Title
Bridging the Gap: Studying Sequence to Product Correlation among Fungal Polyketide Synthases

Permalink
https://escholarship.org/uc/item/8q12b08v

Author
Zabala, Angelica Obusan

Publication Date
2014

Peer reviewed|Thesis/dissertation
Bridging the Gap: Studying Sequence to Product Correlation among Fungal Polyketide Synthases

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

by

Angelica Obusan Zabala

2014
ABSTRACT OF THE DISSERTATION

Bridging the Gap: Studying Sequence to Product Correlation among Fungal Polyketide Synthases

by

Angelica Obusan Zabala

Doctor of Philosophy in Chemical Engineering
University of California, Los Angeles, 2014
Professor Yi Tang, Chair

Polyketide synthases (PKSs) catalyze the production of important pharmaceuticals, including statins and antibiotics. Fungi are some of the most prolific producers of these compounds but still have untapped potential as demonstrated by the abundant secondary metabolic clusters uncovered from the genome mining efforts. In order to better understand the correlation between the fungal Type I PKSs and their products for predictive and engineering purposes, it is necessary to biochemically characterize these enzyme systems. To do so, two distinct approaches were used. In the first bottom-up approach, we demonstrated that we can take the sequence of a cluster and elucidate its products by selective activation. This led to the discovery of new compounds
belonging to the azaphilone group of natural products. We studied key aspects on the biosynthesis of this natural product group, particularly the formation of the characteristic pyrano-quinone core. This helped facilitate the discovery of new azaphilone clusters from the broad fungal genome space. In the second top-down approach, we studied the biosynthesis of the potent protein-transport inhibitor Brefeldin A. By sequencing the producing organism and using detailed bioinformatic analysis, we localized the cluster responsible for its production. More importantly, we provide new insights in the programming of these enigmatic HRPKSs, particularly in the permutative use of the reductive domains and the chain-length control by an in trans thiohydrolase.
The dissertation of Angelica Obusan Zabala is approved.

James Liao

Neil Garg

Yi Tang, Committee Chair

University of California, Los Angeles

2014
# Table of Contents

1 Introduction .......................................................................................................................... 1

1.1 Natural products: Spotlight on Polyketides ........................................................................ 1

1.2 Polyketide synthases .......................................................................................................... 3
  
  1.2.1 Fungal Iterative Polyketide Synthases ......................................................................... 7

1.3 Methods for studying PKS gene clusters in the genomic era ............................................. 15
  
  1.3.1 Genetic manipulation of native host ............................................................................. 15
  
  1.3.2 Heterologous expression of clusters and individual genes ........................................... 17
  
  1.3.3 In vitro reconstitution of protein activity ....................................................................... 18

1.4 Engineering PKSs and tailoring enzymes for NP diversification .................................... 19
  
  1.4.1 Engineering PKSs ....................................................................................................... 19
  
  1.4.2 Engineering tailoring enzymes ................................................................................... 28

1.5 Motivation for study ........................................................................................................... 33

2 Results .................................................................................................................................. 34

2.1 Discovery and Biosynthesis of Azanigerones from *Aspergillus niger* ATCC 1015 ...... 34
  
  2.1.1 Genome mining in *Aspergillus niger* ATCC 1015 ..................................................... 34

  2.1.2 Characterization of a Silent Azaphilone Gene Cluster from *Aspergillus niger* ATCC 1015 Reveals a Hydroxylation-Mediated Pyran-Ring Formation.” .......................... 42

  2.1.3 Heterologous expression of *aza* proteins ........................................................................ 68

  2.1.4 Bioactivity assays on azanigerone A ............................................................................. 73
### 2.2 Biosynthesis of Brefeldin

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1</td>
<td>Significance and activity of Brefeldin A</td>
<td>74</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Genome mining for BFA-producing cluster in <em>Eupenicillium brefeldianum</em> ATCC 58665</td>
<td>75</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Fungal Polyketide Synthase Product Chain-Length Control by Partnering Thiohydrolase</td>
<td>82</td>
</tr>
</tbody>
</table>

### 3 Conclusion

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
</tr>
</tbody>
</table>

### 4 Appendix

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Supplementary information for Section 2.1.2</td>
<td>122</td>
</tr>
<tr>
<td>4.2</td>
<td>Supplementary information for Section 2.1.3</td>
<td>158</td>
</tr>
<tr>
<td>4.3</td>
<td>Supplementary information for Section 2.2.3</td>
<td>162</td>
</tr>
</tbody>
</table>

### 5 References

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>187</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1.1. Polyketide natural products from different sources.................................................. 2
Figure 1.2. Divergence between PKS and FAS functions......................................................... 4
Figure 1.3. Schematic representation of the biosynthetic mechanism of different PKSs.......... 6
Figure 1.4. Biosynthetic scheme for the activity of IPKSs.......................................................... 7
Figure 1.5. Representative compounds produced by HRPKSs.................................................. 8
Figure 1.6. Biosynthesis of lovastatin......................................................................................... 10
Figure 1.7. Representative compounds produced by NRPKSs.................................................. 11
Figure 1.8. Representative compounds produced by tandem HRPKS-NRPKS systems. ........ 13
Figure 1.9. Biosynthesis of the hypothemycin precursor DHZ ............................................... 15
Figure 1.10. Increasing Type III PKS product chain length....................................................... 23
Figure 1.11. Protein engineering of a “universal” glycosyltransferase. ...................................... 30
Figure 2.1. Secondary metabolites produced by A. niger............................................................ 36
Figure 2.2. Breakdown of PKSs found from genome mining of A. niger ATCC 1015. ............... 37
Figure 2.3. Examples of azaphilone compounds isolated from fungi........................................ 44
Figure 2.4. Comparison of the A. niger aza cluster with the A. nidulans afo cluster and their respective products........................................................................................................... 46
Figure 2.5. Transcriptional analysis by RT-PCR of the aza genes in A. niger WT and the activated T1 strain...................................................................................................................... 48
Figure 2.6. Metabolic profiles of the activated A. niger T1 strain on days 2, 4 and 7 and of the WT on day 2................................................................................................................................................. 49
Figure 2.7. Structures of 2 - 7 isolated from the activated A. niger T1 strain. ............................ 50
Figure 2.8. Knockout of azaB demonstrates convergent mode of PKS collaboration. ............... 54
Figure 2.9. In vitro reaction with AzaH. ................................................................. 55

