	100 (99.4)
Multiplicity	6.1 (5.4)	7.4 (7.3)	6.9 (6.7)
<i>CC</i> (1/2)	(0.662)	(0.796)	(0.820)
Refinement			
Resolution (Å)	58.12 - 2.45	42.22 - 2.75	48.44 - 2.85
No. reflections	133794 (12965)	90300 (8650)	79173 (7792)
<i>R</i> _{work} / <i>R</i> _{free}	0.195/0.223	0.184/0.242	0.173/0.230
No. atoms			
Protein	22424	21976	22284
Ligand/ion	N/A	55	220
Water	956	1091	392
B-factors			
Protein	48.57	27.71	24.59
Ligand/ion	N/A	52.88	54.39
Water	47.12	26.09	20.27
R.M.S deviations			
Bond lengths (Å)	0.007	0.013	0.007
Bond angles (°)	1.07	1.17	0.90

Sequence comparison of MccB with other ACC β -subunits



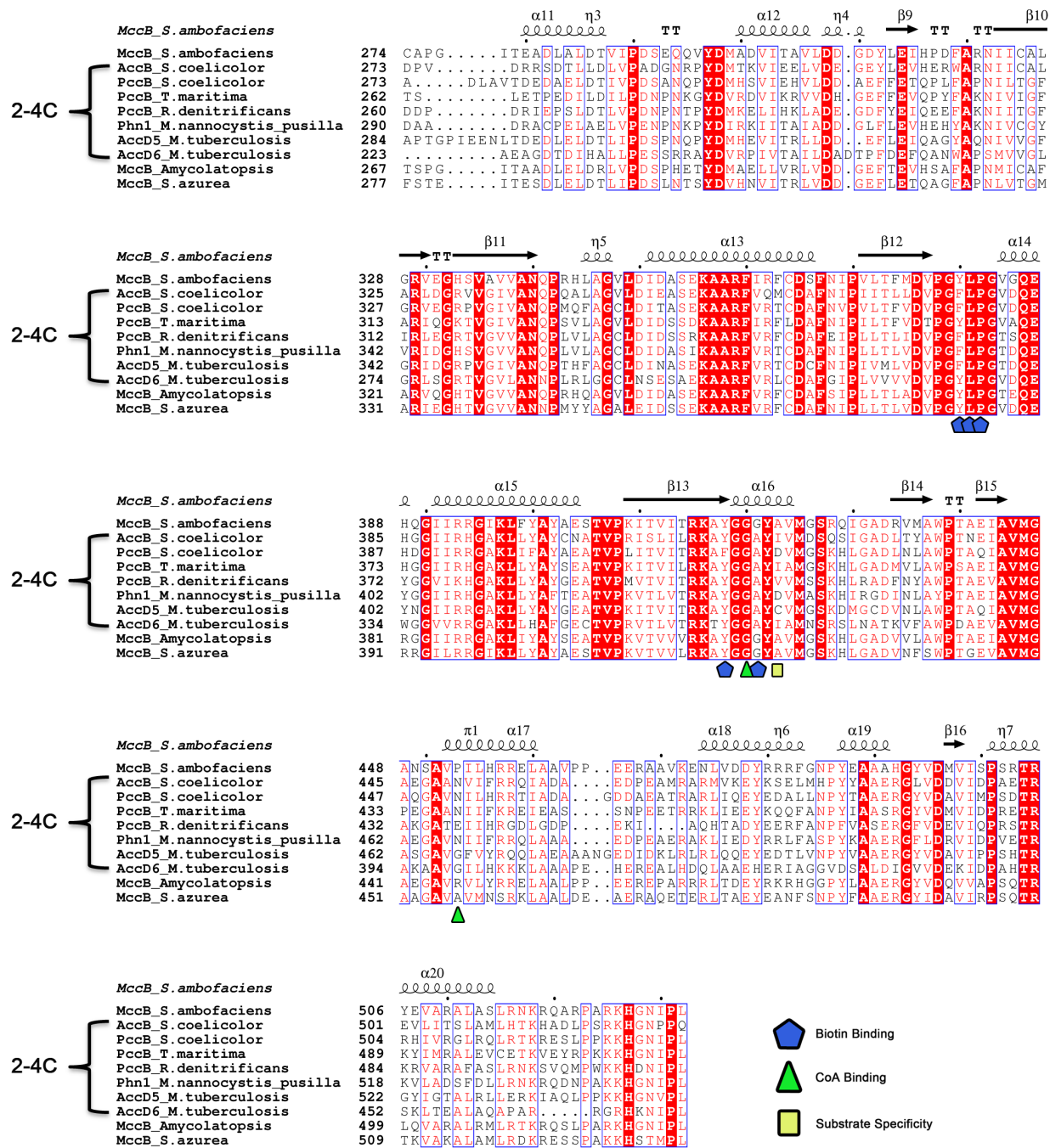


Figure 3.30. Sequence alignment of MccB from *S. ambifaciens* with other ACC β -subunits.

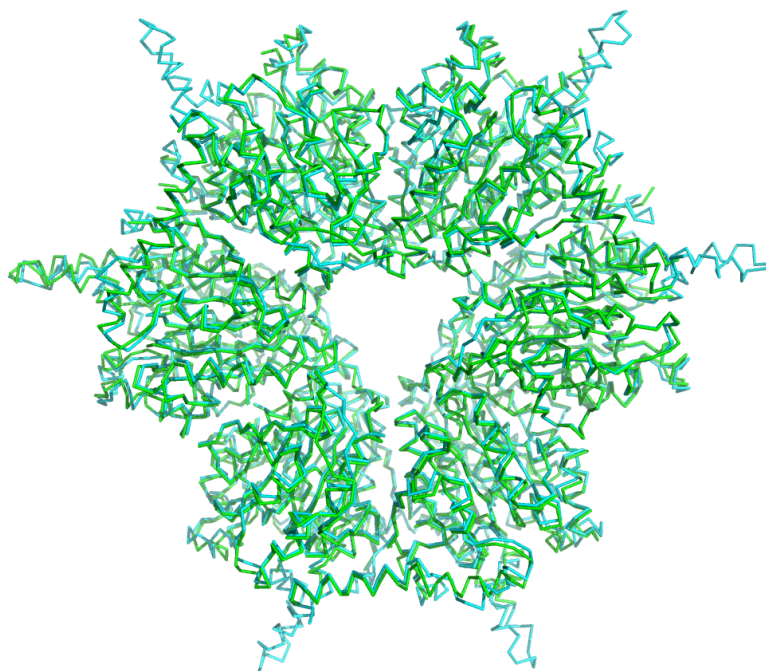


Figure 3.31. Superimposition of PccB from *S. coelicolor* with MccB from *S. ambofaciens*. PccB from *S. coelicolor* (cyan), MccB from *S. ambofaciens* (green), displayed as colored ribbons

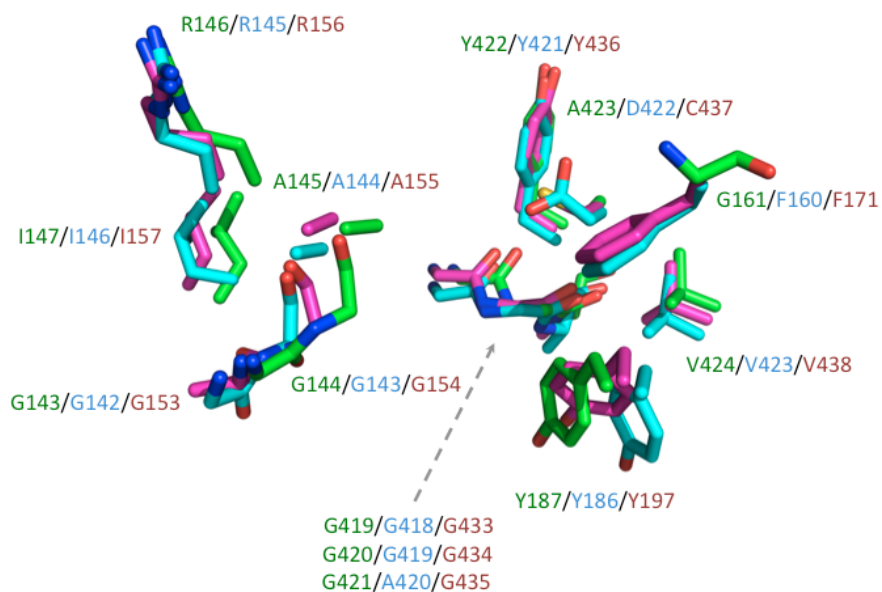


Figure 3.32. Structural comparison of MccB, PccB, and AccD5 active site residues. MccB (PDB code: 5INI), PccB (PDB code: 1XNY), and AccD5 (PDB code: 2A7S) active site residues (colored green, cyan, and magenta, respectively)

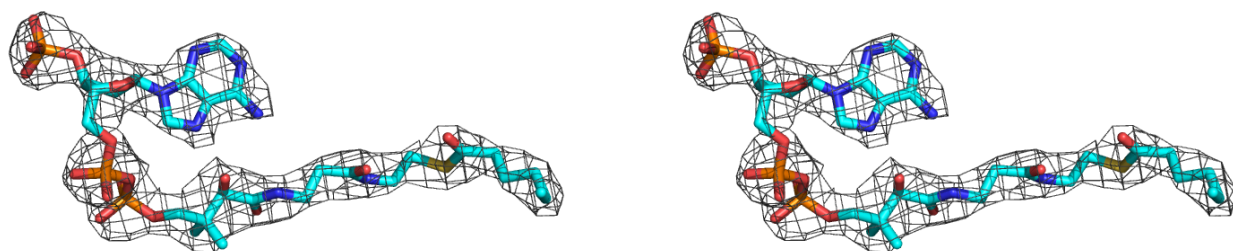


Figure 3.33. Stereo view of hexanoyl-CoA bound to MccB.
 With 2F0-FC SA omit map contoured 1.0 σ at 2.85 Å. Active site residues have been removed for clarity

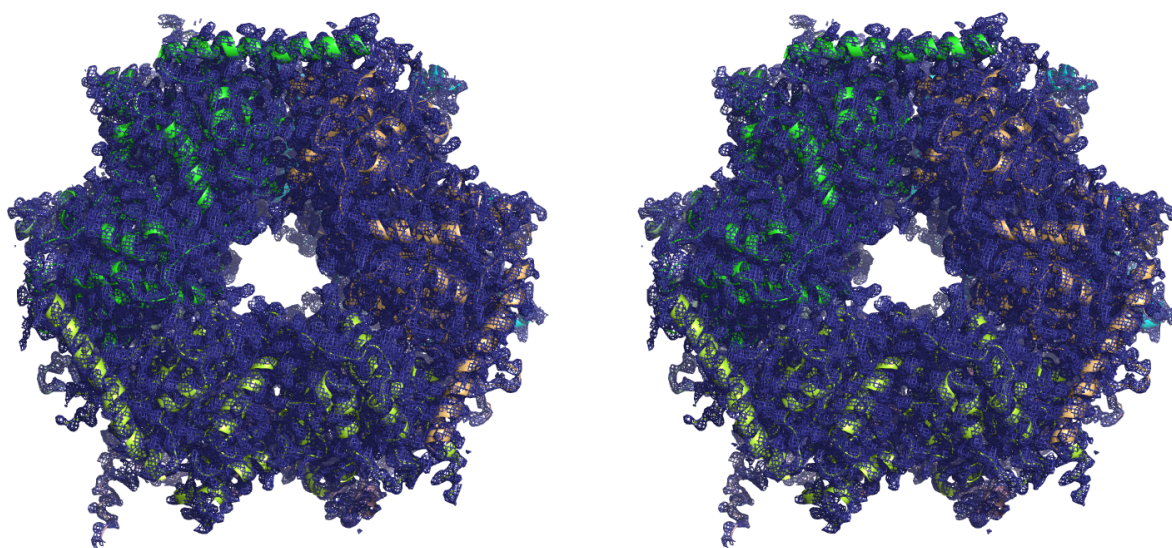


Figure 3.34. Stereo view of the apo MccB hexamer.
 With 2F0-FC map contoured to 1.0 σ at 2.45 Å

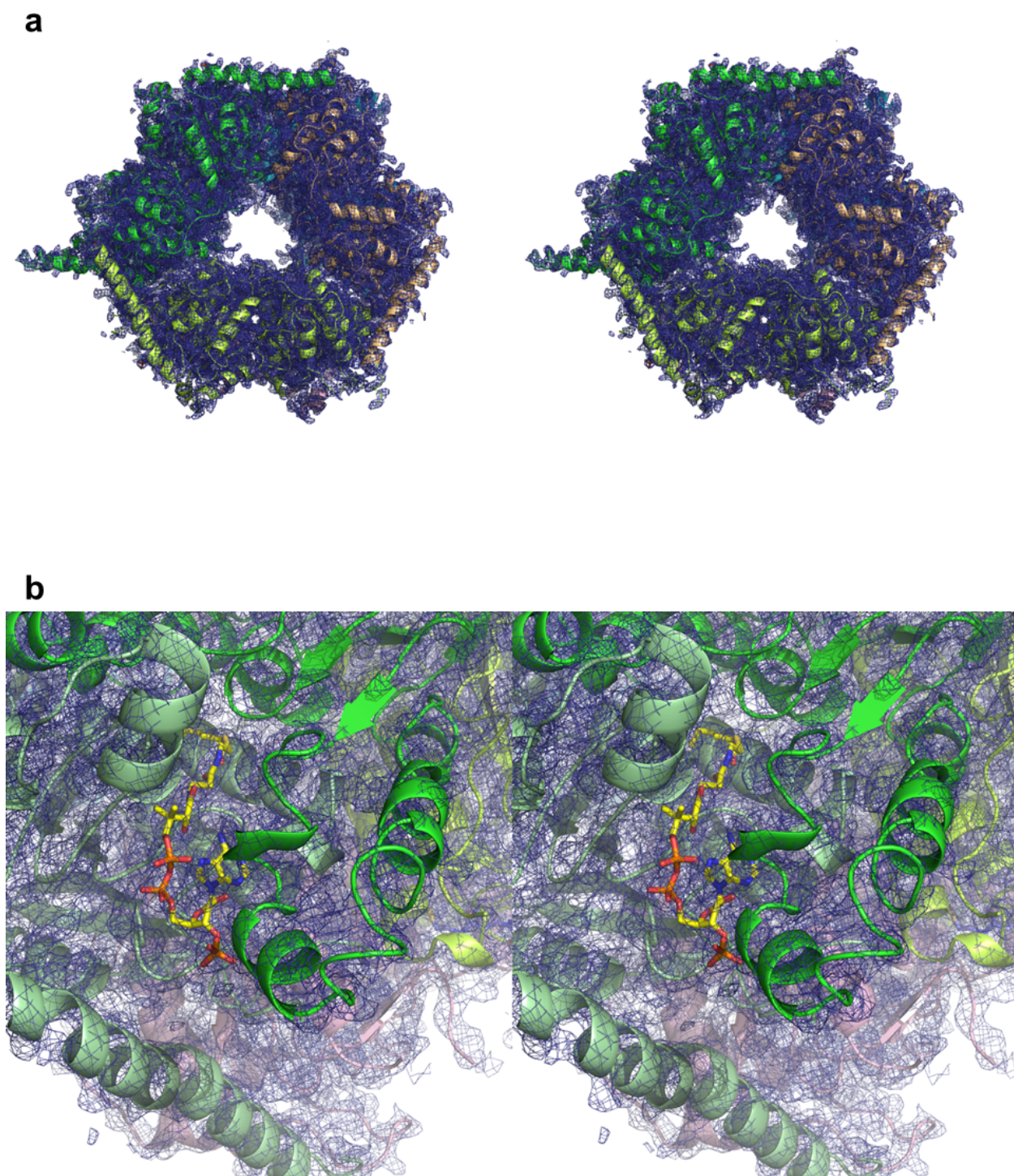


Figure 3.35. Stereo views of MccB hexamer with quadruple binding.

a) Overall architecture of the MccB hexamer bound to four molecules of hexanoyl-CoA and b) the active site pocket at the dimer interface of opposing MccB monomers, with 2F0-FC map contoured to 1.0 σ at 2.85 Å

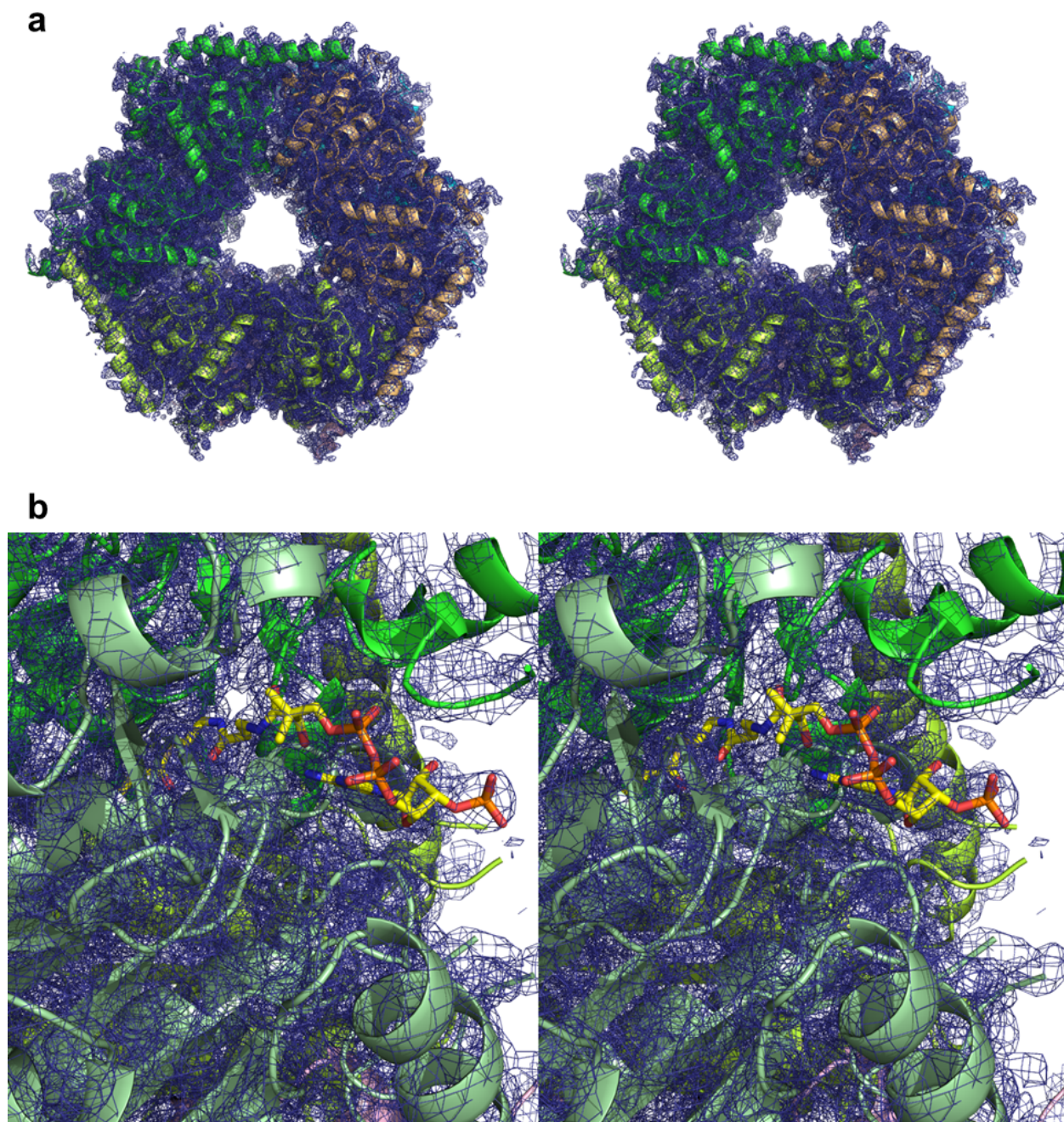


Figure 3.36. Stereo views of MccB hexamer with single binding.

a) Overall architecture of the MccB hexamer bound to one molecule of hexanoyl-CoA and b) active site pocket at the dimer interface of opposing MccB monomers, with 2F0-FC map contoured to 1.0 σ at 2.85 Å.

Synthesis of 8-nonynoic acid 40

Sodium hydride (8.56 g, 356.84 mmol) was added to ethylenediamine (200 ml) under argon and the resulting mixture was heated to 60 °C for 60 min. 3-Nonyl-1-ol (10 g, 71.37 mmol) was added and the resulting deep blue solution was stirred for 4 hours, then cooled to 0 °C and carefully quenched by addition of 2M HCl until the blue color disappeared. The mixture was extracted with EtOAc, and the combined organics were dried (MgSO₄) and concentrated under vacuum. The crude product was purified on silica gel (25% EtOAc in hexanes) to give the product (6.17 g, 61% yield) as a colorless oil.

¹H-NMR (CDCl₃, 700 MHz) δ : 1.31 – 1.44 (6H, HO-(CH₂)₂-CH₂-CH₂-CH₂, m), 1.53 (2H, HO-(CH₂)₅-CH₂, m), 1.57 (2H, HO-CH₂-CH₂, m), 1.96 (1H, H-C $\equiv$ C, t, J = 3.0 Hz), 2.18 (2H, H-C $\equiv$ C-CH₂, td, J = 7.0, 3.0 Hz), 3.64 (2H, HO-CH₂, t, J = 6.5 Hz). ¹³C-NMR (CDCl₃, 175 MHz) δ : 18.35 (CH₂-C $\equiv$ C), 25.59 (HO-(CH₂)₃-CH₂), 28.36 (CH₂-CH₂-C $\equiv$ C), 28.65 (HO-(CH₂)₄-CH₂), 28.86 (HO-(CH₂)₂-CH₂), 32.68 (HO-CH₂-CH₂), 62.95 (HO-CH₂), 68.13 (HC $\equiv$ C), 84.67 (HC $\equiv$ C). HRMS (ESI): *m/z* calculated for C₉H₁₆ONa⁺: 163.1099; Found: 163.1096.

To a solution of oxalyl chloride (5.00 ml, 57.24 mmol) in dry DCM (110 ml), cooled to -78 °C under argon, was added DMSO (7.50 ml, 105.67 mmol) in dry DCM (25 ml) drop wise. The reaction mixture was stirred for 5 min, then 8-nonyn-1-ol (6.17 g, 44.03 mmol) in dry DCM (50 ml) was added and stirring was continued for a further 15 min. Dry triethylamine (26.75 ml, 220.15 mmol) was added and the reaction mixture was allowed to warm to room temperature over 30 min. Water was added and the mixture was extracted with DCM. The combined organics were washed with 2M HCl and saturated NaHCO₃, dried (MgSO₄), and concentrated under vacuum. The crude product was purified on silica gel (10% EtOAc in hexanes) to give the product (3.13 g, 51% yield) as a colorless oil.

¹H-NMR (CDCl₃, 300 MHz) δ : 1.28 – 1.57 (6H, OHC-CH₂-CH₂-CH₂-CH₂-CH₂, m), 1.62 (2H, OHC-CH₂-CH₂, tt, J = 8.0, 8.0 Hz), 1.92 (1H, H-C $\equiv$ C, t, J = 3.0 Hz), 2.17 (2H, H-C $\equiv$ C-CH₂, td, J = 7.0, 3.0 Hz), 2.42 (2H, OHC-CH₂, td, J = 8.0, 2.0 Hz) 9.55 (1H, s, OHC). ¹³C-NMR (CDCl₃, 175 MHz) δ : 18.32 (CH₂-C $\equiv$ C), 21.91 (OHC-CH₂-CH₂), 28.18 (CH₂-CH₂-C $\equiv$ C), 28.40 (OHC-(CH₂)₃-CH₂), 28.61 (OHC-(CH₂)₂-CH₂), 43.82 (OHC-CH₂), 68.28 (HC $\equiv$ C), 84.47 (HC $\equiv$ C), 202.75 (OHC). HRMS (ESI): *m/z* calculated for C₉H₁₄ONa⁺: 161.0937; Found: 161.0939.

To a solution of 8-nonynal (3.13 g, 22.66 mmol) in DMF (30 ml) was added Oxone (15.33 g, 24.93 mmol). The resulting suspension was stirred for 15 min, then poured into water, acidified with 2M HCl and extracted with EtOAc. The combined organics were dried (MgSO₄) and concentrated under vacuum to give the product (3.32 g, 95% yield) as a colorless oil.

¹H-NMR (CDCl₃, 700 MHz) δ : 1.27 – 1.40 (4H, HO₂C-(CH₂)₂-CH₂-CH₂, m), 1.47 (2H, HO₂C-(CH₂)₄-CH₂, tt, J = 7.5, 7.5 Hz), 1.59 (2H, HO₂C-CH₂-CH₂, tt, J = 7.0, 7.0 Hz), 1.87 (1H, H-C $\equiv$ C, t, J = 2.5 Hz), 2.12 (2H, H-C $\equiv$ C-CH₂, td, J = 7.0, 2.5 Hz), 2.29 (2H, HO₂C-CH₂, t, J = 7.5 Hz). ¹³C-NMR (CDCl₃, 175 MHz) δ : 18.33 (CH₂-C $\equiv$ C), 24.51 (HO₂C-CH₂-CH₂), 28.21 (CH₂-CH₂-C $\equiv$ C), 28.30 (HO₂C-(CH₂)₃-CH₂), 28.51 (HO₂C-(CH₂)₂-CH₂), 33.88 (HO₂C-CH₂), 68.27 (HC $\equiv$ C), 84.51 (HC $\equiv$ C), 179.56 (HO₂C). HRMS (ESI): *m/z* calculated for C₉H₁₄O₂Na⁺: 177.0886; Found 177.0884.

Precursor directed biosynthesis of stambomycin analogues 37 and 39

6-azidohexanoic acid **36**, 6-heptynoic acid **38** and 8-nonynoic acid **40** were fed to cultures of *S. ambofaciens* W130 as described in section titled “Incorporation of labeled precursors into the C-26 side chain of the stambomycins”. UHPLC-ESI-QTOF-MS analysis of

the methanolic mycelial extracts, as described in the indicated section, showed that new metabolites with molecular formulae corresponding to stambomycin analogues **37** (Figure 3.31) and **39** (Figures 3.32 and 3.33) were produced.

