Title
Biosynthetic investigations of ansamycin natural products from marine-derived actinomycetes

Permalink
https://escholarship.org/uc/item/22d819pj

Author
Wilson, Micheal Christopher

Publication Date
2011

Peer reviewed|Thesis/dissertation
Biosynthetic Investigations of Ansamycin Natural Products from Marine-Derived Actinomycetes

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

In

Oceanography

by

Micheal Christopher Wilson

Committee in charge:

Bradley S. Moore, Chair
Eric Allen
Pieter Dorrestein
William H. Gerwick
Paul Jensen

2011
The Dissertation of Micheal Christopher Wilson is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair
University of California, San Diego
2011
DEDICATION

To my grandmother, Pat, for whom this would not have been possible. Thank you taking me into your home so that I could have access to a better education. Thank you for instilling in me with the confidence, determination, and passion to achieve my goals. I don’t know where I would have ended up if it were not for your love and generosity.

To Dr. Rashel V. Grindberg for all your support and love and for being my best friend. Thank you for all the adventures (past, present, and future). I love you and I am looking forward the next chapter with you.

To my parents, Mike and Stella. The time spent with you during this process has been great and will always be cherished. Thank you for your support.

To my sister, Nicole, and her children, Calli, Jason, and Magda for sharing this achievement with me and for all your love.

To all my mentors throughout my entire education. Specifically, Linda Culp, Larry Miller, John Baker, Rick Vance, Bill Hamner, Peggy Fong, Dave Jacobs, Sadaf Sehati, and Brad Moore. Thank you for making science fun, stimulating, accessible, challenging, and the greatest way to spend one’s time.

Lastly, to all the scientists that have come before me whose ideas have shaped humanity and our understanding of the Universe in which we live.
# TABLE OF CONTENTS

Signature Page ................................................................................................................. iii
Dedication ......................................................................................................................... iv
Table of Contents .............................................................................................................. v
List of Abbreviations ......................................................................................................... ix
List of Figures ................................................................................................................... xiii
List of Tables ..................................................................................................................... xvii
List of Schemes ................................................................................................................ xviii
Acknowledgements .......................................................................................................... xix
Vita ................................................................................................................................ xxi
Abstract .......................................................................................................................... xxiv

## Chapter 1: Introduction to Natural Products from Actinomycetes .......... 1

1.1: Introduction to Natural Products ..................................................................... 2

1.2: Natural Products from Actinomycetes ......................................................... 6
   1.2.1: Actinomycete Pharmacies .................................................................. 6
   1.2.2: Marine Actinomycetes: An Untapped Resource of New Drug Leads ......................................................... 6
   1.2.3: *Salinispora*: The First Obligate Marine Actinomycetes .......... 9

1.3: Polyketides and Non-Ribosomal Peptides ................................................. 17
   1.3.1: Diversity of PKs and NRPs .............................................................. 17
   1.3.2: Biosynthesis of Polyketides ............................................................ 21
   1.3.3: Biosynthesis of Non-Ribosomal Peptides ...................................... 25

1.4: Ansamycins ................................................................................................. 26
   1.4.1: History and Significance of Ansamycins .................................... 26
   1.4.2: Biosynthesis of Ansamycins .......................................................... 32
   1.4.3: Engineering Ansamycin Analogs ................................................. 37

1.5: References ..................................................................................................... 38
Chapter 2: Shared Biosynthesis of the Saliniketals and Rifamycins in
Salinispora arenicola is Controlled by the sare1259 Encoded
Cytochrome P450 .................................................................................. 57

2.1: Abstract ........................................................................................................ 58

2.2: Introduction .................................................................................................... 58
  2.2.1: Introduction to the Saliniketals and Rifamycins .................................. 58
  2.2.2: Specific Aims .......................................................................................... 61

2.3: Results .......................................................................................................... 62
  2.3.1: Chemical Analysis of S. arenicola CNS-205 and A. mediterranei S699 .................................................. 62
  2.3.2: Bioinformatic Analysis of the S. arenicola CNS-205 Genome ....................................................................... 63
  2.3.3: Inactivation of the AHBA Synthase (rifK) and the Transketolase (rif-orf15) ..................................................... 68
  2.3.4: Identification of Genes Directly Involved in Saliniketal Production .................................................................... 72
  2.3.5: Chemical Complementation of the rifK Mutant with Rifamycin Macrocyclic Intermediates .................................... 78
  2.3.6: Efforts Towards the Functional Characterization of CYP Sare1259 ................................................................ 80

2.4: Discussion ..................................................................................................... 84

2.5: Experimental Section .................................................................................. 89
  2.5.1: General Methods ..................................................................................... 89
  2.5.2: Bacterial Strains, Culture Conditions, and Extraction of Natural Products ................................................................. 89
  2.5.3: Genetic Manipulations ............................................................................ 90
  2.5.4: Isolation of 34a-deoxyrifamycin W from sare1259::aprR Mutant ........................................................................... 91
  2.5.5: Chemical Complementation of S. arenicola Mutants ......................... 91
  2.5.6: Synthesis of $^{15}$N-AHBA ......................................................................... 92
  2.5.7: Stable Isotope Labeling Experiments of Saliniketal A ...................... 93
  2.5.8: Cloning and Overexpression of Recombinant CYP Sare1259 and Putative Related Proteins .......................................... 94
  2.5.9: Enzymatic Assays of the CYP Sare1259 ................................................. 95

2.6: Appendix ...................................................................................................... 97
Chapter 3: Biosynthesis of the Ansalactams from *Streptomyces* sp. CNH-189

3.1: Abstract ..................................................................................... 113

3.2: Introduction .............................................................................. 113
3.2.1: Isolation and Structure Elucidation of Ansalactam A ........ 113
3.2.2: A New Paradigm of Polyketide Extender Unit Biosynthesis .............................................................. 117
3.2.3: Specific Aims ......................................................................... 123

3.3: Results and Discussion .............................................................. 124
3.3.1: Identification of an AHBA Synthase in *S*. sp. CNH-189 ... 124
3.3.2: Stable Isotope Feeding Studies ............................................. 127
3.3.3: Formation of the 2-Isobutyrylmalonate Polyketide Extender Unit .............................................................. 130
3.3.4: Identification of a Partial Ansalactam Biosynthetic Gene Cluster .................................................................. 134
3.3.5: The Isobutyrylmalonyl-CoA Biosynthetic Pathway .......... 142
3.3.6: Efforts Towards the Characterization of the Ansalactam Biosynthetic Gene Cluster ............................................ 145
3.3.7: Engineering *De Novo* Biosynthesis of 36-Methyl-FK506 with Ansalactam Biosynthetic Machinery .................. 147

3.4: Conclusions .............................................................................. 150

3.5: Experimental Methods ............................................................. 159
3.5.1: General Experimental Procedures ................................. 159
3.5.2: Cultivation and Extraction ............................................. 159
3.5.3: Isolation of Ansalactam A ............................................. 160
3.5.4: Reduction of Ansalactam A (4) to Yield Polyol 5 .......... 160
3.5.5: Identification of *Streptomyces* sp. Strain CNH-189 AHBA Synthase .............................................................. 161
3.5.6: Stable Isotope Incorporation Studies ............................ 159
3.5.7: Synthesis of [2,13C]4-methyl-2-pentenoic acid .............. 161
3.5.8: Bioinformatic Identification of the Partial Ansalactam Biosynthetic Gene Cluster ........................................ 162
3.5.9: Progress Towards Inactivation of Ansalactam Biosynthetic Genes .................................................. 163
3.5.10: Mutasynthesis of 36-Methyl-FK506 in Streptomyces sp.
KCTC 11064BP ΔtcsB ................................................. 164
3.6: Appendix ........................................................................................................ 166
3.7: References ........................................................................................................ 171
3.8: Acknowledgements .......................................................................................... 181

Chapter 4: Conclusions and Future Directions ...................................................... 182
4.1: General Conclusions ....................................................................................... 183
4.2: References ........................................................................................................ 187
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenylation domain</td>
</tr>
<tr>
<td>ACP</td>
<td>Acyl carrier protein</td>
</tr>
<tr>
<td>AHBA</td>
<td>3-Amino-5-hydroxybenzoic acid</td>
</tr>
<tr>
<td>AM</td>
<td>Amycolatopsis mediterranei</td>
</tr>
<tr>
<td>anss</td>
<td>Ansalactam</td>
</tr>
<tr>
<td>Apr</td>
<td>Apramycin</td>
</tr>
<tr>
<td>Ar</td>
<td>Aromatase</td>
</tr>
<tr>
<td>AT</td>
<td>Acyltransferase domain</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>C</td>
<td>Condensation domain</td>
</tr>
<tr>
<td>CCR</td>
<td>Crotonyl-CoA reductase/carboxylase</td>
</tr>
<tr>
<td>CLF</td>
<td>Chain length factor</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>Cyc</td>
<td>Cyclase domain</td>
</tr>
<tr>
<td>CYP</td>
<td>Cytochrome P450</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DH</td>
<td>Dehydratase domain</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DUF</td>
<td>Domain of unknown function</td>
</tr>
<tr>
<td>E</td>
<td>Epimerase domain</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
</tbody>
</table>
eMCoA  Ethylmalonyl-CoA
EtBr   Ethidium bromide
EtOAc  Ethyl acetate
ER     Enoylreductase domain
ESI-MS Electro spray ionization mass spectrometry
FAD    Flavin adenine dinucleotide
Fdr    Ferredoxin reductase
Fdx    Ferredoxin
FMN    Flavin mononucleotide
FPLC   Fast protein liquid chromatography
G + C  guanidine + cytosine
gDNA   Genomic DNA
H      Halogenase domain
HBDH   3-Hydroxybutyryl-CoA dehydrogenase
His    Histidine
HMBC   Heteronuclear multiple bond correlation
HPLC   High pressure liquid chromatography
HSQC   Heteronuclear multiple quantum correlation
ibMCoA Isobutyrylmalonyl-CoA
KASIII Ketoacylsynthase type III
Kb     Kilobase
kDa    Kilodalton
KR     Ketoreductase domain
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
<td>Ketosynthase domain</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>MCoA</td>
<td>Malonyl-CoA</td>
</tr>
<tr>
<td>MeCN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>MeMCoA</td>
<td>Methylmalonyl-CoA</td>
</tr>
<tr>
<td>MeOH</td>
<td>Methanol</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MT</td>
<td>Methyltransferase domain</td>
</tr>
<tr>
<td>NAD+</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NF-κB</td>
<td>Nuclear factor κB</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NRP</td>
<td>Non-ribosomal peptide</td>
</tr>
<tr>
<td>NRPS</td>
<td>Non-ribosomal peptide synthetase</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>Ox</td>
<td>Oxidation</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PK</td>
<td>Polyketide</td>
</tr>
<tr>
<td>PKS</td>
<td>Polyketide synthase</td>
</tr>
<tr>
<td>PRICE</td>
<td>Paired read iterative contig extension</td>
</tr>
<tr>
<td>Re</td>
<td>Reductase domain</td>
</tr>
<tr>
<td>rif</td>
<td>Rifamycin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>SA</td>
<td><em>Salinispora arenicola</em></td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>T</td>
<td>Thiolation domain</td>
</tr>
<tr>
<td>TE</td>
<td>Thioesterase domain</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>YMG</td>
<td>Yeast maltose glucose media</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Therapeutically relevant natural products isolated from plants</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Microbial natural products in clinical development for anti-infective (5–9) and antitumor (10–14) activity</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Phylogeny of select marine actinomycetes (blue) and their terrestrial relatives</td>
<td>8</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Structures of natural products isolated from <em>Salinispora tropica</em></td>
<td>10</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Structures of natural products isolated from <em>Salinispora arenicola</em></td>
<td>13</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Structures of natural products isolated from “<em>Salinispora pacifica</em>”</td>
<td>15</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>Diversity of polyketide structures</td>
<td>19</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>Diversity of non-ribosomal peptide structures</td>
<td>20</td>
</tr>
<tr>
<td>Figure 1.9</td>
<td>Examples of PKS systems</td>
<td>24</td>
</tr>
<tr>
<td>Figure 1.10</td>
<td>Example of NRPS system for penicillin antibiotics</td>
<td>25</td>
</tr>
<tr>
<td>Figure 1.11</td>
<td>Benzoic ansamycins isolated from actinomycetes</td>
<td>28</td>
</tr>
<tr>
<td>Figure 1.12</td>
<td>Structures of naphthalenic ansamycins isolated from actinomycetes</td>
<td>30</td>
</tr>
<tr>
<td>Figure 1.13</td>
<td>Biosynthesis of AHBA via the aminoshikimate pathway in <em>Amycolatopsis mediterranei</em> S699</td>
<td>32</td>
</tr>
<tr>
<td>Figure 1.14</td>
<td>Proposed biosynthesis of ansamycin natural products</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Structures of the saliniketals and rifamycins from <em>S. arenicola</em> CNS-205</td>
<td>60</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>HPLC-MS analysis of <em>A. mediterranei</em> S699</td>
<td>63</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Comparison of the rifamycin biosynthetic gene clusters from <em>A. mediterranei</em> S699 and <em>S. arenicola</em> CNS-205</td>
<td>65</td>
</tr>
</tbody>
</table>
Figure 2.4. HPLC-MS analysis of the inactivation of the S. arenicola AHBA synthase and chemical complementation of resulting mutant ....... 69

Figure 2.5. HPLC-MS analysis of SA-rif mutants generated in this study ....... 70

Figure 2.6. Stable isotope labeling pattern observed for 1 and proposed for 7 from the incorporation of [15N]AHBA and [U-13C3]propionate ....... 71

Figure 2.7. HPLC-MS analysis of the chemical complementation of the S. arenicola rif-orf0 mutant strain sare1245::aprR ................. 74

Figure 2.8. Inactivation of the cytochrome P450, sare1259 ..................... 76

Figure 2.9. HPLC-MS analysis of the A. mediterranei rif-orf5 mutant strain MT45025H ......................................................... 77

Figure 2.10. HPLC-MS analysis of the chemical complementation of rifK::aprR with rifamycin biosynthetic intermediates .................. 79

Figure 2.11. SDS-Page gel of Ni2+ affinity purification of recombinantly expressed Sare1259 .......................................................... 81

Figure 2.12. Depiction of Class I cytochrome P450 electron transfer system ... 82

Figure 2.13. Ni2+-NTA affinity purification of putative electron transport enzymes ................................................................. 83

Figure 2.14. Proposed mechanism of saliniketal biosynthesis facilitated by CYP Sare1259 ................................................................. 87

Figure A2.1. 13C NMR spectrum (126 mHz, MeOH-d4) of saliniketal A (1) from [U-13C]propionate feeding study ......................... 100

Figure A2.2. 13C NMR spectrum (126 mHz, MeOH-d4) of saliniketal A (1) from [U-13C]propionate feeding study (127-177 ppm) ......... 101

Figure A2.3. 1H NMR spectrum (600 mHz, DMSO-d6) of 34a-deoxy-rifamycin W (7) from sare1259::aprR mutant ......................... 102

Figure A2.4. gHSQC spectrum (600 mHz, DMSO-d6) of 34a-deoxy-rifamycin W (7) from sare1259::aprR mutant ......................... 103

Figure A2.5. gHMBC spectrum (600 mHz, DMSO-d6) of 34a-deoxy-rifamycin W (7) from sare1259::aprR mutant ......................... 104
Figure 3.1. Ansamycin natural products from actinomycete bacteria ............ 115
Figure 3.2. Structure of Ansalactam A ................................................................. 115
Figure 3.3. Ansalactams B–D isolated as minor products of *Streptomyces* sp. CNH-189 ................................................................. 116
Figure 3.4. Biosynthetic Pathway of ethylmalonyl-CoA ....................................... 118
Figure 3.5. Natural products that incorporate ethylmalonyl-CoA into their structures .................................................................................. 119
Figure 3.6. Natural products that incorporate rare or unique polyketide extender units ................................................................................. 121
Figure 3.7. Identification of an AHBA synthase from *S. sp.* CNH-189 ............ 125
Figure 3.8. $^{13}$C NMR spectrum (125 MHz) of 5 from [1-$^{13}$C]propionate feeding study in methanol-$d_4$ ........................................................................................................ 128
Figure 3.9. $^{13}$C NMR spectrum (125 MHz) of 5 from [1,2-$^{13}$C$_2$]acetate feeding study in methanol-$d_4$ ........................................................................................................ 129
Figure 3.10. $^{13}$C NMR spectrum (125 MHz) of 5 from [1-$^{13}$C]isobutyrate feeding study in methanol-$d_4$ ........................................................................................................ 131
Figure 3.11. $^{13}$C NMR spectrum (125 MHz) of 5 from [2-$^{13}$C]4-Me-2-pentenoic acid feeding study in methanol-$d_4$ ........................................................ 133
Figure 3.12. Gene organization of the ansalactam biosynthetic gene cluster ................................................................................................. 136
Figure 3.13. Biosynthesis of AHBA by the putative ansalactam gene cluster ................................................................................................. 138
Figure 3.14. Nucleotide alignment of the AHBA biosynthetic genes ............... 139
Figure 3.15. Orthologous gene neighborhoods of the putative isobutyrylmalonyl-CoA biosynthetic gene clusters in *Streptomyces* ......................... 142
Figure 3.16. Neighbor-joining tree of CCRs from characterized biosynthetic pathways and uncharacterized genome sequences ......................... 144
Figure 3.17. PCR amplification of genes spanning the 20 Kb partial ansalactam cluster located on pCC2-based fosmids P13C5 and P13F6 used for gene inactivation experiments ........................................... 146

Figure 3.18. Proposed biosynthesis of the allylmalonyl-CoA extender unit of FK506 and generation of FK506 analogs by mutasynthesis ...... 148

Figure 3.19. Proposed scheme for de novo synthesis of 36-methyl-FK506 in the Streptomyces sp. ΔtcsB mutant ............................................... 149

Figure 3.20. Summary of 13C-labeled precursor studies ........................................... 151

Figure 3.21. Proposed biosynthesis of ansalactam A (4) and gene organization of the polyketide region of the ansalactam biosynthetic gene cluster ................................................ 154

Figure 3.22. Structures of divergolides A–D (9–12) and germicidins F (13) and G (14) isolated from Streptomyces sp. HK10576 .............. 156

Figure 3.23. Proposed biosynthesis of isobutyrylmalonyl-CoA in Streptomyces sp. CNH-189 ................................................ 158

Figure A3.1. 13C NMR spectrum (125 MHz) of 5 not enriched from [1-13C]leucine feeding study in methanol-d4 ........................................ 168

Figure A3.2. 13C NMR spectrum (125 MHz) of 5 from [1-13C]isobutyrate feeding study in methanol-d4 (expanded) ............................. 169

Figure A3.3. 13C NMR spectrum (125 MHz) of 5 from [2-13C]4-Me-2-pentenoic acid feeding study in methanol-d4 (expanded) ............. 170
LIST OF TABLES

Table 1.1. Natural products from marine-derived *Streptomyces* sp. ........... 8
Table 1.2. Secondary metabolite gene clusters from *S. arenicola* 
CNS-205 ......................................................................................... 16
Table 1.3. Secondary metabolite gene clusters from *S. tropica* 
CNB-440 ......................................................................................... 17
Table 2.1. Summary of SA-rif cluster and chemotypes of genetically 
mutated strains............................................................................. 66
Table A2.1. Primers synthesized for gene inactivation experiments of 
the SA-rif gene cluster................................................................. 97
Table A2.2. Primers synthesized for verification of double crossover 
mutants of the SA-rif gene cluster................................................ 98
Table A2.3. Primers used for expression of recombinant CYP Sare1259 
and putative associated electron transport enzymes .................... 99
Table 3.1. Proposed functions of ORFs in the ansalactam (4) 
biosynthetic gene cluster................................................................. 137
Table 3.2. Predictive residues for acyltransferase substrate specificity 
of the putative ansalactam PKS ORFs........................................... 141
Table A3.1. Primers synthesized for screening the CNH-189 fosmid 
library ............................................................................................. 166
Table A3.2. Primers synthesized for gene inactivation experiments in 
CNH-189 .......................................................................................... 166
Table A3.3. Primers synthesized for heterologous expression 
of the isobutyrylmalonyl-CoA biosynthetic gene cassette 
in the ΔtcsB mutant ......................................................................... 167
LIST OF SCHEMES

**Scheme 3.1.** Preparation of polyol 5 from ansalactam A (4)............................ 127
ACKNOWLEDGEMENTS

I would like to sincerely thank everyone who has been involved, either directly or indirectly, with my development as a scientist. Specifically, I need to acknowledge the following people:

I would like to acknowledge Professor Bradley S. Moore for his excellent guidance and mentoring of my doctoral training. Thank you for your constant generosity with your time and for giving me the opportunity to initially volunteer in your lab. Thank you for always being available to discuss projects, knowing when to let me struggle on my own, and when to provide guidance. Also, thank you for expressing the importance of balancing hard work and dedication to science while still enjoying all that life has to offer. Honestly, no one could ask for a better mentor.

I need to acknowledge the entire Moore lab, both past and present for not only helping me to develop as a scientist but also for being my friends. Especially, Dr. Michelle Moffitt who trained me as a volunteer and convinced Brad to accept me as a Ph.D. student. I thank Dr. Alessandra Eustáquio and Dr. Andrew Schultz for training me to manipulate Salinispora arenicola. Dr. Tobias A. M. Gulder helped to hone my chemistry skills as well as has become one of my best friends. So many great times were had with the entire Moore lab over the years that it truly is hard to move on.

I would also like to thank the rest of my dissertation committee, Bill Gerwick, Peter Dorrestein, Paul Jensen, and Eric Allen for all the stimulating and
challenging discussions over the years. Especially Paul Jensen, because without his collection trips I would not have had such interesting organisms to work with.

I also would like to thank to hosts of my favorite podcasts; Skeptics Guide to the Universe, Skeptoid, Science Talk, and NPR Science Friday for promoting science and critical thinking to the masses. During most of my time spent in the lab, I was listening to these great shows and they have been instrumental to my development as a scientist. Thank you for combating scientific illiteracy, pseudoscience, religious nonsense, magical thinking, and other forms of idiocy. Keep up the good work.

Chapter 2, in full, is a reprint of the material as it appears in Shared Biosynthesis of the Saliniketals and Rifamycins in *Salinispora arenicola* is Controlled by the *sare1259*-Encoded Cytochrome P450 (2010). Wilson, Micheal C.; Gluder, Tobias A. M.; Mahmud, Taifo; Moore, Bradley S., Journal of the American Chemical Society, 132, 12757-12765. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is a reprint of the material as it appears in Structure and Biosynthesis of the Marine Streptomycete Ansamycin Ansalactam A and Its Distinctive Branched Chain Polyketide Extender Unit (2011) Wilson, Micheal C.; Nam, Sang-Jip; Gulder, Tobias A.M.; Kauffman, Christopher A.; Jensen, Paul R.; Fenical, William; Moore, Bradley S., Journal of the American Chemical Society, 133, 1971-1977. The dissertation author was the primary investigator and author of this paper.
VITA

2006-2011  **Doctor of Philosophy**, Oceanography, Scripps Institution of Oceanography, University of California, San Diego

2001-2004  **Bachelor of Science**, Biochemistry and Marine Biology, University of California, Los Angeles

1999-2001 **Major in Chemistry**, IGETC completion for transfer, Moorpark College

PUBLICATIONS


PRESENTATIONS

2010 M. C. Wilson, T. A. M. Gulder, T. Mahmud, and B. S. Moore. “Insights into the Biosynthesis of the Saliniketals by the Salinispora arenicola Rifamycin Cluster.” General Meeting of the American Society for Microbiology, San Diego, CA, Poster


2009 M. C. Wilson, T. Mahmud, and B. S. Moore, "Biosynthesis of the Saliniketals by the Salinispora arenicola Rifamycin Gene Cluster" 50th Anniversary Meeting of the American Society of Pharmacognosy, Honolulu, HI, Poster

2006 M. C. Wilson, D. Pinkard, and J. Butler. “Habitat Preference and Distribution of the Sebastomus Subgenus of Rockfish Sebastes spp. in the Southern California Bight.” 14th Western Groundfish Conference, Newport, OR, Poster

FELLOWSHIPS AND AWARDS

2011 Life Science Scholarship for “Medicinal Chemistry Intensive” Program, UCSD Extension

2008 National Science Foundation Graduate Research Fellowship, Honorable Mention

2007 National Science Foundation Graduate Research Fellowship, Honorable Mention
2006  T. R. and Edith Folsom Graduate Fellowship, Scripps Institution of Oceanography
2001  Fred Kavli Scholarship for Excellence in Biology, Moorpark College

---

RESEARCH EXPERIENCE

2006-2011  Graduate Student Research Assistant, Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, Laboratory of Prof. Bradley S. Moore

2005-2006  Volunteer Researcher, Scripps Institution of Oceanography, Dr. Bradley S. Moore

2004  Independent Undergraduate Research Project, UCLA Marine Biology Field Quarter, Moorea, Tahiti

2003  Undergraduate Research Assistant, University of California, Los Angeles, Laboratory of Prof. Joan Valentine

---

PROFESSIONAL EXPERIENCE

2005-2006  Biological Technician II, NOAA/NMFS/Southwest Fisheries Science Center, Aquatic Farms, La Jolla, CA

2004-2005  Quality Analytical Laboratory Analyst I, ACT•1 on-site at Amgen Inc., Thousand Oaks, CA

---

ACADEMIC TEACHING EXPERIENCE

2011  Graduate Teaching Assistant, Recombinant DNA Techniques Lab, University of California, San Diego, Dr. Ella Tour

2000  Supplemental Instructor, Biology and Life Sciences, Moorpark College, Dr. Larry O. Miller
ABSTRACT OF THE DISSERTATION

Biosynthetic Investigations of Ansamycin Natural Products from Marine-Derived Actinomycetes

by

Micheal Christopher Wilson

Doctor of Philosophy in Oceanography

University of California, San Diego, 2011

Professor Bradley S. Moore, Chair

Ansamycin polyketides from actinobacteria include the potent antibiotic and anticancer agents rifamycin SV, ansamitocin P-3, and geldanamycin. These natural product macrolactams are characterized by an mC\textsubscript{7}N structural unit derived from the aromatic acid 3-amino-5-hydroxybenzoate, which is carboxy extended by multimodular polyketide synthases utilizing primarily acetate and propionate building blocks prior to macrolactam cyclization. Herein, I report a multidisciplinary investigation of the biosynthesis of two structurally fascinating natural products produced by ansamycin biosynthetic machinery.