Figure 2.10. Proposed biosynthetic pathway for production of azanigerones from the aza cluster. ............................................................................................................... 58

Figure 2.11. Cloning strategy to construct the azaA-harboring plasmid, pAZ103. ............... 69

Figure 2.12. HPLC analysis of Day 3 cultures of S.c. with AzaA and the control strain. Compound 6 is exclusively produced by the AzaA-expressing strain. ......................................................... 70

Figure 2.13. Conversion of 8 to 9 ........................................................................ 71

Figure 2.14. AzaE controls spontaneous cyclization of AzaA product. ................................ 72

Figure 2.15. Possible timing of AzaE activity in controlling the spontaneous cyclization into the naphthoquinone 6........................................................................................................ 73

Figure 2.16. Breakdown of PKS genes found in E. brefeldianum ATCC 58665. ...................... 77

Figure 2.17. Putative biosynthetic pathway for BFA ......................................................... 84

Figure 2.18 Production of BFA correlated with transcription of Contig_286 HRPKS. .......... 87

Figure 2.19. Phylogenetic tree containing BrefPKS and HRPKSs of known products......... 89

Figure 2.20. Transcriptional analysis of genes in Contig_286 determines the putative boundary of the bref cluster ............................................................................................................ 90

Figure 2.21. Minimal PKS assay on Bref-PKS ................................................................ 92

Figure 2.22. In vitro products of Bref-PKS and Bref-TH are acyclic octaketides with variable degrees of β-reduction. ......................................................................................... 93

Figure 2.23. Yeast biotransformation of compounds 12 and 14 to 13 and 15, respectively........ 98

Figure 2.24. Effect of NADPH on the product profile. ...................................................... 99