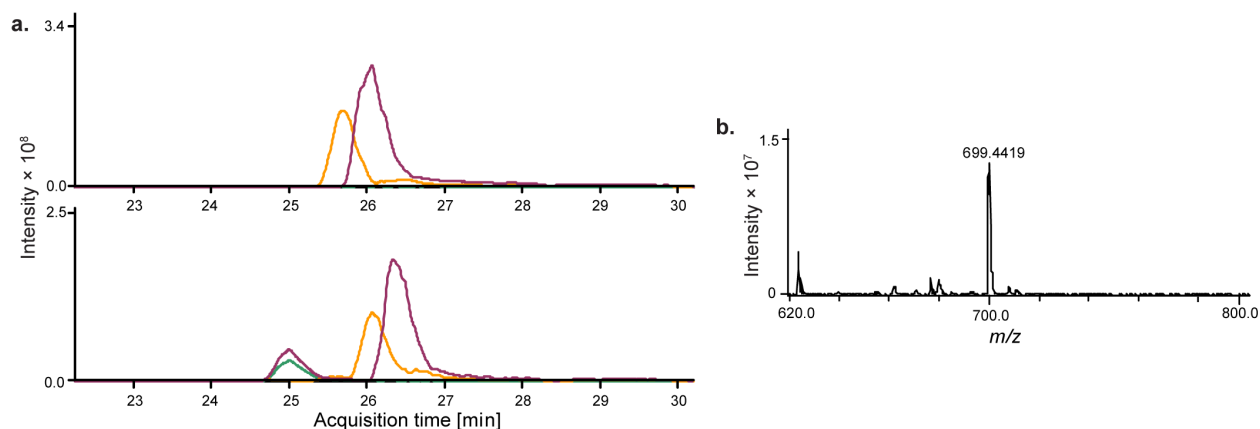


Figure 3.37. Mass spectra of stambomycin analogue **37 feeding studies.**

(a) EICs at $m/z = 699.4$, corresponding to $[M+H+Na]^{2+}$ for stambomycin analogue **37** (green), at $m/z = 692.9$, corresponding to $[M+H+Na]^{2+}$ for stambomycins C/D **3/4** (orange), and at $m/z = 699.9$, corresponding to $[M+H+Na]^{2+}$ for stambomycins A/B **1/2** (purple), from UHPLC-ESI-Q-TOF-MS analyses of methanolic mycelial extracts of *S. ambofaciens* W130 grown in the absence (top trace) and presence (bottom trace) of 5mM 6-azidohexanoic acid. (b) Measured mass spectrum for peak corresponding to stambomycin analogue **37** with a retention time of 25 minutes (calculated for $C_{70}H_{127}N_4NaO_{22}^+$: 699.4414).

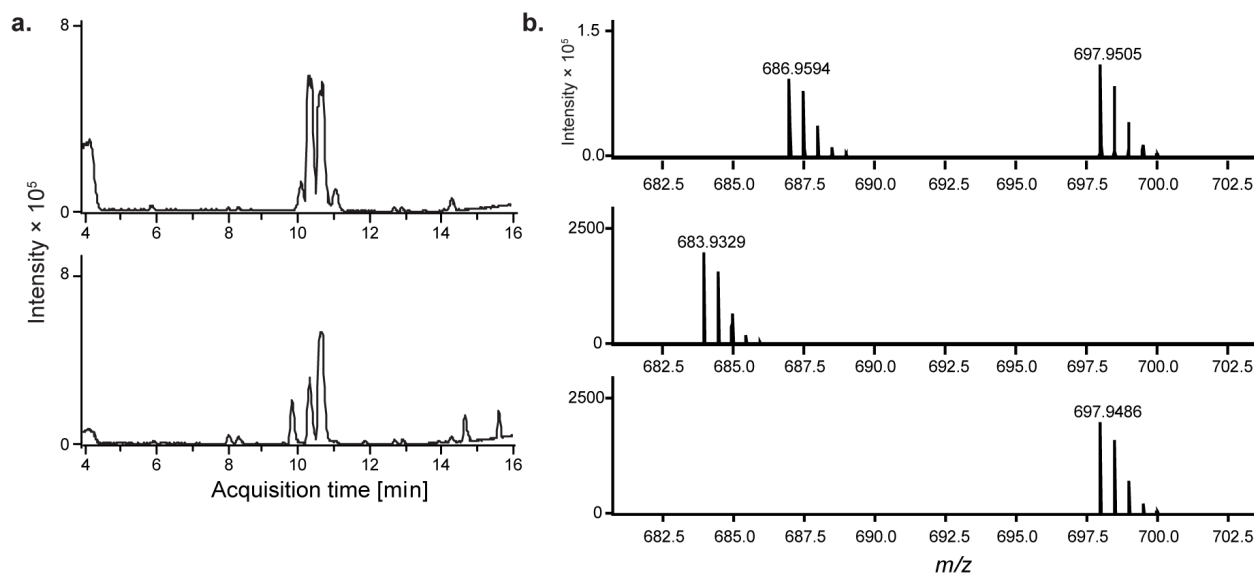


Figure 3.38. Chromatograms of extracts grown in the presence or absence of 6-heptynoic acid.

(a) Base peak chromatograms from UHPLC-ESI-Q-TOF-MS analyses of mycelial extracts of *S. ambofaciens* W130 grown in the presence (bottom panel) and absence (top panel) of 5 mM 6-heptynoic acid. The peaks with retention times of 10.3 minutes and 10.7 minutes correspond to stambomycins C/D **3/4** and stambomycins A/B **1/2**, respectively. A new peak with a retention time of 9.8 minutes (bottom panel), corresponding to a novel stambomycin analogue is observed.

(b) Top panel: Mass spectrum of the stambomycin analogue with a retention time of 9.8 minutes observed in UHPLC-ESI-Q-TOF-MS analyses of *S. ambofaciens* W130 fed with 6-heptynoic acid. Middle panel: Simulated spectrum for $C_{71}H_{126}NNaO_{22}^{2+}$, corresponding to the stambomycin analogue expected to result from direct incorporation of 6-heptynoic acid. Bottom panel: Simulated spectrum for $C_{73}H_{130}NNaO_{22}^{2+}$, corresponding to stambomycin analogue **39**.

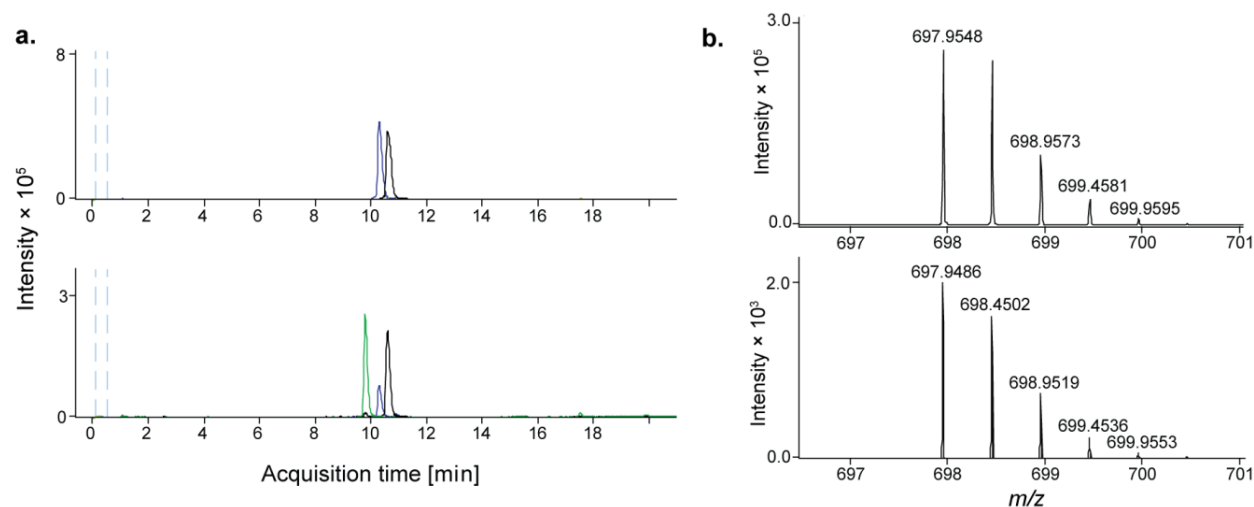


Figure 3.39. Mass spectra of stambomycin analogue 39 feeding studies.

(a) EICs at $m/z = 697.9$, corresponding to $[M+H+Na]^{2+}$ for stambomycin analogue **39** (green), at $m/z = 692.9$, corresponding to $[M+H+Na]^{2+}$ for stambomycins C/D **3/4** (blue), and at $m/z = 699.9$, corresponding to $[M+H+Na]^{2+}$ for stambomycins A/B **1/2** (black), from UHPLC-ESI-Q-TOF-MS analyses of methanolic mycelial extracts of *S. ambofaciens* W130 grown in the absence (top trace) and presence (bottom trace) of 5mM 8-nonynoic acid. (b) Measured (top) and simulated (bottom) mass spectra for the $[M+Na+H]^{2+}$ ion of stambomycin analogue **39**.

Stambomycin analogue **39** was purified from the mycelial extract as described in section titled “Precursor directed biosynthesis of stambomycin analogue **22**” and analyzed by ^1H NMR spectroscopy (Figure 3.32).

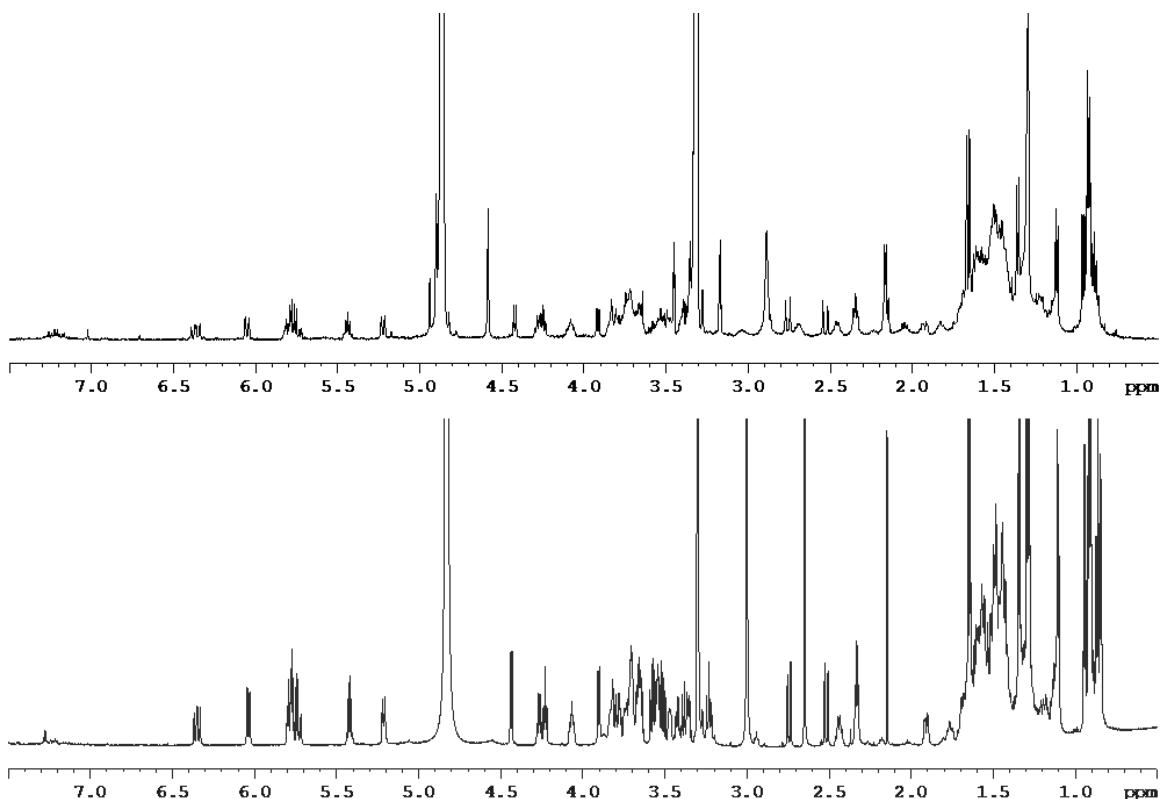


Figure 3.40. Comparison of ^1H -NMR spectra for stambomycin analogue **39 and stambomycins C/D **3/4**.**

^1H -NMR spectra (600 MHz, MeOD), **39** top, **3/4** bottom.

*Reaction of stambomycin analogue **39** with azide-PEG3-biotin conjugate.*

To stambomycin analogue **39** (4.5 mg, 3.28 μmoles) suspended in degassed methanol (600 μL) and butanol (400 μL) under argon was added copper sulfate (0.1 mg, 0.65 μmoles , 0.2 equivalents) and sodium ascorbate (0.65 mg, 3.28 μmoles , 1 equivalent) in degassed water (1 mL), followed by azide-PEG3-conjugate (1.46 mg, 3.28 μmoles , 1 equivalent) in degassed butanol (200 μL). Following 64 hours stirring at 21 $^{\circ}\text{C}$, the solvent was removed *in vacuo* and the biotinylated stambomycin derivative was purified by semi-preparative HPLC on an Agilent C18 column (100 x 21 mm, fitted with a C18 pre-column 10 x 21 mm).

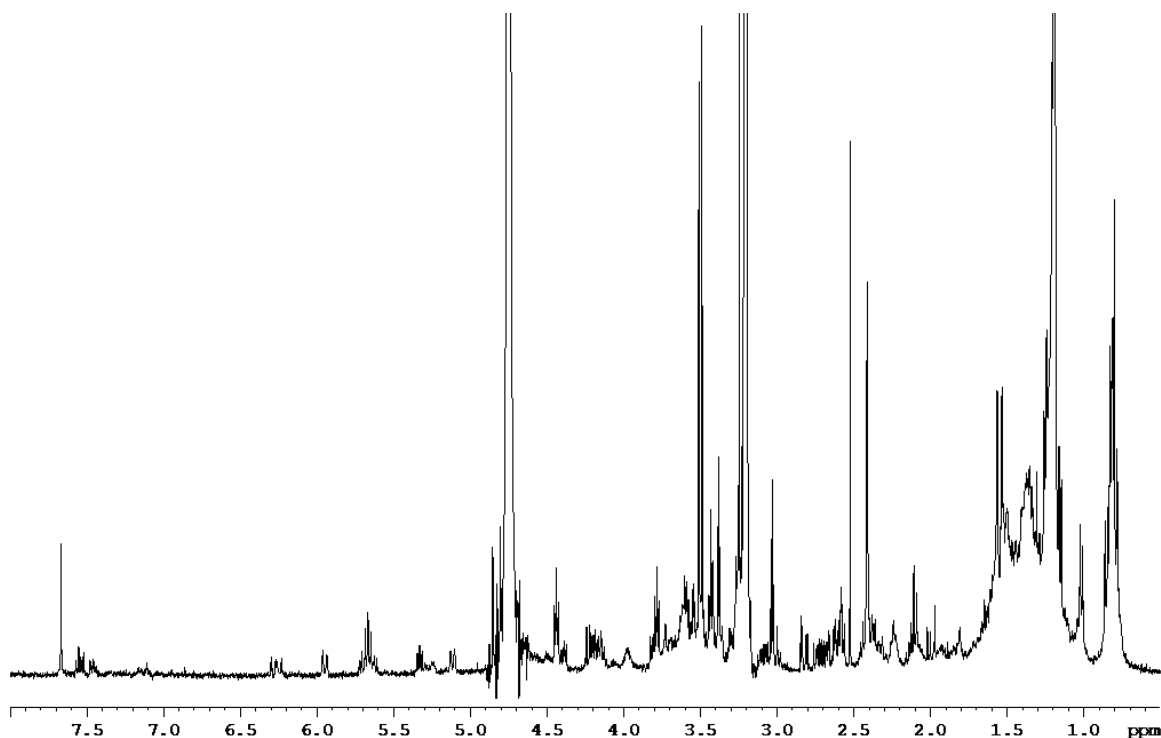


Figure 3.41. NMR spectrum of biotinylated stambomycin analogue derived from the click reaction of stambomycin analogue 39 with the azide-PEG3-biotin conjugate.

^1H -NMR (400 MHz, MeOD). The signal at 7.68 ppm corresponds to the triazole proton and the signals at 4.2 ppm correspond to the protons around the fused thiolane ring of biotin.

Elution conditions were as follows: flow rate, 5 mL min⁻¹; solvent A, water containing 0.1% formic acid; solvent B, methanol containing 0.1% formic acid; gradient initiated after 5 min from 40% B to 100 % B over 25 min. The biotinylated stambomycin derivative eluted between 18.6 mins and 19.2 mins. 2 mg (34%) of the biotinylated stambomycin derivative was recovered and analyzed by HRMS (m/z calculated for $\text{C}_{91}\text{H}_{162}\text{N}_7\text{NaO}_{27}\text{S}^+$: 621.0339 $[\text{M}+\text{H}+2\text{Na}]^{3+}$; Found: 621.0337) and ^1H NMR spectroscopy (Figure 3.29).

3.6 References

1. Hertweck, C. *Angew. Chem. Int. Ed.* 2009, 48, 4688.
2. Laureti, L.; Song, L.; Huang, S.; Corre, C.; Leblond, P.; Challis, G. L.; Aigle, B. *Proc. Natl. Acad. Sci. U. S. A.* 2011, 108, 6258.
3. Song, L.; Laureti, L.; Corre, C.; Leblond, P.; Aigle, B.; Challis, G. L. *J. Antibiot. (Tokyo)* 2014, 67, 71.
4. Eustaquio, A. S.; McGlinchey, R. P.; Liu, Y.; Hazzard, C.; Beer, L. L.; Florova, G.; Alhamadsheh, M. M.; Lechner, A.; Kale, A. J.; Kobayashi, Y.; Reynolds, K. A.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* 2009, 106, 12295.
5. Quade, N.; Huo, L.; Rachid, S.; Heinz, D. W.; Müller, R. *Nat. Chem. Biol.* 2012, 8, 117.
6. Yoo, H.-G.; Kwon, S.-Y.; Kim, S.; Karki, S.; Park, Z.-Y.; Kwon, H.-J. *Biosci. Biotechnol. Biochem.* 2011, 75, 1191.
7. Sandy, M.; Rui, Z.; Gallagher, J.; Zhang, W. *ACS Chem. Biol.* 2012, 7, 1956.
8. Chang, C.; Huang, R.; Yan, Y.; Ma, H.; Dai, Z.; Zhang, B.; Deng, Z.; Liu, W.; Qu, X. *J. Am. Chem. Soc.* 2015, 137, 4183.
9. Mo S.; Kim, D. H.; Lee, J. H.; Park, J. W.; Basnet, D. B.; Ban, Y. H.; Yoo, Y. J.; Chen, S. W.; Park, S. R.; Choi, E. A.; Kim, E.; Jin, Y. Y.; Lee, S. K.; Park, J. Y.; Liu, Y. A.; Lee, M. O.; Lee, K. S.; Kim, S. J.; Kim, D.; Park, B. C.; Lee, S. G.; Kwon, H. J.; Suh, J. W.; Moore, B. S.; Lim, S. K.; Yoon, Y. J. *J. Am. Chem. Soc.* 2011, 133, 976.
10. Mantovani, S. M.; Moore, B. S. *J. Am. Chem. Soc.* 2013, 135, 18032.
11. Miyazawa, T.; Takahashi, S.; Kawata, A.; Panthee, S.; Hayashi, T.; Shimizu, T.; Nogawa, T.; Osada H. *J. Biol. Chem.* 2015, 290, 26994.
12. Erb, T. J.; Berg, I. A.; Brecht, V.; Mueller, M.; Fuchs, G.; Alber, B. E. *Proc. Natl. Acad. Sci. U. S. A.* 2007, 104, 10631.
13. Erb, T. J.; Brecht, V.; Fuchs, G.; Mueller, M.; Alber, B. E. *Proc. Natl. Acad. Sci. U. S. A.* 2009, 106, 8871.
14. Rosenthal R. G.; Ebert, M.-O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J. *Nat. Chem. Biol.* 2014, 10, 50.
15. Gago, G.; Kurth, D.; Diacovich, L.; Tsai, S. C.; Gramajo, H. *J. Bacteriol.* 2006, 188, 477.

16. Diacovich, L.; Peiru, S.; Kurth, D.; Rodriguez, E.; Podesta, F.; Khosla, C.; Gramajo, H. J. *Biol. Chem.* 2002, 277, 31228.
17. Li, Y.; Florova, G.; Reynolds, K. A. *J. Bacteriol.* 2005, 187, 3795.
18. Gago, G.; Diacovich, L.; Arabolaza, A.; Tsai, S. C.; Gramajo, H. *FEMS Microbiol. Rev.* 2011, 35, 475.
19. Takahashi, S.; Toyoda, A.; Sekiyama, Y.; Takagi, H.; Nogawa, T.; Uramoto, M.; Suzuki, R.; Koshino, H.; Kumano, T.; Panthee, S.; Dairi, T.; Ishikawa, J.; Ikeda, H.; Sakaki, Y.; Osada, H. *Nat. Chem. Biol.* 2011, 7, 461.
20. Diacovich, L.; Mitchell, D. L.; Pham, H.; Gago, G.; Melgar, M. M.; Khosla, C.; Gramajo, H.; Tsai, S. C. *Biochemistry*, 2004, 43, 14027.
21. Lin, T. W.; Melgar, M. M.; Kurth, D.; Swamidass, S. J.; Purdon, J.; Tseng, T.; Gago, G.; Baldi, P.; Gramajo, H.; Tsai, S. C. *Proc. Natl. Acad. Sci. U. S. A.* 2006, 103, 3072.
22. Arabolaza, A.; Shillito, M. E.; Lin, T. W.; Diacovich, L.; Melgar, M.; Pham, H.; Amick, D.; Gramajo, H.; Tsai, S. C. *Biochemistry*, 2010, 49, 7367.
23. Sundermann, U.; Bravo-Rodriguez, K.; Klopries, S.; Kushnir, S.; Gomez, H.; Sanchez-Garcia, E.; Schulz, F. *ACS Chem. Biol.* 2013, 8, 443.
24. Yan, Y.; Chen, J.; Zhang, L.; Zheng, Q.; Han, Y.; Zhang, H.; Zhang, D.; Awakawa, T.; Abe, I.; Liu, W. *Angew. Chem. Int. Ed.* 2013, 52, 12308.
25. Zhu, X.; Liu, J.; Zhang, W. *Nat. Chem. Biol.* 2015, 11, 115.
26. Laureti, L.; Song, L.; Huang, S.; Corre, C.; Leblond, P.; Challis, G. L.; Aigle, B. *Proc. Natl. Acad. Sci. U. S. A.* 2011, 108, 6258.
27. Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics*, 2000, John Innes Foundation, Norwich.
28. Osada, H.; Koshino, H.; Isono, K.; Takahashi, H.; Kawanishi, G. *J. Antibiot. (Tokyo)*, 1991, 44, 259.
29. Komatsu, M.; Uchiyama, T.; Omura, S.; Cane, D. E.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.*, 2010, 107, 2646.
30. Miyazawa, T.; Takahashi, S.; Kawata, A.; Panthee, S.; Hayashi, T.; Shimizu, T.; Nogawa, T.; Osada, H. *J. Biol. Chem.* 2015, 290, 26994.
31. Kindly provided by Dr Jean-Luc Pernodet, Université Paris-Sud 11, Orsay, France.

32. Onaka, H.; Taniguchi, S.; Ikeda, H.; Igarashi, Y.; Furumai, T. *J. Antibiot. (Tokyo)*, 2003, 56, 950.
33. Takahashi, S.; Takagi, H.; Toyoda, A.; Uramoto, M.; Nogawa, T.; Ueki, M.; Sakaki, Y.; Osada, H. *J. Bacteriol.* 2010, 192, 2839.
34. Battye, T. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. *Acta Crystallogr. D Biol. Crystallogr.* 2011, 67, 271.
35. Otwinowski, Z.; Minor, W. *Methods Enzymol.* 1997, 276, 307.
36. Evans, P. R.; Murshudov, G. N. *Acta Crystallogr. D Biol. Crystallogr.* 2013, 69, 1204.
37. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* 2007, 40, 658.
38. Emsley, P.; Debreczeni, J. E. *Methods Mol. Biol.* 2012, 841, 143.
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.; Zwart, P. H. *Acta Crystallogr. D Biol. Crystallogr.* 2010, 66, 213.
40. Joosten, R. P.; Joosten, K.; Murshudov, G. N.; Perrakis, A. *Acta Crystallogr. D Biol. Crystallogr.* 2012, 68, 484.
41. Karplus, P. A.; Diederichs, K. *Science* 2012, 336, 1030.

CHAPTER 4: Structural Basis for Condensation Domain Mediated Chain Release During Enacyloxin Biosynthesis

4.1 Abstract

Enacyloxin is an example of a polyene natural product biosynthesized by a type I module *cis/trans*-AT PKS/NRPS hybrid. The biosynthesis of enacyloxins is unusual because it utilizes a condensing enzyme bamb_5915 for product release, whereas most PKS/NRPS hybrids harbor thioesterase or reductase domains for polyketide chain liberation from the enzymatic assembly line. Here we report the crystal structure and functional analysis of bamb_5915, a terminal condensation domain (C domain) from the enacyloxin biosynthesis. The crystal structure of bamb_5915 revealed an unprecedented N-terminal subdomain (COM_{2N}) that is structurally similar to the communication mediating domain (COM) of module C of tubulysin NRPS/PKS. Subsequent structural and bioinformatics analysis of the sole partner enzyme of bamb_5915, PCP_5917, revealed that PCP_5917 contains a complimentary C-terminal COM motif (COM_{2C}). Removal of the COM_{2C} domain of bamb_5917 led to aberrant enzyme activity *in vitro*. These findings strongly suggest that complimentary COM domains are crucial for protein-protein recognition during enacyloxin chain release. A bioinformatic analysis revealed that COM_{2N} and COM_{2C} motifs are conserved across different PKS/NRPS hybrids pointing to a retained structural feature necessary for specific protein interaction in PKS/NRPS hybrid systems.