Saliniketals A and B are unusual bicyclic polyketides from the obligate marine actinomycete Salinispora arenicola that inhibit the cancer
chemopreventative drug target ornithine decarboxylase. The saliniketals are noteworthy because of their structural similarity to the ansa portion of the rifamycin antibiotics, which co-occur in fermentation broth. Analysis of the S. arenicola genome identified a 106 kb rif gene cluster consisting of all the PKS and tailoring genes necessary for the biosynthesis of rifamycin B. No additional type I PKS with domain architecture consistent with saliniketal assembly could be identified, which suggested the saliniketals are products of the S. arenicola rif locus. Using stable isotope incorporation studies, PCR targeted mutagenesis, and chemical complementation with rifamycin pathway intermediates I have shown that the saliniketals and rifamycins share a single biosynthetic locus and a cytochrome P450 facilitates the divergence of the two compound classes.

Additionally, in 2009 the novel ansamycin ansalactam A was isolated by the Fenical laboratory from an extract of a marine sediment-derived actinomycete of the genus Streptomyces. Unique from its predecessors, ansalactam A possesses a spiro-fused γ-lactam moiety and an unusual isobutyrylmalonate polyketide extender unit. Using $^{13}$C-labeled precursors, I have shown that the unique branched chained polyketide extender unit of ansalactam A originates from the condensation of isobutyrate and acetate precursors. Further bioinformatic analysis of the partial ansalactam biosynthetic gene cluster identified a putative three gene cassette (KASII-HBDH-CCR) for the biosynthesis of isobutyrylmalonyl-CoA. Efforts towards the functional characterization of the novel polyketide biosynthetic genes are also discussed within.
Chapter 1:

Introduction to Natural Products From Actinomycetes
1.1: Introduction to Natural Products

Small molecules of pharmaceutical importance are ubiquitous in nature. These so-called natural products (also known as secondary metabolites) are biosynthesized by all three domains of life. Unlike primary metabolites (e.g. fatty acids, amino acids, and carbohydrates), secondary metabolites are generally organism specific and believed to play a role in ecological adaptation. Although the natural function of most of these metabolites are not well understood, many have been shown to be ecologically important for chemical defense, intra– and intercellular signaling, quorum sensing, nutrient sequestration, and many more biological processes.\(^1\)\(^2\) In addition to their innate functions, human’s have been exploiting natural products for their medicinal properties for longer than recorded history, mainly in the form of herbal remedies from plants. For instance, medicinal plants such as *Centaurea solstitialis, Ephedra altissima,* and *Achillea* sp. were found with Neanderthal remains dating back to 60,000 B.C.\(^3\) In recorded history, the use of medicinal plants and animals date back to the ancient Greeks, Egyptians, Mesopotamians, and Chinese.\(^4\)\(^5\) Herbal teas and poultices from yarrow, willow, henbane, and coca have been used as analgesics in many cultures around the world.\(^5\)\(^6\)

With birth of modern pharmacognosy\(^7\) and advances in chemical separation and identification techniques, we developed the ability to purify the active compounds from many of these sources. Therapeutically important compounds isolated from plants include the precursor to aspirin, salicylic acid (1)\(^8\) the antimalarial agent quinine (2)\(^9\) the potent analgesic morphine (3)\(^10\) and
the antitumor drug paclitaxel (4, Figure 1.1). With more than 200,000 biologically active compounds isolated from plants, they are the leading source of naturally derived human therapeutics.

Figure 1.1. Therapeutically relevant natural products isolated from plants.

In the modern era, microbial sources including bacteria, fungi, and microalgae have proven to be an especially prolific source of pharmaceutically relevant natural products with anti-infective agents alone yielding over 66 billion dollars in U.S. pharmaceutical markets in 2007. Since Alexander Fleming’s discovery of penicillin from the mold *Penicillium notatum*, microbial natural products have proven to be an invaluable resource for treating infectious diseases, cancers, inflammation, neurodegenerative diseases, and the list goes on. Microbial anti-infective agents including antibacterials, antivirals, and antifungals prevent millions of human deaths every year and are considered a
major factor in prolonging the human life expectancy by almost 30 years since 1900.\textsuperscript{14} However, as our technology to combat pathogenic organisms evolves, so too does antibiotic resistance.\textsuperscript{16} Though interest in natural products declined during the 1990’s as pharmaceutical companies shifted towards combinatorial synthetic chemistry, they continue to be our best weapon against drug resistant pathogens. This is evident considering approximately 75% of approved drugs between 1981 and 2002 originated from natural products.\textsuperscript{17} In 2007, there were 67 microbially-derived natural products in clinical development as anti-infective, anticancer, neurological, cardiovascular, and immunological agents.\textsuperscript{18} Figure 1.2 depicts some of the chemical diversity of microbial natural products currently being developed as anti-infective (5–9) and anticancer (10–14) therapeutics.

As some of the oldest and most ubiquitous organisms on Earth, bacteria are capable of producing secondary metabolites with seemingly endless variations of structural complexity. Microbial natural products include both ribosomally and non-ribosomally derived peptides, modified amino acids and nucleosides, polyketides, terpenoids, alkaloids, glycosides, and many mixed structural classes. It is proposed that this variability is a major contributing factor for microbial adaptation, allowing them thrive in nearly every ecological niche discovered to date.\textsuperscript{1}
Figure 1.2. Microbial natural products in clinical development for anti-infective (5–9) and antitumor (10–14) activity (adapted from Lam 2007).
1.2: Natural Products from Actinomycetes

1.2.1: Actinomycete Pharmacies

Among the natural product producing bacteria, no group has had a larger impact on human health than those belonging to the order Actinomycetales. These actinobacteria are morphologically diverse yet often filamentous, Gram-positive, and have a high G+C content. The Actinomycetales are also the most highly prolific producers of pharmaceutically relevant natural products, especially in the realm of antibiotics. Since the discovery of actinomycin (54, see Figure 1.8) by Waksman and Woodruff in 1940, there have been more 10,000 bioactive metabolites isolated from actinomycetes with approximately 120 in practical use for human chemotherapy, agriculture, or research. Of the 110 genera encompassed by the order Actinomycetales, none have been shown to be more diverse or biosynthetically fruitful than the *Streptomyces*. With more than 900 described species that produce 80% of all actinomycete bioactive compounds and responsible for the production of such clinically important antibiotics as streptomycin, vancomycin, erythromycin, and tetracycline, the antifungal amphotericin B, anticancer agents mitomycin and bleomycin, and the immunosuppressive compound rapamycin, the therapeutic potential of *Streptomyces* sp. abounds.

1.2.2: Marine Actinomycetes: An Untapped Resource of New Drug Leads

Famous for being one of the most isolated genera of terrestrial soil-dwelling actinomycetes, *Streptomyces* actually inhabit a wide range of biological niches from hyper-arid deserts to deep ocean sediments. An
increasing number of *Streptomyces* species are now known to also thrive as symbionts of plants,\(^{32-33}\) insects,\(^ {34}\) and marine invertebrates.\(^ {35-36}\) As terrestrial habitats have become over exploited and the frequency of discovering novel chemical diversity has decreased, it has become necessary to explore more remote habitats for new chemical entities.\(^ {21,37-39}\) Representing approximately 70% of the Earth’s surface, the world’s oceans are a logical habitat to search for novel chemistry. Actinomycete natural product producers have been known from marine environments since at least the late 1960’s.\(^ {40}\) Marine isolates have been recovered from the genera *Actinomadura*, *Actinosynnema*, *Amycolatopsis*, *Arthrobacter*, *Blastococcus*, *Brachybacterium*, *Corynebacterium*, *Dietzia*, *Frankia*, *Frigoribacterium*, *Geodermatophilus*, *Gordonia*, *Kitasatospora*, *Micromonospora*, *Micrococcus*, *Microbacterium*, *Mycobacterium*, *Nocardioides*, *Nocardiopsis*, *Nonomurea*, *Psuedonocardia*, *Rhodococcus*, *Saccharopolyspora*, *Salinispora*, *Serinicoccus*, *Solwaraspora*, *Streptomyces*, *Streptosporangium*, *Tsukamurella*, *Turicella*, *Verrucosispora*, and *Williamsia*.\(^ {21,41-43}\) However, it was long thought that marine-derived *Streptomyces* (and other actinobacteria) were simply washed into the sea from terrestrial environments.\(^ {41}\) It has since been shown that isolates from several actinomycete genera are indigenous to the marine environment (Figure 1.3).\(^ {37,39,41,44}\) Of these, *Streptomyces* species continue to shine as the most abundant and prolific natural product producers, however, this may be due to the fact that they are much more amenable to cultivation.\(^ {37}\) Some notable natural products from marine-derived *Streptomyces* sp. are listed in Table 1.1.
Figure 1.3. Phylogeny of select marine actinomycetes (blue) and their terrestrial relatives. Neighbor-joining tree was constructed in Geneious\textsuperscript{45} with 1000 bootstrap replications.

Table 1.1. Natural products from marine-derived \textit{Streptomyces} sp.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Chemical Class</th>
<th>Biological Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Allyloxyphenol\textsuperscript{46}</td>
<td>phenolics</td>
<td>antimicrobial, antioxidant</td>
</tr>
<tr>
<td>Albidopyrone\textsuperscript{47}</td>
<td>α-pyrene</td>
<td>cytotoxic</td>
</tr>
<tr>
<td>Aureoverticillactam\textsuperscript{48}</td>
<td>mactolactam</td>
<td>cytotoxic</td>
</tr>
<tr>
<td>Caboxamycin\textsuperscript{49}</td>
<td>benzoazole</td>
<td>antibacterial, cytotoxic</td>
</tr>
<tr>
<td>Cyclomarins\textsuperscript{50}</td>
<td>cyclic peptides</td>
<td>antiinflammatory</td>
</tr>
<tr>
<td>Enterocin\textsuperscript{51}</td>
<td>polyketide</td>
<td>bacteriostatic</td>
</tr>
<tr>
<td>Essramycin\textsuperscript{52}</td>
<td>thiazolopyrimidine</td>
<td>antibacterial</td>
</tr>
<tr>
<td>Mansouramycins\textsuperscript{53}</td>
<td>isoquinolinequinones</td>
<td>cytotoxic</td>
</tr>
<tr>
<td>Marinopyrroles\textsuperscript{54}</td>
<td>bispyrrole</td>
<td>antibacterial, cytotoxic</td>
</tr>
<tr>
<td>ML-449\textsuperscript{55}</td>
<td>macrolactam</td>
<td>cytotoxic</td>
</tr>
<tr>
<td>Salinamides\textsuperscript{56}</td>
<td>bicyclic depsipeptides</td>
<td>antiinflammatory</td>
</tr>
</tbody>
</table>

\textsuperscript{*}adapted from Zotchev 2011\textsuperscript{37}
1.2.3: *Salinispora*: The First Obligate Marine Actinomycetes

Formally described in 2005, the *Salinispora* are the first actinomycetes to be classified as obligate marine. There are two described species, *S. arenicola* and *S. tropica*, and a third, “*S. pacifica*”, awaiting formal classification.\(^{57-58}\) Since their initial isolation from the Bahamas in 1991,\(^ {59}\) the *Salinispora* have been recovered from marine sediments\(^ {58,60-61}\) and sessile marine invertebrates\(^ {35,62}\) throughout the tropics and sub-tropics but exhibit species-specific biogeographical distributions. *S. tropica* has the most limited range having only been isolated from the Bahamas, whereas “*S. pacifica*” has been recovered from Guam, Palau, the Red Sea, and Fiji\(^ {63}\) and *S. arenicola* has an apparent global distribution.\(^ {58}\)

Exploitation of *Salinispora* sp. for biologically active natural products has yielded a wealth of both novel and known chemical entities (Figures 1.4–1.6). One of the most promising drug leads from the *Salinispora* is the potent 20S proteosome inhibitor salinisporamide A (15)\(^ {64}\) from *S. tropica*. Currently in Phase-I clinical trials for the treatment of multiple myeloma, 15 has also recently been shown to have potent antimalarial activity.\(^ {65}\) Additionally, *S. tropica* produces the novel enediyene product, sporolide A (16).\(^ {66}\) Both the immunosuppressant lymphostin (17), previously isolated from *Streptomyces* sp. KY11783,\(^ {67}\) and the polyketide salinilactam A (18) were predicted through bioinformatic genome mining, leading to their eventual isolation.\(^ {68}\)
An even more prolific source of natural products is the widely distributed *S. arenicola*. In addition to also producing 17, *S. arenicola* produces several novel and known polyketides and non-ribosomal peptides. The antibiotic rifamycins (19),69-70 antitumor staurosporines (20),71 and anti-inflammatory cyclomarins (21)50,72 were previously isolated from non- *Salinispora* actinomycetes (Figure 1.5). However, their respective pathway shunt products saliniketal (23, see Chapter 2)70,73 and cyclomarazine (22)72,74 were unknown prior to their isolation from *S. arenicola* (Figure 1.5). While investigating the biosynthesis of cyclomarin, Schultz et al. isolated the cyclomarin precursor N-(1,1-dimethyl-1-allyl)-tryptophan (24), learning that the novel prenylation of tryptophan occurs prior to incorporation into the peptide structure. Also,
CNR059.325 (29) and salinisporamycin$^{75}$ (not shown, see Chapter 2) are predicted to be additional products of the rifamycin biosynthetic pathway based on the naphthoquinone core and similarity to known rifamycin shunt metabolites.$^{76}$ The macrolide arenicolide (25)$^{77}$ and the aromatic polyketides saliniquinone (26)$^{78}$ and arenimycin (28)$^{79}$ all exhibit antitumor activity against the human colon adenocarcinoma cell line HCT-116. The arenamide (27)$^{80}$ lipopeptides exhibit modest cancer chemopreventative and anti-inflammatory activity with no significant antibiotic activity, unlike similar lipopeptides of the daptomycin$^{81}$ class.

Akin to the two formally described species, “S. pacifica” produces several notable natural products. For example, lomaiviticin A (31) was originally isolated from the ascidian-associated “S. pacifica” (Figure 1.6). “S. pacifica” also shares several metabolite families with S. tropica including salinilactam (18) and a novel salinosporamide analog, sallinosporamide K (32), which was originally predicted by mining the draft genome of S. “pacifica” CNT-133.$^{63}$ The salinosporamide gene cluster in S. “pacifica” CNT-133 lacks the genes necessary for the biosynthesis of the chloroethylmalonyl-CoA PKS extender unit of salinosporamide A (15), but instead incorporates a malonyl-CoA to give the novel analog 32.$^{63}$ The pacificanones (33), salinpyrones (34), and related polyketides (35–36) are likely biosynthesized by a single PKS cluster in which promiscuous AT domains and module skipping may play a role in diversification of the polyketide products.$^{82}$ Similarly, the tylosin analogs (37–39) are polyketide macrolides with at least one promiscuous AT domain that selects for either
methylmalonyl-CoA or ethylmalonyl-CoA and then undergo various post-PKS modifications (unpublished). Cyanosporaside A (40) is an enediyne-derived polyketide with modest cytotoxicity against HCT-116 cells. 83

In addition to the wealth of natural products currently known from the Salinispora it has become evident through genome sequencing that their biosynthetic potential far exceeds the number of known compounds. 68, 84 A comprehensive investigation into the completed genomes of S. arenicola CNS-205 and S. tropica CNB-440 revealed at least 49 putative secondary metabolite biosynthetic gene clusters between the two species (Tables 1.2 and 1.3). Approximately 10% of the total DNA of these two species is dedicated to the biosynthesis of secondary metabolites exceeding some of the most prolific and well studied terrestrial actinomycetes. 68 The draft genome of “S. pacifica” CNT-133 has already led to the isolation of the novel salinosporamide analog (32) and partial biosynthetic gene clusters for 16–18, as well as several other pathways homologous to S. arenicola and S. tropica cryptic metabolites (unpublished). Several avenues of research including heterologous expression and genetic manipulation of regulatory genes are currently being explored in the Moore laboratory in an effort to access some of these cryptic natural products.
Figure 1.5. Structures of natural products isolated from *Salinispora arenicola*. 
Figure 1.6. Structures of natural products isolated from "Salinispora pacifica".
<table>
<thead>
<tr>
<th>No.</th>
<th>Cluster name</th>
<th>Biosynthetic class</th>
<th>Product</th>
<th>Biological activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SA nrps1</td>
<td>non-ribosomal peptide</td>
<td>pentapeptide</td>
<td>N/D</td>
</tr>
<tr>
<td>2</td>
<td>SA pksnrps1</td>
<td>polyketide/non-ribosomal peptide</td>
<td>N/D</td>
<td>N/D</td>
</tr>
<tr>
<td>3</td>
<td>SA pks1A</td>
<td>polyketide</td>
<td>9-membered enediyne unit</td>
<td>cytotoxin</td>
</tr>
<tr>
<td>4</td>
<td>SA misc1</td>
<td>aminoacyl tRNA synthetase-derived</td>
<td>amino acid conjugate</td>
<td>N/D</td>
</tr>
<tr>
<td>5</td>
<td>SA bac1</td>
<td>ribosomal peptide</td>
<td>class I bacteriocin (lantibiotic)</td>
<td>antimicrobial</td>
</tr>
<tr>
<td>6</td>
<td>SA pks2</td>
<td>polyketide</td>
<td>N/D</td>
<td>N/D</td>
</tr>
<tr>
<td>7</td>
<td>SA rif</td>
<td>polyketide</td>
<td>rifamycin</td>
<td>antibiotic</td>
</tr>
<tr>
<td>8</td>
<td>SA terp1</td>
<td>terpenoid</td>
<td>diterpene</td>
<td>N/D</td>
</tr>
<tr>
<td>9</td>
<td>SA pks3A</td>
<td>polyketide</td>
<td>10-membered enediyne</td>
<td>cytotoxin/ DNA</td>
</tr>
<tr>
<td>10</td>
<td>SA sid1b</td>
<td>non-ribosomal peptide</td>
<td>yersiniabactin-related</td>
<td>siderophore</td>
</tr>
<tr>
<td>11</td>
<td>SA pks1B</td>
<td>polyketide-associated</td>
<td>modified tyrosine and deoxy sugars</td>
<td>cytotoxin</td>
</tr>
<tr>
<td>12</td>
<td>SA misc2</td>
<td>aminoacyl tRNA synthetase-derived</td>
<td>amino acid conjugate</td>
<td>N/D</td>
</tr>
<tr>
<td>13</td>
<td>SA pks3B</td>
<td>polyketide-related</td>
<td>aryltetrasaccharide unit</td>
<td>cytotoxin/ DNA</td>
</tr>
<tr>
<td>14</td>
<td>SA sta</td>
<td>indolocarbazole</td>
<td>staurosporine</td>
<td>cytotoxin</td>
</tr>
<tr>
<td>15</td>
<td>SA pksnrps2</td>
<td>polyketide/non-ribosomal peptide</td>
<td>N/D</td>
<td>N/D</td>
</tr>
<tr>
<td>16</td>
<td>SA amc</td>
<td>carbohydrate</td>
<td>aminocyclitol</td>
<td>N/D</td>
</tr>
<tr>
<td>17</td>
<td>SA bac2</td>
<td>ribosomal peptide</td>
<td>class I bacteriocin (non-lantibiotic)</td>
<td>antimicrobial</td>
</tr>
<tr>
<td>18</td>
<td>SA pks4</td>
<td>polypeptide</td>
<td>aromatic polypeptide</td>
<td>N/D</td>
</tr>
<tr>
<td>19</td>
<td>SA des</td>
<td>hydroxamate</td>
<td>desferrioxamine</td>
<td>siderophore</td>
</tr>
<tr>
<td>20</td>
<td>SA nrps2</td>
<td>non-ribosomal peptide</td>
<td>tetrapeptide</td>
<td>N/D</td>
</tr>
<tr>
<td>21</td>
<td>SA nrps3</td>
<td>non-ribosomal peptide</td>
<td>dipeptide</td>
<td>N/D</td>
</tr>
<tr>
<td>22</td>
<td>SA pks5</td>
<td>polypeptide</td>
<td>macrolide</td>
<td>N/D</td>
</tr>
<tr>
<td>23</td>
<td>SA bac3</td>
<td>ribosomal peptide</td>
<td>class I bacteriocin (non-lantibiotic)</td>
<td>antimicrobial</td>
</tr>
<tr>
<td>24</td>
<td>SA lym</td>
<td>polypeptide</td>
<td>lymphostatin</td>
<td>immunosuppressant</td>
</tr>
<tr>
<td>25</td>
<td>SA terp2</td>
<td>terpenoid</td>
<td>carotenoid pigment</td>
<td>antioxidant</td>
</tr>
<tr>
<td>26</td>
<td>SA cym</td>
<td>non-ribosomal peptide</td>
<td>cyclomarin</td>
<td>anti-inflammatory</td>
</tr>
<tr>
<td>27</td>
<td>SA pks6</td>
<td>polyketide</td>
<td>phenolic lipids</td>
<td>cell wall lipid</td>
</tr>
<tr>
<td>28</td>
<td>SA nrps4</td>
<td>non-ribosomal peptide</td>
<td>tetrapeptide</td>
<td>N/D</td>
</tr>
<tr>
<td>29</td>
<td>SA terp3</td>
<td>terpenoid</td>
<td>carotenoid pigment</td>
<td>antioxidant</td>
</tr>
<tr>
<td>30</td>
<td>SA pks1C</td>
<td>polypeptide</td>
<td>naphtholic acid unit</td>
<td>cytotoxin/ DNA</td>
</tr>
</tbody>
</table>

*bold corresponds to isolated metabolites (modified from Penn et al. 2009)*

---

**Table 1.2. Secondary metabolite gene clusters from S. arenicola CNS-205**
Table 1.3. Secondary metabolite gene clusters from *S. tropica* CNB-440