Figure 2.25. In vitro reactions with Bref-PKS demonstrate the TH-controlled PKS chain length release. ............................................................................................................. 101
Figure 2.26. Modeling of Bref-TH to known hydrolases allows identification of its active site residues. ................................................................. 103
Figure 2.27. Fma-PKS produces longer polyenes in the absence of cognate Fma-AT. ............ 104
Figure 2.28. Summary of the programming rules exhibited by Bref-PKS. .......................... 109
Figure 2.29. Proposed mechanism for the formation of the cyclopentane ring. ................. 111
Figure 4.1. Mass and UV spectra of compounds 1 to 7 (A - F)............................................. 127
Figure 4.2. Tandem MS spectra of 1...................................................................................... 128
Figure 4.3. Tandem MS spectra of 3...................................................................................... 129
Figure 4.4. Tandem MS spectra of 4...................................................................................... 130
Figure 4.5. 1H NMR Spectrum for 1. Measured in CDCl3 at 500 MHz............................... 131
Figure 4.6. 13C NMR Spectrum for 1. Measured in CDCl3 at 125 MHz............................. 132
Figure 4.7. Figure S5C. 1H-13C HSQC Spectrum for 1. Measured in CDCl3 at 125MHz..... 133
Figure 4.8. 1H-13C HMBC Spectrum for 1. Measured in CDCl3 at 125MHz.................... 134
Figure 4.9. 1H NMR Spectrum for 2. Measured in methanol-d4 at 500 MHz.................... 135
Figure 4.10. 13C NMR Spectrum for 2. Measured in methanol-d4 at 125 MHz................ 136
Figure 4.11. 1H-13C HSQC Spectrum for 2. Measured in methanol-d4 at 125MHz.......... 137
Figure 4.12. 1H-13C HMBC Spectrum for 2. Measured in methanol-d4 at 125MHz.......... 138
Figure 4.13. 1H NMR Spectrum for 3. Measured in CDCl3 at 500 MHz............................ 139
Figure 4.14. 13C NMR Spectrum for 3. Measured in CDCl3 at 125 MHz.......................... 140
Figure 4.15. 1H-13C HSQC Spectrum for 3. Measured in CDCl3 at 125MHz.................. 141
Figure 4.16. 1H-13C HMBC Spectrum for 3. Measured in CDCl3 at 125MHz.................. 142
Figure 4.17. 1H NMR Spectrum for 5. Measured in methanol-d4 at 500 MHz.................. 143
Figure 4.18. 1H NMR Spectrum for 5. Measured in acetone-d6 at 500 MHz.................... 144
Figure 4.19. 13C NMR Spectrum for 5. Measured in methanol-d4 at 125 MHz. ................. 145
Figure 4.20. 13C NMR Spectrum for 5. Measured in acetone-d6 at 125 MHz. .................. 146
Figure 4.21. 1H-13C HSQC Spectrum for 5. Measured in acetone-d6 at 125 MHz. ............. 147
Figure 4.22. 1H-13C HMBC Spectrum for 5. Measured in acetone-d6 at 125 MHz. .......... 148
Figure 4.23. 1H NMR Spectrum for 6. Measured in acetone-d6 at 500 MHz. .................... 149
Figure 4.24. 13C NMR Spectrum for 6. Measured in acetone-d6 at 125 MHz. .................. 150
Figure 4.25. 1H-13C HSQC Spectrum for 6. Measured in acetone-d6 at 125 MHz. .......... 151
Figure 4.26. 1H-13C HMBC Spectrum for 6. Measured in acetone-d6 at 125 MHz. .......... 152
Figure 4.27. 1H NMR Spectrum for 7. Measured in CDCl3 at 500 MHz. ......................... 153
Figure 4.28. 13C NMR Spectrum for 7. Measured in CDCl3 at 125 MHz. ...................... 154
Figure 4.29. 1H-13C HSQC Spectrum for 7. Measured in CDCl3 at 125 MHz. .............. 155
Figure 4.30. 1H-13C HMBC Spectrum for 7. Measured in CDCl3 at 125 MHz. .......... 156
Figure 4.31. Circular dichroism spectra for azanigerones A-F (1,3-7). .......................... 157
Figure 4.32. 1H NMR Spectrum for 9. Measured in CDCl3 at 500 MHz. ....................... 158
Figure 4.33. 13C NMR Spectrum for 9. Measured in CDCl3 at 125 MHz. .................... 159
Figure 4.34. 1H-13C HSQC Spectrum for 9. Measured in CDCl3 at 125 MHz. ............. 160
Figure 4.35. 1H-13C HMBC Spectrum for 9. Measured in CDCl3 at 125 MHz. ............ 161
Figure 4.36. Expression of Bref-PKS and Bref-TH and its mutants. ............................ 162
Figure 4.37. MS and UV spectra of compounds 12 to 17 (A-F). ................................. 163
Figure 4.38. C13-labeling of the product polyketides using the MatB regeneration system. 164
Figure 4.39. Effect of base hydrolysis on product release and post-addition of Bref-TH on
product release. .............................................................................................................. 165
Figure 4.40. MS and UV spectra of compounds 18 to 20 (A-C) .................................... 166
Figure 4.41. C13-labeling of the Fma-PKS heptaketide product ........................................ 167
Figure 4.42. $^1$H spectrum for compound 12. Measured in CD$_3$OD at 500 MHz .................. 168
Figure 4.43. $^{13}$C spectrum for compound 12. Measured in CD$_3$OD at 125 MHz .............. 169
Figure 4.44. HSQC-INEPT135 spectrum for compound 12 ............................................ 170
Figure 4.45. HMBC spectrum for compound 12 ............................................................... 171
Figure 4.46. COSY spectrum for compound 12 ............................................................... 172
Figure 4.47. $^1$H spectrum for compound 13. Measured in CD$_3$OD at 500 MHz ................ 173
Figure 4.48. $^{13}$C spectrum for compound 13. Measured in CD$_3$OD at 125 MHz .............. 174
Figure 4.49. HSQC-INEPT135 spectrum for compound 13 ............................................ 175
Figure 4.50. HMBC spectrum for compound 13 ............................................................... 176
Figure 4.51. $^1$H spectrum for compound 14. Measured in CD$_3$OD at 500 MHz .............. 177
Figure 4.52. $^{13}$C spectrum for compound 14. Measured in CD$_3$OD at 125 MHz .............. 178
Figure 4.53. HSQC-INEPT135 spectrum for compound 14 ............................................ 179
Figure 4.54. HMBC spectrum for compound 14 ............................................................... 180
Figure 4.55. $^1$H spectrum for compound 15. Measured in CD$_3$OD at 500 MHz .............. 181
Figure 4.56. $^{13}$C spectrum for compound 15. Measured in CD$_3$OD at 125 MHz .............. 182
Figure 4.57. HSQC-INEPT135 spectrum for compound 15 ............................................ 183
Figure 4.58. HMBC spectrum for compound 15 ............................................................... 184
List of Tables

Table 2.1. Detailed list of PKS gene clusters found in *A. niger* ATCC 1015........................................... 38
Table 2.2. Putative functions and homologs of the genes in the aza cluster. ............................... 47
Table 2.3. Quality of reads for genome assembly of *E. brefeldianum* ATCC 58665.............................. 76
Table 2.4. Detailed list of PKS gene clusters found in *E. brefeldianum* ATCC 58665..................... 78
Table 2.5. Genes annotated in Contig_286......................................................................................... 90
Table 2.6. 1H NMR (500 MHz, CD$_3$OH) and 13C NMR (125 MHz, CD$_3$OH) data for 12, 13, 14 and 15................................................................................................................................. 95
Table 4.1. NMR data for Azanigerone A (left) and FK17-p2A (right). .............................................. 122
Table 4.2. NMR data for Azanigerone B (left) and Azanigerone D (right)........................................ 123
Table 4.3. NMR data for Azanigerone E (left) and Azanigerone F (right). ....................................... 124
Table 4.4. Primers used in this study (azanigerone)......................................................................... 125
Table 4.5. Primers used in this study (brefeldin A)........................................................................... 185
ACKNOWLEDGEMENTS

Section 1.4 contains reproduced material from Zabala, Angelica O., Cacho, R., Tang, Y., “Protein Engineering towards natural product synthesis and diversification,” Journal of Industrial Microbiology and Biotechnology, Volume 39, Issue 2, February 2012, Page 227, ISSN 1367-5435.


The work in this dissertation was supported by NIH Grants 1R01GM085128 and 1DP1GM106413 to Y.T. and the NIH Biotechnology Training Grant 5T32GM067555 to A.O.Z.

They say it takes a village to raise a child. It also took a village for me to obtain my Ph.D. and I am greatly indebted to a lot of people who have contributed to this work on a professional and personal level. First and foremost, I would like to thank my advisor Prof. Yi Tang for welcoming me into his lab. He allowed me to learn and explore at my own pace, but kept me motivated and excited about my projects. I am grateful that he believed in my abilities and he continued to challenge me to be better in what I do. I also thank him for his continued encouragement and patience as I hit frustrating road blocks along the way, as well as the opportunities he has given me to succeed academically and professionally.