4.2 Introduction

Type I modular polyketide synthases (PKS) are intricate enzymatic assemblies that are responsible for the biosynthesis of a large repertoire of medically and agriculturally relevant

biologically active molecules.[1] Polyketides such as erythromycin, rapamycin and avermectin are pharmaceutically important natural products generated by modular PKSs, and are currently used as antibiotics, immune-suppressants and antiparasitics, respectively.[2, 3] These complex enzyme machines employ multiple strategies to generate chemical diversity through the incorporation of differing starter and extender units, variation in the intramodular α - or β -carbon processing enzymes, the number of modules within the assembly line and the recruitment of auxiliary enzymes *in trans* for supplementary polyketide chain tailoring steps.[4-9] For the past thirty years, the thioesterase (TE) domain has been established as the primary chain release mechanism of polyketide products from modular PKSs; however, advances in genome mining has led to the discovery of many hybrid modular PKS natural product pathways, such as the enacyloxin pathway, that do not encode TE domains and instead alternative enzymes are utilized for product release.[10-12]

Enacyloxins are a class of unusual polyene antibiotics whose biosynthesis has recently been proposed. The enacyloxin (*ena*) gene cluster was found to encode a unique type I modular PKS biosynthetic machinery with unusual features. For example, *bamb_5916* encodes a PKS module without the key acyl transferase domain (AT) which, unlike typical *cis*-AT PKSs modules, require a dissociative, *trans*-acting acyl transferase (*trans*-AT) for the loading of malonyl-based extender units to acyl carrier protein (ACP) for subsequent chain extensions. The enacyloxin PKS also contains non-extending ketosynthase domains (KS₀) for intermodule substrate transfer.[12, 13] In addition, the enacyloxin PKS utilizes non-canonical nonribosomal peptide synthetase (NRPS), mainly a condensation domain (*bamb_5915*), for the release of a fully grown polyketide chain. Normally, in modular polyketide biosynthesis, a TE domain is present in the last PKS module and is responsible for offloading the product from the assembly

line to biosynthesize a macrolactone.[14] In the ena pathway, however, a full length polyketide chain is first forwarded from the ACP of module-5916 to the KS₀ domain, and then transferred to the upstream PCP domain (PCP_5917). The resultant polyketide chain is released from PCP_5917 through a condensation reaction with shikimate that is catalyzed by bamb_5915, yielding a linear enacyloxin intermediate (Figure 4.1).[12]

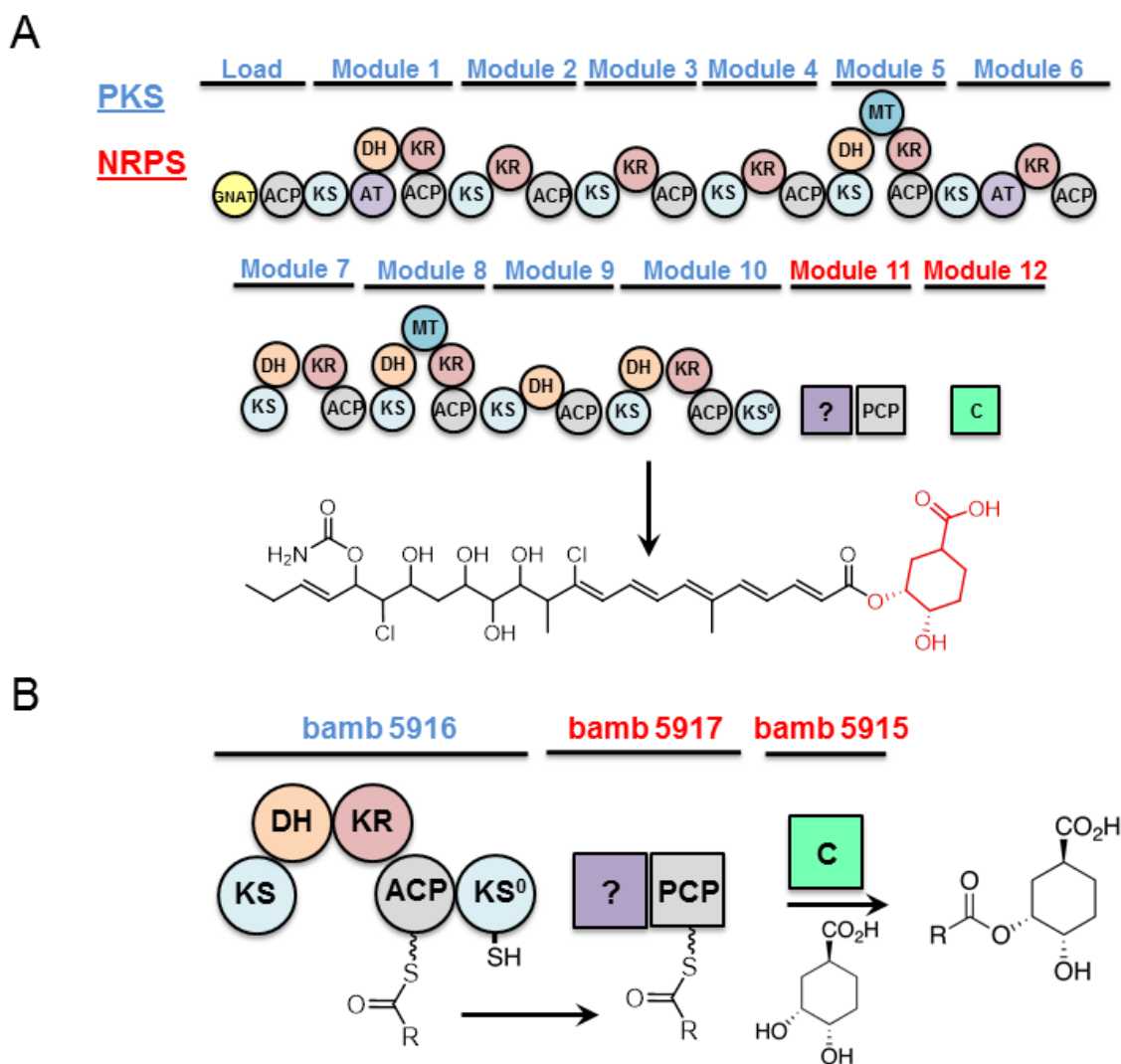


Figure 4.1. The overall architecture of enacyloxin PKS/NRPS hybrid complex (ena PKS).
 (A) Organization of the PKS (blue) and NRPS (red) modules within the enacyloxin gene cluster. Ena PKS produces an intermediate that is tailored by downstream enzymes to yield enacyloxin.

(B) An inspection of the last three modules of ena PKS. During enacyloxin biosynthesis, ACP_5916 forwards a mature polyketide chain to KS₀ 5916 where it is subsequently donated to PCP_5917. Product release is achieved through a C domain mediated condensation (bamb_5915) of shikimate with acyl-PCP_5917 yielding a linear enacyloxin intermediate.

The utilization of PKS, NRPS and shikimate biosynthetic machineries for the generation of a single natural product is uncommon to natural product biosynthesis, and it requires a tightly controlled choreography on protein-protein and protein-substrate interactions between compatible partner enzymes. An understanding of the structural basis for successful communication between enzymatic assemblies from differing biosynthetic pathways is essential for rational combinatorial biosynthesis.[15, 16] How different enzyme pairs are recruited in a synchronized manner that promotes productive biosynthesis remains ambiguous, which hampers current synthetic biology for the generation of new “unnatural” natural products.[17]

Here we report a series of experiments that elucidates the molecular ballet of the enzyme pairs involved in enacyloxin product release. Using *in vitro* reconstitution of the enacyloxin chain release machinery, X-ray crystallography, and bioinformatics analysis of the terminal modules of ena PKS, we show that successful inter-protein communication between hybrid (PKS and NRPS) biosynthetic assemblies is achieved through conserved structural motifs at the protein-protein interface inherent to complimentary partner enzymes. The crystal structure of the condensing enzyme bamb_5915 revealed an unprecedented N-terminal type 2-communication mediating domain (COM_{2N}) that is crucial for specific protein recognition during enacyloxin biogenesis.[18] Subsequent structural and bioinformatics analysis of PCP_5917 revealed a complimentary C-terminal COM domain (COM_{2C}) that is responsible for inducing productive contacts with bamb_5915. Successful *in vitro* chain release assays with bamb_5915 in the presence of non-endogenous holo-acyl-PCPs harboring a COM_{2C} domain demonstrated that

complimentary COM domain adapters are essential for productive inter-module communication during the final steps of enacyloxin biosynthesis. An examination of multiple modular PKS hybrid biosynthetic pathways using bioinformatics revealed that complimentary COM motifs are prevalent in nature and may serve as universal tools for engineered chain release or controlled enzymatic couplings of various acceptor molecules to a mature polyketide chain for the formation of new natural products.

4.3 Results and Discussion

Overall Structure

To determine the structural basis for enacyloxin product release, the crystal structure of ena condensation domain (hereafter referred to as bamb_5915) was solved to 2.6 Å using multiwavelets anomalous diffraction with selenomethionine derivatized protein (Figure 4.2A). A data set for native bamb_5915 was later obtained and refined to 1.9 Å and used for detailed structural analysis. Bamb_5915 crystallized in the P2₁ space group as a dimer, with two monomers in the asymmetric unit. With the notable exception of an N-terminal subdomain, the bamb_5915 architecture shares a high degree of structural similarity to previously determined crystal structures of NRPS condensation domain (C domain) such as CDA-C1, VibH and TycC5-PCP (PCP-C di-domain), with Cα backbone RMSDs of 3.8, 3.9, and 4.1 Å, respectively (Figure 4.3).[19-22] The structure of bamb_5915 can be dissected into two components: the N-term subdomain (the N domain), and the canonical C-term domain (the C domain) that consists of two chloramphenicol acyltransferase (CAT) folds, which form a V-shaped pseudodimer.

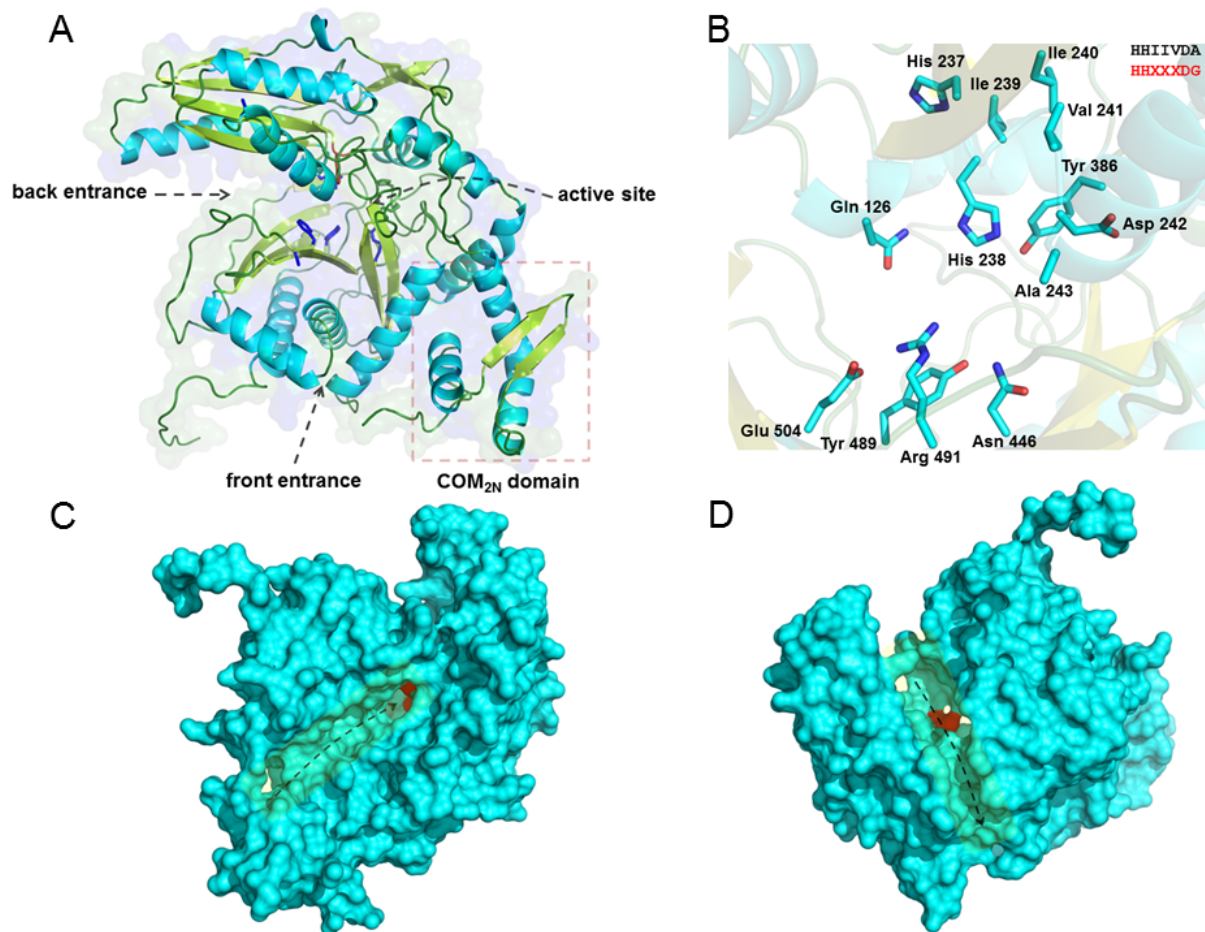


Figure 4.2. The crystal structure of bamb_5915.

(A) Tertiary structure of bamb_5915. (B) The active site of bamb_5915. The canonical C domain active site residues are sequentially and structurally conserved. (C) Surface representation of bamb_5915 with the active site and the front entrance of the substrate binding tunnel highlighted. (D) A snap shot of the back entrance of the substrate binding tunnel of bamb_5915.

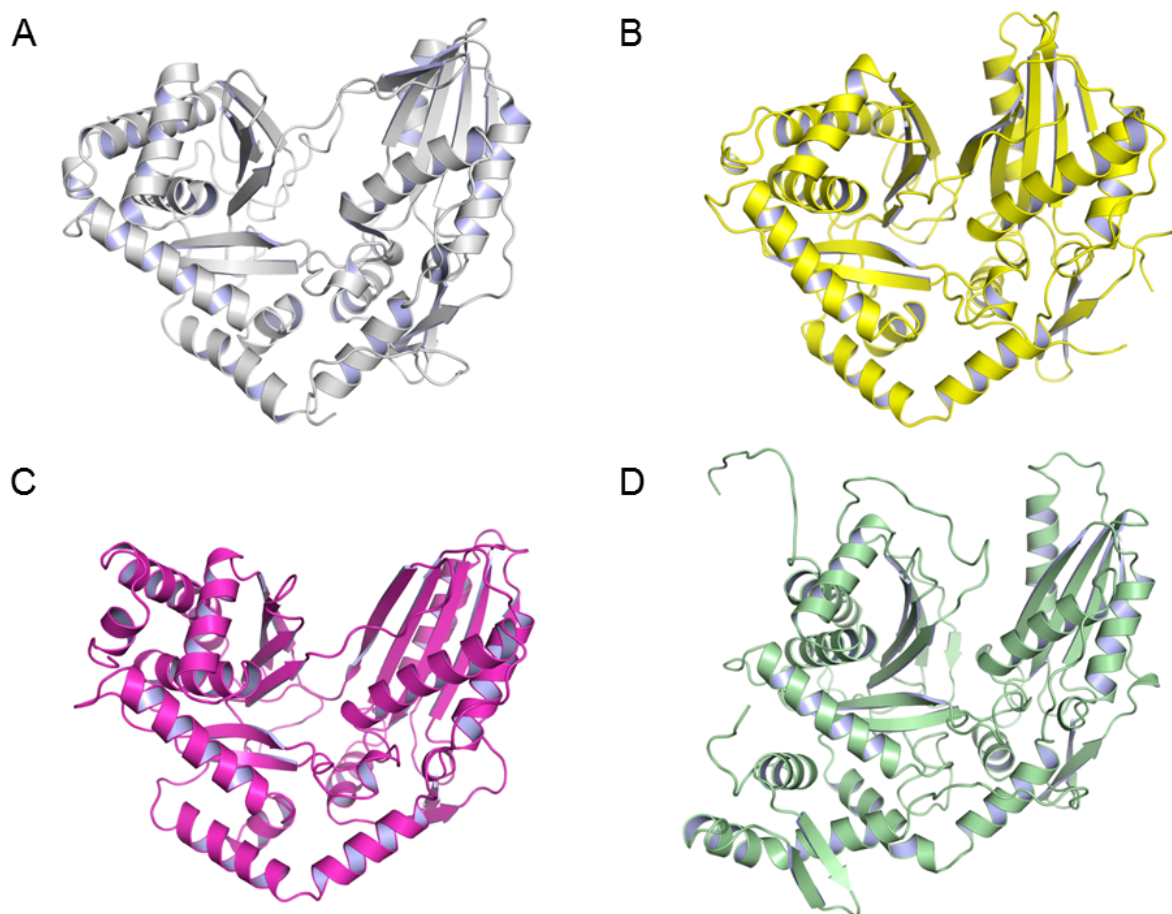


Figure 4.3. A structural comparison of bamb_5915 to other C domain structures.
 (A) TycC5-PCP with PCP domain removed for clarity. (B) CDA-C1. (C) VibH. (D) bamb_5915.

The two CAT folds will be referred to as N-terminal CAT domain and C-terminal CAT domain throughout this chapter. Each CAT fold can be further described as two copies of an α - β - α sandwich that consists of two α -helices flanking a central β -sheet. The active site consensus HHXXDG motif of the C domain is conserved for bamb_5915; it is located at the center of the two CAT folds. A long, 30 Å substrate channel runs through the center of the C domain, allowing access to the active site from two separate faces that could easily accommodate a mature PCP-tethered polyketide substrate (Figure 4.2D). The structural similarity of the overall

fold between the C domain of bamb_5915 and other known CON domains enables us to critically compare their active site and correlate the region-specificity and substrate specificity with their structural components, which will be discussed in the following sections.

Structural Comparison of bamb_5915 with Known C domain Structures

A Dali structural alignment was performed for the C domain of bamb_5915, which revealed that bamb_5915 shares the highest degree of structural similarity with SgcC5 (PDB code:4ZXW), a condensing trans-esterase from the C-1027 biosynthetic pathway with a C α RMSD of 3.7 Å and a Z score of 28.5.[23, 24] Both SgcC5 and bamb_5915 are standalone enzymes that catalyze transesterification reactions; however, only bamb_5915 accepts linear polyketide substrates.[25] Although both maintain the protein fold typical of the C domain, there are some striking structural differences. For example, there is a large structural perturbation in the loop region between β 12 and β 13 in bamb_5915 that encompasses residues 418-451, which is shifted 19.5 Å downward from the corresponding loop region in SgcC5 (Figure 4.4A). Further, in SgcC5 the loop region that spans between residues 371-401 contains a β -hairpin formed by strands β 12 and β 13, which is absent in the bamb_5915 structure (Figure 4.4B). Consequently, bamb_5915 adopts more of an open confirmation in comparison to SgcC5, which may aid bamb_5915 in binding a long linear polyketide substrate.

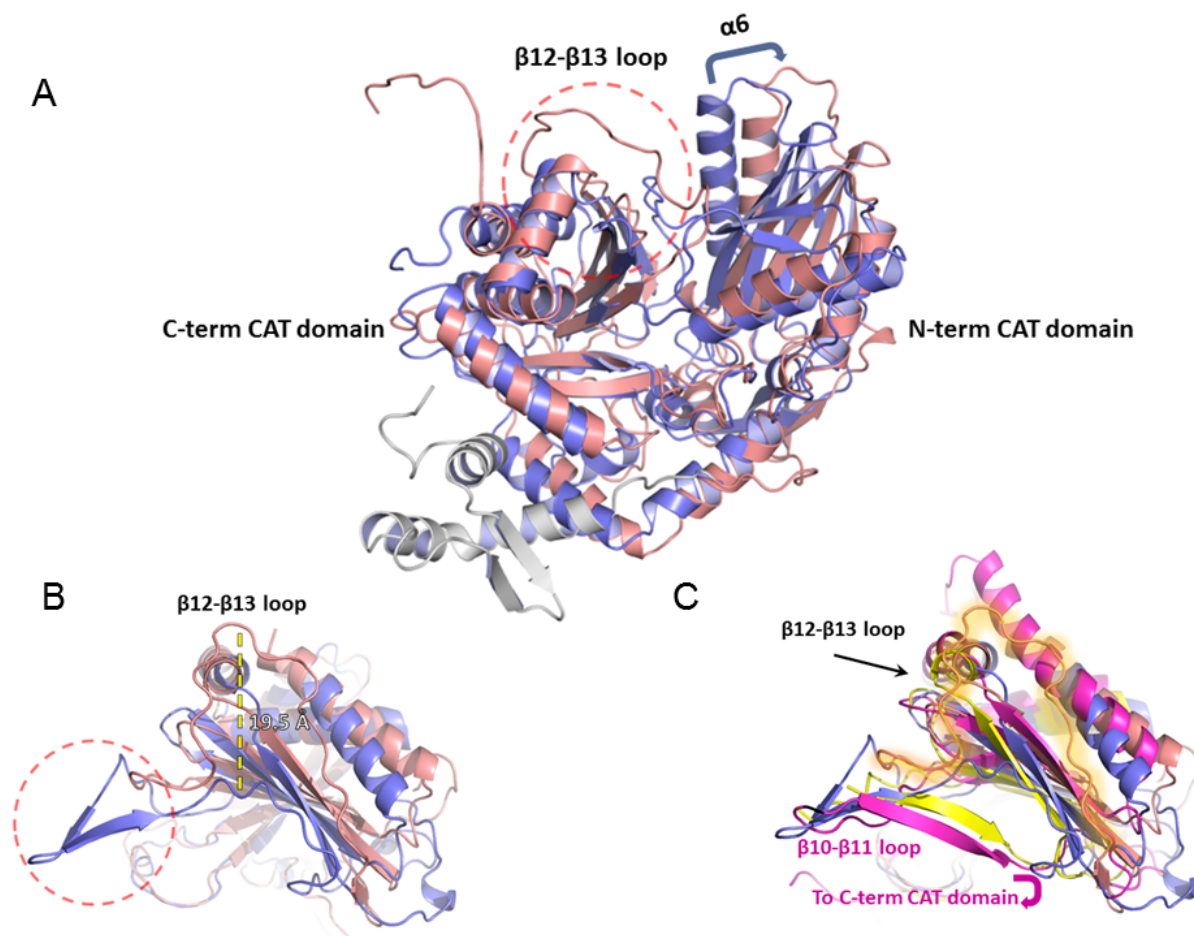


Figure 4.4. Superimposition of bamb_5915 with SgcC5.

(A) Structural overlay of bamb_5915 (pink) emphasizing structural differences in the $\beta 12$ - $\beta 13$ loop and $\alpha 6$ regions. (B) The $\beta 12$ - $\beta 13$ loop regions of bamb_5915 is shifted 19.5 Å upwards with respect to SgcC5 (dark blue). (C) In VibH and CDA-C1, the long $\beta 12$ - $\beta 13$ loop found in bamb_5915 is absent and instead is replaced by a 10 amino acid linker that links $\beta 10$ of the N-term CAT domain to $\beta 11$ of the C-term CAT domain.

In other C domains such as CDA-C1, VibH and TycC5, the long extended $\beta 12$ and $\beta 13$ loop region found in bamb_5915 and SgcC5 is missing, and is replaced by a 10-residue loop that connects $\beta 10$ of N-terminal CAT domain with $\beta 11$ of C-terminal CAT domain (Figure 4.4C). This is significant, because the loop region that spans residues 418-451 in bamb_5915 defines part of the substrate-binding pocket.

Based on the above structural comparison of bamb_5915 with known C domains, it is within reason to suggest that the absence or movement of this particular loop could leave the protein in a more open confirmation with its active site exposed to the solvent. Besides the region discussed above, additional structural deviations can be found in the $\alpha 6$ region of bamb_5915, which in SgcC5 is shifted 5.6 Å away and tilted 20 degrees clockwise with respect to bamb_5915. The most structurally conserved portion of bamb_5915 resides in the active site, which overlays well with the majority of C domain structures (Figure 4.4A). Although the bamb_5915 core retains a conserved structural topology to other C domains, it is unique in that it harbors a higher abundance of unstructured regions as compared with other C domain homologues. Because bamb_5915 is not directly positioned within the interior of an intact NRPS module like the majority of C domains, the increased flexibility of bamb_5915 may be important for *in trans* PCP binding and substrate turn over.

Structural Analysis of the N-terminal Domain of bamb_5915

While the C domain core is retained for bamb_5915, the most salient feature of bamb_5915 is the unexpected presence of an N-terminal subdomain (the N domain). Bioinformatic analysis revealed that the first 100 amino acid residues of bamb_5915 shared no sequence homology with annotated domains across multiple protein structure prediction databases. Indeed, multiple molecular replacement attempts with C domain homologues with sequence similarities greater than 30% failed due to the additional presence of the N-terminal subdomain of bamb_5915. The crystal structure of bamb_5915 reveals that its N domain adopts an $\alpha\beta\beta\alpha\alpha$ topology, where $\alpha 1$ is directly connected to an antiparallel β -hairpin formed by strands $\beta 1$ and $\beta 2$ that is fused to two additional downstream helices that form a V shape (Figure 4.5D).

The N domain is connected to the core C domain through a 16-residue $\alpha 3$ - $\alpha 4$ linker that spans from residues 50-66. Helices $\alpha 1$ and $\alpha 3$ of the N domain pack up against helices $\alpha 9$ and $\alpha 12$ of the core C domain. This configuration is stabilized via hydrogen bond contacts between Thr2-Glu380, Arg263-Glu380, Tyr51-Ser383, and the carbonyl oxygen of Ala57-Arg253.

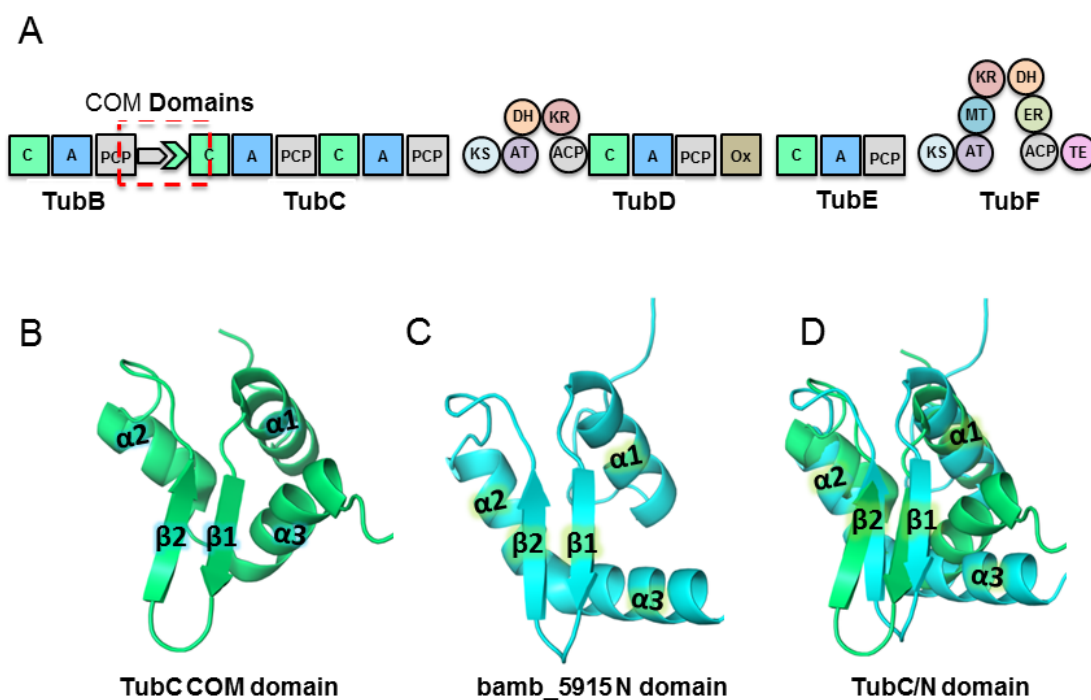


Figure 4.5. COM domains involved in tubulysin and enacyloxin biosynthesis.

(A) The domain architecture of tubulysin PKS/NRPS. (B) The solution state structure of the TubCdd COM domain. (C) The crystal structure of the bamb_5915 N domain. (D) A structural overlay of TubCdd COM domain with the N domain of bamb_5915.