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Class</th>
<th>Product</th>
<th>Biological activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ST pks1</td>
<td>polyketide</td>
<td>10-membered enediyne</td>
<td>cytotoxin</td>
</tr>
<tr>
<td>2</td>
<td>ST nrps1</td>
<td>non-ribosomal peptide</td>
<td>dipeptide</td>
<td>N/D</td>
</tr>
<tr>
<td>3</td>
<td>ST sal</td>
<td>polyketide/non-ribosomal peptide</td>
<td>salinosporamide</td>
<td>cytotoxin</td>
</tr>
<tr>
<td>4</td>
<td>ST pks2</td>
<td>polyketide</td>
<td>glycosylated decaketide</td>
<td>N/D</td>
</tr>
<tr>
<td>5</td>
<td>ST amc</td>
<td>carbohydrate</td>
<td>aminocyclitol</td>
<td>N/D</td>
</tr>
<tr>
<td>6</td>
<td>ST bac1</td>
<td>ribosomal peptide</td>
<td>class I bacteriocin</td>
<td>antimicrobial</td>
</tr>
<tr>
<td>7</td>
<td>ST pks3</td>
<td>polyketide</td>
<td>aromatic polyketide</td>
<td>N/D</td>
</tr>
<tr>
<td>8</td>
<td>ST desb</td>
<td>hydroxamate</td>
<td>desferrioxamine</td>
<td>siderophore</td>
</tr>
<tr>
<td>9</td>
<td>ST sid2</td>
<td>non-ribosomal peptide</td>
<td>yersiniabactin-related</td>
<td>siderophore/iron chelation</td>
</tr>
<tr>
<td>10</td>
<td>ST spo</td>
<td>polyketide</td>
<td>sporolide</td>
<td>N/D</td>
</tr>
<tr>
<td>11</td>
<td>ST slm</td>
<td>polyketide</td>
<td>salininactam</td>
<td>N/D</td>
</tr>
<tr>
<td>12</td>
<td>ST sid3</td>
<td>non-ribosomal peptide</td>
<td>siderophore</td>
<td>siderophore/iron chelation</td>
</tr>
<tr>
<td>13</td>
<td>ST sid4</td>
<td>non-ribosomal peptide</td>
<td>siderophore</td>
<td>siderophore/iron chelation</td>
</tr>
<tr>
<td>14</td>
<td>ST bac2</td>
<td>ribosomal peptide</td>
<td>class I bacteriocin</td>
<td>antimicrobial</td>
</tr>
<tr>
<td>15</td>
<td>ST lym</td>
<td>polyketide/non-ribosomal peptide</td>
<td>lymphostin</td>
<td>immunosuppressant</td>
</tr>
<tr>
<td>16</td>
<td>ST terp1</td>
<td>terpenoid</td>
<td>carotenoid pigment</td>
<td>antioxidant</td>
</tr>
<tr>
<td>17</td>
<td>ST pks4</td>
<td>polyketide</td>
<td>phenolic lipids</td>
<td>cell wall lipid</td>
</tr>
<tr>
<td>18</td>
<td>ST nrps2</td>
<td>non-ribosomal peptide</td>
<td>tetrapeptide</td>
<td>N/D</td>
</tr>
<tr>
<td>19</td>
<td>ST terp2</td>
<td>terpenoid</td>
<td>carotenoid pigment</td>
<td>antioxidant</td>
</tr>
</tbody>
</table>

*bold corresponds to isolated metabolites (modified from Penn et al. 2009)*

### 1.3: Polyketides and Non-ribosomal Peptides

#### 1.3.1: Diversity of PKs and NRPs

Polyketides (PKs) and non-ribosomal peptides (NRPs) are among the two most highly diverse and prolific families of microbial natural products. The diversity of PK structures is too numerous to list here, but some representatives include polyethers such as brevetoxin (41) and monensin (42), polyenes such as amphoterocin (43), a wide variety of macrolides (44–46), dibenzofurans like usnic acid (47), furanocoumarins such as aflatoxin (48), anthraquinone antibiotics such as the tetracyclines (49), polyene lactones like...
discodermolide (50),\textsuperscript{92} and the list goes on (Figure 1.7). As one might assume, the diversity of polyketide structures is paralleled by an equally impressive breadth of biological activities and molecular targets. For example the antibiotics, erythromycin (44)\textsuperscript{93} and the tetracyclines (49)\textsuperscript{94} inhibit translation by binding to the 50s and 30s ribosomal subunits, respectively. Brevetoxins (41) are neurotoxins that inhibit sodium ion channels.\textsuperscript{95} Aflatoxins (48)\textsuperscript{91} are hepatotoxins commonly found in food supplies of developing countries. Discodermolide (50) is an anticancer agent that stabilizes microtubule bundle formation.\textsuperscript{92}

Like PKs, NRPs are extremely diverse in both structure and biological activity. NRPs can be either linear or cyclic, contain a variety heterocycles, and many are comprised of non-proteinogenic or modified amino acids (Figure 1.8). Mixed NRP structures are also commonplace such as glycopeptides that incorporate sugars onto the NRP backbone, lipopetides that incorporate fatty acids, or mixed PK/NRPs (Figure 1.8). Perhaps the most well known NRP natural products are the β-lactam antibiotics of the penicillin (51)\textsuperscript{96} family. Other important NRPS-encoded antibiotics include zwittermicin (52),\textsuperscript{97} actinomycin (54),\textsuperscript{98} the glycopeptide vancomycin (58)\textsuperscript{23} and the lipopeptide surfactin (60).\textsuperscript{99} Cyclosporin (59)\textsuperscript{100} is an important immunosuppressive agent used during organ transplants. The siderophores yersiniabactin (56)\textsuperscript{101} and enterobactin (57)\textsuperscript{102} are used in nature for chelating iron from the environment.
Figure 1.7. Diversity of polyketide structures.
Figure 1.8. Diversity of non-ribosomal peptide (NRP) structures
1.3.2: Biosynthesis of Polyketides

PKs and NRPs are both biosynthesized in an assembly-line fashion in which small monomeric precursors are incorporated into the growing oligomer by iterative condensation reactions. The biosynthetic enzymes that perform these reactions for PKs are referred to as polyketide synthases (PKSs) and for non-ribosomal peptides they are called non-ribosomal peptide synthetases (NRPSs). PKs are biosynthesized in much the same way as fatty acids through decarboxylative Claisen condensation reactions of acyl thioesters. Carbonyl-CoA PK starter units are quite diverse and can be derived from sources such as fatty acids, benzoic acid, and amino acids. PK extender units however, are not nearly as diverse. The most common PK building blocks are derived from malonyl-CoA and methylmalonyl-CoA, however ethylmalonyl-CoA derived extender units are well represented. Propylmalonyl-CoA\textsuperscript{104-106}, chloroethylmalonyl-CoA\textsuperscript{107}, hexylmalonyl-CoA\textsuperscript{104,106-110}, allylmalonyl-CoA\textsuperscript{111}, and isobutyrylmalonyl-CoA\textsuperscript{112-113} have also been reported but are extremely rare.\textsuperscript{114} Additionally, there are four rare acyl carrier protein (ACP)-linked extender units: hydroxylmalonyl-ACP, methoxymalonyl-ACP, aminomalonyl-ACP, and glycerol-ACP.\textsuperscript{115-116}

PKSs are divided into three types. Type I or modular PKSs are large multifunctional proteins in which distinct active sites referred to as domains are organized into repeating modules and catalyze a specific reaction during polyketide extension and modification. Typically, a minimum module consists of an acyltransferase (AT) which selects the incoming extender unit, a
ketosacyl synthase (KS) that performs the decarboxylative Claisen condensation, and an acyl carrier protein (ACP) which covalently tethers the incoming extender units and growing polyketide chain via a flexible phosphopantethienyl arm.\textsuperscript{117-119} Additional optional domains determine the degree to which the ketide is reduced and can include a ketoreductase (KR), dehydratase (DH), enoylreductase (ER). In many cases the last module also consists of a thioesterase (TE) domain that releases the linear polyketide. For the vast majority of type I PKS biosynthetic gene clusters studied to date, the genes for biosynthesis are organized in a colinear fashion with each building block corresponding to a specific module.\textsuperscript{117} However several non-canonical examples exist that include module or domain skipping, stuttering, or the utilization of trans-AT domains.\textsuperscript{120} Type I PKSs can also be subdivided into non-iterative (multimodular) PKSs (as discussed above and typical to bacteria) and iterative (monomodular) type I PKSs in which a single module catalyzes multiple condensation reactions on the growing polyketide (typical in fungi). Most PKS gene clusters also include a suite of tailoring enzymes such as oxidoreductases, cytochrome P450s, (amino-, prenyl-, glycosyl-, and acyl-) transferases, and many more. Some examples of polyketides assembled by multimodular type I PKSs include monensin (42), amphoterocin (43), erythromycin (44), and FK506 (46).

Type II PKSs differ in that each catalytic domain is represented as a discrete protein and the set of proteins act iteratively to produce the fully extended polyketide.\textsuperscript{121-122} A minimal type II PKS system consists of a KS\textsubscript{α} that performs the condensation reaction, a KS\textsubscript{β} also termed chain length factor (CLF),
and an ACP. Additional catalytic subunits may include KRs, cyclases (Cyc), and aromatases (Ar).\textsuperscript{118} Type II PKSs are almost exclusive to bacteria and are responsible for aromatic PKs such as the tetracyclines (49).

Type III PKSs are also known as chalcone synthases originally known only from plants.\textsuperscript{123} These enzymes are multifunctional in which they select the starter unit and perform both the extension and folding of the final polyketide.\textsuperscript{118} Additionally, PK chain elongation by type III PKSs is iterative with the enzyme active site pocket determining the chain length.
Figure 1.9. Examples of PKS systems. Mutlimodular type I PKS for erythromycin (A) and iterative types I, II, and III (B). (from Hertweck 2009)
1.3.3: Biosynthesis of Non-Ribosomal Peptides

NRPS enzymes and biosynthetic gene clusters are reminiscent of multimodular type I PKS systems in which large multifunctional enzymes consisting of discrete domains catalyze the elongation and modification of the growing peptide in a collinear fashion.\textsuperscript{117,124} A minimal NRPS consists of a condensation (C) domain that catalyzes the peptide bond, an adenylation (A) domain that selects the incoming amino acid forming an acyl adenylate that is incorporated into the growing peptide chain, and a thiolation (T) domain that tethers incoming adenylates or the growing peptide. Also akin to PKSs, optional domains such as reductases (Re), cyclases (Cyc), epimerases (E), halogenases (H), and methyltransferases (MT) may be present.\textsuperscript{117,124} In microbial systems, mixed PKS/NRPS systems are fairly common such as the rifamycin\textsuperscript{125} biosynthetic gene cluster discussed below.

\textbf{Figure 1.10.} Example of NRPS system for penicillin antibiotics (from Felnagle et al. 2008)\textsuperscript{126}
1.4: Ansamycins

1.4.1: History and Significance of Ansamycin Natural Products

The ansamycin family of polyketides was first described by Sensi and co-workers in 1959 with the isolation of several rifamycins (19) from a terrestrial actinomycete later classified as *Amycolatopsis mediterranei*.69 These natural product macrolactams are characterized by a distinct mC7N structural unit biosynthetically derived from the aromatic acid 3-amino-5-hydroxybenzoate (AHBA), which is carboxy extended by multimodular type 1 PKSs utilizing primarily acetate and propionate building blocks.127-130 Two ansamycin variations incorporating either benzyl (Figure 1.11) or naphthyl (Figure 1.12) ring system joined at two nonadjacent positions by an aliphatic handle (ansa) have emerged based on the AHBA-derived structural unit.

The benzoic ansamycins such as geldamycin (61),131-132 herbimycin (62),133-135 and ansamitocin (63)136 (Figure 1.11) mainly exhibit antitumor properties by inhibiting the heat-shock protein 90 (Hsp90) complex. The naphthalenic ansamycins rifamycin (19),69,127,137-138 naphthomycin (70),139 and rubradirin140-141 (71, Figure 1.2) on the other hand, are most commonly used as antimicrobial therapies against gram-negative and some gram-positive bacteria. For example, the semisynthetic derivatives of rifamycin SV, rifampicin142-144 and rifabutin,145-146 are clinically approved for the treatment of tuberculosis and HIV/AIDS related mycobacterial infections. Additional antitumor30,125,147-150 and anti-inflammatory151 properties have also been reported for ansamycins. Because of their potency, wide-range of pharmacological properties, and
increasingly rapid evolution of antibiotic resistant bacteria, both chemical and genetic screening methods have been developed to specifically identify ansamycin producing bacteria.$^{152-155}$

Three new families of ansamycins were reported in early 2011, all from *Streptomyces* species.$^{30,32,112}$ Two of these *Streptomyces* spp. strains are associated with the marine environment. The ansalactams ($72)^{112}$ were isolated from a marine sediment derived *Streptomyces* sp., while the divergolides ($64, 65, 74, \text{ and } 75)^{32}$ are from an endophytic *Streptomyces* sp. cultured from the mangrove tree *Bruguiera gymnorrhiza*. Even more interesting is that these two novel ansamycin families both incorporate an unprecedented polyketide extender unit into their ansa backbone (see discussion below and Chapter 3).$^{32,112-113}$ The chaxamycins ($76 \text{ and } 77$) on the other hand were isolated from a *Streptomyces* sp. collected from hyper-arid soil from a Chilean desert.$^{30}$ The chaxamycin producing strain was first identified by PCR screening for ansamycin biosynthetic genes, demonstrating the power of genetic screening as a drug discovery tool.
Figure 1.11. Benzoic ansamycins isolated from actinomycetes.
Figure 1.2. Structures of naphthalenic ansamycins isolated from actinomycetes.
1.4.2: Biosynthesis of Ansamycins

Early investigation into the origin of the ansamycin mC7N structural unit with $^{14}$C-labeled AHBA identified that the chromophore is derived from the unusual naturally occurring aromatic amino acid.$^{156-157}$ Through *in vivo* mutagenesis studies in *A. mediterranei*.$^{158-160}$ and heterologous expression in both *S. coelicolor*.$^{158}$ and *E. coli*,.$^{161-162}$ it was shown that AHBA biosynthesis proceeds via the recently described aminoshikimate pathway from UDP-glucose.$^{158-160,162}$ It was concluded that nine enzymes encoded by *rifG-N* are required for the biosynthesis of AHBA in *A. mediterranei* (Figure 1.13), most of which share homologs in all known AHBA gene clusters.$^{163}$ The AHBA synthase encoded by *rifK* which catalyzes both the second and final step of AHBA biosynthesis has become the definitive molecular target for identifying putative ansamycin producers.$^{152}$
Once synthesized, the AHBA starter unit is then primed by the adenylation (A) domain of a mixed PKS/NRPS encoded by rifA and extended along a multimodular type I PKS with primarily malonyl-CoA and methylmalonyl-CoA (Figure 1.14). Ansamycins that incorporate less common extender units into the polyketide structures include kanglemycin (73) and the recently discovered divergolides (64, 65, 74, and 75). Both incorporate an ethylmalonyl-CoA unit into their ansa backbone (Figure 1.12). Benzoic ansamycins 61–63 are extended with one or more methoxymalonyl-ACP derived units (Figure 1.11). Even more remarkable was the co-discovery earlier this year of a novel isobutyrylmalonyl-CoA derived extender unit from the ansalactams (72).
and divergolides \((64, 65, 74, \text{ and } 75)\).\(^\text{32,113}\) Not only are these the only ansamycins to incorporate a isobutyrylmalonyl-CoA extender unit, but also the only known polyketides with this unusual moiety (see Chapter 3 for discussion).

For the naphthalenic ansamycins, the second ring of the naphthalene core is formed following hydroxylation of the AHBA moiety of the ACP-bound tetraketide product of module 3 by an FAD-dependent monooxygenase (Figure 1.14A–B).\(^\text{165,170}\) Inactivation of the rifamycin and naphthomycin homologs, \textit{rif-orf19}\(^\text{170}\) and \textit{nat2},\(^\text{165}\) respectively, led to abolishment of the final ansamycin product and accumulation of the truncated tetraketide product, desacetyl-SY4b.\(^\text{165,170}\) Homologs to \textit{rif-orf19} and \textit{nat2} are also present in some benzoic ansamycin pathways and have been shown to also catalyze hydroxylation of the AHBA derived chromophore in 61,\(^\text{171}\) 62,\(^\text{166}\) and macbecin.\(^\text{172}\)

Following elongation, an amide synthase (e.g. \textit{rifF} in the rifamycin pathway) catalyzes the release and cyclization of the linear polyketides to form the macrolactam (Figure 1.14).\(^\text{76,173}\) Most, if not all, post-PKS ansamycin intermediates are then further modified by various tailoring enzymes including but not limited to cytochrome P450s, acetyltransferases, methyltransferases, and halogenases.\(^\text{70,165-166,170,172,174-175}\)
Figure 1.14. Proposed biosynthesis of ansamycin natural products. Polyketide elongation and cyclization by the (A) rifamycin/salinketal pathway (adapted from Floss and Yu 2005),\textsuperscript{125} (B) naphthomycin pathway (from Wu et al. 2011),\textsuperscript{165} and (C) ansamitocin pathway (adapted from Yu et al. 2002).\textsuperscript{169}
1.4.3: Engineering Ansamycin Analogs

Most ansamycins used as therapeutics are modified derivatives of the natural products (e.g. rifampicin, rifabutin, and 17AAG). Historically unnatural derivatives of ansamycins have been developed for use as pharmaceuticals by semisynthesis. Hundreds of rifamycin analogs including the clinically important rifampicin and rifabutin were generated through semisynthesis, mainly by varying substituents on C-3 of the naphthoquinone core.¹²⁵ However recent advances in genetic engineering and heterologous expression have facilitated the in vivo mutasynthesis of highly bioactive ansamycin analogs. For example, inactivation of the AHBA synthase in both the geldanamycin (61)¹⁷⁶-¹⁷⁷ and ansamitocin (63)¹⁷⁸ producing Streptomyces strains followed by chemical complementation with substituted AHBA derivatives led to the synthesis of highly potent analogs of each. Combined, these strategies could prove to be an attractive means to develop new pharmaceuticals with maximum efficacy and minimum side-effects.

In this dissertation, I report my extensive efforts towards the elucidation of the biosynthesis of two atypical ansamycin compounds. Using a multidisciplinary approach including stable isotope feeding studies, genetic manipulation, and bioinformatics, I have revealed an unprecedented multifunctional cytochrome P450 and new polyketide extender unit biosynthetic pathway. Hopefully, the identification and partial characterization of these new biocatalysts will lay the foundation for bioengineering polyketide (ansamycin) analogs of pharmaceutical importance.
1.5: References


(22) Jones, D.; Metzger, H. J.; Schatz, A.; Waksman, S. A. Control of Gram-Negative Bacteria in Experimental Animals by Streptomycin. Science 1944, 100, 103-105.


(163) Floss, H. G.; Yu, T.-W.; Arakawa, K. The biosynthesis of 3-amino-5-hydroxybenzoic acid (AHBA), the precursor of mC7N units in ansamycin and mitomycin antibiotics: a review. *J Antibiot* 2011, 64, 35-44.


Chapter 2:

Shared Biosynthesis of the Saliniketals and Rifamycins in *Salinispora arenicola* is Controlled by the *sare1259* Encoded Cytochrome P450
2.1: Abstract

Saliniketals A and B are unusual polyketides from the marine actinomycete *Salinispora arenicola* that inhibit ornithine decarboxylase induction. The structural similarities between the saliniketals and the ansa chain of the potent rifamycin antibiotics, which co-occur in the fermentation broth, suggest a common origin between the two compound classes. Using PCR-directed mutagenesis, chemical complementation studies, and stable isotope feeding experiments, we showed that the saliniketals are byproducts of the rifamycin biosynthetic pathway diverging at the stage of 34a-deoxyrifamycin W. Our results suggest that a single enzyme, the cytochrome P450 monooxygenase encoded by *sare1259*, catalyzes multiple oxidative rearrangement reactions on 34a-deoxyrifamycin W to yield both the saliniketal and rifamycin structural classes.

2.2: Introduction

2.2.1: Introduction to the Saliniketals and Rifamycins

Saliniketals A (1) and B (2), originally isolated by Fenical and co-workers in 2007, are novel bicyclic polyketides produced by the obligate marine actinomycete *Salinispora arenicola* (Figure 2.1). These inhibitors of ornithine decarboxylase induction, an effective treatment strategy for cancer chemoprevention, harbor unusual structural features that have inspired three total syntheses to date. The saliniketals contain an unprecedented 1,4-dimethyl-2,8-dioxabicyclo[3.2.1]octane ring system with a polyketide side-chain at C11 that terminates in a primary amide. The global structure of 1 and 2 is also
notable because of the striking similarity, including the exact stereochemistry, to the ansa chain of the potent rifamycin antibiotics (3–6), which co-occur in the fermentation broth (Figure 2.1).\textsuperscript{1} While the rifamycin polyketides are assembled from the aromatic starter unit 3-amino-5-hydroxybenzoic acid (AHBA) with two acetate and eight propionate extender units,\textsuperscript{6} the presence of the primary amide on the saliniketals suggested that they are not biosynthetic shunt products nor degradation products of rifamycin because cleavage of the C–N bond of the AHBA-derived aromatic starter unit was unlikely.\textsuperscript{1} Instead, the saliniketals were proposed to originate from a three carbon starter unit that is extended by a disparate polyketide synthase (PKS) with two acetate and five propionate units to form the ketide with the exact stereochemistry as in 3–6.\textsuperscript{1} However, prior to this study, this hypothesis had not been explored experimentally.
Figure 2.1: Structures of the saliniketals (1–2) and rifamycins (3–6) from S. arenicola CNS-205. The numbering of 1–6 is consistent with prior publications.\(^1,42\)

The rifamycins were first identified by Sensi and co-workers in 1959 from a terrestrial actinomycete that would eventually be classified as *Amycolatopsis mediterranei*.\(^6,7\) In 2006, Kim et al. first reported the production of rifamycins B (3) and SV (4) from a strain of *Salinispora* isolated from an Australian marine sponge.\(^8\) Later, Jensen and co-workers showed that the *Salinispora* exhibit species-specific secondary metabolite production trends in which saliniketal and
unspecified rifamycins along with the staurosporines were observed in all examined strains of *S. arenicola*, thus comprising the *S. arenicola* core chemotype. Bioinformatic analysis of the Palauan *S. arenicola* CNS-205 genome and polymerase chain reaction (PCR) based screening experiments confirmed genes predicted to be involved in the biosynthesis of rifamycins. The high sequence identity between genes in the *S. arenicola* rifamycin cluster (SA-*)rif* and genes from the well characterized *A. mediterranei* S699 rifamycin biosynthetic gene cluster (AM-*rif*) suggests that the *rif* locus has undergone horizontal gene transfer.

2.2.2: Specific Aims

In the present study, the biosynthetic relationship between the rifamycins and saliniketals was explored by a multidisciplinary approach involving bioinformatic analysis, *in vivo* mutagenesis, chemical complementation, and stable isotope incorporation studies. In doing so, the following questions were addressed: Does the structural similarity between the saliniketals and rifamycins originate from a common biosynthetic machinery? If so, are the saliniketals shunt products of the rifamycin polyketide assembly or molecules selectively generated from a rifamycin-type biosynthetic precursor by a distinct enzyme in the pathway? Herein we report that the saliniketals are unexpected products of the central *rif* pathway intermediate 34a-deoxyrifamycin W (7, see Figure 4) in which the cytochrome P450 monooxygenase (CYP) Sare1259 is responsible for dual oxidative rearrangement reactions that lead to the formation of the mature
rifamycins 3–6 and the truncated saliniketals 1–2. An effort to functionally characterize the CYP Sare1259 in vitro was also attempted, however, a suitable assay was never achieved despite successful overexpression and isolation of soluble protein.

2.3: Results

2.3.1: Chemical Analysis of S. arenicola CNS-205 and A. mediterranei S699.

Organic extracts of S. arenicola CNS-205 and A. mediterranei S699 were analyzed by liquid chromatography mass spectrometry (LC-MS) to directly compare their rif chemistry with known saliniketal and rifamycin chemical standards. In addition to 1 and 2, S. arenicola CNS-205 produces a mixture of rifamycin SV (4), 27-O-demethyl-25-O-desacetylrifamycin SV (5) and 27-O-demethylrifamycin SV (6) when cultured in A1 production media (Figure 2.1). *Amycolatopsis mediterranei* S699, on the other hand, primarily produces rifamycin B (3). After a closer inspection of extracts from the fermentation broth of A. mediterranei S699 grown in YMG liquid media, we learned that 1 and 2 are not unique to S. arenicola but are also minor components of the original rifamycin producer (Figure 2.2). The co-production of both compound classes by these distantly related actinomycetes suggested that the saliniketals and rifamycins may arise from a common biosynthetic pathway.
Figure 2.2. HPLC-MS analysis of *A. mediterranei* S699. (A) Extracted ion chromatogram for all masses corresponding to the m/z of the sodium adducts of the saliniketals (1–2) and rifamycins (3–8) produced by *A. mediterranei* S699. (B) MS fragmentation pattern for 1 from *A. mediterranei* S699 (top panel) and authentic standard (bottom panel).