I also want to offer my sincerest gratitude to my committee members for their time and helpful discussions that have made me improve as a scientist. I thank Prof. Monbouquette for
allowing me to be a part of the NIH Biotechnology Training Program that has opened a lot of opportunities for me, and for the words of wisdom he imparted along the way; Prof. Liao for my first molecular biology classes that have guided me in my training and for always opening up his lab to us; and Prof. Garg for his chemical insights on my projects and for always being available whenever I needed someone to verify my mechanisms. Additionally, I want to extend my thanks to the people at LS9 where I did my internship, especially to Dr. Andreas Schirmer, who was a great mentor to me during my brief stay there. I also want to thank the people in the Chemical Engineering Department, including John Berger, Sara Reubelt, Miguel Perez and Alain de Vera for taking care of other issues for us so we didn’t have to worry about them.

Many thanks to the people from our lab who have imparted the skillset I needed to get started on my research. Dr. Hui Zhou first taught me basic molecular biology techniques including how to clone and express proteins; Dr. Yit Heng Chooi gave me my initial lessons in bioinformatics, gene annotation and fungal genetics, and patiently helped me with revising my papers; Dr. Wei Xu showed me how to use the NMR machine and how to interpret the spectra; and Dr. Jackie Winter provided technical insights on the azaphilone project. I also want to thank David Choi for his hard work and dedication in helping me finish my experiments in lab.

On a personal note, I would like to thank a long list of friends from the lab – Dr. Sherry Xue Gao, Dr. Anuradha Biswas, Jingjing Wang, Lene Petersen, Dr. Nidhi Tibrewal, Dr. Hsiao-Ching Lin, Dr. Ryan Peng Wang, Bo Wang, Dr. Yanran Li, Leo Chen, Linda Vien, Dr. Lauren Pickens, Dr. Wenbing Yin, Dr. Sameh Soliman, Dr. Youcai Hu, Dr. Xuming Mao, Ralph Cacho, Anthony Denicola, Diego Gavia, Carly Bond, Dr. Kangjian Qiao, Dr. Woncheol Kim, Dr. Qian Liu and Dr. Roslinah Hussain – for all the help and support during my graduate studies. I will forever treasure the good laughs we’ve shared and memories we have made in lab and beyond. I
am also greatly indebted to my friends in UCLA Shiva Gojgini, Jimmy Lafontaine-Rivera, Brenda Gonzalez, Paul Opgenorth, Eugenia Zah, Xiaqian Li, Candy Sio Si Wong, Suwei Zhu and Bonhye Koo; my friends from San Diego Aerang Kim Dacanay, Joey Dacanay, Jennifer Inouye, Laurina Lazarte, Bernadette Pacheco and Prof. Pao Chau; and my best friends from high school Kristine Tolod and Joan Freya. I was blessed to have known these people who have rooted for me from the beginning.

I want to give my sincerest gratitude to Chris Bautista for his patience, love and support. His tirelessness in listening to me as I rant about how my experiments failed (even if he didn’t understand what they meant) allowed me to keep my sanity and helped me put things in proper perspective. I don’t know if I could have made it through grad school without him in my life.

Most importantly, I would like to thank my family. My parents Manuel and Josefina have supported me unconditionally and never once questioned my decision to go back to grad school nor doubted my abilities. My dad is my biggest inspiration, having conquered a lot of hurdles in his young life to get a good education and start a comfortable life for us. He dreamt of becoming a doctor and I hope to have fulfilled a part of it. My mom never stopped taking care of us, even when we’re old enough to do things for ourselves. She epitomizes selflessness, and her love and prayers have really kept me going. My two sisters Michelle and Kristine have been the greatest support system anyone could ever have. They have been my loudest cheerleaders and my best friends, who understand me deeply and know even my tiniest quirks. As we grow older, our sisterly bond has only gotten stronger as we face new experiences and conquer new challenges.

Lastly, I would like to dedicate this thesis to my niece Sophia, who inspired me to be a better person from the moment I witnessed her birth. I hope I could help make the world a little better for her.
VITA

1986
Born, Camarines Norte, Philippines

2008
B.S. Chemical Engineering
University of California – San Diego
San Diego, CA

2010 – 2012
National Institute of Health Biotechnology Training Grant
University of California – Los Angeles
Los Angeles, CA

2010 – 2013
Teaching Assistant
Department of Chemical and Biomolecular Engineering
University of California – Los Angeles

2010 – 2014
Research Assistant
Department of Chemical and Biomolecular Engineering
University of California – Los Angeles

PUBLICATIONS AND PRESENTATIONS


Zabala, Angelica O. and Tang, Y., “Thiohydrolase involved in Brefeldin A biosynthesis controls polyketide chain length,” Society of Industrial Microbiology and Biotechnology-Symposium on Biotechnology of Fuels and Chemicals, Clearwater Beach, FL, April 28th-May1st 2014 (poster presentation)

Zabala, Angelica O. and Tang, Y., “Biosynthesis of azaphilones,” American Chemical Society National Conference, New Orleans, LA, April 7th-11th 2012 (oral presentation)


Zabala, Angelica O. and Tang, Y., “Genome mining and its role in solving the mystery of azaphilone biosynthesis,” UCLA Biotechnology Training in Biomedical Sciences and Engineering Program Third Annual Symposium, June 15th, 2012 (oral presentation)

Zabala, Angelica O., Tang, Y. “Discovery of azaphilones from genome mining of Aspergillus niger ATCC 1015,” UCLA Biotechnology Training in Biomedical Sciences and Engineering Program Second Annual Symposium, June 15th, 2011 (poster presentation)
1 Introduction

1.1 Natural products: Spotlight on Polyketides

Natural products (NPs) are still an indispensable resource for bioactive leads that eventually achieve pharmaceutical relevance. In a review published in 2007, it was reported that more than 70% of anticancer agents and antiinfectives derive from non-synthetic sources, with over 40% coming from small molecule secondary metabolites or their mimics (Newman and Cragg 2007). Additionally, among the new compounds in various stages of clinical trials, natural products are similarly well-represented in various therapeutic areas, as in oncology, bacterial infections or metabolic problems (Butler 2008).