Structural alignments of bamb_5915 to other C domain structures show that the N-term subdomain and the linker loop of bamb_5915 are completely unconserved. When comparing bamb_5915 to other protein structures, however, a Dali search revealed that the N domain of bamb5915 shares high structural homology with TubCdd (PDB code: 2JUG), a C-terminal type 2 communication mediating domain (COM2 domain) found in module C of the tubulysin

NRPS/PKS, with a C α RMSD of 3.57 Å (Figure 4.5D).[18, 23] COM2 domains have been shown to be important for intermodule communication and assembly in PKS and NRPS/PKS hybrid systems by serving as protein docking platforms. Several groups have shown that the grafting of COM2 analogs onto non-cognate PKS or NRPS partner enzymes has led to productive biosynthesis between otherwise orthogonal protein pairs.[26, 27] Further, disruption of COM2 docking domain interfaces through mutagenesis or domain excision has been shown to inhibit successful inter-protein communication and abolish biosynthesis.[28] The Weissman group demonstrated that the β -hairpin of TubCdd contains residues important for intermodular assembly in the tublysin biosynthesis.[18] The N-term domain of bamb_5915 shares only 14 % sequence similarity with TubCdd, yet it adopts a very similar fold, although there are some subtle differences. α -helix one, for example, is longer in bamb_5915 and is tilted 45 degrees with respect to that of helix α 1 of TubCdd. In addition, the linker region between the α 2 and α 3 helices in TubCdd is replaced with a longer α 2 helix in bamb_5915 that connects directly to α 3. Finally, the β -hairpin moiety of bamb_5915 is offset by 45 degrees with respect to TubCdd. Nevertheless, the striking structural resemblance of the N domain of bamb_5915 to the COM2 domain strongly suggests that this structural feature is crucial for specific protein-protein interactions with its partner enzyme, the upstream PCP_5917. Sequence alignments of the N domain of bamb_5915 (hereafter referred to as COM_{2N}) revealed that this particular structural motif is conserved across various *Burkholderia* strains as well as other NRPS/PKS hybrid systems (Figure 4.6).

Gene	Natural product	Species (example)	Bioactivity
aerB,J	aeruginosin	<i>Microcystis aeruginosa</i>	protease inhibitor
ajuL	ajudazol	<i>Chondromyces crocatus</i>	antifungal
apdC,D	anabaenopeptilide	<i>Anabaena</i> sp. 90	protease inhibitor?
aptD	anabaenopeptin	<i>Anabaena</i> sp. 90	protease inhibitor
bacB	bacitracin	<i>Bacillus subtilis</i>	antibiotic
barG	barbamide	<i>Lyngbya majuscula</i>	molluscicidal
blmVIII	bleomycin	<i>Streptomyces verticillus</i>	anti-tumour
cmdC	chondramide	<i>Chondromyces crocatus</i>	anti-cancer
SCO7683	coelibactin	<i>Streptomyces coelicolor</i>	
colG	columbamide	<i>Moorea bouillonii</i>	
crpC	cryptophycin	<i>Nostoc</i> sp. ATCC 53789	anti-tumour
ociC	cyanopeptin oci	<i>Planktothrix agardhii</i>	
mcnD,E	cyanopeptolin	<i>Microcystis aeruginosa</i>	
ctaG	cystothiazole	<i>Cystobacter fuscus</i>	antibiotic
didC	didemnin	<i>Tistrella mobilis</i>	anti-cancer
dkxG,J	DKxanthene	<i>Myxococcus xanthus</i>	antioxidative
bamb_5915	enacyloxin	<i>Burkholderia ambifaria</i>	antibacterial
etlC	entolysin	<i>Pseudomonas entomophila</i>	hemolytic
epoB	epothilone	<i>Sorangium cellulosum</i>	anticancer
eqbG	equibactin	<i>Streptococcus equi</i>	anticancer
hctE,F	hectochlorin	<i>Moorea porducens</i>	antifungal
leuA,E	leupyrrin	<i>Sorangium cellulosum</i>	antifungal
lynE,F	lyngbyabellin	<i>Moorea bouillonii</i>	
massC	massetolide	<i>Pseudomonas fluorescens</i>	antibiotic
melG	melithiazol	<i>Melittangium lichenicola</i>	antifungal
mleC,D	microginin	<i>Microcystis aeruginosa</i>	treat hypertension
psm3b	micropeptin	<i>Microcystis aeruginosa</i>	chymotrypsin inhibitor
mseI	microsclerodermins	<i>Sorangium cellulosum</i>	antifungal
mxmA	myxalamid	<i>Myxococcus xanthus</i>	antibiotic
mtaD,G	myxothiazol	<i>Stigmatella aurantiaca</i>	antifungal
ncyE,F	nannocystin	<i>Nannocystis</i> sp. ST201196	anticancer
ncpA	nostocyclopeptide	<i>Nostoc</i> sp. ATCC53789	
nosC	nostopeptolide	<i>Nostoc punctiforme</i>	cell-differentiation
ofaC	orfamide	<i>Pseudomonas protegens</i>	insecticide
poaC	poaeamide	<i>Pseudomonas poae</i>	antifungal
puwF	puwainaphycin	<i>Cylindrospermum alatosporum</i>	
pchF	pyochelin	<i>Pseudomonas aeruginosa</i>	antifungal
depD	romidepsin	<i>Chromocacterium violaceum</i>	anticancer
sesB	sessilin	<i>Pseudomonas</i> sp. CMR12a	(anti-fungal?)
syfB	syringafactin	<i>Pseudomonas syringae</i>	biosurfactant
sypB	syringopeptin	<i>Pseudomonas syringae</i>	antimicrobial
tImVII	tallysomyacin	<i>Streptoalloteichus hindustanus</i>	(anticancer)
taaB,C	tolaasin	<i>Pseudomonas costantinii</i>	mushroom toxin
tubC	tubulysin	<i>Cystobacter</i> sp. SBCb004	anti-mitotic
vbXO	vibroxin	<i>Vibrio rhizosphaerae</i>	antibiotic
viscC	viscosin	<i>Pseudomonas fluorescens</i>	antibiotic
sven_0517	watasemycin	<i>Streptomyces</i> sp. TP-A0597	antibiotic
wlpC	WLIP	<i>Pseudomonas putida</i>	antibacterial
xtlC	xantholysin	<i>Pseudomonas putida</i>	antimicrobial
zbmVIIIa	zorbamycin	<i>Streptomyces flavoviridis</i>	anti-tumour

Figure 4.6. Biosynthetic genes from characterized pathways utilizing COM_{2N} domains.

Docking of bamb_5915 with PCP_5917

Because COM domains function as an adaptor for the complimentary protein pairs, the presence of COM_{2N} located at the downstream bamb_5915 merited the search for a COM domain within its only partner enzyme, PCP_5917. Significantly, bioinformatics analysis of PCP_5917 revealed an unannotated 11 amino C-terminal peptide that is conserved only across gene clusters that harbor a PCP upstream of a termination module containing a COM_{2N} domain. These findings encouraged us to determine the solution structure of PCP_5917 to examine the topology of a possible novel C-terminal COM domain (COM_{2C}) attached to PCP_5917. Interestingly, the solution structure of PCP_5917 revealed that unlike the COM_{2N} domain of bamb_5915, the COM_{2C} domain of PCP_5917 is completely unstructured and adopts an extended configuration (Figure 4.7).

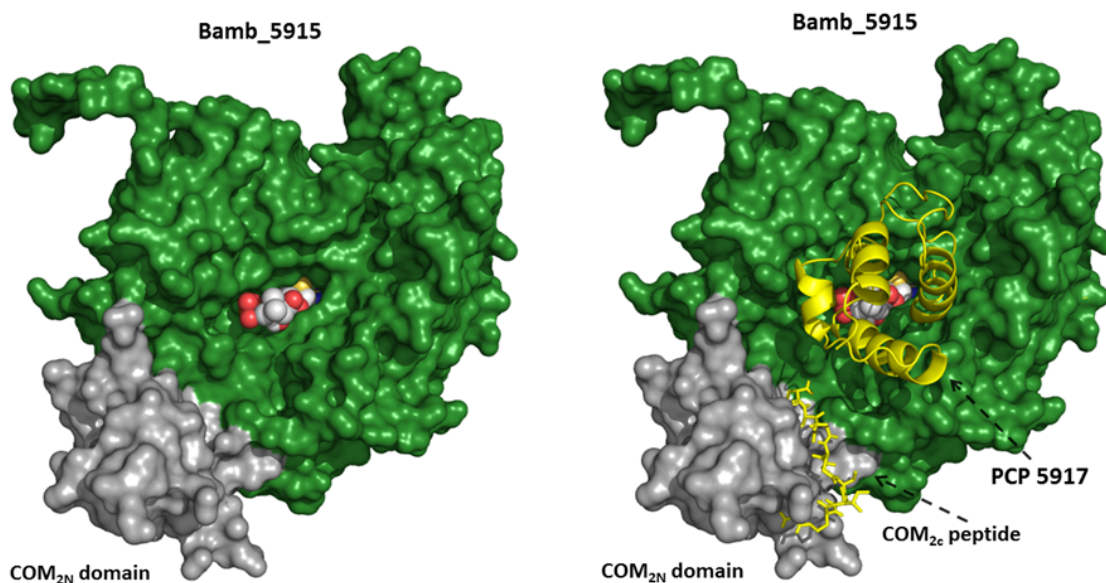


Figure 4.7. Docking of bamb_5915 with PCP_5917.

The top docking solutions place the COM_{2C} of PCP_5917 near the COM_{2N} domain of bamb_5915. This structural configuration would allow the phosphopantetheine arm of PCP_5917 to access the active site of bamb_5915.

Bioinformatic Analysis of COM_{2N} and COM_{2C} Domains

The existence of two unprecedented COM_{2N} and COM_{2C} motifs appended to different enzyme pairs raised the question: are these domains essential for protein recognition? Protein-protein docking simulations were performed with the crystal structure of bamb_5915 and the solution structure of PCP_5917 to assess whether the COM_{2N} and COM_{2C} can make favorable contacts that would aid in protein complex formation (Figure 4.7). The top docking solutions placed the C-terminus of the PCP in close proximity to the COM_{2N} of bamb_5915, with helix 2 of PCP in proximity to the substrate binding tunnel of bamb_5915. These *in silico* docking solutions indicated that bamb_5915-PCP_5917 recognition is reliant on the presence of both COM_{2N} and COM_{2C} domains.

In vitro Reconstitution of Ena Chain Release Machinery

The chain termination from enacyloxin PKS-NRPS was reconstituted *in vitro* using PCP-COM_{2C} (from bamb_5917) and bamb_5915 to assess the role of COM domains within their biochemical context. The condensation reaction was performed using substrate mimics, acetyl-PCP_5917, bamb_5916 and shikimate, and the product (acetylated shikimate) was detected by LC-MS (Figure 4.8B). The reaction containing the wild-type enzymes present three peaks in an extracted ion chromatogram for the ion [M+Na]⁺ = 239.052 (Figure 4.9), where M denotes acetylated shikimate. Each peak is proposed to correspond to acetylation of each hydroxyl within the cyclohexene moiety, either by the condensation domain, or subsequent acetyl rearrangement. The role of the COM_{2C} was examined by performing the condensation assay with a C-terminal truncated form of PCP, where the COM_{2C} domain is not present. No acetylated shikimate was observed for the assay with the truncated enzyme (Figure 4-8C), which is proposed to arise from

significantly decreased efficiency due to the removal of the communication-mediated catalysis phenomenon. Inhibition of the condensation reaction with wild-type enzymes could be achieved by supplementing a surfeit of COM₂C peptide (Figure 4-8D), contributing towards competitive COM₂C binding.

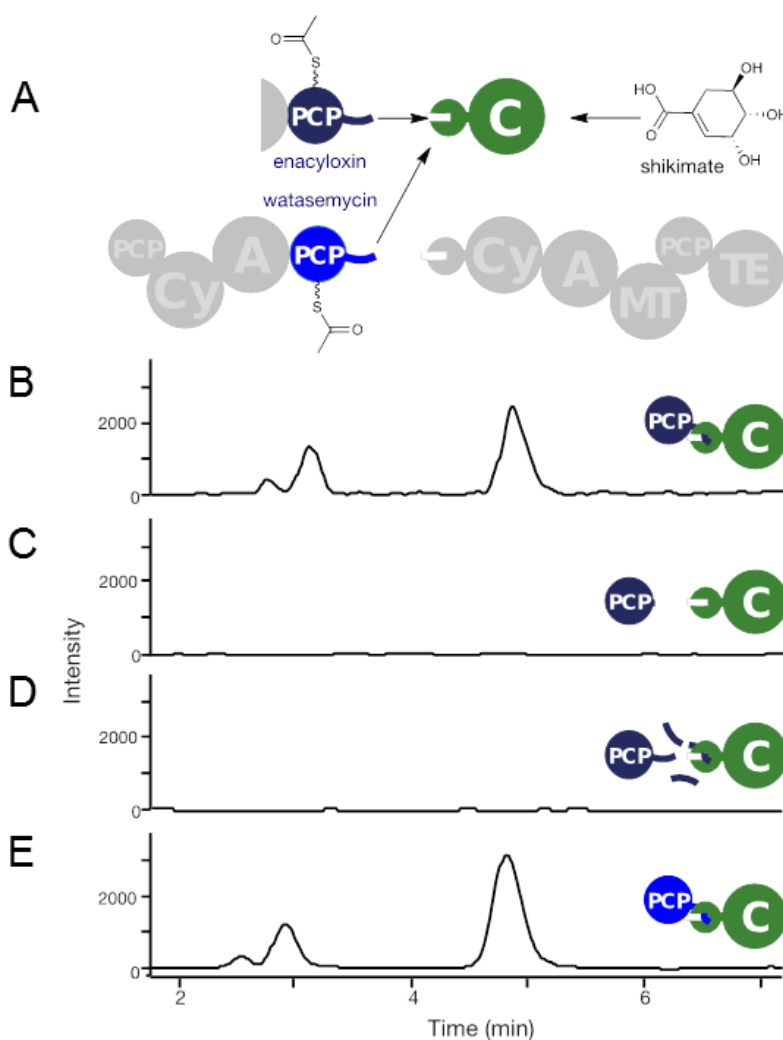


Figure 4.8. *In vitro* reconstitution of enacyloxin chain release machinery.

(A) Schematic of condensation reaction of acetyl-PCP and shikimate in enacyloxin PKS-NRPS, and crosstalk using watasemycin acetyl-PCP. (B-E) Extracted ion chromatograms for the COM₂N-C catalyzed condensed product, acetylated shikimate, [M+Na]⁺ 239.0526 (B) Wild-type acetyl-PCP-COM₂C. (C) C-terminal truncated (nine amino acids) form of acetyl-PCP. (D) Wild-type acetyl-PCP-COM₂C supplemented with a tenfold equivalent of COM₂C (11 amino acids). (E) Wild-type watasemycin acetyl-PCP-COM₂C.

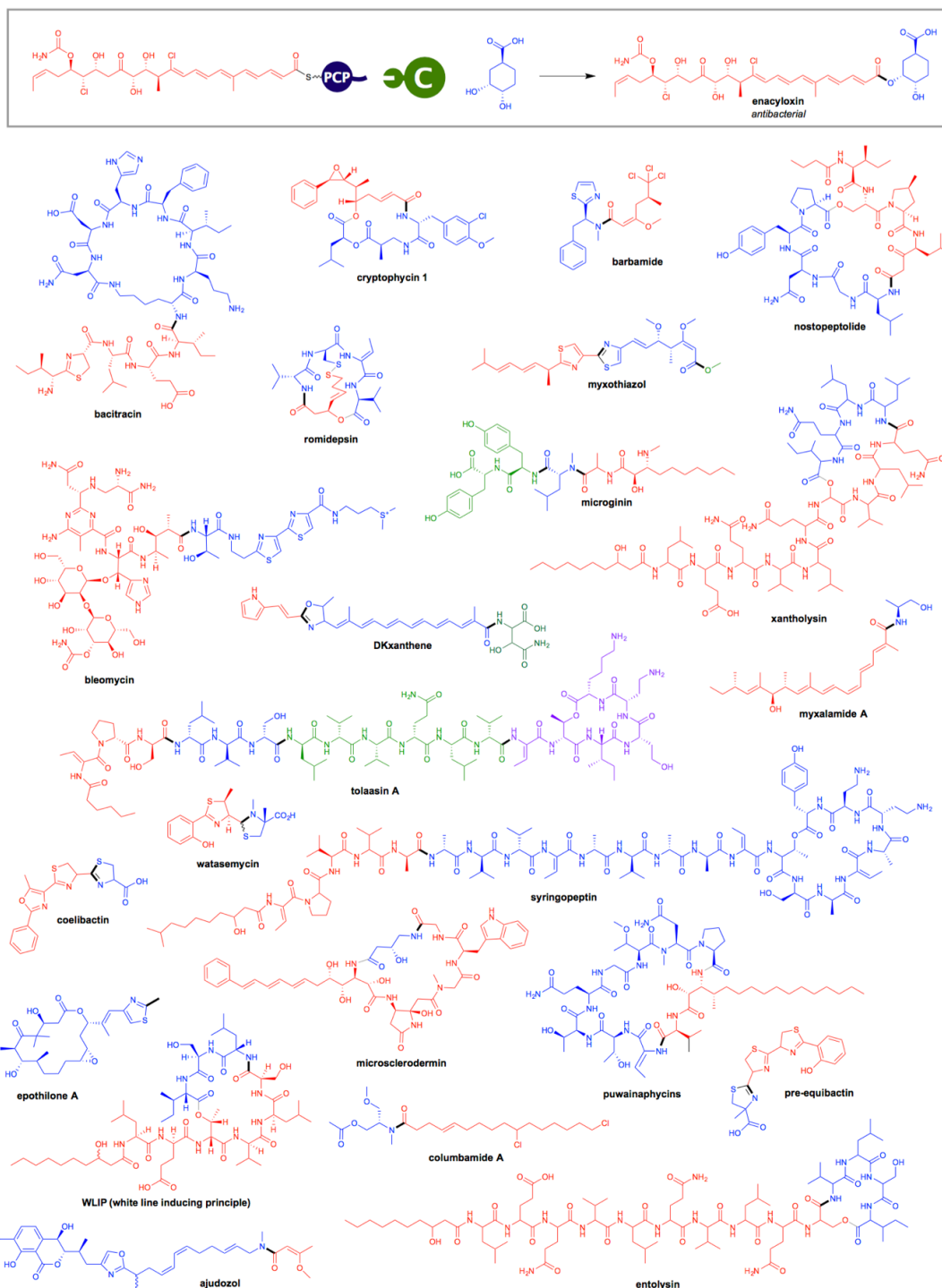


Figure 4.9. Structures of natural products utilizing COM junctions.

COM junction bonds in bold. Acceptor molecules and donor molecules highlighted as blue and red, respectively. Chemical moieties derived from differing sets of biosynthetic machineries are coloured green.

A set of COM_{2N} sequences were next collected to seed the development of a hidden Markov model (HMM), which was subsequently updated iteratively from positive results. This novel HMM, provides the opportunity to identify COM_{2N} domains within biosynthetic pathways, or any given amino acid sequence. The HMM model leads to the identification of natural products employing at least one COM_{2N} domain, demonstrating its widespread use from a diverse set of genera (Figure 4.9). The HMM of COM_{2N} was used to identify an example of a COM2 junction in between a PCP and a cyclisation domain within watasemysin NRPS from *S. venezuelae*. This biosynthetic pathway was chosen to explore crosstalk and promiscuity of COM2 domains with the corresponding COM2 junction from enacyloxin PKS-NRPS. Acetyl-PCP-COM_{2C} domain from watasemycin NRPS (sven_0515) was incubated with bamb_5915 from enacyloxin PKS-NRPS and shikimate to assay whether the COM domains, and their adjacent enzymes, could engage compatible interactions to perform the condensation or acetate and shikimate. The biosynthesis of acetylated shikimate was observed over the hybrid watasemysin-enacyloxin (Figure 4.8E) COM2-boundary, indicating an inherent promiscuity of these particular pair of COM2 domains

Biological Significance

The generation of new pharmaceuticals through “mixing and matching” of enzyme partners from differing biosynthetic pathways is a major goal in natural products research. However, successful rationale combinatorial biosynthesis requires an in-depth understanding of the molecular rules that govern productive protein-protein interactions during natural product biogenesis, which have not been fully elucidated.

The work reported here greatly expands upon synthetic biology for the generation of new natural products. The structure of bamb_5915 is the first condensing enzyme structure solved with an intact NRPS type 2 COM domain (COM_{2N}) appended to its N-terminus which otherwise could not be predicted using bioinformatics. In addition, the solution NMR structure of PCP_5917 was revealed to have a complimentary C-terminal COM domain COM_{2C} necessary for interacting with COM_{2N}-C. *In vitro* chain release assays in the absence of COM_{2C} resulted in complete inhibition of bamb_5915 condensing activity, demonstrating that these COM domains do in fact mediate productive protein-protein during in enacyloxin chain release. The presence of COM_{2N} and COM_{2C} domains in myriad biosynthetic pathways points to a pair of conserved structural motifs that can serve as an adapter for otherwise incompatible partner enzymes. This hypothesis was further tested by the incubation of bamb_5915 with non-cognate PCPs harboring COM_{2C} that resulted in productive biosynthetic crosstalk.

Taken together, these results lay the framework for the generation of new natural products through the direct enzymatic coupling of a mature polyketide chains from biosynthetic pathway with differing acceptor molecules from another. The use of compatible COM domains will greatly aid in the engineering of non-cognate PKS/NRPS systems for the production of novel pharmaceuticals. Future work with deciphering the role of the COM_{2N} domain of bamb_5915 and the engineering of different natural product pathways will be discussed in Chapter 6.

4.4 Materials and Methods

Expression and Purification of Bamb_5915

E. coli BL21(DE3) cells harboring pET151-bamb_5915 were grown in 4 X 1L of LB medium in the presence of 50 µg/mL carbenicillin until $A_{600} = 0.6-0.7$ at 37 °C and 180 rpm. The cultures were induced with 0.5 mM IPTG and the cells were cooled and incubated at 18°C for approximately 16-18 hours. Cells were harvested via centrifugation at 5500 RCF for 10 minutes, resuspended in lysis buffer (50 mM Tris·HCl pH 8.0, 300 mM NaCl, 20 mM imidazole and 10 % glycerol) and sonicated. The resulting cell lysate was clarified by centrifugation (21000 RCF for 60 minutes). The supernatant was incubated with 5 ml of nickel NTA resin pre-equilibrated with lysis buffer (Bio-Rad) and gently stirred for 1 hour at 4 °C. The cell lysate was then transferred to a fritted gravity flow column, the flow through was collected, and the nickel resin was washed 2X with 50 mL of lysis buffer. Elution of bamb_5915 was accomplished through the addition of 7 X 5 mL elutions containing 40-450 mM Imidazole in lysis buffer. Individual fractions were analyzed by SDS PAGE for purity prior to being collected and dialyzed against Q column buffer (50 mM Tris pH 8.0 and 1 mM DTT) at 4 °C for 16 hours. Post dialysis, Bamb_5915 was concentrated to a final volume of 5 mL using a 30,000 Da molecular weight cutoff centrifugal filter (Millipore) and passed through a High Trap Q column (GE Healthcare). Bamb_5915 was eluted using a linear gradient of 0-1 M NaCl in Q column buffer and dialyzed against 50 mM Tris·HCl pH 8.0 and 1 mM DTT. Post Q column treatment, bamb_5915 was concentrated to 5-10 mg/mL and was injected onto a Superdex 200 column preequilibrated with 50 mM Tris·HCl and 1 mM DTT. Protein concentrations were determined by Bradford assay (Bio-Rad). Fractions containing the highest purity ($\geq 90\%$) were pooled, concentrated to 10 mg/mL and flash frozen in liquid nitrogen and stored at -80 °C.

Expression and Purification of SeMet Derived bamb_5915

SeMet derived bamb_5915 was produced in B121(DE3) cells. LB cultures supplemented with 50 µg/L kanamycin (4 X 1L) were grown at 37 °C at 180 rpm until A₆₀₀ reached 0.6. The cells were then pelleted at 5500 rcf for 10 minutes at 4 °C and washed three times by suspension with 50 mL M9 minimal media containing 50 µg/L carbenicillin. The washed cells were evenly transferred to (4 X 1L) flasks containing M9 minimal media with 50 µg/L carbenicillin and the following additives: lysine, phenylalanine, and threonine (100 mg each); valine, isoleucine, and leucine (50 mg each); and L-selenomethionine (80 mg). Protein expression proceeded overnight (16-18 hours) at 18 °C after the addition of 0.4 mM IPTG. The following day, cells were collected via centrifugation at 5500 rcf for 12 minutes, resuspended in 50 mM Tris·HCl pH 8.0, pelleted and stored at -80 °C. The purification of SeMet bamb_5915 followed the same procedure as outlined for native bamb_5915. The incorporation of selenium into bamb_5915 was verified using MALDI-TOF mass spectrometry.

Crystallization of bamb_5915

Frozen aliquots of either native or SeMet bamb_5915 were thawed on ice prior to crystal screens. Crystals were obtained using the hanging drop vapor diffusion method by mixing 2 µL of bamb_5915 at 10 mg/ml with 2 µL of 0.2 M di-ammonium phosphate and 20% PEG 3,350 at 25 °C. Large rectangular crystals of native bamb_5915 grew between 7-14 days. SeMet derived bamb_5915 crystals were procured by mixing 2 µL of SeMet protein at 7-8 mg/mL with 0.17-0.21 diammonium phosphate and 22-25 % PEG 3,350 at 25 °C. SeMET bamb_5915 crystals grew after a month in the dark. Crystals were harvested, cryoprotected in well solution and flash frozen in liquid nitrogen prior to data collection.

Data collection, Processing and Structure Determination

Datasets for SeMet crystals of bamb_5915 were collected at SSRL beamline 12-2. The MAD data were collected at 0.9176, 0.9793, and 0.9796 Å corresponding to the remote, peak and inflection respectively. For MAD data collection, two individual data sets for peak, remote and inflection were collected, consisting of frames 1-138 (Set 1) and 139-360 (Set 2). The individual data sets were processed in the $P2_1$ space group using Mosflm and scaled together using Aimless (CCP4 suites).[29, 30] Phasing was accomplished using PHENIX Autosol.[31] A preliminary model was constructed using Autobuild that was further improved through repetitive rounds of refinement and model rebuilding using PHENIX Refine and COOT.[32, 33] Data corresponding to native bamb_5915 were collected at ALS beamline 8.2.1 to 1.9 Å. The native data was processed and scaled using Mosflm.[29] The higher resolution native structure was determined using molecular replacement (Phaser) using the lower resolution model of SeMet bamb_5915 as a search model.[34] The native bamb_5915 structure was refined using a combination of PHENIX Refine and COOT.[32, 33]

PCP_5917: Expression, Purification, and Generation of Hexyl-holo-PCP

The expression and purification of apo PCP_5917 was the same for that of bamb_5915 except after nickel NTA purification the protein was dialyzed against protein storage buffer (50 mM Tris·HCl pH 8.0, 50 mM NaCl, 10 % glycerol and 1 mM DTT). The protein was concentrated to 10 mg/mL using a 10,000 Da molecular weight cutoff centrifugal filter (Millipore), flash frozen in liquid nitrogen and stored at -80 °C.