### 2.3.2: Bioinformatic Analysis of the *S. arenicola* CNS-205 Genome.

In early 2007, the completion of the *S. arenicola* CNS-205 genome sequencing project yielded a 5,786,361 bp genome (CP000850) with at least 30 distinct secondary metabolite gene clusters.\(^\text{10}\) Initially excluding the rifamycin biosynthetic gene cluster (SA-rif), none of the remaining twelve polyketide synthase (PKS)-associated pathways had a domain architecture consistent with
the predicted assembly of the saliniketals. We therefore were left with no other viable option but the 92 kb SA-rif cluster for saliniketal biosynthesis. The SA-rif pathway is comprised of 39 open reading frames (ORFs), 33 of which are homologous to genes in the AM-rif cluster with identities >50% (Table 2.1). The SA-rif cluster is organized similarly to the AM-rif cluster with a distinct 10 module PKS operon (rifA–E), an AHBA biosynthetic sub-cluster, and a 24 kb tailoring and regulation region (Figure 2.3). All genes determined essential for the biosynthesis of rifamycin B in A. mediterranei S699 share a homolog in the SA-rif cluster.6,12-17 There are, however, nine genes in the tailoring region of the AM-rif cluster that do not have homologs in the SA-rif cluster (rifO, rif-orf2, rifP, rifQ, rif-orf3, rif-orf4, rif-orf8, rif-orf11, rif-orf17, and rif-orf37) and conversely, the SA-rif cluster contains six genes without counterparts in AM-rif (sare1240, sare1241, sare1258, sare1261, sare1276, and sare1278). The unique S. arenicola genes are annotated as a multicopper oxidase (sare1241), a Rieske-domain protein (sare1258), a FAD-dependent oxidoreductase (sare1278), two hypothetical proteins (sare1240 and sare1276), and a ferrodoxin-domain containing protein (sare1261). The regions upstream and downstream of the SA-rif locus also differ from the AM-rif cluster in that none of the transport, membrane protein, ribosomal protein, or RNA polymerase genes are shared. Homologs of many of these genes are, however, located elsewhere in the S. arenicola genome (data not shown). Interestingly, SA and AM rif share genes related to aminodeoxysugar biosynthesis (sare1263–7 corresponding to rif-orf7, rif-orf6, rif-orf9, rif-orf10, and rif-orf18, respectively) while no glycosylated rifamycins have been reported from
either strain. The co-production of the closely related glycosylated compound, tolypomycin Y, with rifamycins B (3) and O by *Amycolatopsis tolypomycina* \(^{18}\) suggests that these genes may be silent in SA-rif, as proposed for the aminodeoxysugar locus in AM-rif.\(^{19}\)

*Amycolatopsis mediterranei* S699 (AM-rif)

*Salinispora arenicola* CNS-205 (SA-rif)

**Figure 2.3.** Comparison of the rifamycin biosynthetic gene clusters from *A. mediterranei* S699 (AF040571) and *S. arenicola* CNS-205 (CP000850).
Table 2.1. Summary of SA-rif cluster and chemotypes of genetically mutated strains.
<table>
<thead>
<tr>
<th>SA-rif gene</th>
<th>annotation</th>
<th>AM-rif homolog</th>
<th>% identity</th>
<th>chemotype of mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sare_1240</td>
<td>hypothetical protein</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Sare_1241</td>
<td>multicopper oxidase type 2</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Sare_1242*</td>
<td>oxidoreductase domain protein</td>
<td>rifS</td>
<td>76</td>
<td>No production of 1–8</td>
</tr>
<tr>
<td>Sare_1243*</td>
<td>oxidoreductase domain protein</td>
<td>rifT</td>
<td>57</td>
<td>No production of 1–8</td>
</tr>
<tr>
<td>Sare_1244</td>
<td>hypothetical protein</td>
<td>orf35</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Sare_1245*</td>
<td>cytochrome P450</td>
<td>orf0</td>
<td>82</td>
<td>No production of 1–8</td>
</tr>
<tr>
<td>Sare_1246</td>
<td>AMP-dependent synthetase and ligase</td>
<td>rifA</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Sare_1247</td>
<td>beta-ketoacyl synthase</td>
<td>rifB</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Sare_1248</td>
<td>beta-ketoacyl synthase</td>
<td>rifC</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Sare_1249</td>
<td>beta-ketoacyl synthase</td>
<td>rifD</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Sare_1250</td>
<td>acyltransferase</td>
<td>rifE</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Sare_1251</td>
<td>N-acetyltransferase</td>
<td>rifF</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Sare_1252</td>
<td>hypothetical protein</td>
<td>orf1</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Sare_1253</td>
<td>3-dehydroquinate synthase</td>
<td>rifG</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Sare_1254</td>
<td>3-deoxy-7-phosphoheptulonate synthase</td>
<td>rifH</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Sare_1255*</td>
<td>AHBA synthase</td>
<td>rifK</td>
<td>86</td>
<td>No production of 1–8</td>
</tr>
<tr>
<td>Sare_1256</td>
<td>oxidoreductase domain protein</td>
<td>rifL</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Sare_1257</td>
<td>putative phosphatase</td>
<td>rifM</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Sare_1258</td>
<td>rieske [2Fe-2S] domain protein</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Sare_1259*</td>
<td>cytochrome P450</td>
<td>orf5</td>
<td>71</td>
<td>7</td>
</tr>
<tr>
<td>Sare_1260*</td>
<td>cytochrome P450</td>
<td>orf16</td>
<td>73</td>
<td>WT</td>
</tr>
<tr>
<td>Sare_1261</td>
<td>DUF1271, ferredoxin domain</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Sare_1262*</td>
<td>acetyltransferase, PapA5</td>
<td>orf20</td>
<td>58</td>
<td>1, 2, &amp; 6</td>
</tr>
<tr>
<td>Sare_1263*</td>
<td>protein of unknown function</td>
<td>orf7</td>
<td>66</td>
<td>WT</td>
</tr>
<tr>
<td>Sare_1264*</td>
<td>DegT/DnrJ/EryC1/StrS aminotransferase</td>
<td>orf6</td>
<td>86</td>
<td>WT</td>
</tr>
<tr>
<td>Sare_1265*</td>
<td>aminotransferase class-III</td>
<td>orf9</td>
<td>79</td>
<td>WT</td>
</tr>
<tr>
<td>Sare_1266*</td>
<td>oxidoreductase domain protein</td>
<td>orf10</td>
<td>70</td>
<td>WT</td>
</tr>
<tr>
<td>Sare_1267*</td>
<td>NDP-hexose 2,3-dehydratase</td>
<td>orf18</td>
<td>79</td>
<td>WT</td>
</tr>
<tr>
<td>Sare_1268</td>
<td>monoxygenase FAD-binding</td>
<td>orf19</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Sare_1269</td>
<td>type II thioesterase</td>
<td>rifR</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Sare_1270</td>
<td>regulatory protein LuxR</td>
<td>orf36</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Sare_1271</td>
<td>methyltransferase type 11</td>
<td>orf14</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Sare_1272*</td>
<td>transketolase domain protein</td>
<td>orf15A</td>
<td>80</td>
<td>No production of 1–8</td>
</tr>
<tr>
<td>Sare_1273*</td>
<td>transketolase central region</td>
<td>orf15B</td>
<td>70</td>
<td>No production of 1–8</td>
</tr>
<tr>
<td>Sare_1274*</td>
<td>shikimate dehydrogenase</td>
<td>rifl</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Sare_1275</td>
<td>ROK family protein</td>
<td>rifN</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Sare_1276</td>
<td>DoxX family protein</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Sare_1277</td>
<td>3-dehydroquinate dehydratase</td>
<td>rifJ</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Sare_1278</td>
<td>FAD-dependent oxidoreductase</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

* indicates genes inactivated in this study, Abbreviations: (WT) wild-type chemistry observed (compounds 1–2 and 4–6).
2.3.3: Inactivation of the AHBA Synthase (rifK) and the Transketolase (rif-orf15).

The bioinformatic analysis of the SA and AM rif clusters did not support the previously proposed mechanism for saliniketal formation involving a distinct three-carbon primer unit intercepting the rifamycin synthase in lieu of the aromatic precursor AHBA. Since the saliniketals are not unique to S. arenicola but are also produced by A. mediterranei, we explored the biosynthetic scenario that the saliniketals are directly related to the rifamycins as either shunt or degradative products. In order to initially probe their putative biosynthetic relationship, we employed a PCR-directed gene replacement methodology to disrupt the AHBA synthase encoding gene rifK (sare1255) with an apramycin resistance cassette (acc(3)IV) as described previously for cyclomarin biosynthesis in S. arenicola.20 As predicted, inactivation of rifK in S. arenicola completely abolished the production of all rifamycins. In addition, the saliniketals were also not produced in the rifK mutant (rifK::aprR), providing unequivocal evidence that these natural products share a common biosynthetic origin from AHBA (Figure 2.4A–B). This conclusion was further substantiated through the successful chemical complementation of rifK::aprR with synthetically prepared AHBA that restored both saliniketal (1–2) and rifamycin (3–6) production to wild-type levels (Figure 2.4C). Similar results were also observed when I inactivated, with a single replacement, the SA-rif transketolase A (sare1272) and B (sare1273) subunits corresponding to rif-orf15A and rif-orf15B from AM-rif, respectively (Figure 2.5H). Based on in vitro reconstitution of genes involved in
AHBA biosynthesis with the *E. coli* transketolase (*tktA*), Orf15 is predicted to catalyze the conversion of 3-amino-3-deoxy-D-fructose-6-phosphate to imino-erythrose-4-phosphate during the biosynthesis of AHBA. To the best of our knowledge, these are the first experimental data to confirm the function of Orf15 in vivo.

**Figure 2.4.** HPLC-MS analysis of the inactivation of the *S. arenicola* AHBA synthase (*rifK, sare1255*) and chemical complementation of the resulting mutant (*rifK::aprR*). Extracted ion chromatogram (EIC) for masses corresponding to the m/z of the sodium adducts of the saliniketals (1–2) and rifamycins (4–6) from *Salinispora arenicola* (A) wild-type, (B) *rifK::apr R*, and (C) *rifK::apr R* mutant chemically complemented with synthetically prepared AHBA (25 mM).
To explore whether the nitrogen atom of saliniketal’s primary amide originates from AHBA or from an alternate source, we synthesized and then administered \[^{15}N\]AHBA to production cultures of the rifamycin deficient mutant rifK::apr\(^R\). We observed by MS similar isotopic enrichments of the saliniketal and rifamycin biosynthetic products as in the synthetic \[^{15}N\]AHBA precursor of 25 atom\% \(^{15}N\), thereby confirming that the primary amide nitrogen of the saliniketals
is derived from the “unlikely” C–N bond cleavage of the AHBA unit (Figure 2.6). These observations also suggested that 1 and 2 are not shunt products formed during polyketide extension but rather originate from a mature rifamycin macrocyclic product since the priming AHBA nitrogen atom ultimately resides at the terminating carbonyl C1 group of saliniketal.

Figure 2.6. Stable isotope labeling pattern observed for 1 and proposed for 7 (see White et. al 1974) from the incorporation of [15N]AHBA and [U-13C3]propionate.
2.3.4: Identification of Genes Directly Involved in Saliniketal Production.

Once we established that 1–6 share a single post-PKS assembly biosynthetic precursor, we methodically inactivated genes in the SA-rif cluster with homologous representatives in AM-rif to determine which are involved with the production of 1 and 2 and to elucidate the point at which the rifamycin and saliniketal pathways diverge. We inactivated four genes with predicted functions based on characterization or inactivation of their corresponding AM-rif homologs (sare1245, sare1259, sare1260, and sare1262 corresponding to rif-orf0, rif-orf13/rif-orf5, rif-orf16, and rif-orf20, respectively).

We first explored rif-orf20, which encodes an acetyltransferase that catalyzes the acetylation of the C25 hydroxyl of 6 to form 5. Based on an early mechanistic proposal for the formation of 1, we envisioned the assimilation of the 25-O-acetyl unit of 4 or 5 into the bicyclic ring of 1 at C16/C17. Deletion of sare1262 yielded a mutant that accumulated the desacetylated 6 without any effect on the production of 1 and 2, suggesting that the pathways must diverge prior to the formation of 6. This result was corroborated through an isotope experiment with [U-13C]propionate that enriched 1 with six intact propionate units and one incomplete propionate unit at C13/C14 (Figure 2.6) corresponding to the formal loss of C34a of 34a-deoxyrifamycin W (7). Importantly, C16/C17 of 1 originate from a propionate unit versus acetate as we originally contemplated.

We next turned our attention to the other three genes with characterized AM-rif homologs (sare1245, sare1259, and sare1260) that code for CYPs. Previous inactivation of the CYP rif-orf16 in A. mediterranei S699 produced a mixture of 3
and 4,\textsuperscript{13} and upon further analysis, we showed that this mutation had no effect on the production of 1 and 2. Similarly, inactivation of the \textit{S. arenicola} homolog, \textit{sare1260} (73\% identity), did not alter the production of 1, 2, and 4–6.

The CYP encoded by \textit{sare1245} shares 82\% identity with \textit{rif-orf0}, and unlike the other \textit{rif} CYPs, they are similarly positioned directly upstream of the rifamycin PKS gene \textit{rifA} rather than in the downstream region that primarily harbors the tailoring and regulatory genes (Figure 2.3). Lee et al. suggest that the AM \textit{rif-orf0} CYP is responsible for the introduction of the C34a hydroxyl group in rifamycin W (8, Figure 2.8A) based on their observation that inactivation of \textit{rif-orf0} in \textit{A. mediterranei} resulted in a mutant that no longer produced rifamycin B yet accumulated proansamycin X.\textsuperscript{25} Our attempt to inactivate the homologous \textit{sare1245} resulted in a mutant that no longer produced either the rifamycins or saliniketals. Subsequent chemical complementation of the \textit{sare1245} mutant (\textit{sare1245::apr}\textsuperscript{R}) with AHBA failed to restore saliniketal/rifamycin production. However, complementation of \textit{sare1245::apr}\textsuperscript{R} with 7 rescued the production of the saliniketals and rifamycins, suggesting that the gene product of \textit{sare1245} is not involved in the post-PKS modification of rifamycin nor the biosynthesis of 1 and 2 (Figure 2.7). Identical results were observed for the inactivation of the \textit{rifS} (\textit{sare1242}) and \textit{rifT} (\textit{sare1243}) homologs, which are located immediately upstream of \textit{sare1245} on the same putative transcriptional operon. This suggested that inactivation of \textit{rif-orf0}, \textit{rifS}, and \textit{rifT} produced a polar effect on the transcription of the rifamycin polyketide genes, \textit{rifA–F}, directly downstream of
sare1245, thus resulting in the observed chemotype for all respective mutant strains.

![HPLC-MS analysis of the chemical complementation of the S. arenicola rif-orf0 mutant strain sare1245::aprR](image)

**Figure 2.7.** HPLC-MS analysis of the chemical complementation of the S. arenicola rif-orf0 mutant strain sare1245::aprR. EIC for all masses corresponding to the m/z of the sodium adducts of the saliniketals (1–2) and rifamycins (3–8) for (A) sare1245::aprR chemically complemented with 7, (B) sare1245::aprR chemically complemented with AHBA (C) sare1245::aprR negative control. (* indicates unrelated compounds that share an m/z ratio with one of the compounds 1–8)

The third and final SA-rif CYP sare1259 is homologous to two AM-rif CYPs, namely orf13 and orf5, at 79% and 71% identity, respectively (Table 2.1). Xu et al. showed, using in vivo mutagenesis, that the gene product of rif-orf13 in A. mediterranei S699 is not involved in rifamycin B (3) production, while inactivation of rif-orf5 yielded a mutant that accumulated the biosynthetic intermediate rifamycin W (8, Figure 2.8A) in lieu of 3. The authors concluded that the CYP encoded by rif-orf5 catalyzes the multistep oxidative conversion of 8 to 6. Further analysis of the AM-rif-orf5 mutant in this study, however, revealed that the mutant also produced 1, 2, 4, and 6 as minor components (Figure 2.9).
With this knowledge in hand, we genetically inactivated *sare1259* in *S. arenicola*. The resulting mutant (*sare1259::apr*) not only lost the biosynthetic ability to produce rifamycins 4–6, as expected, but significantly the saliniketals as well (Figure 2.8B). This suggested that the CYP encoded by *sare1259* is a key enzyme in the biosynthesis of 1 and 2 in *S. arenicola* and confirms our hypothesis that the saliniketals are products of an enzymatic cleavage of a rifamycin macrocyclic intermediate versus a shunt product of the polyketide elongation. Interestingly, *sare1259::apr* accumulated an unexpected product that was confirmed by NMR analysis to be the rifamycin biosynthetic intermediate 34a-deoxyrifamycin W (7) instead of the predicted product 8. The presence of two homologs to *sare1259* in the *AM-rif* pathway may account for the accumulation of 8 in the *AM-rif-orf5* mutant, where the CYP encoded by *rif-orf13* may hydroxylate C34a of 7, and perhaps even perform the conversion of 7 to 1, 2, and 6.
Figure 2.8. Inactivation of the cytochrome P450, sare1259. (A) Structure of rifamycin biosynthetic intermediates 34a-deoxyrifamycin W (7) and rifamycin W (8). (B) EIC for masses corresponding to the m/z of the sodium adducts of the saliniketals (1–2) and rifamycins (4–8) produced by S. arenicola wild-type and sare1259::aprR strains, showing the loss of saliniketal (1–2) production and accumulation of 7 in the mutant.
Figure 2.9. HPLC-MS analysis of the *A. mediterranei* rif-orf5 mutant strain MT45025H. EIC for all masses corresponding to the m/z of the sodium adducts of the saliniketals (1–2) and rifamycins (3–8) produced by MT45025H cultured for 10 days in A1 media at 28°C and 225 rpm.

The mechanism proposed for the function of the CYP encoded by rif-orf5 involves the oxidative cleavage of the C12/C29 double bond followed by rearrangement of the ansa chain to form 6 from 8.\textsuperscript{13} If Sare1259 also performs the same oxidative cleavage of the C29/C12 olefin to yield 1, then C16/C17 of 1 would have to come from an alternate 2-carbon source such as the acetate unit in 5 that we previously ruled out. Analysis of the [U-\textsuperscript{13}C\textsubscript{3}]propionate-labeled 1, however, unequivocally showed that C15/C16/C17 originates from an intact propionate molecule, thereby suggesting that this 3-carbon unit is most likely derived from C12/C13/C29 in 7 (Figure 2.6). Consequently, these data suggest that Sare1259 has dual function to oxidatively cleave 7 at the C12/C29 olefin to yield the rifamycins and at the C11/C12 bond to give the saliniketals. The biosynthesis of 2 may similarly proceed from a 30-hydroxyrifamycin W precursor,
which was previously observed in a recombinant strain of *A. mediterranei*,\(^\text{26}\) to account for its C18 hydroxyl group.

In order to determine if any gene products other than the CYP encoded by *sare1259* are also necessary for the biosynthesis of 1, an additional eight genes (*sare1242, sare1243, sare1263–sare1267, and sare1274*), for which no function has been either confirmed or evaluated in *A. mediterranei* (Table 2.1), were inactivated in *S. arenicola*. While inactivation of *sare1242* (*rifS*) and *sare1243* (*rifT*) created putative polar effects on the downstream PKS resulting in the total loss of rifamycin/saliniketal biosynthesis, inactivation of the six remaining genes in either *S. arenicola* CNS-205 (this study) or *A. mediterranei* S699 (see ref. 13) had no effect on the biosynthesis of 1–8 (Figure 2.5).

### 2.3.5: Chemical Complementation of the *rifK* Mutant with Rifamycin Macrocyclic Intermediates.

In order to confirm 7 as an immediate precursor to the saliniketals, we complemented cultures of * rifK::apr\(^R\) * with successive biosynthetic intermediates from the rifamycin pathway 7→8→6→5→4. Complementation with both 7 and 8 were able to restore production of the saliniketals, while intermediates 4–6 did not rescue their production (Figure 2.10). While the complementation of * rifK::apr\(^R\) * with 8 was able to restore saliniketal production, the bioconversion of 8 to 1 and 2 does not appear to be as efficient as the conversion of 7 to 1 and 2 (Figure 2.10C and D). These findings support our mutagenesis results indicating that the saliniketals are synthesized during the bioconversion of 7 to 6 by Sare1259 with
8 as a potential intermediate. The intermediates that were not able to rescue saliniketal production are presumed to be taken up by the cells based on the observation that the early biosynthetic intermediates can undergo bioconversion into later stage rifamycins (Figure 2.10C–E).

Figure 2.10. HPLC-MS analysis of the chemical complementation of rifK::apr<sup>R</sup> with rifamycin biosynthetic intermediates. EIC for all masses corresponding to the m/z of the sodium adducts of the saliniketals (1–2) and rifamycins (3–8) for (A) wild-type S. arenicola CNS-205, (B) rifK::apr<sup>R</sup> negative control (C) rifK::apr<sup>R</sup> complemented with 7, (D) rifK::apr<sup>R</sup> complemented with 8, (E) rifK::apr<sup>R</sup> complemented with 6, (F) rifK::apr<sup>R</sup> complemented with 5, and (G) rifK::apr<sup>R</sup> complemented with 4. (* indicates unrelated compounds that share an m/z ratio with one of the compounds 1–8)
2.3.6: Efforts Towards the Functional Characterization of CYP Sare1259

To further explore the enzymatic function of Sare1259 in the biosynthesis of 1–8, efforts were carried out towards the in vitro characterization of the CYP encoded by sare1259. The recombinant CYP was produced by cloning sare1259 into the pHis8 expression vector and transforming the construct into E. coli BL21(DE3). Optimal expression was achieved by cultivation of E. coli BL21(DE3)/pHis8sare1259 in Terrific Broth (TB) containing 4% glycerol and kanamycin (50 µg/ml). When an A_{600} of 1.0 was reached, expression was induced by addition of 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) followed by an additional 24 hrs of growth at 16°C. The cells were harvested and the His-tagged protein (Sare1259His8, 49.1 kDa) isolated by Ni^{2+}-affinity column purification (Figure 2.11). Soluble protein fractions were pooled, concentrated, and further resolved by fast protein liquid chromatography (FPLC). The purified soluble protein was concentrated to 18 mg/ml and stored at -20°C until assayed. Characteristic of CYP enzymes, the soluble protein was deep red in color and easily visible with the naked eye due to the presence of the active site heme prosthetic group.
Figure 2.11. SDS-Page gel of Ni\(^{2+}\) affinity purification of recombinantly expressed Sare1259. Each lane corresponds to protein marker (M), insoluble fraction (I), soluble fraction (S), flow through (FT), wash fractions 1, 4, and 8, and elution fractions (E1-7), respectively.

Bacterial class I CYPs homologous to Sare1259 require enzymatic redox partners, ferredoxin (Fdx) and ferredoxin reductase (Fdr), to shuttle two electrons necessary for the CYP catalytic reaction from NAD(P)H (Figure 2.12).\(^{28-29}\) For the majority of bacterial CYPs, the specific redox enzymes have not been explored in detail, however, Fdx and Fdr from spinach (Spinacea oleracea) are commonly utilized as a suitable substitute for in vitro assays.\(^{30-31}\) There are two genes with [2Fe-2S] ferredoxin domains, sare1258 and sare1261, located within the SA-rif biosynthetic gene cluster (Table 2.1). Neither gene shares a homolog within the AM-rif cluster. The gene sare1258 is annotated as a Rieske[2Fe-2S] domain protein which are typically involved in the transfer of electrons from NAD(P)H to nonheme dioxygenase enzymes.\(^ {32-33} \) The gene sare1261 is
annotated as a protein of unknown function DUF1271, however, closer inspection revealed that *sare1261* is homologous to actinomycete Fdx proteins, including as high as 44% sequence identity to Fdx genes in the rifamycin SV (4) producing strain *A. mediterranei* U32 (data not shown).

![Depiction of Class I cytochrome P450 electron transfer system.](image)

**Figure 2.12.** Depiction of Class I cytochrome P450 electron transfer system. Two electrons are transferred from NAD(P)H through a FAD-containing ferredoxin reductase (Fdr) to an iron-sulfur [2Fe-2S] containing ferredoxin (Fdx). Then the electrons transferred from the Fdx to the cytochrome P450 (CYP) providing the redox potential necessary for the oxidation of organic substrates.

To determine if either of the putative electron transfer proteins are associated with CYPSare1259, *sare1258* and *sare1261* were cloned and expressed using similar conditions described for Sare1259. Both Sare1258 and Sare1261 were purified as soluble proteins (Figure 2.13). However, when analyzed on a 16% Novex® Tris-glycine precast gel (Invitrogen), both proteins migrated slower than expected appearing approximately 2 kDa larger than their predicted sizes of 15 and 9 kDa, respectively. The size discrepancy may be due several factors including incomplete denaturation of the proteins by the SDS running buffer. However, since the 2 kDa size difference is consistent for both proteins and sequencing of the pHis8sare1258 and pHis8sare1261 plasmids
revealed that the plasmids were constructed properly, the putative electron transfer enzymes were used for downstream assays.