In a broad sense, NPs could refer to both primary and secondary metabolites. As the name implies, primary metabolites derive from primary metabolism, which are essential processes for survival. On the other hand, secondary metabolites are typically small molecules that are produced by specific organisms that are not indispensable in the survival of the host organism, but could potentially confer them some advantage (e.g. toxins, attractants, coloring agents, etc.). In this field, NPs often refer only to the latter class (Dewick 2009). NPs are classified into several classes, depending on their building blocks. Polyketides derive from acetate units, terpenes are formed by isoprene units, phenylpropanoids trace their origin from the shikimic acid pathway, whereas the alkaloids obtain their characteristic nitrogen groups from amino acid sources (2003).

Polyketides are a particularly diverse group of NPs produced by a variety of organisms from the prokaryotic bacteria to the eukaryotic fungi and plants. Despite being assembled from basic simple acetate subunits, they form a wide variety of product scaffolds, including aromatics, macrolides and fatty acid-like structures. More importantly, they also possess diverse bioactivities,
which are directly used as drugs or are used as inspiration for the development of more potent agents. This group includes different antibiotics, such as the broad spectrum tetracycline from the bacteria *Streptomyces sp*, insecticidal agent abamectin from *S. avermectilis*, potent cholesterol-lowering drug lovastatin from *Aspergillus terreus*, fungistatic agent griseofulvin from *Penicillium aethiopicum*, antibacterial macrolide erythromycin from *Saccharopolyspora erythrea*, anticancer epothilone A from *Sorangium cellulosum* and the antioxidant resveratrol from plants (Figure 1.1).

![Polyketide Natural Products](image)

**Figure 1.1.** Polyketide natural products from different sources.

The first polyketide was discovered by Collie in 1893, with the molecule orcinol. He proposed a mechanism that involved repeated condensation of acetate units and termed this new molecular group as polyketides to reflect the occurrence of the -CH\(_2\)-OH- motif (Collie 1893). The field of polyketide biosynthesis obtained widespread attention and acceptance, however, only in the 1950’s. Arthur Birch demonstrated the incorporation of \(^{14}\)C-labeled acetate on the natural
products orcinol and phloroglucinol in what he termed as “the head-to-tail linkage of acetate units” (Birch and Donovan 1953). With the advent of recombinant DNA techniques, scientists discovered the genetic basis of these polyketides, starting with the characterization of the actinohordin gene cluster from *Streptomyces coelicolor* (Malpartida and Hopwood 1984). In recent years, genome sequencing has provided even better access to the genetic code that dictates the biosynthesis of the polyketides. Indeed, this field is in full upswing, as we continue to discover new polyketides, as well as new clusters and enzymes that synthesize them.

### 1.2 Polyketide synthases

Polyketide synthases (PKSs) are used to denote the enzyme or group of enzymes that catalyze the biosynthesis of polyketides. The elongation could be thought of as a series of Claisen reactions involving malonyl-CoA, a more reactive species generated from the carboxylation of acetyl-CoA in a reaction involving ATP, CO$_2$ and the CO$_2$ carrier, biotin. Polyketide biosynthesis mirrors the enzymology used in fatty acid biosynthesis by the fatty acid synthases (FASs). Similar to FAS, PKS uses three domains during the chain extension: the ketosynthase (KS) domain catalyzes the Claisen condensation between the acyl-thioesters; the malonyl-CoA:ACP transacylase (MAT) loads the malonyl group from malonyl-CoA onto the ACP; while the acyl carrier protein (ACP) serves as the anchor to support the growing chain. Additionally, it can also contain the following β-keto processing enzyme components: ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER), which can successively catalyze the reduction of the newly extended β-ketone to a β-alcohol, dehydration of the β-alcohol into an α,β unsaturated enoyl and finally, reduction of the double bond to a methylene.
Despite the similarity in the enzymology and domain architecture, FASs and PKSs are very distinct. FAS follows a more stringent programming in its biosynthesis of the fatty acids. It selects from a narrow range of starter unit and only catalyzes the formation of specific product chain lengths. More importantly, it uses all the tailoring domains during each extension, resulting in a fully saturated molecule. These characteristics explain why there is less structural diversity among fatty acids relative to the polyketides. However, it is important to note that fatty acids are primary metabolites and are therefore crucial for the survival of the organism, which is why it is important to have high fidelity in their biosynthesis (Hertweck 2009).