Holo-hexyl-PCP was generated *in vitro*. A 500 µL loading reaction containing 1 mM PCP_5917, 12.5 mM MgCl₂, 2 mM hexanoyl-CoA (Sigma), and 0.1 mg/mL SFP was incubated

for 2 hours at 37 °C. The formation of holo-hexyl-PCP was verified using MALDI-TOF mass spectrometry.

Purification of Complexes between bamb_5915 and Apo or Holo-hexyl-PCP_5917

Purified apo or His_{6x}-holo-hexyl-PCP was incubated with polished His_{6x}-bamb_5915 on ice for one hour to aid in complex formation. The resulting protein mixture was injected onto a Superdex 200 column pre-equilibrated with crystal buffer and subject to SDS-PAGE. Both *apo*- and *holo*-hexyl PCP_5917 co-eluted with bamb5915 over a broad range suggesting a transient interaction. Fractions containing His_{6x}-bamb_5915 and *apo* or *holo*-hexyl-PCP_ 5917 were collected and concentrated to 12 mg/mL prior to crystal screens.

Generation of COM_{2N}-C and bamb_5915ΔCOM_{2N} constructs

The regions corresponding to COM_{2N} and bamb_5915ΔCOM_{2N} were amplified by PCR from pET151-bamb_5915 using the primers listed in Table 4.1. The resulting amplimers were cloned into the *Nde*I and *Not*I sites of pET28a (Novagen) to produce pET28a-COM_{2N} and pET28a-bamb_5915ΔCOM_{2N} encoding N-terminal hexahistidine COM_{2N} and bamb_5915ΔCOM_{2N} fusion proteins. Both pET28a-COM_{2N} and pET28a -bamb_5915ΔCOM_{2N} were transformed into BL21(DE3) competent cells for protein expression.

Table 4.1. Primers for the generation of COM_{2N}-C and bamb_5915ΔCOM_{2N}.

bamb_5915ΔCOM_{2N}F	5' GCGATACATATGCGCCAGCTGAACCAGG '3
bamb_5915ΔCOM_{2N}R	5' GCGATACATATGCGCCAGCTGAACCAGG '3
COM_{2N}F	5' GCGCTACATATGACGATTCCCGCCTTGCTG 3'
COM_{2N}R	5' TATATAGCGGCCGCCTAATCGGTGCGGC'3

Expression and Purification of COM_{2N} and bamb_5915ΔCOM_{2N}

Cells containing pET28a-COM_{2N} or pET28a-bamb_5915ΔCOM_{2N} were grown in 4 X 1L of TB containing 50ug/ml kanamycin at 37 °C at 180 rpm until A₆₀₀ reached 0.8-1.0, and protein expression was induced upon the addition of 0.5 mM IPTG followed by overnight growth at 18 °C. Cells were collected via centrifugation at 5500 RCF for 15 minutes, resuspended in lysis buffer (50 mM Tris·HCl pH 8.0, 300 mM NaCl, 20 mM imidazole and 10% glycerol), and sonicated for 8X cycles (1 cycle= 30 seconds on, 1 minute off and 30 seconds on). The resulting cell lysate was clarified via centrifugation at 21000 RCF for 60 minutes and then incubated with nickel NTA resin (Bio-Rad) preequilibrated with lysis buffer at 4 °C for 1 hour. The nickel lysate slurry was applied to a gravity flow column and washed 2X with 50 mL lysis buffer. Protein elution was achieved through the addition of 5 mL fractions containing 40, 100, 150, 250 and 450, mM imidazole in lysis buffer. Individual fractions were assessed for purity via SDS PAGE. Fractions containing pure protein were collected and dialyzed against 50 mM Tris·HCl pH 8.0, 10 % glycerol and 1 mM DTT. Post dialysis protein was concentrated using a 3000 molecular weight cutoff centrifugal filter (Millipore) and subjected to anion exchange chromatography (Hitrap Q column, GE Healthcare). After passage of the protein solution, the column was first washed with 10 mL of Q buffer (50 mM Tris·HCl pH 8.0 and 1 mM DTT) followed by elution with a linear gradient of 0-1 M NaCl in Q column buffer. The polished protein was dialyzed against storage buffer, concentrated; flash frozen in liquid nitrogen, and stored in -80 °C for later use.

Table 4.2. Data Collection and Refinement Statistics

	SeMet bamb_5915				
	Native bamb_5915-1	Native bamb_5915-2	Inflection	Peak	Remote
Wavelength	0.9999	1.00	0.9796	0.9793	0.9176
Resolution range	43.34 - 1.85 (1.916 - 1.85)	51.15 - 2.5 (2.589 - 2.5)	52.07 - 2.65 (2.745 - 2.65)	52.07 - 2.65 (2.745 - 2.65)	52.07 - 2.65 (2.745 - 2.65)
Space group	P 1 2 1 1	P 1 2 1 1	P 1 2 1 1	P 1 2 1 1	P 1 2 1 1
Unit cell	103.66 58.81 107.04 90 104.98 90	103.719 58.87 106.94 90 104.98 90	103.71 58.9484 107.741 90 105.304 90	103.71 58.9484 107.741 90 105.304 90	103.71 58.9484 107.741 90 105.304 90
Total reflections	379075(18921)	86006 (8393)	68845 (6645)	69852 (6845)	69225 (6843)
Unique reflections	106592 (10578)	43519 (4312)	35063 (3425)	35309 (3425)	35348 (3425)
Multiplicity	3.6 (3.5)	2.0 (2.0)	6.1 (5.2)	5.8 (5.2)	6.1 (5.2)
Completeness (%)	0.99 (0.99)	0.98 (0.99)	0.95 (0.94)	0.95 (0.97)	0.95 (0.97)
Mean I/sigma(I)	7.4 (1.3)	22.60 (3.18)	10.18 (4.13)	9.90 (3.64)	8.93 (3.60)
Wilson B-factor	21.29	29.35	28.53	28.53	28.53
R-merge	0.109 (0.716)	0.04651 (0.2581)	0.05018 (0.1575)	0.06179 (0.1933)	0.06811 (0.2238)
CC1/2	0.982 (0.709)	0.996 (0.851)	0.993 (0.912)	0.991 (0.87)	0.988 (0.833)
Reflections used in refinement	104905 (9603)	42837 (4215)		35050 (3425)	35050 (3425)
Reflections used for R-free	1965 (169)	1358 (135)		988 (103)	988 (103)
R-work	0.2361 (0.4687)	0.1928 (0.2401)		0.2027 (0.2558)	
R-free	0.2577 (0.4752)	0.2358 (0.3067)		0.2559 (0.3488)	
Number of non-hydrogen atoms	9065	8716		8324	
macromolecules	8054	8010		7616	
Protein residues	1044	1038		987	
RMS(bonds)	0.009	0.003		0.009	
RMS(angles)	1.24	0.81		1.46	
Ramachandran favored (%)	93	98		94	
Ramachandran allowed (%)	4.4	1.8		3.7	
Ramachandran outliers (%)	2.8	0.48		1.9	
Rotamer outliers (%)	3.3	0.24		4.9	
Clashscore	7.46	6.26		12.31	
Average B-factor	31.84	36.20		27.87	
macromolecules	31.73	36.44		27.54	
solvent	32.74	33.44		31.47	

4.5 References

1. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review*. Nat Prod Rep, 2001. **18**(4): p. 380-416.
2. Song, L., et al., *Cytochrome P450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis*. J Antibiot (Tokyo), 2014. **67**(1): p. 71-6.
3. Benjamin, D., et al., *Rapamycin passes the torch: a new generation of mTOR inhibitors*. Nat Rev Drug Discov, 2011. **10**(11): p. 868-80.
4. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork*. Nat Prod Rep, 2007. **24**(1): p. 162-90.
5. Keatinge-Clay, A.T., *The structures of type I polyketide synthases*. Nat Prod Rep, 2012. **29**(10): p. 1050-73.
6. Laureti, L., et al., *Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens*. Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6258-63.
7. Williams, G.J., *Engineering polyketide synthases and nonribosomal peptide synthetases*. Curr Opin Struct Biol, 2013. **23**(4): p. 603-12.
8. Wilson, M.C. and B.S. Moore, *Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity*. Nat Prod Rep, 2012. **29**(1): p. 72-86.
9. Bretschneider, T., et al., *Vinylogous chain branching catalysed by a dedicated polyketide synthase module*. Nature, 2013. **502**(7469): p. 124-8.
10. Tsai, S.C., et al., *Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel*. Proc Natl Acad Sci U S A, 2001. **98**(26): p. 14808-13.
11. Tsai, S.C. and B.D. Ames, *Structural enzymology of polyketide synthases*. Methods Enzymol, 2009. **459**: p. 17-47.
12. Mahenthiralingam, E., et al., *Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic Burkholderia ambifaria Genomic Island*. Chem Biol, 2011. **18**(5): p. 665-77.

13. Piel, J., *Biosynthesis of polyketides by trans-AT polyketide synthases*. Nat Prod Rep, 2010. **27**(7): p. 996-1047.
14. Dutta, S., et al., *Structure of a modular polyketide synthase*. Nature, 2014. **510**(7506): p. 512-7.
15. Wong, F.T. and C. Khosla, *Combinatorial biosynthesis of polyketides--a perspective*. Curr Opin Chem Biol, 2012. **16**(1-2): p. 117-23.
16. Walsh, C.T., *Polyketide and nonribosomal peptide antibiotics: modularity and versatility*. Science, 2004. **303**(5665): p. 1805-10.
17. Poust, S., et al., *Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform*. Curr Opin Biotechnol, 2014. **30**: p. 32-9.
18. Richter, C.D., et al., *Multienzyme docking in hybrid megasynthetases*. Nat Chem Biol, 2008. **4**(1): p. 75-81.
19. Kittila, T. and M.J. Cryle, *Capturing the Structure of the Substrate Bound Condensation Domain*. Cell Chem Biol, 2016. **23**(3): p. 315-6.
20. Bloudoff, K., D. Rodionov, and T.M. Schmeing, *Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases*. J Mol Biol, 2013. **425**(17): p. 3137-50.
21. Samel, S.A., et al., *Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase*. Structure, 2007. **15**(7): p. 781-92.
22. Keating, T.A., et al., *The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains*. Nat Struct Biol, 2002. **9**(7): p. 522-6.
23. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D*. Nucleic Acids Res, 2010. **38**(Web Server issue): p. W545-9.
24. Lin, S., et al., *Specificity of the ester bond forming condensation enzyme SgcC5 in C-1027 biosynthesis*. Org Lett, 2012. **14**(9): p. 2300-3.
25. Lin, S., S.G. Van Lanen, and B. Shen, *A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis*. Proc Natl Acad Sci U S A, 2009. **106**(11): p. 4183-8.
26. Hahn, M. and T. Stachelhaus, *Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains*. Proc Natl Acad Sci U S A, 2004. **101**(44): p. 15585-90.

27. Buchholz, T.J., et al., *Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains*. ACS Chem Biol, 2009. **4**(1): p. 41-52.
28. Hahn, M. and T. Stachelhaus, *Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides*. Proc Natl Acad Sci U S A, 2006. **103**(2): p. 275-80.
29. Battye, T.G., et al., *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM*. Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 271-81.
30. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.
31. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
32. Emsley, P., et al., *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
33. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard*. Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 61-9.
34. McCoy, A.J., et al., *Phaser crystallographic software*. J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.

CHAPTER 5: Structural and Functional Studies of Enzymes Involved in Pentangular Polyphenol Biosynthesis

5.1 Abstract

Arixanthomycins are pentangular polyphenols (PP) with potent antiproliferative activities that were discovered through the heterologous expression of environmental DNA-derived gene clusters. The biosynthesis of arixanthomycin and other PPs is unusual, because it requires several novel type II polyketide synthase (PKS) enzymes for its complete maturation. Most type II PKSs contain a ketoreductase (KR) that mediates the C7-C12 first ring cyclization and C9-reduction. In contrast, the arixanthomycin (ARX) gene cluster harbors a C11-reducing ketoreductase (ARX 27), a C9-C14 reducing aromatase/cyclase (ARX 19), and an unprecedented C-17 and C-19 reducing KR (ARX 21). While bioinformatics is useful for predicting novel enzymes, the functions of ARX 19, ARX 21 and ARX 27 have yet to be confirmed *in vitro*. Further, the structural features that predispose the ARX biosynthetic enzymes to process atypical poly- β -ketone scaffolds remain unknown. Using *in vitro* reconstitution assays we show that ARX 19 is indeed a bona fide ARO/CYC. We further demonstrate that ARX 19 is independent of C-11 or C-9 ketoreduction, suggesting it is compatible in both reducing and non-reducing PKS pathways. In addition, we solved the crystal structure of ARX 21, the first structure of a C17- and C19 reducing-KR, which, is structurally similar to C-15 reducing KRs. This is the first *in vitro* and structural characterization of type II PKSs involved in PP biogenesis.

5.2 Introduction

Pentangular polyphenols (PP) are an underrepresented subfamily of bacterial aromatic polyketides that exhibit diverse biological activities.[1] Pradimicin, benastatin, and fredericamycin A are examples of PPs that serve as antifungals, chemotherapeutics, and antibiotics, respectively.[2] In *Actinomyces* and other bacteria, aromatic polyketides are generated by type II polyketide synthases (PKSs).[3-5] A prominent feature of type II PKSs is the utilization of a minimal PKS (Min PKS) that consists of a ketosynthase-chain length factor heterodimer (KS-CLF) and an acyl-carrier protein (ACP) for the generation of linear poly- β -ketone chains through repetitive decarboxylative condensations of malonyl-CoA.[6] Once a poly- β -ketone chain has been biosynthesized, it is enzymatically processed in a conserved order.[7, 8] For example, in type II PKS reducing pathways, the mature poly- β -ketone chain appended to the ACP is first reduced at position C-9 by a ketoreductase (KR) followed by cyclization and aromatization by an aromatase/cyclase (ARO/CYC) to establish a set C7-C12 ring pattern.[9] Additional downstream cyclases, ketoreductase and oxygenases further tailor the polyketide intermediate product to maturity.[10] While the biosynthesis of anthracycline and angucycline natural products by type II PKSs have been shown to follow a conserved sequence of enzyme catalyzed transformations, little is known about the order of processing and tailoring events that occur during PP biogenesis.[11, 12] In addition, the gene clusters that encode the PP biogenesis assemblies harbor many unusual features.[1, 13] The PP scaffolds, for example, are derived from some of the longest polyketide chains; whereas most bacterial min PKSs produce polyketide chains that range from 16 - 24 carbons in length, these chains are 26 - 30 carbons in length.[14] Second, PP formation in certain cases follows non-canonical carbonyl reduction at positions C-11, C-17 and C-19, whereas most aromatic KRs reduce the C-9 and C-15 carbonyl

groups.[1, 2, 4, 15, 16] Finally, PP biogenesis requires an additional 15 - 20 tailoring enzymes for complete maturation that include glycosyl transferases, transamidases, and multiple oxygenases, which are found in lower abundance in typical type II PKSs.[4]

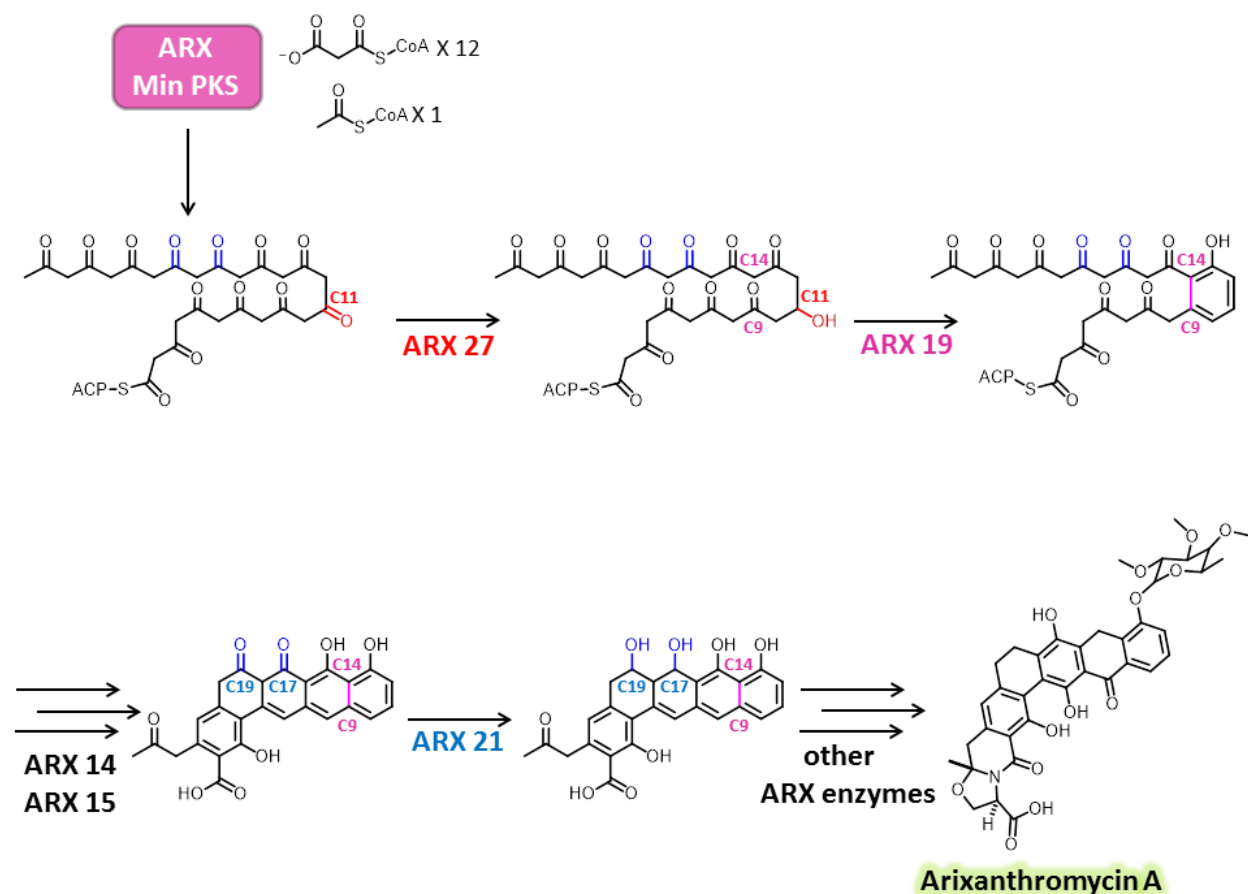


Figure 5.1. The proposed biosynthesis of arixanthromycin A.
Emphasizing key type II PKS processing enzymes.

Advances in genome mining over the past decade have led to the discovery of unusual biosynthetic gene clusters from environmental samples.[17, 18] Heterologous expression of these gene clusters has led to the identification of novel aromatic polyketides from unidentified or unculturable hosts.[14] In particular, the arixanthomycins are a class of recently discovered PPs

that exhibit attractive *in vitro* anti-proliferative activities against bacteria and cancer cells.[1] Further, the arixanthomycins gene cluster (ARX) contains a treasure trove of unusual PKS biosynthetic enzymes which includes an atypical ARO/CYC (ARX 19) that cyclizes a reduced polyketide chain at C9-C14, a novel C-11 reducing KR, and a tailoring C-17 and C-19 reducing KR (ARX 21) (Figure 5.1).[1] How ARX 19 and ARX 21 achieve such unusual feat of polyketide modification remains a mystery.

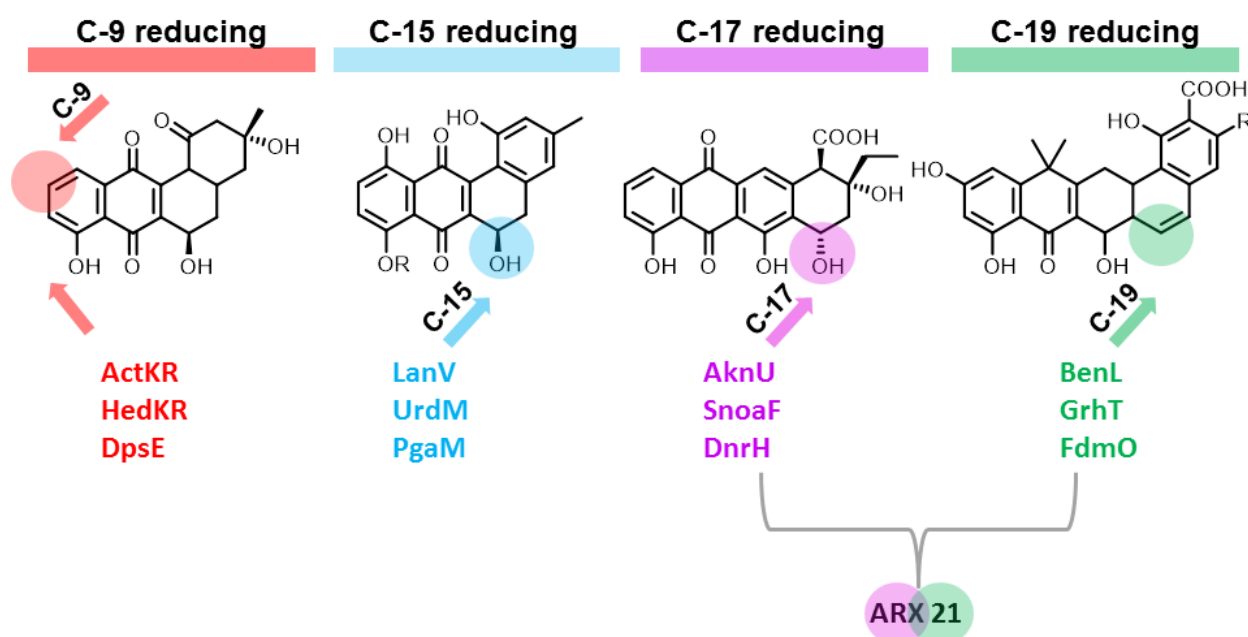


Figure 5.2. A schematic of the differing KR clades and their substrate specificities.
Note, ARX 21 has been predicted to behave as a C-17 and C-19 ketoreductase.

In this work, we present the crystal structure of ARX 21, a C-17 and C-19 reducing KR, and elucidated the molecular basis for its unusual regio-specificity. We also report results from the *in vitro* reconstitution of ARX 19, a reducing monodomain ARO/CYC. We found that the activity of ARX 19 can only be successfully reconstituted with a nonaketide that was supplied by a fungal PKS from the Pks4 gene cluster, which suggested that ARO/CYCs from PP biosynthetic

pathways have a more stringent chain length specificity in comparison to typical ARO/CYCs.[13, 19] In addition, ARX 19 was capable of biasing the cyclization profile of min PKSs in the absence of a KR, demonstrating that its activity in a reducing pathway is not dependent on C-11 ketoreduction. To the best of our knowledge, the crystal structure of ARX 21 represents the first structure of a bifunctional C-19 and C-17 reducing KR (Figure 5.2). Interestingly, ARX 21 shares a high degree of structural similarity to C-15 KRs, which leads to our proposal about the molecular mechanism on the ability of ARX 21 to reduce C-19 and C-17 carbonyl groups.[20] Taken together, the structural and functional studies here will further aid future efforts to implement regio-specific reduction and cyclization to diversify products generated by type II PKSs.

5.3 Results and Discussion

ARX 21 crystallized in the absence of NADPH in the $P2_1$ space group with 8 monomers in the asymmetric unit that forms a dimer of tetramers (Figure 5.3A). The structure was determined to 2.7 Å via molecular replacement using the crystal structure of gluconate 5-dehydrogenase from *Thermotoga maritima* as a search model (PDB code: 1VL8). High quality electron density for the entire polypeptide chain could be resolved in the crystal structure with the exception of poor density corresponding to the $\alpha 8$ - $\alpha 9$ helix-loop-helix lid region encompassing residues 180-195 which, depending on the monomer, was found to be partially or completely disordered. Surprisingly, upon inspection of the active site of ARX 21, partial electron density corresponding to NADPH was present for three of the eight ARX 21 monomers, suggesting that the NADPH cofactor was acquired intracellularly from the *E.coli* expression host. Previous structural studies with KRs have demonstrated that co-factor binding is crucial for the stabilization of the KR loop-

helix-loop region through favorable structural perturbations that may improve crystal packing.[21] To assess if NADPH binding could aid in the modeling of a complete ARX 21 structure, the ARX 21 crystals were soaked in excess NADPH and intensity statistics were collected. Clear electron density for NADPH and the missing residues 180-195 were present in the NADPH soaked crystal structure, providing an intact ARX 21 model that was refined to 2.4 Å (Figures 5.3B and 5.4C). Because the complete backbone of ARX 21 could only be successfully traced when bound to NADPH, the resulting structural analysis will refer to the NADPH-bound structure.

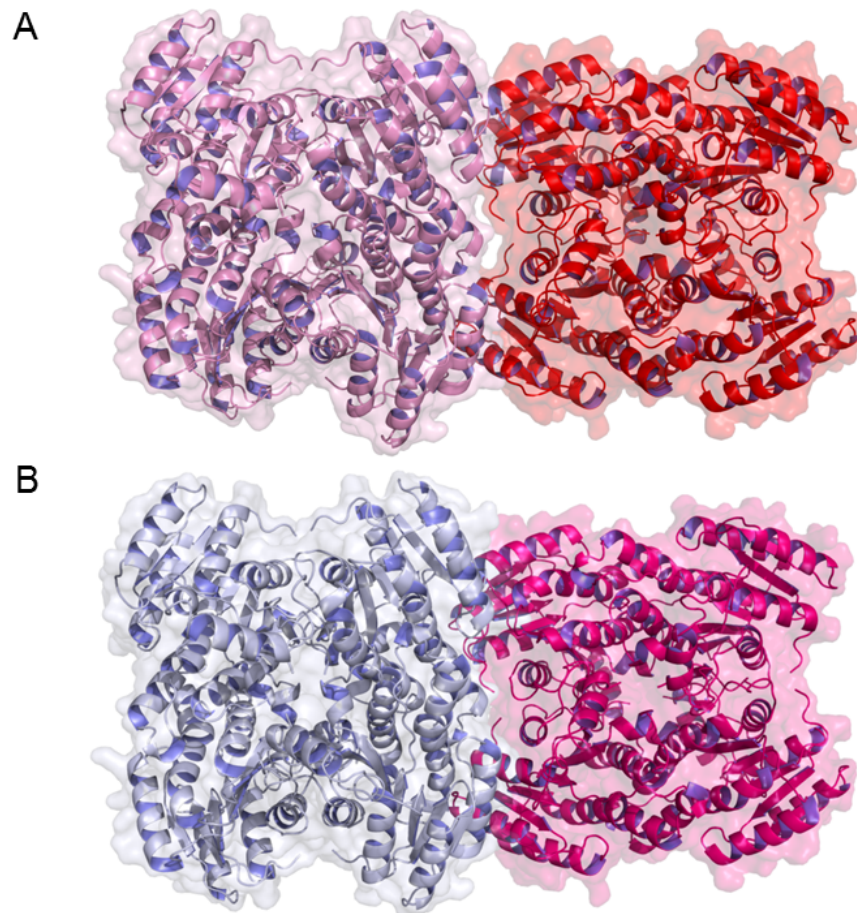


Figure 5.3. The overall architecture of ARX 21.