Figure 2.13. Ni^{2+}-NTA affinity purification of putative electron transport enzymes Sare1258 (top) and Sare1261 (bottom). Recombinant poly-his-tagged proteins, Sare1258 and Sare126, were isolated from overnight culture of E. coli BL21(DE3) transformed with the plasmid pHis8sare1258 and pHis8sare1261, respectively. Polyacrylamide gel electrophoresis was performed on a 16% Novex Tris-glycine precast gel (Invitrogen) with SDS-Tris-glycine running buffer. Each lane corresponds to protein marker (M), insoluble fraction (I), soluble fraction (S), flow through (FT), wash fractions 1, 4, and 8, and elution fractions(E1-5), respectively.
In an attempt to biochemically characterized CYPsare1259, an assay was developed containing at minimum: His-tagged Sare1259, Fdr from spinach, either Fdx from spinach/Sare1258/Sare1261, and substrate (7 or 8) in the presence of NADH and NADPH (see Material and Methods). Despite numerous attempts and a wide variety of assay conditions (molar ratios, buffers, pH, etc.), no oxidation of 7 to 1 or any proposed intermediate (Figure 2.14) was observed in vitro. This is consistent with the efforts by Mahmud and co-workers to characterize rif-Orf5 (Taifo Mahmud, personal communication). Similarly, they successfully isolated soluble protein but were unable to observe conversion of 8 to 6 as predicted.\(^{13}\)

2.4: Discussion

The great structural diversity of naturally occurring rifamycins is largely due to the extensive post-PKS enzymatic tailoring carried out by enzymes encoded by pathway specific genes. After more than 50 years of study, the rifamycin pathway continues to reveal new and unusual biosynthetic potential. The saliniketals are another example of how enzymatic tailoring of natural products by CYPs can produce novel chemical structures through complex chemical reactions.

CYPs are known to catalyze a multitude of diverse oxidative modifications (e.g. hydroxylation, epoxidation, oxidative bond cleavage, or dehydrogenation) on a wide range of molecular substrates, often activating or detoxifying the chemical species.\(^ {34-36}\) The CYP encoded by sare1259 is most closely related to
rif-orf13 and rif-orf5 from the AM-rif cluster as well as other CYPs belonging to the CYP107 family. This superfamily includes a large number of bacterial CYPs, many of which are associated with xenobiotic degradation and secondary metabolite tailoring.\textsuperscript{37-40}

Our data suggest that Sare1259 catalyzes multiple oxidations of 7 at variable positions for its conversion to either the saliniketals or rifamycins. During the biosynthesis of 3, C34a of 7 is proposed to be sequentially oxidized twice and then lost via a decarboxylation of the resulting carboxyl.\textsuperscript{13,41} We hypothesize that the timing of this decarboxylation reaction determines whether the final product of the SA-rif pathway yields the rifamycins (3–6) or the saliniketals (1–2). The current mechanism proposed for the function of the CYP encoded by rif-orf5 in A. mediterranei involves the oxidative cleavage of the C12/C29 double bond of 7 followed by formation of the ketal of 6 and decarboxylation of the C34a carboxyl group.\textsuperscript{13} In contrast, for the production of 1 and 2, we propose that the decarboxylation reaction occurs prior to oxidative bond cleavage, shifting the C12/C29 olefin to the C11/C12 position where it is then cleaved in a similar fashion by Sare1259 (Figure 2.14). Reduction of the resulting C28/C29 olefin and rearrangement of the ansa chain would form the 5-carboxy intermediate of the salinisporamycin A (10), a recently described compound isolated from S. arenicola YM23-082.\textsuperscript{42} Our proposed mechanism is further supported by the [U-\textsuperscript{13}C]propionate tracer experiments and by the isolation of (5-carboxy)salinisporamycin A from fermentations of S. arenicola (J. B. MacMillan, personal communication). The isolation of 10 by Matsuda and co-workers also
suggests the sequence of transformation from 7 to 1 proceeds with the formation of the bicyclic ketal prior to the construction of the primary amide. Mechanisms for this latter reaction may involve an oxidation of the quinone at C2 or reduction of the quinone to the hydroquinone followed by a retro-Michael cleavage of the C–N bond to yield 1 and the respective naphthoquinone (Figure 2.14). Attempts to identify and isolate the resulting naphthoquinone moiety have thus far been unsuccessful. Additionally, we did not identify a rif pathway gene associated with the putative conversion of (5-carboxy)salinisporamycin A to 1, which may suggest that if this reaction is indeed catalyzed by a dedicated enzyme, that it may be encoded outside the rif cluster.

In conclusion, we took a multidisciplinary approach involving a combination of genomics, in vivo mutagenesis, chemical complementation, stable isotope incorporation studies, and chemical analyses to interrogate the biosynthetic origin of the saliniketals. Such a multifaceted approach was required to effectively explore the biosynthesis of the saliniketals that ultimately led to the identification of the Sare1259 CYP as the key biosynthetic gene product that governs the diversification of the rif pathway at the stage of 34a-deoxyrifamycin W (7) leading to the truncated saliniketals and mature rifamycins (Figure 2.14). Though currently unsuccessful, we believe that probing the multifunctional CYP Sare1259 to explore its preferred substrates, reaction profile, and molecular mechanism will likely lead to the discovery new P450 catalyzed chemistry.
**Figure 2.14.** Proposed mechanism of saliniketal biosynthesis facilitated by CYP Sare1259 (intermediates with published structures are numbered).1,26,41-42
2.5: Experimental Section

2.5.1: General Methods.

Low-resolution LC-MS was carried out on a Hewlett-Packard series 1100 LC-MS system in positive ion mode with a linear gradient of 10–90% MeCN at a flow rate of 0.7 ml/min over 24 min on a RP C_{18} column (Phenomenex Luna, 4.6 mm x 100 mm, 5 µm). $^1$H, heteronuclear multiple bond correlation (gHMBC) and heteronuclear single quantum coherence (gHSQC) NMR spectral data were obtained on a Bruker DRX600 spectrophotometer equipped with a 1.7 mm cryoprobe. $^{13}$C-NMR spectral data were obtained on a Varian VX-500 instrument equipped with an Xsens™ cold probe. Rifamycin SV (4) sodium salt was purchased from Sigma Aldrich (St. Louis, MO). Saliniketal A (1) was graciously provided by William H. Fenical.¹ Rifamycin W (7), 27-O-demethyl-24-O-desacetyl rifamycin SV (6, DMDARSV), and 27-O-demethyl rifamycin SV (5, DMRSV) were isolated and confirmed based on literature precedence.¹³⁻¹⁷

2.5.2: Bacterial Strains, Culture Conditions, and Extraction of Natural Products.

Salinispora arenicola strain CNS-205 and Amycolatopsis mediterranei strains S699, MT45025H, and MT1601KH were obtained as described previously.¹³⁻²⁰ All S. arenicola seed cultures were grown in A1 liquid media (10 g starch, 5 g yeast extract, and 2 g peptone per liter of seawater) and production cultures were grown in A1BFeC liquid media (10 g starch, 5 g yeast extract, 2 g peptone, 100 mg KBr, 40 mg Fe₂(SO₄)₃·4H₂O, 1 g CaCO₃ per 1 L seawater) at
28°C and 225 rpm.\textsuperscript{20} \textit{Amycolatopsis mediterranei} S699 seed and production cultures were both grown in YMG liquid media\textsuperscript{43} at 28°C and 225 rpm, unless otherwise noted. Extraction of salinketals and rifamycin compounds was achieved by acidifying the culture broth to pH 2–3 with 1 N HCl and extracting the cultures 3X with equal volumes of EtOAc. The organic fraction was dried under vacuum, resuspended in MeOH, and analyzed by LC-MS. \textit{Escherichia coli} strains EPI300, BW25113/pKD20\textsuperscript{44} and S17-1\textsuperscript{45} used for mutagenesis experiments were grown in Luria-Bertani (LB) media with appropriate antibiotics. \textit{E. coli} BL21(DE3) (Stratagene) used for protein expression.

\textbf{2.5.3: Genetic Manipulations.}

Inactivation of rifamycin and salinketals biosynthetic genes was performed using REDIRECT\textsuperscript{©} PCR targeting technology as described previously for \textit{S. arenicola}.\textsuperscript{20,46} \textit{E. coli} EPI300 (Epicentre) carrying fomids BPPW5227, BPAF1230, and BPAF1361 were provided by the Joint Genome Institute, Walnut Creek, CA. Each of the genes targeted for inactivation was replaced with an apramycin resistance (\textit{acc}(3)/\textit{IV}) cassette by double crossover homologous recombination on the fosmid containing the gene of interest. The mutant fosmid was then transformed into \textit{E. coli} S17-1 and transferred to \textit{S. arenicola} CNS-205 via conjugation. Exconjugates were confirmed by PCR analysis and restriction digest.
2.5.4: Isolation of 34a-Deoxyrifamycin W (7) from the sare1259::aprR Mutant.

The sare1259::aprR mutant was cultured in 8 x 1 L Fernbach flasks of A1BFeC media for 7 days at 28°C while shaking at 225 rpm. XAD-7 resin (30 g) was added to each flask on day 7 and continued to shake for 5 hours before collecting the resin by filtration. The resin was then extracted with 3 x 1 L of acetone and concentrated under vacuum. The residue was separated by silica gel flash chromatography under isocratic conditions with MeOH/CHCl3 (1:9). Silica fractions were further purified by HPLC on a RP C18 column (Phenomenex Luna, 10 mm x 250 mm, 5 µm, MeCN/H2O 56:44, v/v).

34a-Deoxyrifamycin W (7). HPLC ESI-MS: [M+Na]+ 662.3; 1H NMR (600 MHz, DMSO-d6) δ 12.40 (s, 8-OH), 10.54 (br s, 6-OH), 9.59 (s, NH), 7.47 (s, H-3), 6.32 (dd, J = 15.5, 11.2 Hz, H-18), 6.16 (d, J = 10.8 Hz, H-17), 5.99 (dd, J = 15.6, 6.9 Hz, H-19), 4.76 (d, J = 6.5 Hz, 23-OH), 4.35 (s, 21-OH), 4.31 (d, J = 5.8 Hz, 27-OH), 3.84 (d, J = 5.6 Hz, H-27), 3.80 – 3.73 (m, H-25), 3.68 (d, J = 6.5 Hz, 25-OH), 3.35 – 3.28 (m, H-23), 2.40 (dd, J = 15.1, 7.3 Hz, H-28), 2.20 (dd, J = 14.6, 7.2 Hz, H-20), 2.13 (s, H-14), 2.00 (s, H-30), 1.89 (s, H-13), 1.65 (br d, J = 6.8 Hz, H-22), 1.59 – 1.51 (m, H-24), 1.25 – 1.15 (m, H-26), 0.95 (d, J = 7.0 Hz, H-34a), 0.91 (d, J = 6.9 Hz, H-32), 0.81 (d, J = 6.8 Hz, H-31), 0.55 (d, J = 6.7 Hz, H-33), 0.18 (d, J = 6.9 Hz, H-34).

2.5.5: Chemical Complementation of S. arenicola Mutants.

Chemical complementation studies were carried out in 50 ml cultures of the S. arenicola rifK::aprR mutant in A1 production media. Compounds 4–8 (2–6
mg) were individually added in triplicate to 1 day old production cultures of rifK::apr<sup>R</sup>. The cultures were then allowed to grow for 4 more days before being extracted as described above. Bioconversion of the rifamycin intermediates was monitored by HPLC-MS. Retention times, m/z ratios, and UV profiles of the biotransformation products were compared to authentic standards.

2.5.6: Synthesis of <sup>15</sup>N-AHBA.

AHBA was synthesized by Tobias A..M. Gulder by reacting 3,5-dihydroxybenzoic acid (2.0 g, 13 mmol) with NH₄Cl (1.7 g, 32 mmol) and aqueous NH₄OH (14.8N, 6 mL, 89 mmol) in a sealed high-pressure reaction tube at 180°C for 40 hours.<sup>47-48</sup> After cooling, the volatile components of the reaction mixture were removed in vacuo and the remaining solid redissolved in 100 mL 6N HCl. The resulting solution was heated under reflux for 16 hours, filtered, extracted with EtOAc (3 x 40 mL) to remove unreacted starting material, and concentrated to 25 mL. The desired product crystallized from this solution yielding grey crystals, which were collected and recrystallized from 6N HCl to furnish pure AHBA hydrochloride as white crystals in 68% overall yield (1.67 g, 8.8 mmol). The [<sup>15</sup>N]AHBA was prepared following the same synthetic route but substituting NH₄Cl with <sup>15</sup>NH₄Cl. The statistically expected <sup>15</sup>N-enrichment of 25% in the product was verified by mass spectrometric analysis. <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) δ 7.04 (m, 1H), 7.48 (m, 1H), 7.51 (m, 1H). HPLC ESI-MS: AHBA [M+H]<sup>+</sup> 154.1; [<sup>15</sup>N]AHBA [M+H]<sup>+</sup> 154.1 (75%), 155.2 (25%).
2.5.7: Stable Isotope Labeling Experiments of Saliniketal A.

[U-\textsuperscript{13}C\textsubscript{3}]Propionate incorporation experiments were carried out in 2 x 1 L cultures of \textit{S. arenicola} CNS-205 grown in A1BFeC liquid media at 225 rpm and 30°C for 7 days. 50 mg/L of sodium [U-\textsuperscript{13}C\textsubscript{3}]propionate (Cambridge Isotopes Laboratories, Incorporated) was aseptically added to each liter of culture after 48 hrs. The cultures were extracted with EtOAc, and 1 was isolated as described previously\textsuperscript{1} and analyzed by \textsuperscript{13}C-NMR (Figure S5 and S6, Supporting Information).

[\textsuperscript{15}N]AHBA incorporation studies were carried out in 50 mL cultures of \textit{rifK::apr\textsuperscript{R}} in A1BFeC at 225 rpm and 30°C for 5 days. 3 mg of [\textsuperscript{15}N]AHBA was added to each culture after 24 hrs and extracted as described above on day five. Crude extracts were analyzed by HPLC-MS. Incorporation of \textsuperscript{15}N into AHBA and 1 were calculated from \textit{m/z}.\textsuperscript{49} HPLC ESI-MS: saliniketal A (1) [M+Na]\textsuperscript{+} 418.3; [\textsuperscript{15}N]saliniketal A (1) [M+Na]\textsuperscript{+} 418.2 (75%), [M+Na]\textsuperscript{+} 419.2 (25%).

2.5.8: Cloning and Overexpression of Recombinant CYP Sare1259 and Putative Related Proteins.

The CYP \textit{sare1259} gene was amplified from genomic DNA of \textit{S. arenicola} CNS-205 via PCR with forward primer 1259\_ProtF and reverse primer 1259\_ProtR (Appendix Table A2.3). The PCR product and pHis8 vector were digested with \textit{Ncol} and \textit{HindIII}, then ligated overnight at 14°C. The resulting plasmid, pHis8\textit{sare1259}, was transformed into \textit{E. coli} BL21(DE3). A single kanamycin-resistant colony from the transformed cells was grown overnight in LB
medium containing kanamycin (50 µg mL\(^{-1}\)) at 37°C. A 1% inoculum was added to each liter of Terrific Broth (TB) containing 4% glycerol and kanamycin (50 µg mL\(^{-1}\)) and grown at 25°C while shaking. When an \(A_{600}\) of 1.0 was achieved, the temperature was lowered to 16°C and expression was induced by the addition of 0.1 mM isopentyl-β-D-1-thiogalactopyranoside (IPTG). Cultivation continued for an additional 24 hrs before the cells were harvested by centrifugation (6500 x g, 20 min). The cells were disrupted in lysis buffer (50 mM NaH\(_2\)PH\(_4\), 300 mM NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, pH 8.0) by incubation with lysozyme (1 mg ml\(^{-1}\)), followed by sonication. The cytosolic fraction was obtained by centrifugation (20\(^4\) x g, 60 min at 4°C) and applied to a polypropylene column containing 20 mL of Ni\(^{2+}\)-NTA agarose (Qiagen) pre-equilibrated with lysis buffer. The column was washed eight volumes of wash buffer (50 mM NaH\(_2\)PH\(_4\), 300 mM NaCl, 2 mM β-mercaptoethanol, 40 mM imidazole, pH 8.0), and eluted with elution buffer (50 mM NaH\(_2\)PH\(_4\), 300 mM NaCl, 2 mM β-mercaptoethanol, 250 mM imidazole, pH 8.0). The elution fractions were pooled and concentrated to 2.5 mL by centrifugation (GE Healthcare Vivaspin 20 column, MWCO 30 kDa). The concentrated His-tagged protein (2.5 mL in lysis buffer) was desalted on a PD-10 desalting column (GE Healthcare) and further purified by FPLC on a Superdex 200 gel filtration column pre-equilibrated with assay buffer (25 mM HEPES, 100 mM NaCl, pH 7.4). Eluted fractions were analyzed by SDS-PAGE with SimplyBlue SafeStain (Invitrogen). Pure fractions were pooled and concentrated to 18 mg mL\(^{-1}\) by centrifugation (as discussed above) and stored at -20°C in assay buffer with 10% glycerol.
Putative electron transport proteins Sare1258 (Rieske-domain protein) and Sare1261 (ferredoxin) were expressed and purified under similar conditions as Sare1259. Sare1258 was amplified from S. arenicola CNS-205 gDNA using the primers 1258_ProtF and 1258_ProtR, while Sare1261 was amplified using primers 1261_ProtF and 1261_ProtR (Appendix Table A2.3). Both genes were ligated into pHis8 and expressed under the conditions described for Sare1259. Due to the expected sizes of His-tagged Sare1258 (15 kDa) and His-tagged Sare1261 (9 kDa) a Vivaspin 20 column (MWCO 3 kDa) was used to concentrate purified protein. All proteins were stored at -20°C until assayed.

2.5.9: Enzymatic Assays of the CYP Sare1259

In order to functionally characterize Sare1259, an assay was designed to determine if Sare1259 could oxidize 34a-deoxyrifamycin W (7), to yield saliniketal A (1) or any of the proposed intermediates shown in Figure 2.14. The basic assay set-up was comprised of the following: His-Tagged Sare1259 (200 nM–50 µM), 1–5 µM ferredoxin (spinach Fdx or His-tagged Sare1258/Sare1261), 1–5 µM spinach ferredoxin reductase (Fdr), NADH (0.2 U ml⁻¹) and NADPH (2 mM), 0.1–5 mM substrate (7 or 8) dissolved in ether MeOH, IPA, or DMSO, and assay buffer to 100–500 µL. All reactions were performed in 1.5 mL Eppendorf tubes and incubated for 0.5–24 hrs at 30°C. After incubation, reactions were quenched with 1 M HCl and either extracted with two volumes of EtOAc or lyophilized and resuspended in 100 µL of MeOH. 5–10 µL of the MeOH solution was analyzed by LC-MS as described above. Additional variations to the assay protocol
included the use of phosphate buffer (100 mM potassium phosphate, pH 7.4), Tris-HCl buffer (1 M Tris-HCl, pH 7.5), or sodium acetate buffer (1 M NaOAc, pH 5.5) or addition of glucose-6-phosphate (G6P) and G6P dehydrogenase to regenerate NAD(P)H.
### Appendix Table A2.1. Primers synthesized for gene inactivation experiments of the SA–rif gene cluster.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Oligonucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>sare1242KOF</td>
<td>5'-cgtgagccccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1242KOR</td>
<td>5'-tcagccgagccatctctccatcagccgagccatctccagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1243KOF</td>
<td>5'-atgagagttgagccgagttgagccgagctccgacgactggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1243KOR</td>
<td>5'-ccgctccgttccctccgagccgagccgagccgagctccgacgactggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1245KOF</td>
<td>5'-tcctgtccgagccatctctccatcagccgagccatctccatcagccgagctccgacgactggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1245KOR</td>
<td>5'-gcctgtccgagccatctctccatcagccgagccatctccatcagccgagctccgacgactggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1246KOF</td>
<td>5'-gccgacgccagactggggaagggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1246KOR</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1247KOF</td>
<td>5'-gcctgtccgagccagactggggaagggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1247KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1248KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1248KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1249KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1249KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1250KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1250KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1251KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1251KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1252KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1252KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1253KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1253KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1254KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1254KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1255KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1255KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1256KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1256KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1257KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>sare1257KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>orf15KOF</td>
<td>5'-gcctgtccgagccagaggctgctccgagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
<tr>
<td>orf15KOR</td>
<td>5'-gcctgtccgagccagagttcggagagccgaggtgagccgaggtgagccgagttgaggaagggctggaaagcaatgattccggggatccgtcgacc-3'</td>
</tr>
</tbody>
</table>
Appendix Table A2.2. Primers synthesized for verification of double crossover mutants of the SA-\textit{rif} gene cluster.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Oligonucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>sare1242CKF</td>
<td>5'-ccggcaccggtgaactacc-3'</td>
</tr>
<tr>
<td>sare1242CR</td>
<td>5'-aggtgtggccgagaacag-3'</td>
</tr>
<tr>
<td>sare1243CKF</td>
<td>5'-gagttccgctgcttcgcc-3'</td>
</tr>
<tr>
<td>sare1243CR</td>
<td>5'-ggcctcaacctttgcggg-3'</td>
</tr>
<tr>
<td>sare1245CKF</td>
<td>5'-ggaccggccacgaagaatt-3'</td>
</tr>
<tr>
<td>sare1245CR</td>
<td>5'-cggggccggtctctcgaag-3'</td>
</tr>
<tr>
<td>sare1255CKF</td>
<td>5'-cgccgtgcaattcactgga-3'</td>
</tr>
<tr>
<td>sare1255CR</td>
<td>5'-gtgccttcctcggagaacgtt-3'</td>
</tr>
<tr>
<td>sare1258CKF</td>
<td>5'-caacgcggatccggcaggac-3'</td>
</tr>
<tr>
<td>sare1258CR</td>
<td>5'-ctcgccccgtgctctcagat-3'</td>
</tr>
<tr>
<td>sare1259CKF</td>
<td>5'-gaacagccggaccgagatg-3'</td>
</tr>
<tr>
<td>sare1259CR</td>
<td>5'-cggtgaacccgacgaccg-3'</td>
</tr>
<tr>
<td>sare1260CKF</td>
<td>5'-gagcggagggatgtgtggcgc-3'</td>
</tr>
<tr>
<td>sare1260CR</td>
<td>5'-ggagacccacggtcctcagct-3'</td>
</tr>
<tr>
<td>sare1262CKF</td>
<td>5'-acgctccacccgacgcactgg-3'</td>
</tr>
<tr>
<td>sare1262CR</td>
<td>5'-cagttgctcgtgacgcgggt-3'</td>
</tr>
<tr>
<td>sare1263CKF</td>
<td>5'-tggtccggtgttgcctccc-3'</td>
</tr>
<tr>
<td>sare1263CR</td>
<td>5'-gtggtcagctgcagcgggt-3'</td>
</tr>
<tr>
<td>sare1264CKF</td>
<td>5'-acgagttcgggcgtctcag-3'</td>
</tr>
<tr>
<td>sare1264CR</td>
<td>5'-gtgtgacctggtcagttgc-3'</td>
</tr>
<tr>
<td>sare1265CKF</td>
<td>5'-gccccagcacacatccggagg-3'</td>
</tr>
<tr>
<td>sare1265CR</td>
<td>5'-acagctcagctggtgcggg-3'</td>
</tr>
<tr>
<td>sare1266CKF</td>
<td>5'-gtcctcaggtgacgctcgcg-3'</td>
</tr>
<tr>
<td>sare1266CR</td>
<td>5'-ggcatgtgcggccctatcttc-3'</td>
</tr>
<tr>
<td>sare1267CKF</td>
<td>5'-ggaacccggaacagacgacc-3'</td>
</tr>
<tr>
<td>sare1267CR</td>
<td>5'-tacgagctccggcggcggg-3'</td>
</tr>
<tr>
<td>orf15CKF</td>
<td>5'-tgatcaacacagacgcgtg-3'</td>
</tr>
<tr>
<td>orf15CR</td>
<td>5'-ctcgccgctagctcgcggg-3'</td>
</tr>
<tr>
<td>sare1274CKF</td>
<td>5'-agcaccggcttcgacccgggtc-3'</td>
</tr>
<tr>
<td>sare1274CR</td>
<td>5'-tcagggccccggcaccagaac-3'</td>
</tr>
</tbody>
</table>
Appendix Table A2.3: Primers used for expression of recombinant CYP Sare1259 and putative associated electron transport enzymes (Sare1258 and Sare1261).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Oligonucleotide</th>
<th>Restriction Enzyme Recognition Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1259_ProtF</td>
<td>5’ CATG *<strong>CATGG</strong> CATGACCACCACTGCCGAG 3’</td>
<td>Nco I</td>
</tr>
<tr>
<td>1259_ProtR</td>
<td>5’ TACCC <strong>AAGCTT</strong> TCAGACGAGCTTCAGCGG 3’</td>
<td>Hind III</td>
</tr>
<tr>
<td>1258_ProtF</td>
<td>5’ CGG <strong>GGATCC</strong> CATGTTCGTACTGAAG 3’</td>
<td>BamH I</td>
</tr>
<tr>
<td>1258_ProtR</td>
<td>5’ CCC <strong>AAGCTT</strong> TCAGTCGCCAGTGTT 3’</td>
<td>Hind III</td>
</tr>
<tr>
<td>1261_ProtF</td>
<td>5’ CAC <strong>GGATCC</strong> GTGAGGGTGCGCGCG 3’</td>
<td>BamH I</td>
</tr>
<tr>
<td>1261_ProtR</td>
<td>5’ CCC <strong>AAGCTT</strong> TCAGCTGCGAACC 3’</td>
<td>Hind III</td>
</tr>
</tbody>
</table>

* bolded and underlined nucleotides correspond to restriction enzyme binding site.
Appendix Figure A2.1. $^{13}$C NMR spectrum (126 mHz, MeOH-$d_4$) of salniketal A (1) from [U-$^{13}$C]propionate feeding study.
Appendix Figure A2.2. $^{13}$C NMR spectrum (126 mHz, MeOH-$d_4$) of salniketal A (1) from [U-$^{13}$C]propionate feeding study (127-177 ppm)
Appendix Figure A2.3. $^1$H NMR spectrum (600 mHz, DMSO-$d_6$) of 34a-deoxy-rifamycin W (7) from sare1259::apr$^R$ mutant
Appendix Figure A2.4. gHSQC spectrum (600 mHz, DMSO-\textit{d}_6) of 34a-deoxy-rifamycin \textit{W} (\textit{7}) from \textit{sare1259::ap}^{R} mutant.
Appendix Figure A2.5. gHMBC spectrum (600 mHz, DMSO-\textit{d}_6) of 34a-deoxy-rifamycin W (7) from \textit{sare}1259::\textit{apr}^R mutant
2.7: References


(20) Schultz, A. W.; Oh, D. C.; Carney, J. R.; Williamson, R. T.; Udwary, D. W.; Jensen, P. R.; Gould, S. J.; Fenical, W.; Moore, B. S. Biosynthesis and structures


evolution study of a vitamin D3 hydroxylase from *Pseudonocardia autotrophica*. *Biochem Biophys Res Commun* 2009, 385, 170-175.