**Figure 1.2.** Divergence between PKS and FAS functions.

Borrowing from the FAS nomenclature, PKSs are generally categorized into Type I, Type II and Type III (Staunton and Weissman 2001). Type I PKSs are characterized as large megasynthases that contain multiple domains, each catalyzing distinct reactions in the polyketide assembly. They are further categorized as modular or iterative. The Type I iterative PKSs are often
found in fungal and are the subject of this dissertation. They will be discussed in more details in subsequent sections. Modular PKSs are often found in bacterial systems (Figure 1.3A). These enzymes are assembled in modules each containing the minimal extension domains KS, MAT and ACP, along with the optional reductive domain DH, ER and KR that determines the specific tailoring for each extension (Katz 2009; Khosla, Kapur et al. 2009). Because of this, it is easier to predict the backbone of the polyketide product; for instance, the number of extension is determined by the number of modules, whereas the reduction at each extension is determined by the presence or lack thereof of functional tailoring domains. Some of the opportunities for diversity are in selecting the starter and extender unit. Engineering on these enzymes have been performed that include rational domain swaps, domain inactivation and combinatorial biosynthesis (Katz and McDaniel 1999; McDaniel, Thamchaipenet et al. 1999). The prototype for these PKSs is the 6-deoxyerythronolide B (DEBS) that synthesize the macrolide precursor of the antibiotic erythromycin (Figure 1.1). Type II PKSs, on the other hand, are comprised of distinct monofunctional proteins, and are exclusively found in bacteria (Hertweck, Luzhetskyy et al. 2007). Instead of having domains, these PKSs consist of several dissociated extension, reduction and tailoring proteins that collaborate to synthesize the final product. The minimal extension domain consists of two KS proteins (KSα and KSβ) and the ACP that group together in a KSα/KSβ/ACP architecture, with the KSα catalyzing the Claisen condensation while the KSβ (also known as the chain length factor CLF) determining the product chain length (Figure 1.3B). Additionally, tailoring domains such as ketoreductases, aromatases and cyclases are also commonly found clustered with the minimal extension proteins that determine the folding pattern of the reactive polyketide chain. Type II PKSs are known to produce important bioactive molecules including the antibiotic tetracycline and its derivatives (Figure 1.1). Type III PKSs round up the PKS groups.
While also found in bacteria and fungi, they are most commonly seen in plants and do not require an ACP, unlike the other PKS types (Austin and Noel 2003; Abe and Morita 2010). They instead use coenzyme A esters in a single active site to perform the decarboxylation, condensation, cyclization and release of the product (Figure 1.3C). Plant Type III PKSs are also called chalcone/stilbene synthases and are involved in the synthesis of flavanoids, an example of which is resveratrol (Figure 1.1) commonly found in wine that is believed to have anti-inflammatory and anti-oxidant properties.

**Figure 1.3.** Schematic representation of the biosynthetic mechanism of: (A) Type I modular PKS; (B) Type II PKS; and (C) Type II PKS.
1.2.1 Fungal Iterative Polyketide Synthases

Fungal iterative polyketide synthases (IPKSs) belong to the Type I PKS class characterized by having multiple domains in a large polypeptide. Unlike the modular PKSs, however, IPKSs use only one module consisting of the extension domains KS, MAT and ACP, along with the optional β-keto tailoring domains KR, DH and ER (Figure 1.4). Additionally, some IPKS could contain a methyl transferase (MT) domain that transfers a methyl group to the Cα position in the polyketide from a molecule of S-Adenosylmethionine (SAM), as well as some optional releasing domains such as the thioesterase (TE) domain or a reductive release (R) domain. Because of its iterative nature, it is much more difficult to predict the products of these PKSs. From the numerous biosynthetic studies on this enzyme group, we are only beginning to understand how these enzymes are programmed (Cox 2007; Chooi and Tang 2012).

Figure 1.4. Biosynthetic scheme for the activity of IPKSs.
IPKSs are often classified on the basis of the reductive domains that are present in the protein, which is also reflected on the biosynthetic product (Cox 2007). Those that contain the full set of reductive domains KR, DH and ER are referred to as highly-reducing (HRPKSs); those that only lack the ER domain are called partially-reducing PKSs (PRPKSs); whereas those do not contain any of these domains are named non-reducing PKSs (NRPKS). Furthermore, PKSs that form a hybrid polypeptide with an NRPS are referred to as PKS-NRPS. Only the HR- and NR-PKS groups will be discussed extensively in the following sections.

1.2.1.1 Highly-reducing PKS

Highly reducing polyketide synthases (HRPKS) resemble FAS biosynthesis more closely. In fact, the domain organizations of these two enzyme families are very similar (from N- to C-terminus: KS-MAT-DH-ER-KR-ACP). Curiously, unlike FASs, HRPKSs use the reductive domains in different permutations at each extension cycle resulting in a more diverse product scaffold, as shown in the HRPKS products on Figure 1.5.

![Figure 1.5. Representative compounds produced by HRPKSs.](image-url)
Another key difference between HRPKSs and FASs is the absence of a dedicated releasing domain for the former. FASs contain a thioesterase (TE) domain that dictates the chain length of a product. HRPKSs, on the other hand, often rely on an \textit{in trans} interaction with thioesterase/hydrolase-like enzymes that typically cluster with them in the genome. This makes the task of determining the final product even more difficult, as these enzymes are found to highly influence the product chain length (See section 2.2.3 for more details). Most of the products are linear, non-aromatic in nature (as in T-toxin and fumonisin B1 in Figure 1.5), which is in contrast to the NRPKS products that will be discussed in more detail later (Section 1.2.1.2). Others contain a decalin ring (as in lovastatin in Figure 1.5), whose formation is internally programmed in the enzyme but has not been thoroughly explored as of yet. Some, however, become integrated as a starter unit that is further processed by a downstream enzyme (Tandem PKS systems, Section 1.2.1.3).