(A) The structure of Apo ARX 21 with dimeric tetramers colored in pink and red. (B) The structure of NADPH bound ARX 21 with dimeric tetramers colored in grey and magenta.

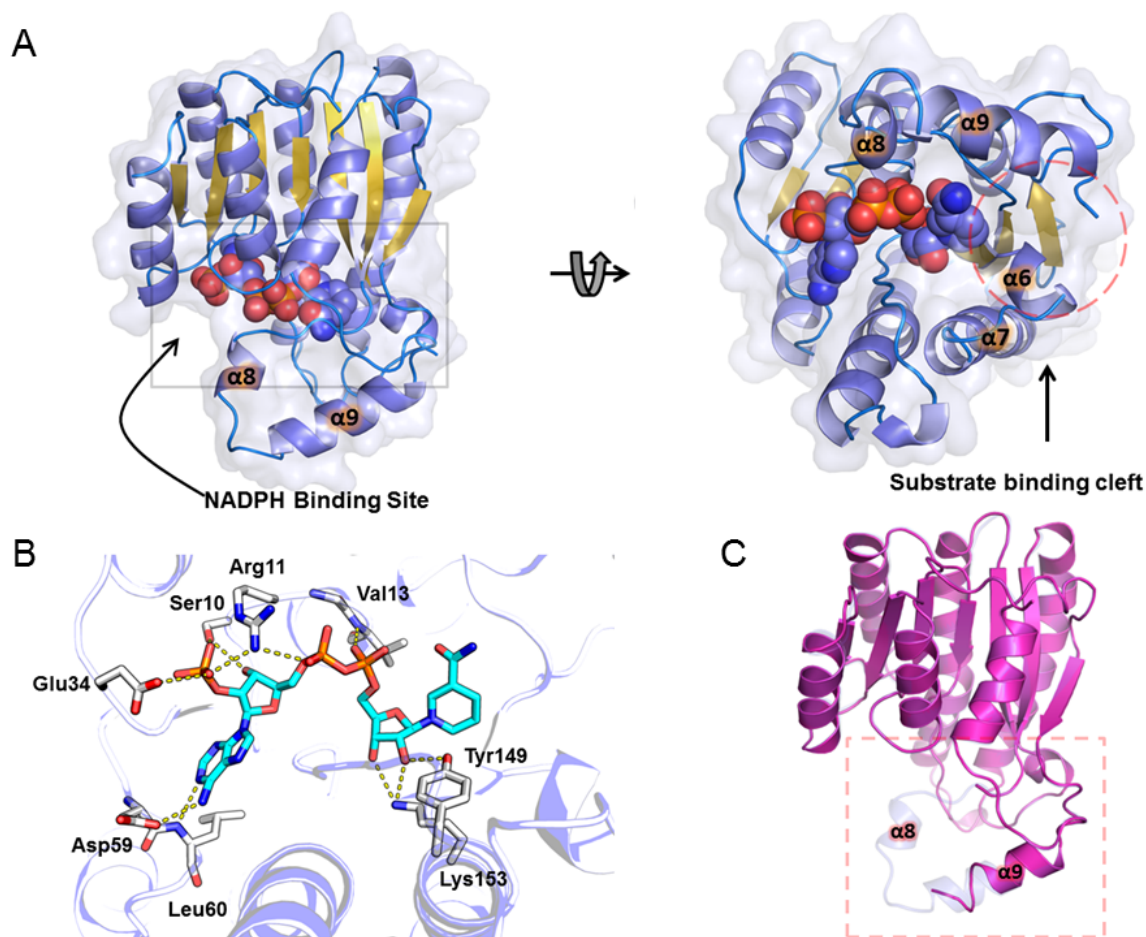


Figure 5.4. The co-factor and proposed substrate binding site of ARX 21.

(A) The tertiary structure of one ARX 21 monomer bound to NADPH, with highlighting both the substrate and cofactor binding sites. The active site is located at a cleft formed by the $\alpha 8$ - $\alpha 9$ flap and the loop region that spans $\beta 4$ and $\eta 1$. (B) Inspection of the residues that encompass the NADPH binding site of ARX 21 with the NADPH co-factor colored as cyan. (C) A structural overlay of an apo and NADP bound ARX 21 monomer. Clear electron density defining the $\alpha 8$ - $\alpha 9$ flap is absent in the apo form of ARX 21.

Overall Fold

Similar to other enzymes that belong to the short chain dehydrogenase/reductase (SDR) superfamily, ARX 21 contains a conserved Rossmann fold that consists of six parallel α helices bundled together in two groups of three, flanking a central seven-stranded β -sheet core. In addition, ARX 21 harbors the canonical SDR TGXXXGXG NADPH binding consensus

sequence and the conserved Ser-Tyr-Lys active site triad. The front of ARX 21's active site is demarcated by a groove that is formed between the α 8- α 9 flap and the loop region that spans from β 4 to η 1 (Figure 5.4). The back of the substrate-binding pocket is formed by the C-terminal end of the central seven-stranded β sheet core and a loop located between β 1 and α 1. The structural topology provides a large cavity that can accommodate both NADPH and a bulky pentacyclic aromatic substrate. In both the apo- and NADPH-bound structures, the active site of ARX 21 is in an open conformation that leaves both the NADPH and substrate binding sites largely exposed to the solvent.

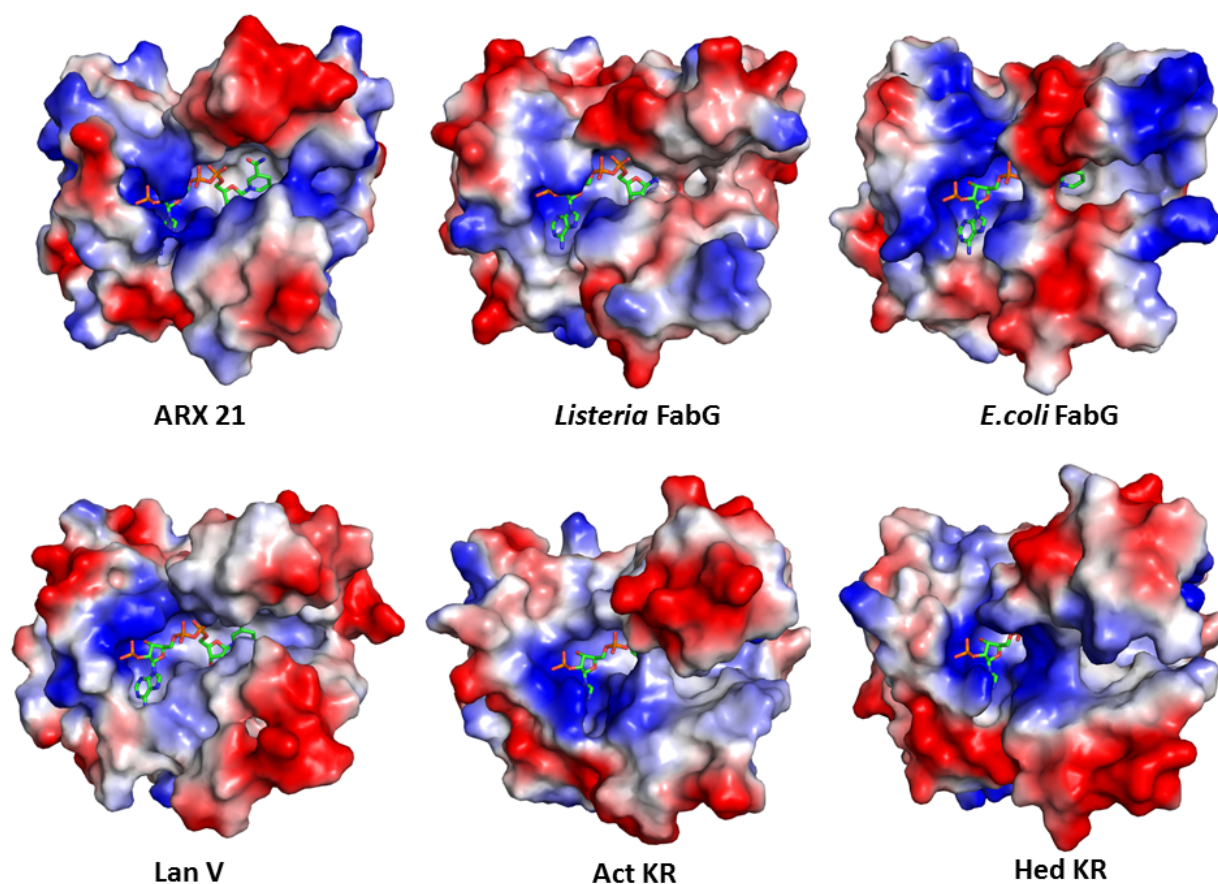


Figure 5.5. Electrostatic potential surfaces of ARX 21 and other KR homologues. Emphasizing the variance of charge distributions and active site accessibility across differing KR clades.

ARX 21 shares a high degree of structural homology but low sequence similarity (less than 34%) to previously determined KR structures involved in type II PKS and FAS biosyntheses. These similar KR structures include: the actinorhodin KR (ActKR), the hedamycin KR (HedKR), LanV, and multiple FabG homologues. Dali structural alignment search revealed that ARX 21 shares the highest structural similarity to *Listeria* FabG (LFabG, PDB code 4JRO) with a Z score of 35.8, followed by an SDR from *Serratia marcescens* BCRC 10948 (SM_SDR, PDB code 4ZGW) with a Z-score of 35.5.[22] Superimposition of the LFabG and SM_SDR structures onto that of ARX 21 revealed that all three have an exposed active site cavity, and ARX 21's pocket is the most accessible one to the solvent (Figure 5.5). The low sequence similarity of ARX 21 to other structurally characterized aromatic KRs merited additional structural comparisons of ARX 21 (a C-17 and C-19 reducing KR) to C-15-reducing KRs from different type II PKSs. ARX 21 shows the highest degree of structural similarity to the crystal structures of two C-15 reducing KRs, LanV and UrdMred, both bound to NADP⁺ and their putative substrates 11-deoxylandomycinone and rebelomycin, respectively, with a C α backbone RMSDs of 1.7 Å and 1.72 Å, respectively.[11, 20, 22, 23] Like LanV and UrdMred, ARX 21 contains a deep active site cavity; however, it is slightly less accessible to the solvent in comparison to the crystal structure of the two C-15 reducing KRs. Given the roles C-15 and C-19 KRs play in the downstream reductive tailoring of angucyclines and pentangular polyphenol intermediates, the crystal structures of the LanV and UrdMred ternary complexes served as a solid framework to compare possible substrate binding modes between KRs with differing substrate reduction sites. In the LanV and UrdMred ternary complex crystal structures, 11-deoxylandomycinone and rebelomycin bind between the α 6- α 7 loop and the α 8- α 9 flap where

the C-15 carbonyl oxygen of their respective substrates is placed within 3.8 Å of the C-4 atom of the nicotinamide ring of NADPH to achieve regiospecific reduction at position C-15. In ARX 21, residues that define the α 6- α 7 loop and α 9 helix are bulkier in comparison to LanV and UrdMred and, consequently, the substrate binding cleft of ARX 21 is narrower (Figure 5.6A). Specifically, a weak salt bridge is formed between R146 and Y202 that constricts the substrate binding cleft of ARX 21, which would lead to protein-substrate steric clashes if an angucycline intermediate were bound in this position. For UrdMred and LanV, R146 and Y202 of ARX 21 correspond to E/Q156 and M/L206, respectively (Figure 5.6). The presence of smaller residues in these positions provides ample space between the α 6- α 7 loop and α 8- α 9 lid to accommodate the tetracyclic aromatic substrates for the C-15 reducing KRs. It is interesting to note that R146 and Y202 are highly conserved amongst annotated C-19-reducing KRs and may be important for substrate recognition and specificity (Figure 5.6). Because ARX 21 acts on large pentacyclic aromatic compound, it is within reason to suggest that substrate binding could cause Y202 and/or R146 to move in a manner that could either allow for entry to the substrate binding region of ARX 21, or make favorable contacts through π - π stacking interactions with an incoming substrate (Figure 5.6A). Alternatively, it is also possible that ARX 21, a C-19 reducing KR, exhibits unique substrate binding modes in comparison to LanV and UrdMred. However, the co-crystal structure of the C-9 reducing ActKR bound to emodin suggests that the substrate binding regions are conserved across all KRs, as exemplified by the emodin binding site that is roughly in the same location as substrates 11-deoxylandomycinone and rebelomycin for LanV and UrdMred, respectively. Therefore, the crystal structure of ARX 21 further shows how KRs can acquire novel substrate and regio-specificities via the simple modification of several amino acids.

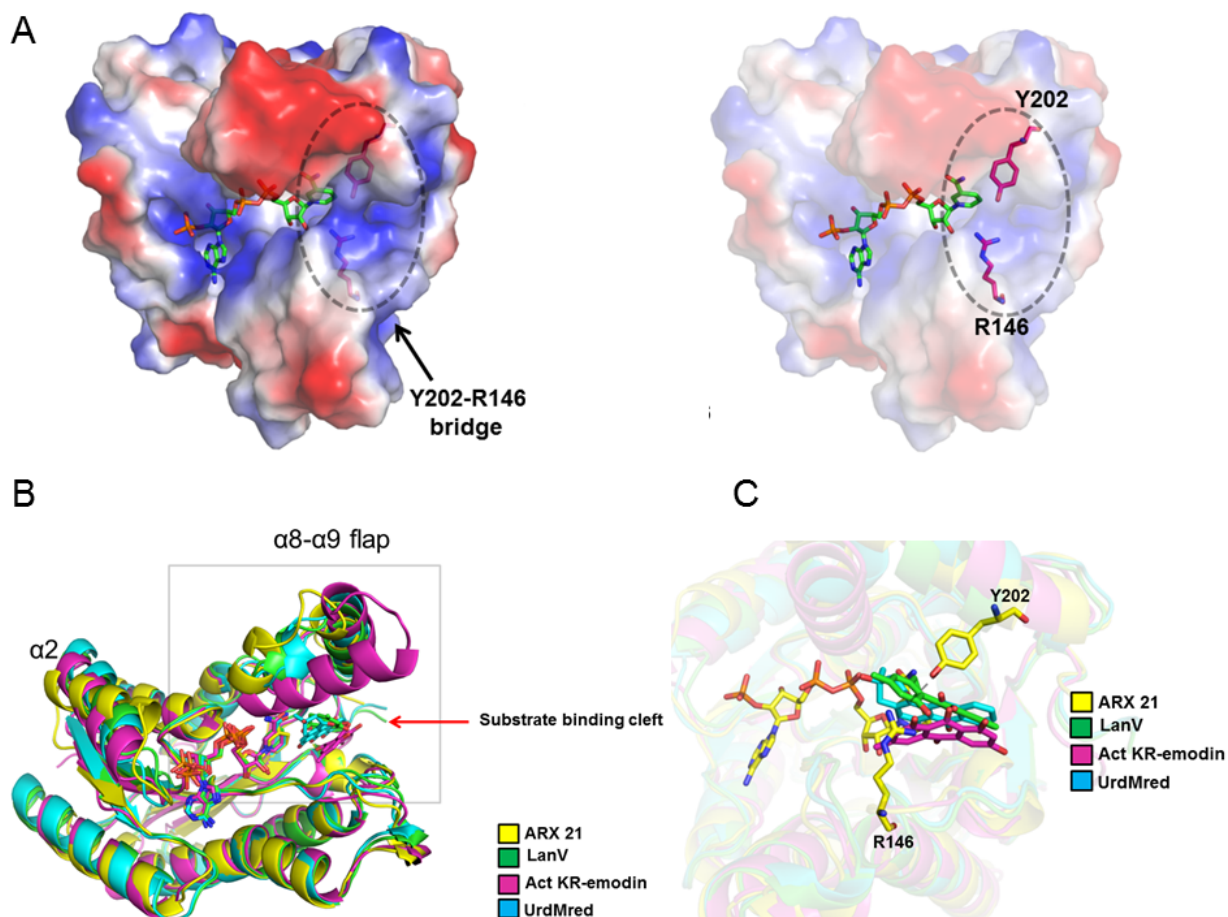


Figure 5.6. Residues proposed to be important for ARX 21 substrate binding and recognition.

(A) An electrostatic surface representation of an NADPH bound ARX 21 monomer. Y202 and R146 block the entrance to the substrate binding cleft. B) Superimposition of NADPH bound ARX 21 to the structure of ActKR (bound with emodin and NADP), LanV-11-deoxylandomycinone-NADP⁺, and rebelomycin-NADP⁺-UrdMred ternary complexes. The models are colored as gold, green, magenta, and cyan, respectively. C) Structural overlay of substrate bound KRs, demonstrating a conserved substrate binding location highlighting Y204 and R146 of ARX 21.

Structure-Sequence-Function Relationship

Due to the high degree of structural similarity of previously determined aromatic KRs with diverse functions, a sequence alignment was performed for ARX 21 against functionally and/or structurally characterized C-19-, C-17-, C-15-, and C-9-reducing KRs. The majority of the

structural deviations correlate with unconserved amino sequences that are localized in the following regions across all of the KR clades: helices $\alpha 2$, $\alpha 3$, $\alpha 8$, and $\alpha 9$, and loop regions that span between $\alpha 2$ - $\beta 3$, $\beta 4$ - $\alpha 4$, $\alpha 6$ - $\alpha 7$, and $\alpha 8$ - $\alpha 9$ (Figure 5.7). The amino acid sequences of the $\alpha 9$ helix are the most variable among all KRs and, consequently, the helix itself shows the greatest levels of structural deviations amongst deposited KR crystal structures with respect to ARX 21.

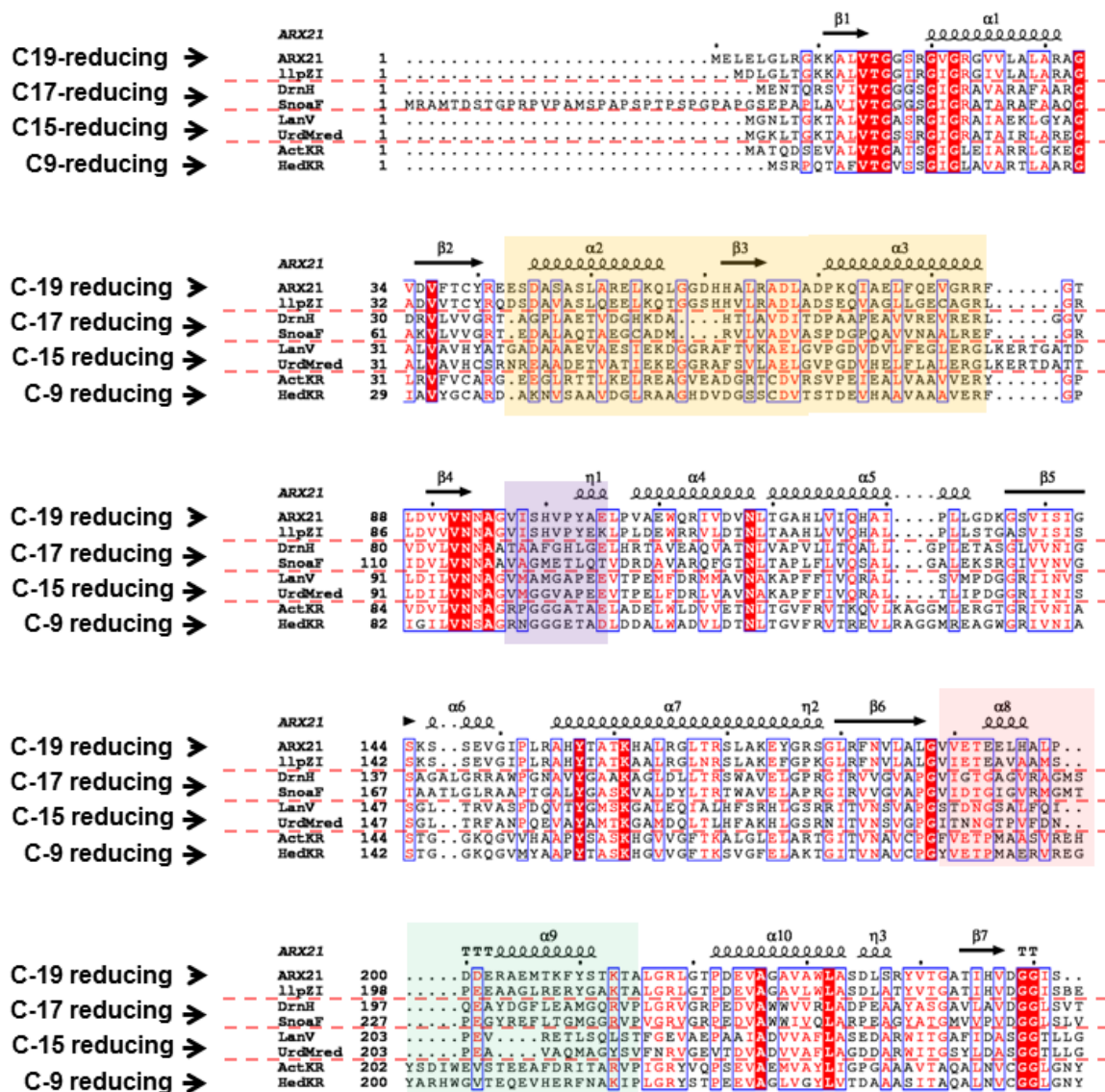


Figure 5.7. Sequence alignment of ARX 21 to functionally characterized C-19, C-17, C-15, and C-9 reducing KRs.

Unconserved amino acid sequences are colored.

In Vitro Reconstitution of ARX 19 Activity

To determine the substrate tolerance of ARX 21, we evaluated its activity *in vitro* using a well-established assay that utilizes the Pks4 minimal PKS (Pks4 Min PKS) and actinorhodin minimal PKS (Act Min PKS), which supply downstream enzymes such as KR with a linear octaketide and nonaketide intermediates, respectively. In both cases, the inclusion of ARX 21 did not lead to the formation of new products. This demonstrated that ARX 21 is not a promiscuous enzyme and is restricted to C-17 and C-19 carbonyl reductions. This also suggests that the lack of ARX 21 activity toward octaketide and nonaketide is a consequence of its inability to process substrates that do not contain a pentangular polyphenol scaffold. Interestingly, lower product yields of the Min PKS were observed when ARX 21 was added to Pks4 Min PKS *in vitro*, indicating that ARX 21 acts as a “molecular decelerator” that not only does not accept the nonaketide from the Min PKS, but also inhibits the chain elongation catalyzed by Min PKS (Figure 5.8A).

The above findings raised the question about the substrate specificity of all ARX modification enzymes such as ARX 19. To systematically evaluate the enzyme specificity, we conducted additional *in vitro* reconstitution assays using the Pks4 and Act Min PKS with the inclusion of ARX 19, the putative reducing C9-C14 ARO/CYC involved in the establishment of the first two rings of arixanthomycin. Significantly, a product profile shift was only observed for Pks4 Min PKS reactions containing ARX 19 (Figure 5.8). This implies that the activity of ARX 19 is dictated by poly- β -ketone chain length which, unlike other functionally characterized ARO/CYCs, cannot tolerate substrates fewer than 18 carbons *in vitro*. Further, these findings show that ARX 19 can successfully process unreduced poly- β -ketone chains in the absence of a C-9 or C-11 KR and therefore C-11 carbonyl reduction is not a mandatory criterion for

productive cyclization *in vitro*. In sum, the result of the *in vitro* assay is consistent with the observation that the structure of ARX 21 shares a degree of similarity with C-15 KR_s, which, do not accept ACP tethered substrates. Future work toward the structural and functional characterization of ARX 21 and ARX 19 is currently being pursued.

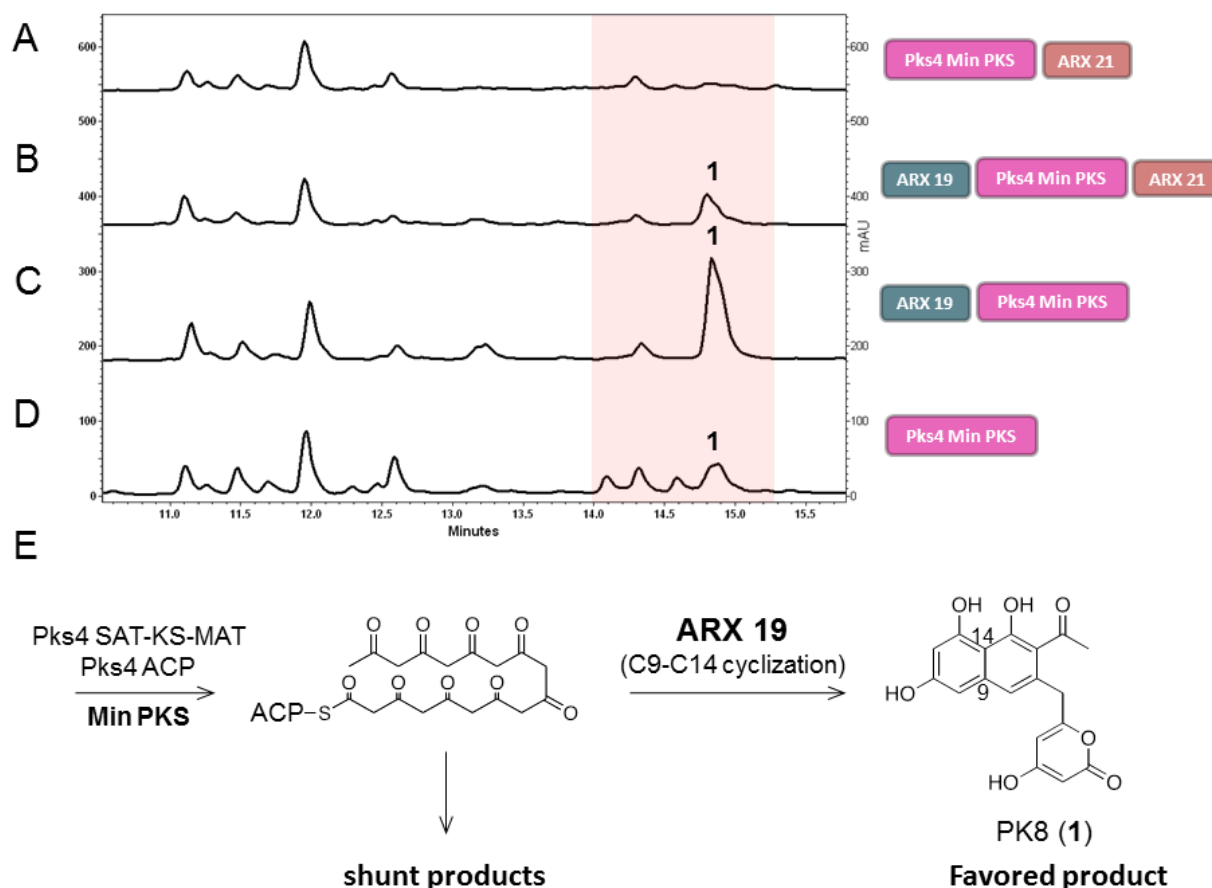


Figure 5.8. HPLC analysis of ARX 19 and ARX 21 activities at 280 nM.