2.8: Acknowledgments

I would like to thank Dr. Tobias A.M. Gulder for the synthesis of $^{15}\text{N}$-AHBA and Drs. Alessandra Eustáqio and Andrew Schultz for training regarding PCR-targeted mutagenesis. We thank William H. Fenical and Paul R. Jensen for kindly providing authentic standards of saliniketal A (1) and *S. arenicola* CNS-205, respectively. We also thank Heinz G. Floss for his comments and assistance in the preparation of the original manuscript.

Chapter 2, in full, is a reprint of the material as it appears in Shared Biosynthesis of the Saliniketals and Rifamycins in *Salinispora arenicola* is Controlled by the *sare1259*-Encoded Cytochrome P450 (2010). Wilson, Micheal C.; Gluder, Tobias A. M.; Mahmud, Taifo; Moore, Bradley S., Journal of the American Chemical Society, 132, 12757-12765. The dissertation author was the primary investigator and author of this paper.
Chapter 3:

Biosynthesis of the Ansalactams from *Streptomyces* sp. CNH-189
3.1: Abstract

The ansalactams are recently discovered ansamycin class polyketides produced by an unusual modification of the polyketide pathway. This new metabolite from a marine sediment-derived bacterium, *Streptomyces* sp. CNH-189, possesses a novel spiro \( \gamma \)-lactam moiety and a distinctive isobutyryl polyketide fragment observed for the first time in this class of natural products. The structure of ansalactam A was defined by spectroscopic methods including X-ray crystallographic analysis. Biosynthetic studies with stable isotopes further led to the discovery of a new, branched chain polyketide synthase extender unit derived from \((E)-4\text{-}methyl\text{-}2\text{-}pentenoic\text{ acid}\). Using next generation sequencing of *Streptomyces* sp. CNH-189, a 35 Kb fragment of the putative ansalactam biosynthetic gene cluster was identified, revealing a novel three gene operon for biosynthesis of the isobutyrylmalonyl-CoA polyketide extender unit. A multidisciplinary investigation into the biosynthesis of ansalactam A and its unusual polyketide extender unit are reported within.

3.2.: Introduction

3.2.1: Isolation and Structure Elucidation of Ansalactam A

Ansamycin polyketides include potent antibiotic and anticancer agents such as rifamycin SV\(^{1-2}\) (1), geldanamycin\(^{3-4}\) (2), and ansamitocin P-3\(^{5-6}\) (3, Figure 3.1). These natural product macrolactams are characterized by a distinct mC\(_{7}\)N structural unit biosynthetically derived from the aromatic acid 3-amino-5-hydroxybenzoate (AHBA), which is carboxy extended by multimodular type 1
polyketide synthases (PKSs) utilizing primarily acetate and propionate building blocks prior to macrolactam cyclization.\textsuperscript{7-9} Two ansamycin variations incorporating either benzyl or naphthyl ring system joined at two nonadjacent carbons by an aliphatic handle (ansa) have emerged based on the AHBA-derived structural unit (as discussed previously in Chapter 1).

As part of an ongoing effort to discover novel marine microbial natural products,\textsuperscript{10-11} the Fenical laboratory at Scripps Institution of Oceanography investigated \textit{Streptomyces} sp. strain CNH-189 (hereafter referred to as CNH-189) isolated from marine sediments retrieved off shore of Oceanside, California. Liquid chromatography-mass spectroscopy (LC-MS) analysis of a whole culture crude extract of CNH-189 revealed several medium molecular weight compounds with an unusual UV chromophore indicative of highly conjugated aromatic molecules. Fractionation of a 60-L culture extract of CNH-189 by standard silica column chromatography, followed by reversed-phase HPLC separation, yielded ansalactam A (4), whose structure was characterized by NMR methods and confirmed by X-ray analysis (Figure 3.2).\textsuperscript{12} In addition to 4, several minor analogs derived from oxidative modifications of post-PKS assembly product were isolated to give ansalactams B–D (4b–d, Figure 3.3).
**Figure 3.1.** Ansamycin natural products from actinomycete bacteria.

**Figure 3.2.** Ansalactam A. (A) Structure of ansalactam A (4) with absolute stereochemistry and (B) ORTEP plot of the final X-ray structure of 4 depicting the absolute configuration.
Figure 3.3. Ansalactams B–D (4b-d, respectively) isolated as minor products of *Streptomyces* sp. CNH-189.

Unlike many naphthalenic ansamycins, ansalactam A (4) has not been shown to display any potent antibiotic activity.\(^{13-14}\) This is most likely due to the fact that 4 is missing three out of the four hydroxyl groups that bind to the active site of the microbial RNA polymerase drug target (see reference 13). However, 4 was active against the nuclear transcription factor NF-κB (IC\(_{50}\) 0.6 µg/ml) and thus may express anti-inflammatory or anticancer properties (Sang-Jip Nam, personal communication). Still, 4 is of more interest to the Moore laboratory because of its unusual biosynthesis rather than its pharmaceutical properties.

There are several features of 4 that make it unique amongst previously reported ansamycins. Notably, the AHBA-derived amino group is involved in a γ-lactam residue with the aliphatic side chain. This lactam ring is spiro fused to the naphthalenic backbone, which is reduced in comparison to other ansamycins. A second unusual structural feature involves the C-22–C-27 aliphatic side chain of
the lactam that suggests the possibility of a novel polyketide biosynthetic building block.

3.2.2: A New Paradigm of Polyketide Extender Unit Biosynthesis

Historically, ansamycin natural products are extended with various combinations of exclusively malonyl-CoA (MCoA) and methylmalonyl-CoA (mMCoA). In the ansalactams however, the terminal extender unit appears to originate from an isobutyrylmalonyl-CoA (ibMCoA) unit, which had not been seen prior to the isolation of 4. The ibMCoA extender unit of 4 is reminiscent of a recently recognized paradigm involving the diversification of polyketide extender units via the reductive carboxylation of $\alpha,\beta$-unsaturated fatty acyl-CoAs by crotonyl-CoA reductase/carboxylase (CCR) enzymes.\(^{15-18}\)

CCRs have long been known to play an important role in fatty acid biosynthesis and degradation.\(^{19-20}\) First isolated from Streptomyces collinus, CCR was shown to catalyze the reduction of crotonyl-CoA to butryl-CoA.\(^{21}\) More than a decade later, Alber and colleagues showed that in addition to the reductase activity, CCRs also catalyzed the carboxylation of crotonyl-CoA to (2S)-ethylmalonyl-CoA in the presence of CO\(_2\) (Figure 3.4).\(^{22}\) The involvement CCRs in polyketide biosynthesis was first experimentally confirmed in Steptomyces cinnamonensis for the biosynthesis of monensin A (Figure 3.5).\(^{23}\) Though eMCoA can originate from several routes,\(^{18}\) the S. cinnamonensis CCR was shown by gene inactivation and heterologous expression to be critical for the flux of eMCoA
to the monensin PKS despite being located elsewhere in the genome with genes associated with primary metabolism.\textsuperscript{23}

\textbf{Figure 3.4.} Biosynthetic Pathway of ethylmalonyl-CoA. (adapted from Alber 2011).\textsuperscript{24}
Figure 3.5. Natural products that incorporate ethylmalonyl-CoA into their structures.
Since the original isolation of the *S. collinus* CCR it has become apparent that CCRs are ubiquitous among *Streptomyces* many other actinobacteria, with *Saccharopolyspora erythraea* being a notable exception.\textsuperscript{18,25-27} Many species have been shown to contain more than one *ccr* in their genomes. One copy is proposed to function in primary metabolic pathways, while a second copy is often located within the boundaries of secondary metabolite biosynthetic gene clusters.\textsuperscript{17} For example, the gene clusters of lasalocid,\textsuperscript{28} oligomycin,\textsuperscript{29-30} kirromycin,\textsuperscript{31} FK520,\textsuperscript{32} rosaramicin,\textsuperscript{33} tautomycetin,\textsuperscript{34} leptomycin,\textsuperscript{35} coronatine,\textsuperscript{36} and midecamycin\textsuperscript{37} all contain a CCR for the biosynthesis of eMCoA polyketide extender units (Figure 3.5).

More recently, pathway specific CCR homologs have been shown to expand PKS extender unit diversity beyond eMCoA in the salinisporamide\textsuperscript{16-17} and FK506\textsuperscript{15,38} biosynthetic gene clusters. Not only CCRs, but an entire suite of biosynthetic genes responsible for the formation of the unusual chloroethylmalonyl-CoA and allylmalonyl-CoA PKS extender units of salinosporamide and FK506 are located within their respective gene clusters (Figure 3.5).\textsuperscript{15,17} Alternatively, the hexylmalonyl-CoA extender units of cinnabaramide\textsuperscript{39} and thuggacin\textsuperscript{40} originates from CCR-mediated reductive carboxylation of octenyl-CoA supplied from fatty acid biosynthetic pool. The reveromycin\textsuperscript{41-42} biosynthetic gene cluster also contains a pathway specific CCR that appears to have a relaxed substrate specificity for a variety of unusual PKS extender units (Shunji Takahashi, personal communication).
Figure 3.6. Natural products that incorporate rare or unique polyketide extender units.
Though completely unique when ansalactam (4) was first reported, the ibMCoA PKS extender unit was co-discovered by Hertweck and colleagues in the divergolides, ansamycin natural products from an endophytic *Streptomyces* species isolated from mangroves (Figure 3.6).\(^4\) Additionally, this endophytic *Streptomyces* species was also shown to produce new germicidin analogs which also contain an ibMCoA derived moiety (Figure 3.5).\(^4\) Similar to the results of our current study, Hertweck and colleagues identified a pathway specific suite of genes responsible biosynthesis of this unusual branched-chain extender unit.\(^4\) This gene cassette located in the divergolide biosynthetic gene cluster supplies ibMCoA to both the divergolides and germicidins.\(^4\) The divergolide gene cluster additionally contains a second CCR proposed to be involved in the biosynthesis of the eMCoA derived PKS extender unit of the divergolides.\(^4\)

In addition to naturally expanding the diversity of the polyketide backbone, CCRs provide a fortuitous opportunity to engineer novel unnatural polyketide analogs through mutasynthesis.\(^15-16,45\) Moore and colleagues showed that inactivation of the chlorinase (SalL) that catalyzes the first step of biosynthesis of chloroethylmalonyl-CoA yields a mutant that no longer produces salinosporamide A and accumulates the eMCoA substituted analog salinosporamide B.\(^17,46\) Subsequent feeding studies with a variety of \(\alpha,\beta\)-unsaturated fatty acyl-CoA precursors yielded a suite of fluoro-, bromo-, propyl-, and hexyl-substituted unnatural salinosporamide analogs.\(^16\) Similar results were observed with mutasynthesis of the FK506\(^15\) and cinnabaramide\(^39\) pathways. The apparent relaxed substrate specificity of both CCRs and acyltransferase (AT) domains that
accept unusual PKS extender units provides a simplistic and rapid method for generating combinatorial biosynthetic libraries that can be used in structure activity relationship assays.

3.2.3: Specific Aims

To address whether 4 is indeed a new member of the ansamycin structure class and to explore the concept of a new PKS building block involved in 4 assembly, we explored its biosynthesis. Utilization of stable isotope incorporation studies led to the discovery of a new, branched-chain PKS extender unit for polyketide assembly observed for the first time in this class of natural products. Subsequently, whole-genome sequencing of CNH-189 revealed a 35 Kb fragment of the putative ansalactam (4) biosynthetic gene cluster and led to the discovery of a novel three gene operon that may be responsible for the biosynthesis of ibMCoA. Herein, we report the comprehensive investigation into the biosynthesis 4 and discuss efforts towards the functional characterization of the ansalactam pathway and heterologous expression of ansalactam biosynthetic machinery.
3.3: Results and Discussion

3.3.1: Identification of an AHBA Synthase in S. sp. CNH-189

To determine whether *Streptomyces* sp. strain CNH-189 had the biosynthetic capacity to synthesize 4, we first employed a molecular approach and tested for the presence of a gene-encoded AHBA synthase, which is indicative of this structure class. Using degenerate PCR primers designed for the high throughput identification of AHBA synthases,\textsuperscript{39,47} we amplified a ~755 bp product from the genomic DNA (gDNA) of strain CNH-189. BLAST\textsuperscript{p}\textsuperscript{48} analysis of the amplicon showed greatest identity (92\%) to ShnS (accession AAL77217), an AHBA synthase from an uncharacterized naphthalenic ansamycin pathway in *Streptomyces hygroscopicus* 17997.\textsuperscript{49} The most similar characterized homolog is that of RifK (81\% identity, accession AAA75105) from the rifamycin pathway in *Amycolatopsis mediterranei* S699.\textsuperscript{50} Further phylogenetic analysis revealed that the CNH-189 AHBA synthase claded with AHBA synthases from characterized naphthalenic ansamycin gene clusters (Figure 3.7), which is consistent with the structure of 4.
Figure 3.7. Identification of a naphthoquinone-type AHBA synthase from S. sp. CNH-189. PCR amplification of AHBA synthase genes from *Streptomyces* sp. strain CNH-189 (lane 2), *Salinispora arenicola* strain CNS-205 (lane 3, positive control), *S. tropica* strain CNB-440 (lane 4, negative control), and no template control (lane 5) using degenerate primers designed by Huitu et al.\textsuperscript{47} The 755 bp amplicon from CNH-189 was aligned in a neighbor-joining tree with AHBA synthases from characterized biosynthetic gene clusters of benzoquinone and naphthoquinone ansamycins as well as uncharacterized AHBA synthases from genome sequencing projects and the AHBA synthase involved in the biosynthesis of mitomycin, a nonansamycin compound derived from AHBA.
1. GeneRuler® 1kb ladder
2. *Streptomyces* sp. CNH-189
3. *S. arenicola* CNS-205
4. *S. tropica* CNB-440
5. No template control
3.3.2: Stable Isotope Feeding Studies

To explore the origin of the carbon backbone of 4, we next administered $^{13}\text{C}$-labeled precursors to strain CNH-189. Due to peak broadening of key carbon signals (C-20, 20-Me, C-21, and C-22) in the $^{13}\text{C}$ NMR spectrum of 4, the $^{13}\text{C}$-enriched natural product was reduced with NaBH$_4$ to 5 to provide sharp NMR signals for isotope analyses (Scheme 3.1). Based on biosynthetic precedence in the rifamycin$^{13,51-52}$ and naphthomycin$^{53}$ series, we predicted that the polyketide backbone of 4 is assembled from malonate, six methylmalonates, and possibly 2-isobutyrylmalonate as a new PKS extender unit. As expected, sodium [1-$^{13}\text{C}$]propionate enriched six carbons at $\sim$17% corresponding to the carboxy position of each predicted propionate-derived extender unit (Figure 3.8). Similarly, sodium [1,2-$^{13}\text{C}$]acetate was assimilated intact into the predicted C-5/C-11 unit ($^2J = 48$ Hz). Additionally, we measured a second incorporated acetate unit at C-22/C-23 ($^2J = 46$ Hz), suggesting acetate is incorporated into the C-1/C-2 position of the putative 2-isobutyrylmalonate extender unit (Figure 3.9).

Scheme 3.1. Preparation of polyol 5 from ansalactam A (4).
Figure 3.8. $^{13}$C NMR spectrum (125 MHz) of 5 from [1-$^{13}$C]propionate feeding study in methanol-$d_4$. 
Figure 3.9. $^{13}$C NMR spectrum (125 MHz) of 5 from $[1,2^{-13}C_2]$acetate feeding study in methanol-$d_4$. 
3.3.3: Formation of the 2-Isobutyrylmalonate Polyketide Extender Unit.

We first hypothesized that the non-traditional extender unit of 4 was derived from L-leucine, which would be consistent with incorporation of acetate at C-22/C-23. However, neither [1-$^{13}$C]leucine or [1,2-$^{13}$C]leucine could be incorporated (Appendix Figure A3.1). Consequently, we next postulated that the 2-isobutyrylmalonate unit may instead originate from the condensation of isobutyrate and acetate (via malonate) and that the resulting olefinic diketide would be utilized as the terminal PKS building block of 4. Support for this proposal was first provided by the specific incorporation of sodium [1-$^{13}$C]isobutyrate at C-24 (4%) (Figure 3.10).
Figure 3.10. $^{13}$C NMR spectrum (125 MHz) of 5 from [1-$_{13}$C]isobutyrate feeding study in methanol-$d_4$. 
Recently, a new paradigm has surfaced for the assembly of 2-substituted malonyl-CoAs in polyketide biosynthesis in which α,β-unsaturated acyl-CoAs are reductively carboxylated with crotonyl-CoA reductase/carboxylase (CCR) homologs. Examples include ethylmalonate\textsuperscript{22,54} (i.e., monensin A and tylosin), chloroethylmalonate\textsuperscript{55} (salinosporamide A), propylmalonate\textsuperscript{56} (salinosporamide E), allylmalonate\textsuperscript{15,38} (FK506), and putatively 2-(2-methylbutyryl)malonate\textsuperscript{57} (polyoxypeptin A). In the case of the unique FK506 extender unit allylmalonyl-CoA, a discrete diketide synthase initiates its biosynthesis via the intermediate \((E)-2\text{-pentenyl-acyl carrier protein prior to CCR-mediated reductive carboxylation by TcsC. A similar pathway may also operate in 4 biosynthesis to give 2-isobutyrylmalonyl-CoA via \((E)-4\text{-methyl-2-pentenyl-CoA. To further explore this scenario, we synthesized }[2-^{13}\text{C}]\text{(E)-4-methyl-2-pentenoic acid (see Experimental Methods for details) and administered it to strain CNH-189. The resultant 4 was reduced and analyzed by }^{13}\text{C NMR to reveal a specific enrichment at C-22 (2\%, Figure 3.11), thereby providing further support that the six-carbon C-22–C-27 fragment in 4 may be derived from 2-isobutyrylmalonyl-CoA via CCR-mediated biochemistry.}
Figure 3.11. $^{13}$C NMR spectrum (125 MHz) of 5 from [2-$^{13}$C]4-Me-2-pentenoic acid feeding study in methanol-$d_4$. 
3.3.4: Identification of a Partial Ansalactam Biosynthetic Gene Cluster.