Among the HRPKS products, the biosynthesis of lovastatin from the filamentous fungi \textit{Aspergillus terreus} has been extensively explored. Its biosynthesis includes two HRPKSs: a diketide synthase LovF and a nonaketide synthase LovB, whose activities have been reconstituted in vitro (Ma, Li et al. 2009; Xie, Meehan et al. 2009). Despite the striking similarity in the domain arrangement, these two enzymes synthesize two very distinct products. LovB catalyzes \(~35\) catalytic steps in producing the dihydromonacolin L intermediate (offloaded by a thioesterase LovG), whereas LovF performs a single extension to afford to the diketide product that is eventually released and transferred by the acyltransferase LovD (Figure 1.6) (Ma, Li et al. 2009; Xie, Meehan et al. 2009; Xu, Chooi et al. 2013). What accounts for this difference is still unknown, although from a phylogenetic tree of Type I PKSs, these two enzymes have been shown to clade separately (Kroken, Glass et al. 2003).
While numerous bioactive compounds seem to be derived fully or in part from an HRPKS product, only a few systems have been biochemically characterized. Besides the lovastatin PKSs, other HRPKS genes that have been linked to their products include the solanapyrone PKS (sol1) (Kasahara, Miyamoto et al. 2010), fumonisin PKS (fum5) (Proctor, Desjardins et al. 1999), two T-toxin PKSs (PKS1 and PKS2) (Baker, Kroken et al. 2006) and depudecin PKS (dep5) (Wight, Kim et al. 2009). Some of the problems that hamper these studies include the difficulty in product detection due to lack of a strong chromophore of the product polyketide. With the exception of polyenes that have the conjugated double bonds, most reduced linear molecules are not UV-sensitive, thereby requiring other means for analysis, such as a sensitive MS detection. Additionally, the necessity for an accessory enzyme to release the product adds complexity to studying these systems. As with other Type I PKSs, HRPKSs are also large polypeptides and getting soluble expression in a heterologous host could also post some challenges. However, to better understand the programming rules of these HRPKSs for potential engineering and better
prediction during genome mining, it is crucial to continue exploring and characterizing these megasynthases.

1.2.1.2 Non-reducing PKS

Non-reducing PKSs are responsible for the fungal aromatic compounds, as shown on Figure 1.7. NRPKSs do not contain the trio of β-processing domains that results in a reactive poly-β-keto chain that cyclizes intramolecularly to form the fused aromatic rings found in the final structure of the compounds (Cox 2007; Chooi and Tang 2012). Instead, the typical domain organization for most NRPKSs is (from N- to C-terminus): SAT-KS-MAT-PT-ACP-CLC/TE/R (SAT=starter unit;ACP transacylase; PT=product template; CLC=Claisen cyclase; TE =thioesterase; R=reductive).

![Viridicatumtoxin](image1)

![YWA1](image2)

![Aflatoxin](image3)

![Bikaverin](image4)

![Citrinin](image5)

![Emodin](image6)

**Figure 1.7.** Representative compounds produced by NRPKSs.

Unlike HRPKSs, more is known about the programming of the NRPKSs. For instance, the functions of each domain and how they influence the PKS product have been thoroughly investigated (Cox 2007; Chooi and Tang 2012). Most NRPKSs begin with an SAT domain coined
by Townsend and coworkers in their study of the norsolorinic acid synthase (precursor to the mycotoxin aflatoxin shown in Figure 1.7) (Crawford, Dancy et al. 2006). From the dissection of the domains and the radiolabeled transacylase assay, they showed that the N-terminus domain, which showed structural similarity to the MAT domain, was responsible for accepting the hexanoyl starter unit from the dedicated FAS and transferring it to the ACP. Later studies demonstrated that the SAT domain could select for more advanced starter units (as in the tandem systems that will be discussed in the next section 1.2.1.3), but also for the basic acetate unit, such as in WA PKS responsible for the production of the pigment YWA1 shown in Figure 1.7) (Crawford, Vagstad et al. 2008).

With regards to chain length control, several studies have alluded to the involvement of the KS domain; particularly, the size of the KS cavity dictates how many extension rounds will be catalyzed (Chooi and Tang 2012). This was supported by the ability of the minimal KS-MAT didomain with the ACP from GfPKS4 involved in bikaverin biosynthesis to produce the correct chain length nonaketide product (Zhang, Li et al. 2008). Additionally, feeding of longer chained precursors to this PKS did not result in longer products; rather the number of extension was instead reduced to yield the same product chain length (Ma, Zhan et al. 2007).

One of the interesting features among NRPKSs is the cyclization pattern. There are three common modes of cyclization among the fungal NRPKS products: C2-C7, C4-C9 and C6-C11 (Zhou, Li et al. 2010). Using dissected domains of the norsolorinic acid synthase PksA, it was shown that the core region between the KS-MAT domains and the ACP influences the cyclization pattern; this was termed as the product template (PT) domain (Crawford, Thomas et al. 2008). Since its discovery, it has also been demonstrated that the PT domains of different NRPKSs clade in the phylogenetic tree in a manner that could predict the product folding. Rational domain swap
of the asperthecin AptA C6-C11 PT into the bikaverin PKS4 C2-C7 PT resulted in the expected switch in the cyclization mode of the latter (Li, Image et al. 2010). The structural basis for its biosynthetic mechanism has also been explained with the high-resolution crystal structure of the norsolorinic acid PksA monodomain (Crawford, Korman et al. 2009).

The final important component of the NRPKSs is the releasing domain. Unlike the HRPKS, NRPKSs often contain a dedicated releasing domain. Thioesterase domains (TE) could release the product either through an intramolecular Dieckmann condensation as in the pigment YWA1 PKS ((Fujii, Watanabe et al. 2001)) or by an intramolecular lactonization (typically among tandem systems, section 1.2.1.3). Alternatively, a reductive (R) domain could also offload the product as an aldehyde, as demonstrated by the Cox group with the 3-methylorcinaldehyde synthase (Bailey, Cox et al. 2007). Other mechanisms for release include the use of in trans β-lactamase to hydrolyze the product or to catalyze a Claisen condensation to produce a tetracycline scaffold (Li, Chooi et al. 2011).

1.2.1.3 Tandem PKSs

Figure 1.8. Representative compounds produced by tandem HRPKS-NRPKS systems.