(A) ARX 21 in the presence of the Pks4 Min PKS. (B) ARX 19 and ARX 21 in the presence of the Pks4 Min PKS. (C) ARX 19 in the presence of the Pks4 Min PKS. (D) Pks4 Min PKS alone. (E) A schematic of the Pks4 MIN PKS based cyclization assay. In the absence of a C9-C14 ARO/CYC the Pks4 Min PKS generates a nonaketide intermediate that affords spontaneous shunt products with myriad cyclization patterns. The inclusion of a C9-C14 ARO/CYC leads to a product profile shift favoring shunt products with C9-C14 cyclization patterns that can be detected using HPLC.

Biological Significance

The rational combination of different PKSs for the generation of novel therapeutics is a major goal for natural products researchers and synthetic biologists. However, efforts toward controlled combinatorial synthesis have been hampered by a lack of structural and biochemical knowledge that clearly define which sets of protein-protein and protein-substrate interactions allow productive biogenesis. An understanding of compatible enzymes pairs from otherwise orthogonal biosynthetic pathways could allow for the introduction of foreign biosynthetic machinery into non-endogenous metabolic pathways for the formation of new natural products.

In this study, we present the crystal structure of ARX 21, a novel C-17 and C-19 reducing KR, and we successfully conducted *in vitro* reconstitution of ARX 19, an usual C9-C14 reducing ARO/CYC. The crystal structure of ARX 21 is representative of the first C-17 and C-19-reducing KR. The exposed substrate pocket of ARX 21 bears a stronger resemblance to the C-15-reducing KRs, LanV and UrdMred. Superimposition of ARX 21 with the ternary complex structures of LanV, UrdMred, and ActKR revealed a conserved substrate binding location that is positioned between the $\alpha 6$ - $\alpha 7$ loop and the $\alpha 8$ - $\alpha 9$ flap. Although the structure of ARX 21 reported in this work is in its “substrate free” form, the conservation of the substrate binding site across multiple KR clades strongly suggested that the ARX 21 substrate is positioned in a similar location during C-17 and C-19 ketoreduction. Further, *in vitro* reconstitution demonstrated that ARX 21 is an enzyme with high specificity, and it does not accept the regular substrates of a typical KR such as octaketide or nonaketide. The successful *in vitro* reconstitution of the ARO/CYC activity of ARX 18 with the nonaketide producing Pks4 Min PKS demonstrated that the cyclization capacity of ARX 19 is similar to other C9-C14 cyclizing ARO/CYCs. Further, the ARX 19 activity toward a non-reduced polyketide chain indicates that C-11 or C-9 carbonyl reduction is

not required for ARX 19 mediated C9-C14 cyclization. We are now attempting to determine the crystal structure of ARX 19 to elucidate the molecular basis for C9-C14 cyclization during PP biogenesis.

In summary, this work has laid the framework for the *in vitro* reconstitution of the entire ARX pathway. Because the ARX pathway harbors unusual type II PKS biosynthetic machinery, a structural and biochemical investigation of the individual enzyme components involved in PP biogenesis will greatly aid in rational combinatorial biosynthesis of different type II polyketides.

5.4 Materials and Methods

The genes encoding ARX 19 and ARX 21 were amplified from BAC-AZ1076/33/378 (BAC containing the arixanthomycin gene cluster) using primers listed in Table 1. The amplified PCR products were double-digested with NdeI and NotI (New England Biolabs), and then ligated onto pET28c (Novagen) for the generation of ARX 21 and ARX 19 constructs with N-terminal His_{6x} tags. All cloning steps were carried out with NovaBlue competent *E.coli* cells (Millipore). DNA sequences were verified using Genewiz automated sequencing.

Table 5.1. Primers used to generate pET28-ARX 21 and pET28-ARX19.

ARX21F	5' TCGCTGCATATGGAAGCTGGAAGCTCGGC 3'
ARX21R	5' GCTATAGCGGCCGCTCAGGAAATTCCTCC 3'
ARX19F	5' GCTATACATATGGTGGCGCACACCGA 3
ARX19R	5'CATATAGCGGCCGCCTAGCCGACCGC '3

Expression and Purification of ARX 21

The resulting N-terminal His_{6x} ARX 21 construct was transformed in BL21(DE3) *E.coli* competent cells (Novagen) for protein over-expression. Cell cultures harboring the His_{6x} ARX 21 plasmid were grown in terrific broth (2 X 1L) supplemented with 50 µg/mL kanamycin until A₆₀₀ reached 0.6 - 0.8. The temperature was then lowered to 18 °C and protein expression was induced upon the addition of 0.5 mM IPTG. After incubation at 18 °C for 14 - 18 hours, cells were harvested via centrifugation at 5,000 rpm, pelleted, flash frozen, and stored at -80 °C. Cell pellets were thawed on ice and resuspended in lysis buffer (50 mM Tris·HCl pH 8.0, 20 mM imidazole pH 8.0, 300 mM NaCl, and 10 % glycerol). The cell suspension was sonicated (6 X 30 second cycles) and the cell debris was separated by centrifugation at 14,000 rpm for 1 hour at 4 °C. The clarified cell lysate was mixed with 5 mL of nickel NTA resin (Thermo Fischer Scientific) for 50 minutes at 4 °C. The nickel slurry was then transferred to a gravity flow column, the flow through volume was collected and the resin was washed with 100 mL of lysis buffer (2 X 50 mL washes). Elution of ARX 21 was accomplished via the addition of 5 mL fractions containing increasing concentrations of imidazole in lysis buffer (40-450 mM). Protein purity was assessed by SDS-PAGE. Fractions containing ARX 21 were pooled and dialyzed against storage buffer (50 mM Tris·HCl pH 8.0, 50 mM NaCl, 1 mM DTT, and 15 % glycerol). Dialyzed protein was concentrated to 15 - 16 mg/mL using a 10,000 MWCO centrifugal filter device (Millipore) and buffer exchanged into crystallization buffer (50 mM Tris·HCl pH 8.0 and 1 mM DTT) using a PD-10 desalting column (GE Healthcare) prior to crystal screens.

Expression and Purification of ARX 19

The ARX-19-pET28 construct was used to transform BL21(DE3) cells for protein overproduction. 4 X 1 L cultures were grown in TB containing 50 µg/mL kanamycin until O.D₆₀₀ reached 0.8 - 1.0 at 37 °C and 180 RPM. Protein expression was induced by the addition of 0.5 mM IPTG; the cells were then cooled and incubated at 18 °C for approximately 16 - 18 hours. Cells were harvested via centrifugation 5100 RPM for 20 minutes and resuspended in 50 ml lysis buffer (50 mM Tris·HCl pH 8.0, 20 mM imidazole pH 8.0, 300 mM NaCl, and 10 % glycerol). The cells were lysed through two rounds of microfluidization and the cell debris was removed through centrifugation at 14,000 rpm at 4 °C for one hour. The clarified cell lysate was incubated with 10 mL of nickel NTA resin (Bio-Rad) while being gently stirred at 4 °C for one hour. The nickel-lysate slurry was then transferred to a gravity flow column where it was washed with 2 X 50 mL of lysis buffer followed by the addition of 10 mL of 40 mM imidazole in lysis buffer. Elution of ARX 19 was achieved through the inclusion of 3 X 5 mL fractions containing 100 mM imidazole in lysis buffer followed by the subsequent addition of 2 X 5 mL fractions containing: 150, 250, and 450 mM imidazole in lysis buffer, respectively. The resulting fractions were subject to SDS-PAGE analysis prior to being pooled and dialyzed against (Tris·HCl pH 8.0, 50 mM NaCl, and 10 % glycerol). ARX 19 was then concentrated to 5 mg/mL using a 10,000 MWCO centrifugal filter device (Millipore) and injected onto a Superdex 200 column (GE Healthcare) pre-equilibrated with 50 mM Tris·HCl pH 8.0, 50 mM NaCl and 1 mM DTT. Pure ARX 19 was concentrated to 13 mg/mL, flash frozen in liquid nitrogen and stored at -80 °C for later use.

Crystallization and Data Collection

Small dagger shaped ARX 21 crystals were obtained from combining 4 μ L of protein solution at 7-10 mg/mL with 4 μ L of well solution containing 50 mM Tris-HCl pH 8.5, 0.2 M sodium acetate, and 30-32 % PEG 4000 using the vapor diffusion method. ARX 21 crystals formed after 2 - 5 days at 20 °C. ARX 21 crystals were harvested and flash frozen in liquid nitrogen. For the NADPH-bound ARX 21 structure, *apo* crystals were soaked in the crystallization solution containing 20 mM NADPH (Sigma) for 20-30 seconds, then flash frozen in liquid nitrogen. Diffraction data for both *apo* and NADPH-soaked ARX 21 crystals were collected at the Advanced Light Source (ALS) beamline 821 at the Lawrence Berkeley National Laboratory.

Phasing, Model building and Refinement

Datasets for both the *apo* and NADPH-bound forms of ARX 21 were indexed, integrated and scaled using Mosflm.[24] Initial phases were obtained through molecular replacement (Phaser) using the crystal structure of gluconate 5-dehydrogenase from *Thermotoga maritima* as a search model (PDB code: 1VL8) as a search model.[25] PHENIX AutoBuild was used to construct an initial model of ARX 21.[26] The final ARX 21 models were obtained through multiple rounds of model building and refinement using COOT and Phenix.REFINE.[27, 28]

For the NADPH bound ARX 21 structure, LigandFit was used to identify electron density corresponding to NADPH and Elbow was used to generate ligand restraints.[28] The final NADPH-bound ARX 21 models were obtained through multiple rounds of model building and refinement using COOT and Phenix.REFINE.[27, 28]

ARX 21/19 Reconstitution Assays

The fungal Pks4 minimal PKS was used to assay ARX 19 and ARX 21 activity *in vitro*. Two hundred fifty μL *in vitro* reactions with the following final concentrations with or without 100 μM ARX 19 and or 160 μM ARX 21, 50 μM *holo*-Pks4 ACP, 37 μM Pks4 SKM (SAT-KS-MAT tridomain), 2.5 mM malonyl-CoA, and 2.5 mM in 50 mM potassium phosphate pH 7.5 were used to assess the activities of ARX 19 and ARX 21. The reactions were incubated overnight for approximately 15 - 17 hours at 20 °C. The following day, the reactions were extracted 3 times with 300 μL of ethyl acetate : methanol : acetic acid (94 : 5 : 1). The organic layer was collected and dried using a speed vacuum concentrator. The resulting concentrate was then resuspended in 25 μL of DMSO and subject to reverse phase HPLC analysis as previously described.[9].

To test for the chain length specificity of ARX 19, additional assays were performed using the actinorhoidin minimal PKS assay.[8, 9] with or without the addition of 50 μM of ARX 19 and or ARX 21, respectively. The reactions were extracted and analyzed via HPLC as described above.

Expression Purification of Pks4 SKM, Pks4 ACP, Sfp and Act KS/CLF

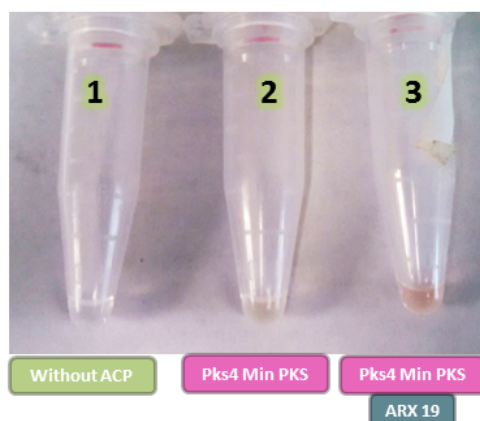
The purification of Pks4 SKM, Pks4 ACP, and Sfp was performed as previously described. In all cases, purified protein was dialyzed against 50 mM Tris·HCl pH 7.5, 50 mM NaCl, 10 % glycerol and 1 mM DTT at 4 °C. The dialyzed protein was concentrated, flash frozen in liquid nitrogen, and stored at -80 °C for later use. The expression and purification of Act KS/CLF was conducted as previously described.[9]

Generation of Holo Pks4 and ActACPs

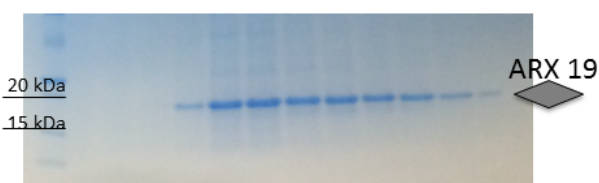
Holo Pks4 ACP was generated via *in vitro* phosphopantetheinylation of apo-Pks4 ACP. A 400 uL ACP loading reaction was setup using the following final concentrations: 375 μ M apo pks4 ACP, 10 μ M of Sfp, 1 mM CoA, and 12.5 mM MgCl₂ 50 mM potassium phosphate pH 7.0. The reaction was incubated at 37 °C for 2.5 hours. The formation of holo-Pks4 ACP was verified using MALDI TOF mass spectrometry.

The holofication of ActACP was achieved *in vivo*. A plasmid encoding ActACP was transformed into BAP1 cells that have been genetically engineered to natively express SFP and ensure complete ACP phosphopantetheinylation. *Holo* ActACP was purified as described for ARX 21 and dialyzed against 50 mM Tris·HCl pH 8.0, 50 mM NaCl and 1 mM DTT. The dialyzed ActACP was then concentrated, flash frozen and stored in -80 °C.

A



B



C

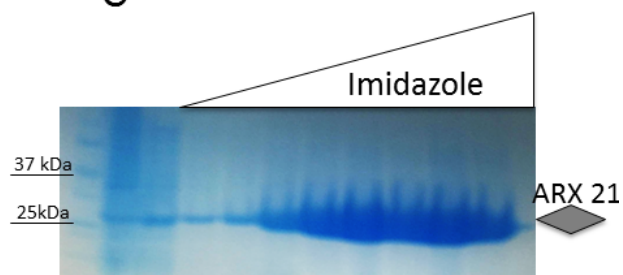


Figure 5.9. *In vitro* reconstitution and protein purification of ARX 19 and ARX 21.

(A) A visual inspection of Pks4 Min PKS *in vitro* reconstitution reactions after 18 hours. (1) Without Pks4 ACP (negative control), (2) Pks4 Min PKS, (3) Pks4 Min PKS and ARX 19. (B) SDS PAGE gel of pure ARX 19. (C) SDS PAGE gel of pure ARX 21.

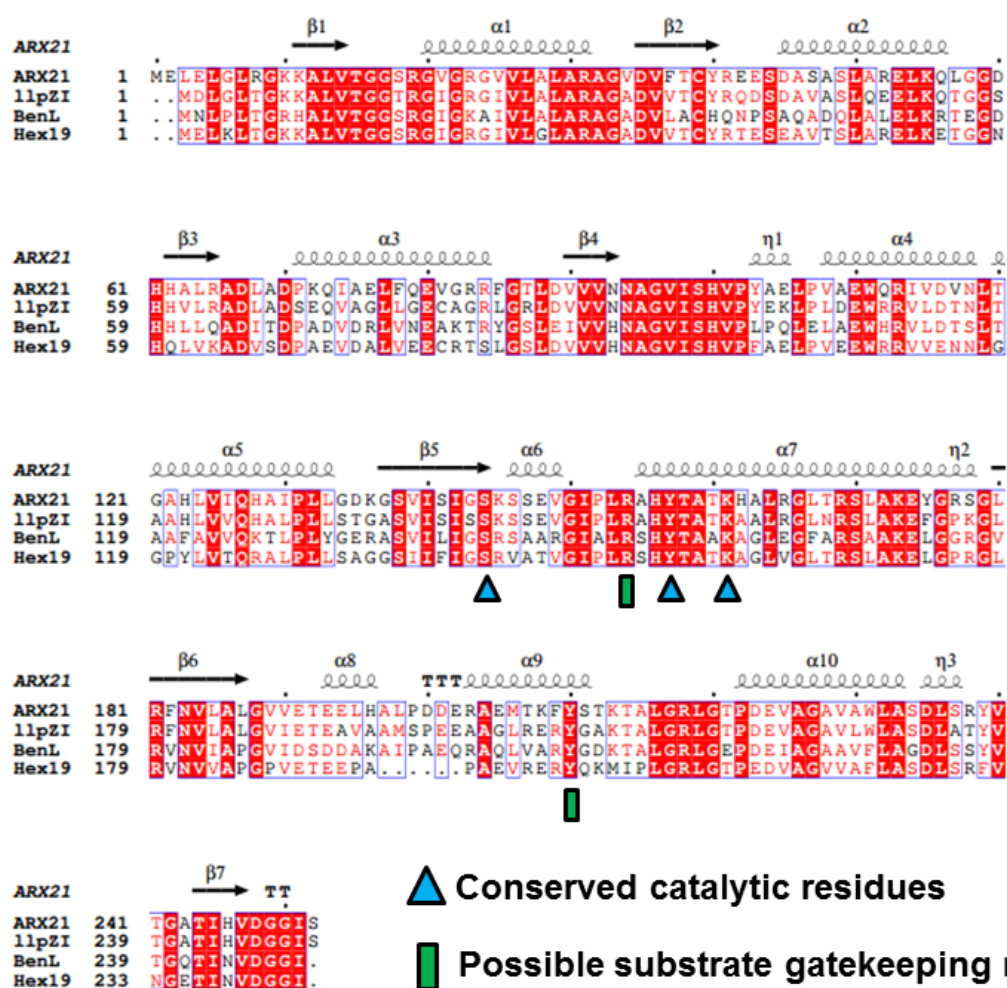


Figure 5.10. Sequence alignment of ARX 21 to functionally characterized C-19 reducing KRs.

Y204 and R146 along with essential catalytic residues are conserved across all C-19 KRs.

Table 5.2. Data collection and refinement statistics.

Structure:	apo ARX 21	NADPH-ARX 21
Wavelength	0.9794	1.000
Resolution range	52.67 - 2.6 (2.693 - 2.6)	86.3 - 2.3 (2.382 - 2.3)
Space group	P 1 21 1	P 1 21 1
Unit cell	105.38 86.6499 107.32 90 108.97 90	105.34 86.54 106.97 90 108.8 90
Total reflections	104658 (10172)	157573 (15055)
Unique reflections	54082 (5326)	79667 (7695)
Multiplicity	1.9 (1.9)	2.0 (2.0)
Completeness (%)	0.96 (0.95)	0.98 (0.96)
Mean I/sigma(I)	6.66 (2.94)	10.70 (3.98)
Wilson B-factor	26.52	28.22
R-merge	0.03741 (0.1316)	0.03323 (0.1514)
R-meas	0.05291 (0.1861)	0.04699 (0.2141)
CC1/2	0.997 (0.944)	0.998 (0.954)
CC*	0.999 (0.985)	0.999 (0.988)
Reflections used in refinement	52666 (4935)	79631 (7692)
Reflections used for R-free	1959 (191)	1990 (198)
R-work	0.1717 (0.1942)	0.1739 (0.1933)
R-free	0.2363 (0.2993)	0.2290 (0.2632)
CC(work)	0.958 (0.932)	0.966 (0.938)
CC(free)	0.910 (0.852)	0.923 (0.841)
Number of non-hydrogen atoms	14678	15662
Macromolecules	14167	14844
ligands	96	384
Protein residues	1919	1984
RMS(bonds)	0.008	0.017
RMS(angles)	1.11	1.46
Ramachandran favored (%)	96	96
Ramachandran allowed (%)	3.3	2.9
Ramachandran outliers (%)	1.2	1.6
Rotamer outliers (%)	2.8	3.5
Clashscore	8.93	9.47
Average B-factor	30.74	34.30
macromolecules	30.64	34.15
ligands	46.51	37.87
solvent	30.65	36.16

5.5 References

1. Kang, H.S. and S.F. Brady, *Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols*. ACS Chem Biol, 2014. **9**(6): p. 1267-72.
2. Lackner, G., et al., *Biosynthesis of pentangular polyphenols: deductions from the benastatin and griseorhodin pathways*. J Am Chem Soc, 2007. **129**(30): p. 9306-12.
3. Hertweck, C., *Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology*. Trends Biochem Sci, 2015. **40**(4): p. 189-99.
4. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review*. Nat Prod Rep, 2001. **18**(4): p. 380-416.
5. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork*. Nat Prod Rep, 2007. **24**(1): p. 162-90.
6. Javidpour, P., et al., *The determinants of activity and specificity in actinorhodin type II polyketide ketoreductase*. Chem Biol, 2013. **20**(10): p. 1225-34.
7. Ames, B.D., et al., *Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides*. Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5349-54.
8. Ames, B.D., et al., *Structural and biochemical characterization of Zhul aromatase/cyclase from the R1128 polyketide pathway*. Biochemistry, 2011. **50**(39): p. 8392-406.
9. Caldara-Festin, G., et al., *Structural and functional analysis of two di-domain aromatase/cyclases from type II polyketide synthases*. Proc Natl Acad Sci U S A, 2015. **112**(50): p. E6844-51.
10. Meurer, G., et al., *Iterative type II polyketide synthases, cyclases and ketoreductases exhibit context-dependent behavior in the biosynthesis of linear and angular decapolyketides*. Chem Biol, 1997. **4**(6): p. 433-43.
11. Patrikainen, P., et al., *Tailoring enzymes involved in the biosynthesis of angucyclines contain latent context-dependent catalytic activities*. Chem Biol, 2012. **19**(5): p. 647-55.
12. Tian, J., et al., *Discovery of pentangular polyphenols hexaricins A-C from marine Streptosporangium sp. CGMCC 4.7309 by genome mining*. Appl Microbiol Biotechnol, 2016. **100**(9): p. 4189-99.

13. Zhou, H., Y. Li, and Y. Tang, *Cyclization of aromatic polyketides from bacteria and fungi*. Nat Prod Rep, 2010. **27**(6): p. 839-68.
14. Kang, H.S. and S.F. Brady, *Mining soil metagenomes to better understand the evolution of natural product structural diversity: pentangular polyphenols as a case study*. J Am Chem Soc, 2014. **136**(52): p. 18111-9.
15. Javidpour, P., et al., *Structural and biochemical studies of the hedamycin type II polyketide ketoreductase (HedKR): molecular basis of stereo- and regiospecificities*. Biochemistry, 2011. **50**(34): p. 7426-39.
16. Javidpour, P., et al., *Structural and biochemical analyses of regio- and stereospecificities observed in a type II polyketide ketoreductase*. Biochemistry, 2011. **50**(21): p. 4638-49.
17. Reddy, B.V., et al., *Natural product biosynthetic gene diversity in geographically distinct soil microbiomes*. Appl Environ Microbiol, 2012. **78**(10): p. 3744-52.
18. Handelsman, J., et al., *Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products*. Chem Biol, 1998. **5**(10): p. R245-9.
19. Zhang, W., Y. Li, and Y. Tang, *Engineered biosynthesis of bacterial aromatic polyketides in Escherichia coli*. Proc Natl Acad Sci U S A, 2008. **105**(52): p. 20683-8.
20. Patrikainen, P., et al., *Structure-based engineering of angucyclinone 6-ketoreductases*. Chem Biol, 2014. **21**(10): p. 1381-91.
21. Price, A.C., et al., *Structure of beta-ketoacyl-[acyl carrier protein] reductase from Escherichia coli: negative cooperativity and its structural basis*. Biochemistry, 2001. **40**(43): p. 12772-81.
22. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D*. Nucleic Acids Res, 2010. **38**(Web Server issue): p. W545-9.
23. Paananen, P., et al., *Structural and functional analysis of angucycline C-6 ketoreductase LanV involved in landomycin biosynthesis*. Biochemistry, 2013. **52**(31): p. 5304-14.
24. Battye, T.G., et al., *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM*. Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 271-81.
25. McCoy, A.J., et al., *Phaser crystallographic software*. J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
26. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard*. Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 61-9.

27. Emsley, P., et al., *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
28. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

CHAPTER 6: Conclusions and New Directions

6.1 Overview

Natural products are secondary metabolites that have long been utilized for medicinal and agricultural purposes.[1-3] Polyketides and nonribosomal peptides are two distinct groups of significant secondary metabolites that exhibit diverse biological activities and are important for drug lead discovery.[4-6] Avermectin (insecticide and antiparasitic), lovastatin (anticholesterol), and doxorubicin (chemotherapeutic) are salient examples of polyketide therapeutics.[7-10] Similarly, surfactin, vancomycin, and teixobactin are representatives of medicinally relevant nonribosomal peptides.[11-13] Both polyketides and nonribosomal peptides are biosynthesized from simple malonate or amino acid precursors by multienzyme complexes known as polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs), respectively.[3, 14] PKSs and NRPSs are complex biosynthetic machines that utilize various tactics to achieve natural product structural diversity. Both PKSs and NRPSs use a wide assortment of differing extender and starter units and employ variation in the types and numbers of processing enzymes during natural product biogenesis.[15-18] While the biosynthetic framework for PKs and NRPs is understood, structural knowledge that would allow for the *de novo* programming of these biosynthetic assemblies for the generation of new products is lacking.[19, 20] The work presented in this dissertation sheds insight on the structures and functions of enzymes involved in polyketide and non-ribosomal peptide biosynthesis that will aid in the development of engineered PKS or NRPS natural products.

6.2 Chapter 2: WhiE

The introduction of a set cyclization motif is a means of achieving natural product structural diversity.[16] The aromatase/cyclases (ARO/CYC) are responsible for establishing first and/or second rings during type II PKS biogenesis.[21] While representatives of all ARO/CYC clades have been characterized, structural details about how these enzymes fold their poly- β -ketone substrate to yield specific cyclization patterns remain unclear.[22] To gain structural insight into ARO/CYC substrate-protein interactions, I solved the co-crystal structures of WhiE, a C9-C14 ARO/CYC, bound to four distinct pantetheine-linked polyketide analogues that include tri-, hexa-, hepta- and octa-ketide mimics.[23] Because natural tri-, hexa-, hepta- and octa-ketide substrates are highly reactive, these stable analogues served as excellent tools for probing protein-substrate interactions that are arduous to capture.[24] In the tri-, hexa-, and hepta-ketide mimic bound structures, the probes are either resolved along the exterior of substrate binding pocket or adopted unproductive binding modes in the interior of the WhiE ARO/CYC. These structural observations with the short chain mimics correlates well with the observed incapacity of WhiE to successfully cyclize poly- β -ketone chain fewer than 16 carbons in lengths. In contrast, in the octaketide-bound structure, the probe was found to bind in a productive motif that placed the C9 and C14 of the polyketide substrate within 4.5 Å of one another and posed for first-ring cyclization. These findings strongly suggest that substrate chain length is an essential criterion for successful ARO/CYC-mediated cyclization, because WhiE can only properly orientate the polyketide mimic with longer than 16 carbons. Moreover, R69, the predicted key catalytic residue, made hydrogen bonds with the hexa-, hepta-, and octa-ketide analogues. Given its non-specific nature, R69 may play a fundamental role in binding the substrate, but not guiding chain length specificity of ARO/CYCs. This is the first application of

pantetheine-linked probes to investigate the chain length control of any type II PKS enzymes. Further, an understanding of the protein-substrate interactions that occur during cyclization is necessary to introduce specific cyclization motifs into foreign PKS pathways to produce new natural products with unnatural cyclization patterns. We are in the process of submitting a manuscript that outlines these findings to PNAS.