In order to identify the ansalactam biosynthetic gene cluster, the whole genome of strain CNH-189 was sequenced using next generation Solexa\textsuperscript{58} multiplex sequencing on a Genome Analyzer IIx (Illumina) platform. From the sequencing data, 2.8 million 65 bp paired-end reads were obtained providing ~300 fold coverage of the genome. An initial assembly using the bioinformatic software package, Geneious,\textsuperscript{59} yielded more than 10\textsuperscript{3} contigs ranging from 65–24,000 bp (data not shown). Using the sequence previously obtained for the CNH-189 AHBA synthase as a query, BLASTn\textsuperscript{48} analysis of the contig sequences identified a 20 Kb contig comprising a complete set AHBA biosynthetic genes,\textsuperscript{60-61} an N-terminal fragment (369 bp) of a PKS, several additional homologs to the ansamycin biosynthetic genes, three hypothetical proteins, a transposase, and a three gene cassette (crotonyl-CoA reductase/carboxylase, type III ketoacylsynthase, and a 3-hydroxybutyryl-CoA dehydrogenase) for the putative biosynthesis of isobutyrylmalonyl-CoA (Figure 3.11). Using a newly developed (unpublished) assembly algorithm termed Paired Read Iterative Contig Extension (PRICE), our collaborators further extended our partial ansalactam biosynthetic gene cluster an additional 15 Kb providing sequence for the terminal PKS module and three additional noncontiguous PKS module fragments (Figure 3.11). As our assembly currently stands we have 23 open reading frames (ORFs) spanning a 35 Kb contig. Most of the ORFs (15/23) are homologous to genes in naphthalenic biosynthetic gene
clusters, such as rifamycin$^{8,62-64}$ and napthomycin$^{65}$ and are thus named according to their corresponding homolog (Table 3.1).
**Figure 3.12.** Gene organization of the partial putative ansalactam biosynthetic gene cluster. The 35 Kb partial ansalactam cluster was assembled from a 20 Kb contig constructed utilizing the Geneious assembler and extended with an additional 15 Kb using PRICE. Genes are tentatively named to correspond to functionally characterized homologs from the rifamycin and naphthomycin gene clusters. A1*, A2*, and A3* are fragmented PKS modules due to probable errors in the assembly.
<table>
<thead>
<tr>
<th>Polypeptide</th>
<th>Amino Acids</th>
<th>Proposed Function</th>
<th>Nearest Homolog</th>
<th>Identity/Similarity</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>AnsA1</td>
<td>1214 PKS (KS AT DH)</td>
<td>NapA_Module 2, S. hygroscopicus JCM4427</td>
<td>82/87</td>
<td>ABB86419</td>
<td></td>
</tr>
<tr>
<td>AnsA2</td>
<td>624 PKS (ACP* KS)</td>
<td>NapA_Module 3, S. hygroscopicus JCM4427</td>
<td>72/78</td>
<td>ABB86419</td>
<td></td>
</tr>
<tr>
<td>AnsA3</td>
<td>1232 PKS (AT DH KR ACP)</td>
<td>NapA_Module 1, S. hygroscopicus JCM4427</td>
<td>70/78</td>
<td>ABB86419</td>
<td></td>
</tr>
<tr>
<td>AnsE</td>
<td>1784 PKS (KS AT DH KR ACP)</td>
<td>NapE, S. hygroscopicus JCM4427</td>
<td>79/87</td>
<td>ABB86423</td>
<td></td>
</tr>
<tr>
<td>AnsF</td>
<td>284 Amide synthase</td>
<td>NapF, S. hygroscopicus JCM4427</td>
<td>46/63</td>
<td>ABB86424</td>
<td></td>
</tr>
<tr>
<td>AnsG</td>
<td>357 3-Dehydroquinase synthase</td>
<td>ShnQ, S. hygroscopicus 17997</td>
<td>90/94</td>
<td>AAN63814</td>
<td></td>
</tr>
<tr>
<td>AnsH</td>
<td>469 AminoDAP synthase</td>
<td>NapH, S. hygroscopicus JCM4427</td>
<td>75/79</td>
<td>ABB86413</td>
<td></td>
</tr>
<tr>
<td>AnsI</td>
<td>278 Aminoquinate dehydrogenase</td>
<td>NapI, S. hygroscopicus JCM4427</td>
<td>81/86</td>
<td>ABB86414</td>
<td></td>
</tr>
<tr>
<td>AnsK</td>
<td>387 3-Amino-5-hydroxybenzoic acid synthase</td>
<td>AHBA synthase, S. hygroscopicus 17997</td>
<td>92/95</td>
<td>AAL7217</td>
<td></td>
</tr>
<tr>
<td>AnsL</td>
<td>367 Oxidoreductase</td>
<td>ShnO, S. hygroscopicus 17997</td>
<td>82/90</td>
<td>AAN63811</td>
<td></td>
</tr>
<tr>
<td>AnsM</td>
<td>233 Phosphatase</td>
<td>ShnP, S. hygroscopicus 17997</td>
<td>85/90</td>
<td>AAN63810</td>
<td></td>
</tr>
<tr>
<td>AnsN</td>
<td>311 AHB kinase</td>
<td>AHB kinase, S. hygroscopicus 17997</td>
<td>86/92</td>
<td>AAR32671</td>
<td></td>
</tr>
<tr>
<td>Ans1</td>
<td>176 MarR family transcriptional regulator</td>
<td>MarR, S. sviceus ATCC29083</td>
<td>67/81</td>
<td>ZP_06917736</td>
<td></td>
</tr>
<tr>
<td>Ans2</td>
<td>293 3-Hydroxybutyl-CoA dehydrogenase</td>
<td>3-HBDH, Nocardia farcinica FM 10152</td>
<td>53/69</td>
<td>YP_121445</td>
<td></td>
</tr>
<tr>
<td>Ans3</td>
<td>346 Ketoacyl synthase type III-FabH</td>
<td>3-oxoacyl-(ACP) synthase III, Streptomyces sp. Mg1</td>
<td>55/70</td>
<td>ZP_05002626</td>
<td></td>
</tr>
<tr>
<td>Ans4</td>
<td>453 Crotonyl-CoA reductase/carboxylase</td>
<td>Crotonyl-CoA reductase, Frankia sp. EUN1f</td>
<td>68/81</td>
<td>ZP_0641404</td>
<td></td>
</tr>
<tr>
<td>AnsJ</td>
<td>148 Amidehydroquininate dehydratase</td>
<td>Amidehydroquininate dehydratase, S. hygroscopicus 17997</td>
<td>90/96</td>
<td>AAR32674</td>
<td></td>
</tr>
<tr>
<td>AnsR</td>
<td>252 Thioesterase</td>
<td>Thioesterase, Streptomyces sp. CS</td>
<td>75/84</td>
<td>ADM46371</td>
<td></td>
</tr>
<tr>
<td>Ans19</td>
<td>592 FAD-dependent monoxygenase</td>
<td>Putative hydroxylase, S. hygroscopicus 17997</td>
<td>86/92</td>
<td>AAR32673</td>
<td></td>
</tr>
<tr>
<td>Ans5</td>
<td>211 Conserved hypothetical protein</td>
<td>Transcriptional modulator of MarE toxin, MazF, Streptosporangium roseum DSM 43021</td>
<td>79/84</td>
<td>YP_00340166</td>
<td></td>
</tr>
<tr>
<td>Ans6</td>
<td>133 Transposase</td>
<td>Transposase IS4 family protein, S. lividans TK24</td>
<td>90/96</td>
<td>ZP_05523008</td>
<td></td>
</tr>
<tr>
<td>Ans7</td>
<td>242 Conserved hypothetical protein</td>
<td>Conserved hypothetical protein, S. sviceus ATCC29083</td>
<td>77/80</td>
<td>ZP_06915554</td>
<td></td>
</tr>
<tr>
<td>Ans8</td>
<td>315 Hypothetical Protein</td>
<td>Hypothetical Protein SCAB_76731, S. scabei 87.22</td>
<td>90/90</td>
<td>YP_003493180</td>
<td></td>
</tr>
</tbody>
</table>
As mentioned above, the 35 Kb partial putative ansalactam gene cluster contains an entire suite of genes, *ansGHILMN* and *ansJ*, required for the biosynthesis of AHBA\textsuperscript{50,60-61,63,66} (Figure 3.13). Interestingly, the organization of the AHBA biosynthetic genes from the putative ansalactam cluster are syntenic with the rifamycin cluster in *A. mediterranei* S699 and the *nap* cluster in *S. hygroscopicus* JCM4427 with *ansGHILMN* forming a single operon and *ansJ* 4.3kb downstream of *ansN* (Figure 3.14). This further supports the conclusion that this cluster encodes the biosynthesis of a naphthalenic ansamycin (ansalactam) since it has been shown in the ansamitocin\textsuperscript{67} (3), geldanamycin\textsuperscript{4,49,68} (2), and herbimycin\textsuperscript{68} gene clusters that AHBA biosynthesis genes in benzoic ansamycin clusters are organized differently or missing entirely.

![Figure 3.13. Biosynthesis of AHBA by the putative ansalactam gene cluster. (adapted from Floss et al. 2011).\textsuperscript{60}](image)
There are four ORFs that code for PKS genes in the partial putative ansalactam cluster (Table 3.1). However, only the N-terminal PKS ansE encodes a complete PKS module which includes a ketosynthase (KS), acyltransferase (AT), dehydratase (DH), ketoreductase (KR), and acyl carrier protein (ACP). AnsE protein is homologous (79% identity) to the N-terminal PKS, NapE, of the uncharacterized putative naphthalenic ansamycin pathway, nap, from S. hygroscopicus (Table 3.1). AnsE also shares 52% identity to the N-terminal module (module 10) of RifE (accession YP_003762843) from the rifamycin (rif) cluster in A. mediterranei U32.

The other three putative PKS genes, ansA1-3, do not contain the minimum domains necessary to produce a functional PKS (KS-AT-ACP). Directly upstream of ansE, ansA3 encodes for AT–DH–KR–ACP domains and is most identical (70%) to module 1 of NapA (accession ABB86419) from S. hygroscopicus JCM4427 (Table 3.1). AnsA3 lacks a KS domain necessary to perform the claisen condensation of two extender units during polyketide elongation (see Chapter 1). Directly upstream of ansA3, ansA2 encodes a
truncated ACP and KS domain. The introduction of a stop codon following the KS domain and the fact that AnsA2 is most similar to module 3 of NapA (72%) indicates that the current assembly may have compressed the PKS region of the ansalactam cluster and positioned fragments of noncontiguous modules adjacent to one another. This is also evident with ansA1, which shares 82% identity with module 2 of NapA and encodes for a truncated ACP, KS, AT, and DH domains (Table 3.1). Similarly, when ansA1-3 are aligned with either the rif or naphthomycin (nat) PKS genes, the corresponding homologous modules are noncontiguous (data not shown). In addition, much of the intergenic region flanking ans1-3 aligns most similarly to protein coding regions of the nap, rif, and nat PKS genes.

PKS genes have proven a long standing challenge for genome assembly algorithms due to their highly repetitive multimodular architecture. These challenges are greatly increased with the short read lengths of the Genome Analyzer (Illumina) platform (65 bp versus 500 bp for the traditional shot-gun sequencing). Though the PRICE algorithm currently being developed by our collaborators is specifically designed to address such issues, it is apparent that some hurdles must still be overcome.

Regardless of the sequencing errors, the fragmented PKS genes are indicative that we have identified the putative ans gene cluster. Using conserved binding residues of the AT domains, we predict that the substrate specificity of AnsA1 and AnsA3 are malonate and methylmalonate, respectively. As expected, AnsE has a binding motif most similar to AT domains that accept uncommon
PKS extender units (i.e. ethylmalonate or hydroxylmalonate, Table 3.2). The prediction that the AT domain of AnsA1 is specific for the single malonate that is incorporated into the structure of 4 lends further evidence that the current assembly of the PKS region of the putative ans cluster is misassembled, as we would expect this module to be located >20 Kb upstream of ansE.

**Table 3.2.** Predictive residues for acyltransferase substrate specificity of the putative ansalactam PKS ORFs.

<table>
<thead>
<tr>
<th>AT Domain</th>
<th>Conserved Residues</th>
<th>Predicted Substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>AnsA1</td>
<td>QQGHSGIGRHTQV</td>
<td>malonate</td>
</tr>
<tr>
<td>AnsA2</td>
<td>RQGHSQGRSHTNV</td>
<td>methylmalonate</td>
</tr>
<tr>
<td>AnsE</td>
<td>QQGHSQGRGHTNV</td>
<td>ethylmalonate/hydroxymalonate</td>
</tr>
</tbody>
</table>

There are three additional ORFs in the putative ans cluster that share homology to genes in the rif and nat biosynthetic pathways. The amide synthase, ansF, is homologous to rifF and natF which has been shown in the rif pathway to perform the offloading and intramolecular cyclization of the linear PKS product to form the characteristic ansamycin macrolactam (see Chapter 1 for details).\(^{70}\) Also, ans19 encodes for an FAD-dependent 3-(3-hydroxyphenyl)-proprionate hydroxylase that is homologous to rif-orf19 and nat2. Through gene inactivation and complementation studies, it has been shown that rif-orf19\(^{71}\) and nat2\(^{65}\) are involved in the formation of the naphthalene ring and are essential to the biosynthesis of naphthalenic ansamycins. The type II thioesterase, ansR, is predicted to participate as a housekeeping thioesterase when PKS modules are misprimed with the wrong extender unit as shown for RifR in the rif pathway.\(^{72}\)
3.3.5: The Isobutyrylmalonyl-CoA Biosynthetic Pathway

The 35 kb partial ansalactam cluster contains a three gene cassette ans234 not seen in ansamycin biosynthetic gene clusters prior to this study (Figure 3.15). These genes encode for a 3-hydroxybutyryl-CoA dehydrogenase (HBDH), a ketoacylsynthase type III (KASIII), and a crotonyl-CoA reductase/carboxylase (CCR), respectively (Table 3.1), and are transcribed in the opposite direction of the PKS and AHBA genes (Figure 3.12). This three gene cassette is predicted to facilitate the biosynthesis of the ibMCoA polyketide extender unit of 4. Though this cluster had not previously been reported to be involved with secondary metabolism, it appears in the draft genomes of at least four other *Streptomyces* organisms (Figure 3.15). Only the cassettes from *S. roseosporus* appear to be situated near genes for secondary metabolism but to the best of our knowledge has not been linked to the production of known compounds.

Figure 3.15. Orthologous gene neighborhoods of the putative isobutyrylmalonyl-CoA biosynthetic gene clusters in *Streptomyces*. 
The *ans3* encoded KASIII is most similar (55% identity) to a KASIII in *Streptomyces sp.* Mg1 (Table 3.1) that is also clustered between a CCR and HBDH (Figure 3.15). KASIII proteins are related to FabH-like 3-oxoacyl-ACP synthases which are essential to the biosynthesis of branched-chain fatty acids in *Streptomyceses.*\(^7\) Short branched-chain acyl-CoA substrates often derived from amino acid catabolism, such as isobutyryl-CoA and isovaleryl-CoA, are primed by KASIII and condensed with a malonyl-CoA extender unit.\(^73-74\) This mechanism supports the results of our stable isotope feeding experiments, in which the ibMCoA of 4 extender unit was shown to be derived from the condensation of an isobutyrate and malonate unit (as discussed above).

The HBDH of the putative ibMCoA cassette shares greatest identity (53%) with an HBDH from *Nocardia farcinica* (Table 3.1). HBDHs belong to the short-chain dehydrogenase family of oxidoreductases and catalyzes the reduction acetoacyl-CoA to 3-hydroxybutyryl-CoA during the biosynthesis of eMCoA.\(^22,75\) Thus, we proposed it would similarly reduce the 4-dimethyl-acetoacyl-CoA product of the KASIII condensation reaction during ibMCoA biosynthesis.

The CCR of the of the putative ibMCoA cassette is most similar (68%) to a CCR from *Frankia* sp. EUN1f that is adjacent to an uncharacterized type 1 modular PKS (Table 3.1). Phylogenetic analysis of the CCR revealed that it clades with CCRs involved with the biosynthesis of non-eMCoA PKS extender units (Figure 3.16). These include CCRs from the salinosporamide and FK506 pathways for the biosynthesis of chloroethylmalonyl-CoA and allylmalonyl-CoA, respectively.
Figure 3.16. Neighbor-joining tree of CCRs from characterized biosynthetic pathways and uncharacterized genome sequences. Red colored nodes indicate CCRs either confirmed or predicted to be involved in the biosynthesis of unusual PKS extender units. Blue nodes indicate pathway specific CCRs involved in the biosynthesis of ethylmalonyl-CoA.
3.3.6: Efforts Towards the Characterization of the Ansalactam Biosynthetic Gene Cluster.

In order to genetically confirm the putative ansalactam biosynthetic gene cluster and investigate the function of individual pathway enzymes, we targeted three genes (ansK, ans3, and ans4) for inactivation using REDIRECT technology.\textsuperscript{76} As discussed previously, ansK is an AHBA synthase and is essential for the biosynthesis of ansamycins.\textsuperscript{50,77} Thus, we would expect the abolishment of all ansalactam products in the ansK\textsuperscript{−} mutant. The genes ans3 and ans4 were targeted to confirm their functional roles in the biosynthesis of the ibMCoA extender unit. We predicted that inactivation of either of these two genes might lead to the production of new ansalactam analogs with either MCoA, mMCoA, or eMCoA providing the terminal extender unit, similar to what has been observed when the genes required for the biosynthesis of CCR-mediated extender units were inactivated in the salinosporamide\textsuperscript{55} and FK506\textsuperscript{15} pathways.

To inactivate ansK, ans3, and ans4, a pCC2Fos-based (Epicentre) fosmid library was constructed from CNH-189 gDNA. In order to ensure adequate coverage of the genome, 1,344 single clones were transferred into fourteen 96-well plates. The fosmid library was screened by PCR using primer pairs specific for the three genes of interest as well as primers that amplified the terminal ends of the 20 Kb partial putative ansalactam gene cluster from the original assembly (see Appendix Table A3.1 for primer sequences). Two fosmids that contained all three genes of interest were identified on plate 13, P13C5 and P13F6 (Figure 3.17). Unfortunately, neither fosmid was amenable to end sequencing. Thus, the
fosmid boundaries were never determined. The fosmids were however confirmed by sequencing amplified PCR products from the library screening primers (data not shown).

**Figure 3.17.** PCR amplification of genes spanning the 20 Kb partial ansalactam cluster located on pCC2-based fosmids P13C5 and P13F6 used for gene inactivation experiments. Each lane corresponds to (1) ansE, (2) ansK, (3) ans4, (4) ans2, (5) ans3, (6) ans8, (7) negative control, and (L) 1 Kb plus ladder (Invitrogen). See appendix Table A3.1 for primer sequences and expected product sizes.
Using λ Red recombineering in *E. coli* BW25113/pKD20, each targeted gene (*ansK*, *ans3*, and *ans4*) was individually replaced with the apramycin resistance cassette, *acc(3)IV*, on both P13C5 and P13F6. Despite successful transfer of apramycin resistance, the mutant fosmids could not be confirmed by PCR or restriction digest (data not shown). Analysis of both fosmids by agarose gel electrophoresis and restriction digests suggest that an unexpected recombination event occurred, preventing successful inactivation of any of the targeted genes. Short of reconstructing the fosmid library, each step of the experiment was repeated several times without success.

### 3.3.7: Engineering De Novo Synthesis of 36-Methyl-FK506 with Ansalactam Biosynthetic Machinery

In parallel to carrying out the gene inactivation studies, we attempted to heterologously express the putative ibMCoA gene cassette in a Δ*tcsB* mutant strain of the FK506 producer, *Streptomyces* sp. KCTC 11064BP. *tcsB* encodes a ketosynthase (KS) didomain that catalyzes the first step of biosynthesis of the allylmalonyl-CoA extender unit of FK506 (6, Figure 3.18A). Inactivation of *tcsB* yielded a mutant deficient in the production of 6 but instead accumulated the ethylmalonyl-CoA extended FK520 (7, Figure 3.18B). When 18 L of Δ*tcsB* in R2YE media were supplemented with 4-methylpentanoic acid (20.9 g) the mutant produced 1.2 mg of 36-methyl-FK506 (8, Figure 3.18C). Subsequent activity assays showed that 8 had an ~20% greater effect on neurite outgrowth of human neuroblastoma cells treated with growth factor than the natural product, 6. This
presented fortuitous opportunity to both functionally characterize the CNH-189 putative ibMCoA pathway by heterologous expression and to simultaneously engineer de novo synthesis of a potent FK506 analog.

**Figure 3.18.** Proposed biosynthesis of the allylmalonyl-CoA extender unit of FK506 and generation of FK506 analogs by mutasynthesis. Biosynthetic schemes for (A) the biosynthesis of FK506 (6) by wild-type *Streptomyces* sp. KCTC 11064BP (adapted from Mo et al. 2011), (B) inactivation of *tcsB* to yield FK520 (7), and (C) cultures of ΔtcsB supplemented with 4-methyl-pentanoic acid to yield 36-methyl-FK506 (8).
To heterologously express the putative ibMCoA pathway in the ΔtcsB mutant, two pSET152-based plasmids were constructed. For the first plasmid, pMW01, the ibMCoA cassette was spliced with the constitutive promoter, *ermE*⁺, by splicing by overlap extension (SOE) PCR before being cloned into pSET152 (Figure 3.19). The second plasmid, pMW02, was constructed by cloning the putative ibMCoA cassette with an additional 200 bp of the *ans4* native promoter region into pSET152. Each plasmid was introduced to the ΔtcsB mutant via conjugation with *E. coli* ET12567/pUZ8002. Successful integration of the ibMCoA cassette was confirmed by PCR.

![Diagram](image)

**Figure 3.19.** Proposed scheme for *de novo* synthesis of 36-methyl-FK506 in the *Streptomyces* sp. ΔtcsB mutant. The 3.3 Kb ibMCoA pathway was spliced with the *ermE*⁺ promoter by PCR and cloned into pSET152. The construct was transferred to the ΔtcsB mutant via conjugation.
After six days of cultivation in R2YE medium, ethyl acetate extracts of the exconjugates were analyzed by electrospray ionization liquid chromatography mass spectroscopy (ESI-MS) and high performance liquid chromatography (HPLC). Unfortunately, no observable differences were detected between the mutants and negative control (ΔtcsB) (data not shown). Due to time constraints, it was not possible to verify if the pathway was indeed being transcribed.

3.4: Conclusions

The isolation and structural determination of ansalactam A (4) marks not only the discovery of a new ansamycin natural product but also a novel PKS building block, isobutyrylmalonyl-CoA. Utilizing traditional stable isotope incorporation studies, we showed that the ibMCoA extender unit of 4 is derived from the condensation of isobutyrate and malonate units, indicating that ibMCoA biosynthesis is analogous to fatty acid biosynthesis rather than amino acid degradation as originally proposed (Figure 3.20).
Figure 3.20. Summary of $^{13}$C-labeled precursor studies showing the biosynthetic origin of the carbons of the linear polyketide structure of 4 compared to well characterized ansamycins, rifamycin and naphthomycin.
Further investigation of the CNH-189 genome led to the identification of a 35 Kb contig containing a partial putative ansalactam biosynthetic gene cluster, consisting of all the genes necessary for the biosynthesis of the AHBA starter unit and a putative ibMCoA pathway. Unfortunately, we did not identify a complete set of PKS genes necessary for the biosynthesis of 4. In retrospect, we discovered that this may be largely due to the preparation of the CNH-189 sequencing library. Unknown to us at the time, we learned that the standard PCR reaction of the Illumina sequencing is inherently biased against high G+C organisms, such as Streptomyces.\textsuperscript{79-80} De novo assembly of the S. coelicolor genome using PRICE showed that sequencing coverage dropped to zero approximately three times per 10 Kb across the genome, mainly in areas where the G+C exceeds 75% (Michael Fischbach, personal communication). In order to address these issues, we may repeat the sequencing of the CNH-189 genome using updated protocols optimized for high G+C organisms.\textsuperscript{79} More immediately, we may expand and address the sequencing errors of our current assembly by primer walking. However, this might prove tedious as we would expect an additional 35–60 Kb of sequence information to achieve the entire ansalactam cluster.

For now we can predict, based on the highly conserved domain architecture of naphthalenic ansamycin gene clusters,\textsuperscript{13,52,65} that the ans PKS region will span approximately 50 Kb and consist of four to five modular proteins with a loading module and eight extension modules (Figure 3.21).
Figure 3.21. Proposed biosynthesis of ansalactam A (4) and gene organization of the polyketide region of the ansalactam biosynthetic gene cluster. Solid red arrows indicate predicted PKS fragments from the current 35 Kb partial putative ansalactam cluster according to best BLAST analysis against known ansamycin pathways. White and red arrows indicate the predicted organization of ansalactam PKS genes based on organization of characterized and homologous ansamycin gene clusters.
Unfortunately, attempts to confirm the putative ansalactam pathway experimentally failed. Regardless, the striking homology to the characterized biosynthetic gene clusters of rifamycin and naphthomycin support the putative assignment of the ansalactam biosynthetic gene cluster. Additionally, the most convincing evidence comes from the recent discovery of divergolides A–D (9–12, Figure 3.21), novel ansamycin natural products from an epiphytic *Streptomyces* species (strain HK10576) isolated from a mangrove tree.\(^\text{43}\) Coincidentally, while we were isolating and exploring the biosynthesis of 4, Hertweck and co-workers isolated and began biosynthetic studies on 9–12 which contain the same novel ibMCoA derived PKS extender unit. In addition, they identified two new germicidin analogs that also incorporate the ibMCoA derived structural unit from the same *Streptomyces* (Figure 3.22).\(^\text{44}\)
Figure 3.2. Structures of divergolides A–D (9–12) and germicidins F (13) and G (14) isolated from *Streptomyces* sp. HK10576.
Through shotgun sequencing of the HK10576 genome, Hertweck and co-workers identified both the divergolide and germicidin biosynthetic gene clusters.\textsuperscript{44} Similar to the putative ansalactam pathway, the divergolide biosynthetic gene cluster contains a three gene cassette ($ccr$-$kasIII$-$hbdh$) for the formation of ibMCoA, whereas no such genes are located near the germicidin cluster.\textsuperscript{44} Interestingly, Hertweck and colleagues were also unable to successfully inactivate the ibMCoA biosynthetic genes in HK10576.\textsuperscript{44} Instead, they choose heterologously express the ibMCoA cassette in *Streptomyces albus* along the type III PKS from the germicidin pathway. Only when expressed together did the heterologous host, *S. albus*, produce the ibMCoA derived germicidins (13 and 14).\textsuperscript{44} The authors concluded that the three gene ($ccr$-$kasIII$-$hbdh$) operon from *S. sp.* HK10576 is responsible for the biosynthesis of IbMCoA from isobutyryl-CoA and malonyl-CoA precursors (Figure 3.23). This is consistent with our stable isotope incorporation studies and bioinformatic proposal of the CNH-189 ibMCoA pathway.
Figure 3.23. Proposed biosynthesis of isobutyrylmalonyl-CoA in Streptomyces sp. CNH-189 (adapted from Xu et al. 2011).

In conclusion, we successfully deduced the biosynthesis of a novel ansamycin natural product and its unusual PKS extender unit through a multidisciplinary approach that included stable isotope incorporation studies and bioinformatic analysis. Efforts to functionally characterize the ibMCoA pathway are still being explored and will hopefully lead to the engineering of novel bioactive compounds. With the recent discovery of the divergolides (9–12) and germicidins F (13) and G (14), along with the identification of ibMCoA-like operons in at least four other Streptomyces species, we have a new target for genome mining of unusual PKS natural products. Additionally, successful heterologous expression of the ibMCoA pathway may also provide useful tool for combinatorial biosynthesis of novel “unnatural” products.
3.5 Experimental Methods

3.5.1 General Experimental Procedures.

The optical rotations were measured using a Rudolph Research Autopol III polarimeter with a 10-cm cell. UV spectra were recorded in a Varian Cary UV-visible spectrophotometer with a path length of 1 cm and IR spectra were recorded on a Perkin-Elmer 1600 FT-IR spectrometer. $^1$H and 2D NMR spectra data were recorded at 500 or 600 MHz in DMSO-$d_6$, pyridine-$d_5$, or methanol-$d_4$ solution containing Me$_4$Si as internal standard on Varian Inova spectrometers. $^{13}$C NMR spectra for structure determination were acquired at 75 MHz on a Varian Inova spectrometer. $^{13}$C NMR spectra for $^{13}$C-labeling studies were acquired at 125 MHz on a Varian VX-500 spectrophotometer equipped with an XSens cold probe. High resolution ESI-TOF mass spectra were provided by the Scripps Research Institute, La Jolla, CA or by the mass spectrometry facility at the UCSD Department of Chemistry. Low resolution LC-MS data were measured using a Hewlett-Packard series 1100 LC-MS system with a reversed-phase C$_{18}$ column (Phenomenex Luna, 4.6 mm × 100 mm, 5 $\mu$m) at a flow rate of 0.7 mL/min.

3.5.2 Cultivation and Extraction.

The bacterium (strain CNH-189) was isolated from a near-shore marine sediment collected off Oceanside, California. It was identified as a Streptomyces sp. based on 16S rRNA gene sequence analysis (accession number HQ214120). It was cultured in sixty 2.8 L Fernbach flasks each containing 1 L of A1
production media (10 g starch, 4 g yeast extract, 2 g peptone, 1 g CaCO\textsubscript{3}, 40 mg Fe\textsubscript{2}(SO\textsubscript{4})\textsubscript{3}•4H\textsubscript{2}O, 100 mg KBr) and shaken at 230 rpm at 27 °C. After seven days of cultivation, sterilized XAD-16 resin (20 g/L) was added to adsorb the organic products, and the culture and resin were shaken at 215 rpm for 2 hours. The resin was filtered through cheesecloth, washed with deionized water, and eluted with acetone. The acetone was removed under reduced pressure, and the resulting aqueous layer was extracted with EtOAc (3 x 500 mL). The EtOAc-soluble fraction was dried in vacuo to yield 4.5 g of crude extract.

3.5.3 Isolation of Ansalactam A.

The crude extract (4.5 g) was fractionated by open column chromatography on silica gel (25 g), eluted with a step gradient of CH\textsubscript{2}Cl\textsubscript{2} and MeOH. The dichloromethane/methanol 50:1 fraction contained a mixture of ansalactams, which were purified by reversed-phase HPLC (Phenomenex Luna C-18 (2), 250 × 100 mm, 2.0 mL/min, 5 μm, 100 Å, UV = 210 nm) using an isocratic solvent system from 65% CH\textsubscript{3}CN to afford ansalactam A (4, 70.0 mg).

\textbf{Ansalactam A (4):} yellow oil; [α]\textsubscript{D}\textsuperscript{21} -78 (c 0.5, MeOH); IR (KBr) \nu_{\text{max}} 3354, 2962, 1691, 1331, 1025, 746 cm\textsuperscript{-1}; UV (MeOH) \lambda_{\text{max}} (log ε) 225 (4.4), 250 (4.3), 290 (4.2), 340 (3.2) nm; \textsuperscript{1}H and 2D-NMR (500 MHz, methanol-d\textsubscript{4}), see Table 1; HRESIMS [M+H]\textsuperscript{+} m/z 546.2852 (C\textsubscript{33}H\textsubscript{40}NO\textsubscript{6}, calcd [M+H]\textsuperscript{+} 546.2850).

3.5.4 Reduction of Ansalactam A (4) to Yield Polyol 5.

Ansalactam A (4, 20 mg) was dissolved in 2 mL of dry MeOH. Twenty milligrams of NaBH\textsubscript{4} was added to the solution, and the reaction mixture was
stirred for 2 hours. The reaction was quenched by 2.5% aqueous NH₄Cl solution and then the mixture was extracted with EtOAc. The solvent was removed in vacuo and the residual material was purified by reversed-phase HPLC (Phenomenex Luna 5 µ C18 (2) 100 Å, 250 x 100 mm, 2.0 mL/min, UV detection at 210 nm) using an isocratic solvent system with 65% CH₃CN in H₂O. The polyol 5 (13.0 mg) was obtained with a retention time of 18 min.

3.5.5 Identification of *Streptomyces* sp. Strain CNH-189 AHBA Synthase.

Genomic DNA from *Streptomyces* sp. strain CNH-189, *S. arenicola* strain CNS-205, and *S. tropica* strain CNB-440 was extracted and purified using standard protocols. The 3-amino-5-hydroxybenzoic acid (AHBA) synthase gene (accession number HQ219709) was amplified from gDNA from each strain using primers and PCR conditions described by Huitu et al. The ~755 bp PCR-product from strain CNH-189 was recovered by gel electrophoresis and purified using a DNA fragment purification kit (Qiagen). The resulting DNA fragment was sequenced at both ends by Seqxcel (San Diego, CA). A neighbor joining distance tree was constructed in MEGA4 using aligned partial sequences (695 bp) from the NCBI database comprising AHBA synthase genes from characterized ansamycin biosynthetic gene clusters and from genome sequencing projects with no confirmed metabolite.

3.5.6 Stable Isotope Incorporation Studies.

For each of the stable isotope enrichment experiments, 20 mL of CNH-189 starter culture was inoculated into 2 X 1 L of A1 production media in 2.8 L Fernbach flasks and allowed to grow for 24 hrs before the addition of labeled
substrate – sodium [1-13C]propionate, 50 mg/L; sodium [1,2-13C]acetate, 100 mg/L; sodium [1,2-13C]leucine, 100 mg/L; sodium [1-13C]leucine, 50 mg/L; [1-13C]isobutyrate, 30 mg/L; [2-13C]4-Me-2-pentenoic acid, 50 mg/L – dissolved in H2O. Cultures were grown for an additional 96 hrs before 4 was extracted, purified, and reduced to 5 for 13C NMR analysis as described above. 13C-enrichment was calculated as described by Bringmann et al. 83

3.5.7 Synthesis of [2-13C]4-Methyl-2-pentenoic Acid.

[2-13C]4-Methyl-2-pentenoic acid was synthesized from [2-13C]malonic acid and isobutyraldehyde by Doebner condensation according to reference 23. The spectroscopic data obtained was in good agreement with the literature: δH (CDCl3, 500 MHz) 7.06 (1H, dd, J = 15.7, 6.5 Hz, CHCOOH), 5.63 (1H, d, J = 15.7 Hz, CHCH(CH3)2), 2.50 (1H, m, CH(CH3)2), 1.09 (6H, d, J = 6.7 Hz, CH(CH3)2). 84

3.5.8: Bioinformatic Identification of the Partial Ansalactam Biosynthetic Gene Cluster

To identify the ansalactam biosynthetic gene cluster, gDNA from CNH-189 was isolated according to standard protocols 81 and prepared for sequencing on a Genome Analyzer IIx (Illumina) platform according to the manufacture’s protocol. 2,832,824 paired-end 65 bp reads were generated providing approximately 300 fold coverage of the CNH-189 genome. The initial assembly of the genome was carried out utilizing default parameters for a medium sensitivity assembly an 10,000 contig cut-off in the Geneious bioinformatics software package. 59 BLASTn analysis of the 10,000 contigs using the previously identified AHBA
synthase, *ansK*, as a query identified a 100% identity match to the 20 Kb contig, contig 10. Contig 10 was extended an additional 15 kb on the C-terminal end by our collaborators using the Paired-Read Iterative Contig Extension (PRICE) assembler (http://derisilab.ucsf.edu/software/price/index.html). ORFs were assigned using Geneious and manually annotated using BLAST\textsuperscript{48} analysis.

### 3.5.9: Progress Towards Inactivation of Ansalactam Biosynthetic Genes

Efforts to inactivate *ansK*, *ans3*, and *ans4* from the putative ansalactam biosynthetic gene cluster were carried out using REDIRECT technology\textsuperscript{76}. Briefly, a Copy Control pCC2-based (Epicentre Biotechnologies) fosmid library was constructed in *E. coli* EPI-300 following the manufacture’s protocol from gDNA isolated from CNH-189. 1,344 clones were screened by PCR using DNA primers designed to amplify a 450–500 bp fragment of genes *ansK*, *ans2*, *ans3*, and *ans4* (Appendix Table A3.1). Two fosmids, P13C5 and P13F6, carrying the three genes of interest were identified and confirmed by sequencing of PCR products from the three targeted genes. The two fosmids were isolated from the EPI-300 cells using a Miniprep Kit (Qiagen) and transformed into *E. coli* BW25113/pKD20 carrying the λ Red genes for recombination. Following the REDIRECT protocol, the three targeted genes were individually replaced in both P13C5 and P13F6 via homologous recombination with the apramycin resistance cassette, *acc(3)IV*, from pIJ773 (see Appendix Table A3.2 for gene inactivation and confirmation primers). Unfortunately, we were never able to confirm successful gene replacement in the BW25113 cells most likely due to an unexpected
recombination event in the fosmids that prevented the mutation (data not shown). Despite several attempts to repeat the experiment, no ansalactam mutants were generated.

3.5.10: Mutasynthesis of 36-Methyl-FK506 in *Streptomyces* sp. KCTC 11064BP ΔtcsB

In an attempt to metabolically engineer the production of 34-methyl-FK506 *de novo*, we integrated the three gene cassette for isobutyrylmalonyl-CoA biosynthesis from the putative *ans* cluster into the genome of a mutant strain of *Streptomyces* sp. KCTC11064BP, ΔtcsB, deficient in the production of FK506. To do this, two pSET152 constructs were generated. For the first construct, pMW01, the *ermE* and 3.3 Kb ibMCoA cassette, *ans432*, were spliced together by splicing by overlap extension (SOE) PCR. The 3.3 Kb ibMCoA cassette was PCR amplified from CNH-189 gDNA with Phusion High Fidelity DNA polymerase (New England Biolabs) and primers ibMCass_FP and ibMCass_RP. While the *ermE* promoter was PCR amplified from the vector pUWL201 with the primers ermEp_FP and ermEp_RP. A 35 bp sequence overlap was added was added to ibMCass_FP and ermE_RP to allow for splicing. SOE PCR using Expand High Fidelity DNA polymerase (Roche) was then used to splice the *ermE* and ibM cassette together using primers ermEp_FP and ibMCass_RP yielding a 3.6 bp product that was subsequently cloned into pSET152 at the *EcoR* I and *Xba I* sites to generate pMW01. For second construct, pMW02, the ibM cassette including a 200 bp native promoter region located upstream of *ans4* was
amplified from the gDNA of CNH-189 and cloned into pSET152. See Appendix Table A3.3 for a list of all primers used for heterologous expression of the ibMCoA cassette.

pMW01 and pMW02 were introduced into the ΔtcsB mutant via conjugation with *E. coli* ET12567/pUZ8002 and exconjugates were selected in the presence of apramycin (50 µg/mL) using the standard protocol. Successful integration of the ibMCoA cassette into the genome of the ΔtcsB mutant was verified by PCR using primers KASIII_CKF and KASII_CKR (Appendix Table 3.1).

To evaluate the production of 36-methyl-FK506, 25 mL starter cultures of tryptic soy broth (TSB) with 50 µg/mL apramycin and 25 µg/mL nalidixic acid were inoculated with 500 µL of a spore stock generated from either heterologous expression mutant using the standard protocol. After 3 days of fermentation while shaking at 30°C, a 10% inoculum of starter culture was transferred to 50 mL of R2YE media and allowed to shake at 30°C for an additional 6 days. The cultures were then extracted 3X with EtOAc and dried *in vacuo*. The crude extract was analyzed by LC-MS and HPLC using the same conditions described above for the ansalactams.
3.6 Appendix

Appendix Table A3.1. Primers synthesized for screening the CNH-189 fosmid library.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Oligonucleotide</th>
<th>Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>189LIB_NapE_F</td>
<td>ACGTGGGCTCGCCGGACGCT</td>
<td>470</td>
</tr>
<tr>
<td>189LIB_NapE_R</td>
<td>CGCTGCCGCCAGCACACTGAC</td>
<td></td>
</tr>
<tr>
<td>189LIB_AHBA_F</td>
<td>CTGCAGGACCACGCCCACGC</td>
<td>483</td>
</tr>
<tr>
<td>189LIB_AHBA_R</td>
<td>AGCAAGCCAGCAGCCGGACGA</td>
<td></td>
</tr>
<tr>
<td>189LIB_HBDH_F</td>
<td>AACCCGCGCCCCTCTTGCGG</td>
<td>517</td>
</tr>
<tr>
<td>189LIB_HBDH_R</td>
<td>GCGGACCCTTCGCCGCCTTC</td>
<td></td>
</tr>
<tr>
<td>189LIB_KASIII_F</td>
<td>GGCCTGCCCCGCTCAACCA</td>
<td>450</td>
</tr>
<tr>
<td>189LIB_KASIII_R</td>
<td>CGCGCCTACGCGGACCAGGA</td>
<td></td>
</tr>
<tr>
<td>189LIB_CCR_F</td>
<td>GGGCAGCCCAGCAGCCTTCT</td>
<td>452</td>
</tr>
<tr>
<td>189LIB_CCR_R</td>
<td>TCCTGGCTCGGAGCCGCGCC</td>
<td></td>
</tr>
<tr>
<td>189LIB_HYP_F</td>
<td>AGGCCGCTCGCTACCCAGGG</td>
<td>504</td>
</tr>
<tr>
<td>189LIB_HYP_R</td>
<td>CCGGGCTTCCGTGGCGAGCA</td>
<td></td>
</tr>
</tbody>
</table>

Appendix Table A3.2. Primers synthesized for gene inactivation experiments in CNH-189

<table>
<thead>
<tr>
<th>Inactivation Primers</th>
<th>Oligonucleotide</th>
<th>Verification Primers</th>
<th>Oligonucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHBA_KO_FP</td>
<td>GGCCGAGAAATGTCCGAAGTGAGATTTGAGATCATAGATTCGCGGATCCGCTGACC</td>
<td>189AHBA_CKF</td>
<td>TGGGGTGGTGAGCCGGCCGAG</td>
</tr>
<tr>
<td>AHBA_KO_RP</td>
<td>CGGCCGCGAATGTCCGAAGTGAGATTTGAGATCATAGATTCGCGGATCCGCTGACC</td>
<td>189CCR_CKF</td>
<td>GCCTGCCGAGGAGCACCGAG</td>
</tr>
<tr>
<td>KASIII_KO_FP</td>
<td>GCCGTACGCGACGACACGGTACGGGCGGCCGGATGCCGACCATGATTCGCGGATCCGCTGACC</td>
<td>189CCR_CKR</td>
<td>GCCTGCCGACGACACGGTACGGGCGGCCGGATGCCGACC</td>
</tr>
<tr>
<td>KASIII_KO_RP</td>
<td>GTTCACGCGAGGAGCACGGTACGGGCGGCCGGATGCCGACCATGATTCGCGGATCCGCTGACC</td>
<td>189KASIII_CKF</td>
<td>CTGGCGAGGAGCACGGTACGGGCGGCCGGATGCCGACC</td>
</tr>
<tr>
<td>CCR_KO_FP</td>
<td>GGGGCTGAGGAGGAGCACGGTACGGGCGGCCGGATGCCGACC</td>
<td>189KASIII_CKR</td>
<td>CTGGCGAGGAGCACGGTACGGGCGGCCGGATGCCGACC</td>
</tr>
</tbody>
</table>
**Appendix Table A3.3.** Primers synthesized for heterologous expression of the isobutyrylmalonyl-CoA biosynthetic gene cassette in the *tcsB* mutant.

<table>
<thead>
<tr>
<th>pMW01 Construct Primers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ibMCass_FP</td>
<td>ACAATCGTGCCGGTTGTTAGGATCCAGCGATGTCACCTCTCCGATGCA</td>
</tr>
<tr>
<td>ibMCass_RP</td>
<td>AAGCTTGGGCTGAGGCAGACTCTTAGAGGGTCAGGATCCGGAGGACTCGT</td>
</tr>
<tr>
<td>ermEp_FP</td>
<td>CAGCTATGACATGATTACGAATTCGATATCGTACGAGCCCAACCGAGCC</td>
</tr>
<tr>
<td>ermEp_RP</td>
<td>GCACCACGCATGCATCGGAGAGTGACATCTCGTGGATCCTACCA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pMW02 Construct primers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ibMCassNP_FP</td>
<td>CAGCTATGACATGATTACGAATTCGATATCGACCGCGCCGCCGCCGCC</td>
</tr>
<tr>
<td>ibMCass_RP</td>
<td>AAGCTTGGGCTGAGGCAGACTCTTAGAGGGTCAGGATCCGGAGGACTCGT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pSET152 Sequencing Primers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pSET_SEQF</td>
<td>GCGGGCAGTGAGGCAACGCA</td>
</tr>
<tr>
<td>pSET_SEQR</td>
<td>CGCCTTGCACATCCCCCT</td>
</tr>
</tbody>
</table>

Restriction sites for *EcoR* I (GATTA) and *Xba* I (TCTAGA) are underlined.
Appendix Figure A3.1. $^{13}$C NMR spectrum (125 MHz) of 5 not enriched from [1-$^{13}$C]leucine feeding study in methanol-$d_4$.
Appendix Figure A3.2. $^{13}$C NMR spectrum (125 MHz) of 5 from [1-$^{13}$C]isobutyrate feeding study in methanol-$d_4$ (expanded).
Appendix Figure A3. $^{13}$C NMR spectrum (125 MHz) of 5 from [2-$^{13}$C]4-Me-2-pentenoic acid feeding study in methanol-d$_4$ (expanded).

$^{13}$C NMR of unlabeled 5 (40-47 ppm)
3.7: References


(60) Floss, H. G.; Yu, T.-W.; Arakawa, K. The biosynthesis of 3-amino-5-hydroxybenzoic acid (AHBA), the precursor of mC7N units in ansamycin and mitomycin antibiotics: a review. *J Antibiot* **2011**, *64*, 35-44.


3.8: Acknowledgements

I thank Dr. Sang-Jip Nam for providing ansalactam A authentic standard and for a productive collaboration on constructing the published manuscript. I also thank Dr. Tobias A. M. Gulder for synthesizing the [2-^{13}C]4-methyl-pentenoic acid for the stable isotope incorporation studies, Dr. Peter Bernhardt for providing CNH-189 gDNA for pCC2 library construction and leading the whole genome sequencing of CNH-189, and Dr. Leonard Kaysser for many discussions regarding troubleshooting the gene inactivation and heterologous expression studies. Lastly, I thank Kelle Freel for providing the CNH-189 strain, 16S sequence, and CNH-189 gDNA used to amplify the AHBA synthase.

Chapter 3, in part, is a reprint of the material as it appears in Structure and Biosynthesis of the Marine Streptomycete Ansamycin Ansalactam A and Its Distinctive Branched Chain Polyketide Extender Unit (2011) Wilson, Micheal C.; Nam, Sang-Jip; Gulder, Tobias A.M.; Kauffman, Christopher A.; Jensen, Paul R.; Fenical, William; Moore, Bradley S., Journal of the American Chemical Society, 133, 1971-1977. The dissertation author was the primary investigator and author of this paper.
Chapter 4: Conclusions and Future Directions
4.1: General Conclusions

As long as humans exist we will strive to live longer and be healthier as a species. Despite our efforts, infectious diseases, cancers, and aging will continue to present hurdles that must be overcome to maximize those goals. Similar to organisms that produce biologically active metabolites, our ability to exploit and harness the therapeutic value of these small molecule natural products has provided an adaptive advantage towards achieving longer and healthier lives. However, overuse and misuse of antibiotics and other drugs has also selected for drug resistant pathogens that are evolving rapidly to carve out their own ideal biological niche. As more than half of all approved drugs are derived from or modeled after natural products,¹ we must have a detailed knowledge of our natural product weapons to be able to combat these resistant pathogens. This includes knowing their mechanism of action, their molecular targets, how are they biosynthesized, and how can they be engineered to function more efficiently. This all begins by discovering or developing new chemical entities.

There are several literal and virtual landscapes by which natural product drug discovery can be pursued. As the rate of discovery of new drug classes decreased and the repetitive isolation of known compounds increased, the pharmaceutical industry shifted away from natural product based drug discovery and moved towards the generation and screening of high-throughput combinatorial libraries.² Still, natural products shine as the most promising therapeutics with ~50% of new chemical entities between the year 2000 – 2006 coming from natural sources and another 30% of therapeutics categorized as
biologicals or vaccines. The continued success rate of natural products drug
discovery largely relies on exploring untapped and rare habitats for biologically
diverse natural product producers. In 1951, Bergmann’s discovery of the unusual
nucleosides spongouridine and spongothymidine from the marine sponge
Chyptotethya crypta introduced natural products chemists to the oceans as an
valuable resource for new potential drug leads. Since then, more than 20,000
compounds have been isolated from marine organisms. Though, most marine-
derived natural products are reported from marine invertebrates (mostly
sponges), it is now largely believed that many of these compounds instead
originate from symbiotic or associated microbial populations. As continually
proven with their terrestrial relatives, marine-derived actinomycetes are prolific
producers of potent pharmaceutically relevant natural products. Found in the
oceans as free living sediment-dwelling bacteria or associated with marine
invertebrates and fishes as on land, actinomycetes are ubiquitous. Both
the common Streptomyces and more rare actinomycetes are underrepresented
using current culturing methodologies. Thus, emphasis must be placed on
development of diverse fermentation conditions in order to access many of the
less well represented taxa and subsequently, their elusive chemistry.

One virtual landscape upon which natural product drug discovery has
great potential is bioinformatics-guided mining of microbial genomes. Whole
genome sequencing of more than 151 Actinobacteria has been completed to
date, most of which belong to the order Actinomycetales. Having access to
total.

Entire genomes have not only shown that the number of known natural products

represents only a small fraction of the true biosynthetic potential of these organisms but also provides the map towards accessing the bulk of microbial medicinal chemistry.\textsuperscript{21,23} To take full advantage of genome-guided drug discovery, we must also know how to turn on or upregulate the expression of orphan pathways for which no compounds have yet been associated.\textsuperscript{24} Alternatively, heterologous expression of whole pathways in organisms amenable to large scale fermentation is another attractive avenue to access this hidden chemistry.\textsuperscript{25} But, first it is important to know how natural products are biosynthesized \textit{de novo}.

To this end, this dissertation portrays my contributions towards the understanding the biosynthesis of ansamycin natural products from marine-derived actinomycetes. Based on structural similarities between the ornithine decarboxylase inhibitors, saliniketals A and B, and the well-known rifamycin antibiotics, I explored their mutual biosynthesis in the obligate marine actinomycete, \textit{Salinispora arenicola} CNS-205. Using both traditional isotope labeling studies and modern gene inactivation experiments, I demonstrated that the saliniketals are in fact shunt products of the rifamycin pathway. I also showed that the saliniketals are products of the original rifamycin producer, a terrestrial actinomycete named \textit{Amycolatopsis mediterranei}. Through extensive gene inactivation and chemical complementation studies, I showed that a cytochrome P450 (CYP) encoded by \textit{sare1259} facilitates a branching of rifamycin biosynthesis to yield either the saliniketals or the mature rifamycins. Since \textit{in vitro} characterization of the CYP has so far been unsuccessful, it remains uncertain
the extent to which Sare1259 catalyzes the conversion of 34a-deoxyrifamycin W to the saliniketals. If Sare1259 is the only enzyme necessary for the bioconversion, it will represent unprecedented multifunctional CYP chemistry and could prove to be an interesting biocatalyst for engineering novel compounds. Hopefully, continued efforts towards the characterization of this remarkable enzyme will be pursued.

In addition to exploring the biosynthesis of the saliniketals, I investigated the unusual biosynthesis of the novel ansamycin, ansalactam A, from a marine-derived *Streptomyces* collected offshore of Oceanside, CA. Similar to the saliniketals, ansalactam A was most interesting for its unusual structural features rather than its biological activity. Ansalactam A incorporates a novel polyketide extender unit derived from isobutyrylmalonyl-CoA (ibMCoA) that was unprecedented at the time of original structural elucidation and preliminary biosynthetic investigations. Through stable isotope feeding experiments, I showed that this novel extender unit is derived from the condensation of isobutyrate and malonate precursors. We proposed that a dedicated ketoacylsynthase (KAS) performed the condensation of the two precursors followed by reductive carboxylation of the reduced diketide by a crotonyl-CoA carboxylase/reductase (CCR) homolog. Genome sequencing of *Streptomyces* sp. CNH-189 indeed revealed a partial ansalactam biosynthetic gene cluster with an curious three gene cassette consisting of a KAS type III, 3-hydroxybutyryl-CoA dehydrogenase (HBDH), and a CCR. While investigating the biosynthesis of ansalactam A, Hertweck and colleagues identified two additional natural products
which incorporate ibMCoA and additionally discovered and heterologously expressed the ibMCoA biosynthetic gene cassette.\textsuperscript{26} Despite not being able to, as of yet, experimentally confirm the role of these genes in the ansalactam cluster, the work by the Hertweck laboratory suggests that our original proposal and bioinformatic assignment of the ibMCoA biosynthetic genes are scientifically sound.

The work presented in this dissertation represents the alignment of several disciplines to achieve a comprehensive understanding of natural product (ansamycin) biosynthesis. The marriage of chemistry, molecular biology, genetics, and informatics along with exploration of diverse habitats provides the optimal toolkit for combating infectious diseases, cancer, and numerous human ailments. It is my hope that the research presented here has laid the foundation for understanding the biosynthesis of a pharmaceutically important class of compounds and leads towards the bioengineering new chemical entities for use as potential therapeutics.

4.2. References