6.3 Chapter 3: MccB

The incorporation of various extender units is a strategy often used by PKSs to generate chemical diversity.[25] Acyl-CoA carboxylases (ACCase) are biotin-dependent enzymes responsible for the generation of an assortment of malonyl-CoA based extender units that are used for polyketide chain propagation.[26] Typically, ACCases are limited to the carboxylation of short chain acyl-CoAs such as acetyl-, propionyl-, and butyryl-CoA to yield malonyl-, methylmalonyl-, and ethylmalonyl-CoA extender units, respectively. In comparison, longer, more extravagant extender units are biosynthesized from NADH dependent crotonyl-CoA carboxylases/reductase (CCRs).[27, 28] Stambomycins are a novel class of 51 membered glycosylated natural products that harbor unusual side chains in position C26 that are derived from hexyl-, heptyl-, and octyl-, nonyl-malonyl-CoA building blocks.[29] Interestingly, the stambomycin gene cluster does not encode CCR, but instead contains an unusual ACCase β subunit thought to be involved in the production of extender units that results in the long alkyl side chain at position C-26.[29] In collaboration with the Challis group at the University of Warwick, using feeding experiments with radiolabeled substrate precursors, we demonstrated that the stambomycin ACCase β subunit (MccB) is responsible for non-conical ACCase mediated production of long-chain extender. This is the first account of a CCR independent

pathway for the formation of uncommon PKS extender units. In addition, I solved the *apo* and hexanoyl-CoA bound structures of MccB to decipher the structural basis for MccB-mediated extender unit formation. The MccB crystal structure revealed that MccB contains a large active site pocket that is capable of binding myriad acyl-CoAs. The co-crystal structure of MccB bound to hexanoyl-CoA led to the identification of residues that expanded the substrate specificity of MccB towards long chain acyl-CoAs that are absent in canonical ACCase β subunits. For example, both the β subunit of propionyl-CoA carboxylase (PccB) from *S. coelicolor* and β the subunit of acetyl CoA carboxylase (AccB) contain a Phe residue at residue 161 that greatly restricts the dimensions of their acyl binding pocket.[30] In contrast, Phe161 becomes Gly161 in MccB, which greatly expands the substrate cavity of MccB, thus accommodating larger substrates. The information obtained from the MccB structures enabled us to conduct feeding experiments with azide and alkyne tagged extender units that led to the generation of novel stambomycins containing azide and alkyne moieties. Subsequent bioinformatic analysis of putative MccB homologues revealed that Gly161 is conserved across ACCase β subunits involved in PKS pathways that utilize long chain extender units. These findings indicate that formation of long chain extender units by ACCases are more prevalent in nature than originally thought and may serve as new biosynthetic route for engineering tagged natural products. A manuscript that encompasses this work has been submitted to Nature Communications and is currently under revision with favorable reviews.

6.4 Chapter 4: Bamb_5915

Enacyloxins are uncommon type I modular polyketides that are biosynthesized by an atypical *cis/trans*-AT PKS/NRPS hybrid (Ena PKS).[31] Unlike canonical modular biosynthesis

where a mature polyketone chain is hydrolyzed from a termination module containing a TE domain, Ena PKS utilizes a standalone NRPS condensation domain (C-domain) for product release.[3, 32] During the final steps of enacyloxin biosynthesis, a mature polyketide chain is passed from the ACP of module 5916 to the KS0, where it is subsequently transferred to module 5917 that contains a free standing PCP domain. The polyketide chain linked to the PCP is released through an intermolecular transesterification reaction between shikimate and the PCPs thioester substrate catalyzed by the C-domain (bamb_5915).[31] Interestingly, bamb_5915 only shows activity towards PCP-linked substrates and is inactive towards ACPs from other PKS modules in the Ena biosynthetic pathway. To determine the structural basis for C-domain mediated chain release during enacyloxin biosynthesis, I solved the crystal structure of bamb_5915. While bamb_5915 bears a strong resemblance to previously determined C-domain structures, the crystal structure of bamb_5915 revealed an unexpected N-terminal subdomain that is unconserved across condensing enzymes which, due to low sequence similarity to other known proteins, is difficult to predict using bioinformatics tools. Subsequent structural analysis of the N-term subdomain of bamb_5915 (COM_{2N}) reveals that it shares a high structural similarity and low sequence similarity to a previously determined type 2 communication mediating domain (COM₂) from the tublysin biosynthetic pathway, which is thought to be important for the recognition between PKS and NRPS, as well as the oligomerization and assembly of the mix-type megasynthase.[33] These observations lead to the identification of a complementary COM domain motif in the C-terminus of PCP_5917, the downstream partner protein of bamb_5915. Removal of the C-terminal COM domain of PCP_5917 (COM_{2C}) resulted in failed communication between PCP and bamb_5915, as monitored by an assay on the chain release of substrate-PCP. These findings encouraged us to search for other natural product pathways that

contain complimentary COM domains pairs near the termination modules that harbor chain releasing C-domains. *In vitro* chain release assays were performed with WT bamb_5915 and non-endogenous PCPs that had COM_{2C} motifs, and the inclusion of COM_{2C} resulted in productive biosynthesis that is otherwise not possible. Collectively, these findings suggest that complimentary COM motifs (COM_{2N} and COM_{2C}) could serve as universal platforms for the generation of new natural products through the direct coupling of a polyketide chain to an assortment of nucleophilic acceptor molecules. Moreover, the existence of these COM domains in mixed PKS/NRPS pathways suggests that these structural motifs may aid in future engineering of mixed PKS/NRPS that can recognize each other using the complementing COM motifs, which is a breakthrough for the ongoing effort to achieve rational combinatorial biosynthesis of PKS/NRPS pathways. We are preparing the manuscript aiming for Nature.

6.5 Chapter 5: ARX Pathway

Arixanthromycins are a group of pentangular polyphenols (PP) whose biosynthetic pathway was discovered through genome mining.[34] Arixanthromycin biosynthesis is unique in that it is a type II PKS pathway that first requires C-11 carbonyl reduction by ARX27 prior to C9-C14 cyclization by ARX19. To our knowledge, this is the first account of a C9-C14 ARO/CYC involved in a reducing PKS pathway. Furtherm the ARX gene cluster contains an array of additional tailoring enzymes that oxidize or reduce the PP scaffold regio-specifically. ARX 21, for example, is the first annotated KR that is responsible for di-ketoreduction at positions C-17 and C-19. I cloned and heterologously expressed ARX 19 and ARX21 in *E. coli*. I then solved the crystal structure of ARX 21, the first non-traditional KR that reduces not at C9 but at C-17 and C-19. A structural comparison of ARX 21 to other PKS KRs demonstrated that

ARX 21 shares the highest degree of structural similarity to that of LanV, a C15-reducing KR from the landomycin biosynthetic pathway. Both ARX 21 and LanV share a large substrate binding grove that can accommodate a large, aromatic substrate. The crystal structures therefore correlate well with the observed unique substrate specificity of ARX 21. In addition, I was able to reconstitute the activity of ARX 19 (but not ARX 21) *in vitro*. Interestingly, ARX 19 was capable of cyclizing a nonaketide in the absence of a C-9 or C-11- reducing KR, indicating that poly- β -ketone reduction is not a crucial determinant for its cyclization activity. We will submit this work to ACS Chemical Biology very shortly.

6.6 New Directions

Gaining A Clearer Glimpse of Cyclization in Action

Although we reported the co-crystal structures of WhiE bound to a set of pantetheine-linked polyketide mimics, which were useful for deciphering the substrate binding modes of ARO/CYCs, there is still room for improvement in terms of probe design. For example, both the hepta- and octa-ketide mimics contain rigid isoxazole rings that restrict the probe flexibility in the substrate pocket of WhiE; therefore, additional steric constraints may limit the binding motifs of these isoxazole probes. Clearly, a new probe design is necessary to remove the steric constraints of the isoxazole rings. If we truly desire to capture cyclization in action, we need to design and synthesize longer and more flexible polyketide mimics that are structurally and chemically similar to poly- β -ketone substrates. In collaboration with the Vanderwal lab at UC Irvine, we are currently developing the next generation of polyketide mimic probes. We are currently in the process of synthesizing a library of long-chain polyketide mimics (16-20C) that feature oxetanes moieties as carbonyl isosteres (Figure 6.1). Oxetanes have been used in

medicinal chemistry as carbonyl isosteres and are more flexible chain than isoxazole rings.[35]

These improved probes will be helpful for testing the hypothesis gleaned in Chapter 2. They will further capturing ARO/CYC in action. These new probes can also be chemoenzymatically conjugated to carrier proteins to study ARO/CYC-ACP interactions. Finally, we will solve other ARO/CYC structures bound with these isosteric probes, so that we can compare the binding motifs of different clades of ARO/CYCs.

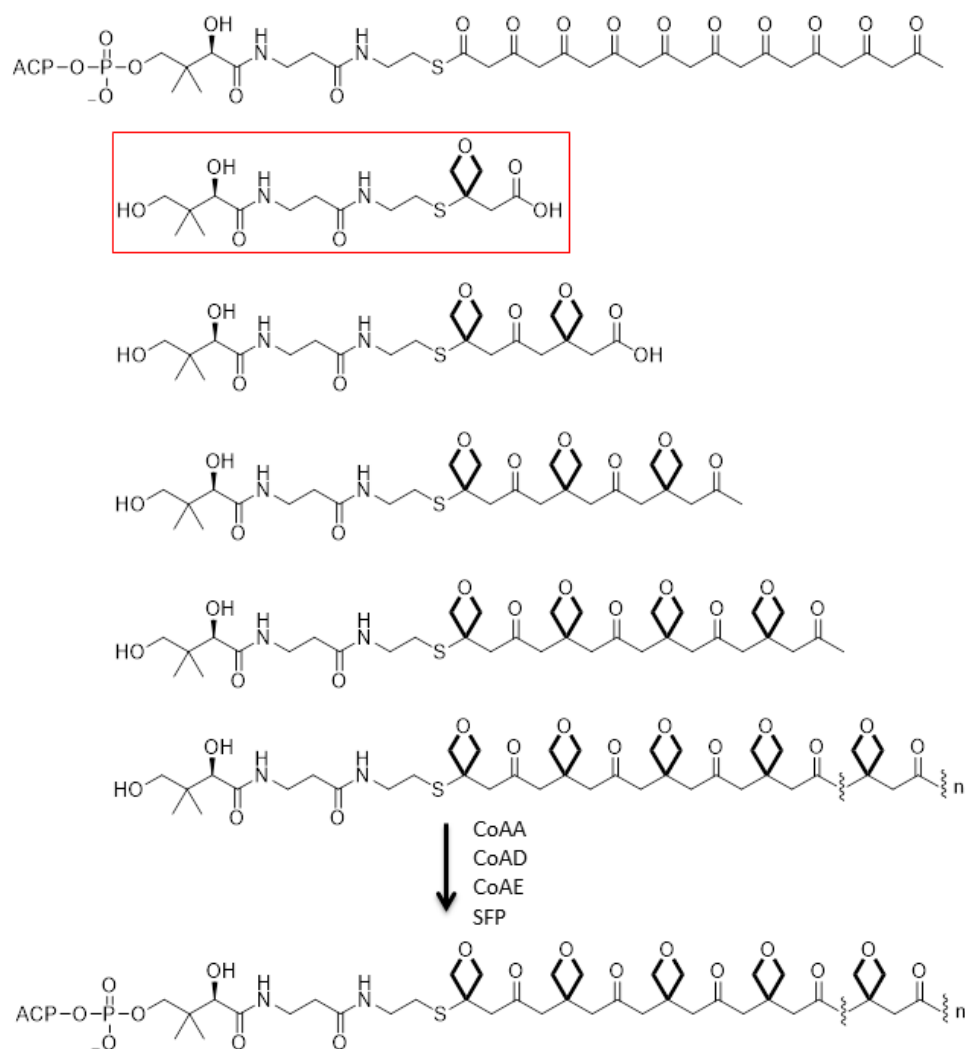


Figure 6.1. Examples of polyketide isosteres currently being developed. The diketide probe outlined in red has been synthesized.

Solving the Crystal Structure of AT₁₂ of Stambomycin PKS for Engineered Extender Unit Incorporation

In Chapter 3, we identified a novel ACCase β subunit from the stambomycin gene cluster that can generate long chain extender units *in vivo*. While unusual extender biosynthesis is interesting, the molecular basis for extender unit selection and incorporation into a growing polyketide chain by the AT domain within module 12 of the stambomycin PKS (AT₁₂) remains unknown. It would be of great interest to solve the crystal structure of (AT₁₂) from the stambomycin PKS and compare the residues that define its active site to ATs from other systems that accept small chain extender units. The information gleaned from such a structural comparison could be utilized to engineer ATs to accept and install long chain extender units into polyketide natural products. Moreover, because AT₁₂ of stambomycin PKS can incorporate alkyne or azide tagged extender units into a growing polyketide chain, we will continue to biosynthesize “clickable” natural products through structure based engineering of ATs from distinct PKS pathways, which will have a high impact on metabolomics of polyketides.

Engineering Universal Bamb₅₉₁₅ Mediated Chain Release

In Chapter 4, we reported that complimentary COM motifs are necessary for productive protein-protein interactions between PCPs and C-domains. This finding opens endless possibilities. Because the chain release activity of bamb₅₉₁₅ is independent of chain length, one could imagine engineering the chain-release of product from different natural pathways by grafting the COM_{2C} motif onto different ACPs or PCPs upstream of bamb₅₉₁₅. Essentially, this would allow for the implementation of chain termination at any given step during PKS or NRPS biosynthesis. Also, because there are other bamb₅₉₁₅ homologues that utilize different

nucleophilic acceptors for chain release, this could be a way to couple polyketide intermediates with a wide variety of differing chain releasing acceptors. Nevertheless, the structural basis for the compatibility between COM_{2C} and COM_{2N} remains a mystery. A crystal structure of bamb_5915 in complex with PCP_5917 is still needed to address how COM_{2C}-COM_{2N} recognize one another at the molecular level, which is crucial for future PKS engineering efforts.

In addition, the importance of COM_{2N} domain for the recognition between bamb_591t and PCP_5917 has yet to be fully validated. I have generated constructs of bamb_5915 standalone COM_{2N} (N-terminal subdomain alone) and bamb_5915 ΔCOM_{2N} (C-domain core alone) that could be used in *in vitro* assays or NMR HSQC titration experiments to assess if the presence of COM_{2N} is essential for productive protein-protein interactions with PCP_5917. I have expressed and purified both COM_{2N} and bamb_5915 ΔCOM_{2N}, and have sent these constructs to the Challis lab to be tested (Figure 6.2). Because the addition of excess COM_{2C} peptide inhibited productive biosynthesis between PCP_5917 and Bamb_5915, it would be interesting to attempt a complimentary experiment with COM_{2N}. If the addition of COM_{2N} does inhibit WT bamb_5915, this would strongly suggest that its presence is important for PCP binding. Chain release experiments with bamb_5915 ΔCOM_{2N} could be performed that would yield similar information. Finally, because the COM_{2N} domain expresses relatively well and is stable, I could envision HSQC titration experiments between COM_{2N} and PCP_5917 to determine the exact docking sites of PCP_5917 for COM_{2N}, if a PCP_5917/bamb_5915 crystal cannot be obtained.

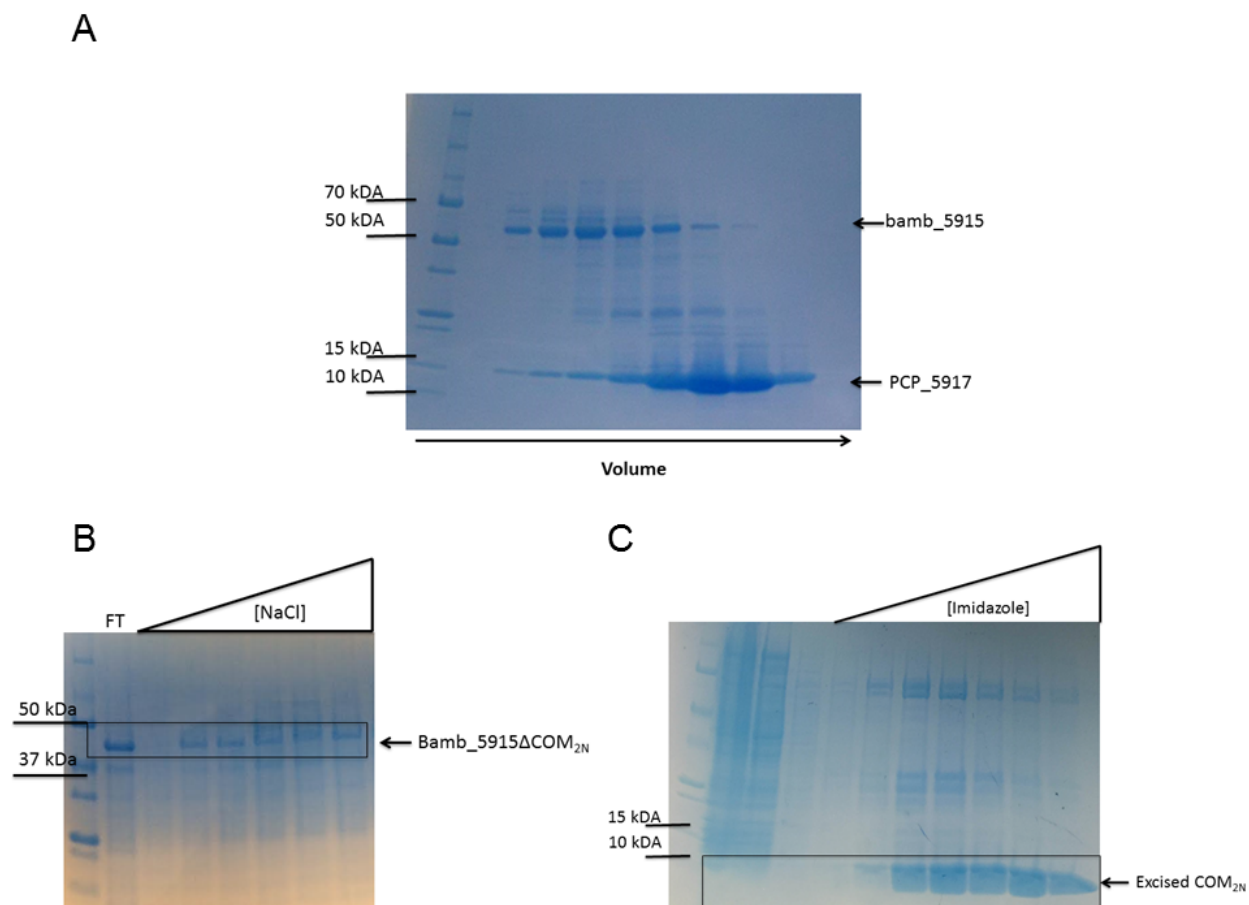


Figure 6.2. The purification of the hexanoyl-PCP_5917/bamb_5915 complex, COM_{2N} and bamb_5915ΔCOM_{2N}.
 (A) SDS PAGE of purified hexanoyl-PCP_5917/bamb_5915 complex post gel filtration. (B) SDS PAGE of COM_{2N} post nickel NTA purification. (C) SDS PAGE of bamb_5915ΔCOM_{2N}.

Structural and Functional Characterization of the Enzymes Involved in the Biosynthesis of Pentangular Polyphenols

In Chapter 5, we reported the crystal structure of ARX 21. Although this structure did identify features that were unconserved across different clades of KRs, the structural basis for C-17 and C-19 carbonyl reduction remains unclear. A co-crystal structure of ARX 21 bound to a substrate analogue or a substrate intermediate would be useful for understanding how ARX 21

achieves both the C-17 and C-19-ketoreduction. Although ARX 21 is the first tailoring KR that is predicted to be involved in the diketone-reduction of both C-17 and C-19, its activity has not been experimentally validated. In collaboration with the Brady lab, we will heterologously express the ARX PKS components in the absence of ARX 21 to determine if ARX 21 is responsible for C17 and C19 carbonyl reduction *in vivo*. If knocking out ARX 21 does indeed lead to an accumulation of unreduced products, these intermediates could be used for co-crystallization with ARX 21. With an ARX 21-substrate co-crystal structure and *in vivo* functional data, the final step would be to run molecular dynamic simulations to propose a model for the region-specificity of ARX 21.

We have a promising start with ARX 19; however, a crystal structure of this novel C9-C14 ARO/CYC is still needed to elucidate the molecular basis for its predicted C-11 dehydration and C9-C14 regiospecific cyclization during the biosynthesis of arxanthromycins. A structural comparison of ARX 19 to nonreducing C9-C14 and reducing C7-C12 ARO/CYCs would be useful to identify active site residues that are essential for the putative aromatase and cyclase activities of ARX 19. It is interesting to note that in reducing PP pathways, the removal of the C11 (ARX 27) reducing KR leads to aberrant biosynthesis *in vivo*. However, *in vitro* assays with ARX 19 in the presence of the Pks4 Min PKS demonstrated that ARX 19 can still influence the formation of C9-C14 products in the absence of a KR. In addition, ARX 27, one of the first predicted C11-reducing KRs, should be cloned and assayed with various min PKSs to assess whether it is capable of reducing the C11 position of a given type II polyketide substrate. Finally, *In vitro* assays that include both ARX 19 and ARX 27 could be performed to test the ARO/CYC activity of ARX 19 in the presence of a C11-reducing KR. With an ARX 19 or ARX 27 crystal

structures, active site mutations could be used to identify residues that are important for C11-dehydration, C9-C14 cyclization, as well as C11-reduction.

6.7 References

1. Challis, G.L. and J.H. Naismith, *Structural aspects of non-ribosomal peptide biosynthesis*. Curr Opin Struct Biol, 2004. **14**(6): p. 748-56.
2. Du, L., C. Sanchez, and B. Shen, *Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules*. Metab Eng, 2001. **3**(1): p. 78-95.
3. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review*. Nat Prod Rep, 2001. **18**(4): p. 380-416.
4. Bentley, R. and J.W. Bennett, *Constructing polyketides: from collie to combinatorial biosynthesis*. Annu Rev Microbiol, 1999. **53**: p. 411-46.
5. Finking, R. and M.A. Marahiel, *Biosynthesis of nonribosomal peptides1*. Annu Rev Microbiol, 2004. **58**: p. 453-88.
6. Koehn, F.E. and G.T. Carter, *The evolving role of natural products in drug discovery*. Nat Rev Drug Discov, 2005. **4**(3): p. 206-20.
7. Davey, S., *2015 Nobel Prize in Physiology or Medicine: Punishing parasites*. Nat Chem, 2015. **7**(12): p. 949.
8. Tobert, J.A., *Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors*. Nat Rev Drug Discov, 2003. **2**(7): p. 517-26.
9. Crawford, J.M. and C.A. Townsend, *New insights into the formation of fungal aromatic polyketides*. Nat Rev Microbiol, 2010. **8**(12): p. 879-89.
10. Barragry, T.B., *A review of the pharmacology and clinical uses of ivermectin*. Can Vet J, 1987. **28**(8): p. 512-7.
11. Tanovic, A., et al., *Crystal structure of the termination module of a nonribosomal peptide synthetase*. Science, 2008. **321**(5889): p. 659-63.
12. Haslinger, K., et al., *X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis*. Nature, 2015. **521**(7550): p. 105-9.

13. Ling, L.L., et al., *A new antibiotic kills pathogens without detectable resistance*. Nature, 2015. **517**(7535): p. 455-9.
14. Felnagle, E.A., et al., *Nonribosomal peptide synthetases involved in the production of medically relevant natural products*. Mol Pharm, 2008. **5**(2): p. 191-211.
15. Hertweck, C., *Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology*. Trends Biochem Sci, 2015. **40**(4): p. 189-99.
16. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork*. Nat Prod Rep, 2007. **24**(1): p. 162-90.
17. Keatinge-Clay, A.T., *The structures of type I polyketide synthases*. Nat Prod Rep, 2012. **29**(10): p. 1050-73.
18. Strieker, M., A. Tanovic, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics*. Curr Opin Struct Biol, 2010. **20**(2): p. 234-40.
19. Wong, F.T. and C. Khosla, *Combinatorial biosynthesis of polyketides--a perspective*. Curr Opin Chem Biol, 2012. **16**(1-2): p. 117-23.
20. Keller, U. and F. Schauwecker, *Combinatorial biosynthesis of non-ribosomal peptides*. Comb Chem High Throughput Screen, 2003. **6**(6): p. 527-40.
21. Ames, B.D., et al., *Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides*. Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5349-54.
22. Caldara-Festin, G., et al., *Structural and functional analysis of two di-domain aromatase/cyclases from type II polyketide synthases*. Proc Natl Acad Sci U S A, 2015. **112**(50): p. E6844-51.
23. Lee, M.Y., B.D. Ames, and S.C. Tsai, *Insight into the molecular basis of aromatic polyketide cyclization: crystal structure and in vitro characterization of WhiE-ORFVI*. Biochemistry, 2012. **51**(14): p. 3079-91.
24. Shakya, G., et al., *Modeling linear and cyclic PKS intermediates through atom replacement*. J Am Chem Soc, 2014. **136**(48): p. 16792-9.
25. Chan, Y.A., et al., *Biosynthesis of polyketide synthase extender units*. Nat Prod Rep, 2009. **26**(1): p. 90-114.
26. Arabolaza, A., et al., *Crystal structures and mutational analyses of acyl-CoA carboxylase beta subunit of Streptomyces coelicolor*. Biochemistry, 2010. **49**(34): p. 7367-76.

27. Wilson, M.C. and B.S. Moore, *Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity*. Nat Prod Rep, 2012. **29**(1): p. 72-86.
28. Tong, L., *Structure and function of biotin-dependent carboxylases*. Cell Mol Life Sci, 2013. **70**(5): p. 863-91.
29. Laureti, L., et al., *Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens*. Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6258-63.
30. Diacovich, L., et al., *Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity*. Biochemistry, 2004. **43**(44): p. 14027-36.
31. Mahenthiralingam, E., et al., *Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic Burkholderia ambifaria Genomic Island*. Chem Biol, 2011. **18**(5): p. 665-77.
32. Tsai, S.C., et al., *Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel*. Proc Natl Acad Sci U S A, 2001. **98**(26): p. 14808-13.
33. Richter, C.D., et al., *Multienzyme docking in hybrid megasynthetases*. Nat Chem Biol, 2008. **4**(1): p. 75-81.
34. Kang, H.S. and S.F. Brady, *Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols*. ACS Chem Biol, 2014. **9**(6): p. 1267-72.
35. Wuitschik, G., et al., *Oxetanes in drug discovery: structural and synthetic insights*. J Med Chem, 2010. **53**(8): p. 3227-46.