From the widespread genome mining of different fungi, one commonly observed phenomenon is the clustering of two PKSs. In some cases, two HRPKSs group together, as in the case of LovB and LovF. However, it is also very typical to see an HRPKS and an NRPKS work together in the synthesis of a product. These so-called tandem systems are associated with the
production of some bioactive molecules, such as the resorcylic acid lactones hypothemycin (kinase-inhibitor) and radicicol (Hsp90-inhibitor), as well as the cytotoxic asperfuranone (Figure 1.8). In terms of enzyology, these systems demonstrate an added level of complexity. It is especially interesting to investigate the specific interaction between the two PKSs. Oftentimes, they collaborate in a sequential manner. The HRPKS produces a reduced precursor that is then released and transferred to the NRPKS. Detailed biochemical studies on the resorcylic acid lactones have been performed to show that this release and transfer must be mediated by the SAT domain, as shown in Figure 1.9. In this case, the HRPKS Hpm8 catalyzes six extension cycles to produce the reduced hexaketide precursor before the SAT domain of the NRPKS Hpm3 takes it and extends the chain for three additional unreduced cycles (Zhou, Qiao et al. 2010). The PT domain of Hpm3 then catalyzes the regiospecific C2-C7 cyclization to afford the aromatic resorcylate core before the chain is transferred covalently into the TE domain for the release and macrolactonization. A similar mechanism is in place for radicicol (also shown in Figure 1.8) (Zhou, Qiao et al. 2010). In asperfuranone biosynthesis, while the mode of collaboration is also sequential, the products are considerably different; its HRPKS AfoG synthesizes a triketide (instead of hexaketide) and the NRPKS AfoE contains an R domain instead of the lactonizing TE (Chiang, Szewczyk et al. 2009).
1.3 Methods for studying PKS gene clusters in the genomic era

In the advent of the sequencing era, it becomes easier to investigate PKSs and their clusters. Some of the commonly used techniques are discussed below that provide insight to the function and mechanism of these enzymes. It should be noted that all these methods require the sequence of the organism to be implemented.

1.3.1 Genetic manipulation of native host

Perhaps the most widely used technique in correlating the gene sequence to the product is by genetic manipulation of the production host. (Van Lanen and Shen 2008) This starts with a good preliminary hypothesis through a logical retrobiosynthetic analysis on the compound of interest. The clustering of biosynthetic genes prove very useful in this task of selecting or narrowing down
gene clusters. To verify, the main biosynthetic gene responsible for producing the scaffold, or an accessory enzyme, could be inactivated or deleted by specific integration of a knockout cassette containing a resistance marker gene into the specified locus. Such technique could be straightforward, as the abolishment of the production of the final compound could be correlated to the inactivation or deletion of the gene being studied. Oftentimes, it could also lead to the accumulation of isolatable intermediate that could further provide useful insights as to the enzyme’s function. This technique is becoming commonplace, as was demonstrated in the viridicatumtoxin and griseofulvin biosynthesis by *Penicillium aethiopicum* (Chooi, Cacho et al. 2010). The individual gene deletion in these systems has enabled the authors to propose a pathway for the production of the final compound (Chooi, Wang et al. 2012; Cacho, Chooi et al. 2013).

The success of these experiments depend on the ease of genetic manipulation of the fungal host. Gene knockout experiments have been demonstrated on a variety of fungal species, including *Aspergillus, Fusarium* and *Penicillium*. Typical methods include the PEG-mediated or the *Agrobacterium tumefaciens*-mediated protocols (Meyer, Ram et al. 2010). Some strains have also been engineered in order to improve the efficiency of homologous recombination by deleting the often dominant non-homologous end joining mechanism performed by the Ku70-Ku80 heterodimer (Jiang, Zhu et al. 2013).

More recently, it has been demonstrated that clusters can be specifically activated by overexpression of cluster-specific fungal transcription factors (Van Lanen and Shen 2008). These could be used on orphan clusters that are silent under current laboratory conditions. In this case, the activator is placed under the control of a strong promoter and the plasmid is transformed as before ectopically. This strategy has been fruitful in the discovery of compounds such as the PKS-
NRPS product aspyridone and asperfuranone (Bergmann, Schumann et al. 2007; Chiang, Szewczyk et al. 2009).

1.3.2 Heterologous expression of clusters and individual genes

When the native host is not amenable to genetic manipulation, it is also a good strategy to move the cluster to a more genetically friendly host. Among filamentous fungi, *Aspergillus nidulans* and *Aspergillus oryzae* have both been well-studied; hence, the techniques for their genetic manipulation are well-established and developed. Additionally, because they are also known producers of natural products, they already have the required accessory enzymes to make the protein active (for instance the ppant transferase), as well as the necessary precursor small molecules in the cell. From this strategy, two clusters have been heterologously reconstituted in *A. oryzae* – the monacolin K cluster from *Monascus pilosus* and the terrequinone cluster from *Aspergillus nidulans* (Sakai, Kinoshita et al. 2012). Meanwhile, our lab has recently reconstituted the activity of an orphan cluster from a dermatophyte into *A. nidulans* (Yin, Chooi et al. 2013). In order to more easily identify the correct integrants, the new cluster is integrated into the *wA* locus known to code for a pigment. The spores turn white instead of the characteristic green color when the PKS is disrupted. In a slightly different approach, Oakley et al. subjected all the biosynthetic genes of asperfuranone from *A. terreus* in an *alcA* inducible promoter and added them one-by-one into the host *A. nidulans* (Chiang, Oakley et al. 2013). This helped them clarify the individual functions of the enzymes. One of the things that could improve success of this method is the presence of a cluster-specific regulator. This could be placed under a strong promoter in order to selectively activate the cluster of interest. Moreover, competing dominant pathways should also be deleted in order to direct the flux towards the new heterologous cluster and to make analysis and detection of new products easier.
In some cases, heterologous expression of specific enzymes into a heterologous host is also beneficial, particularly of the PKS genes. This has been successfully performed both in a fungal host, such as \textit{A. nidulans}, as well as in an engineered yeast (\textit{Saccharomyces cerevisiae} BJ5464-NpgA). It provides a way of isolating the immediate product of the main biosynthetic enzyme before it gets tailored by the accessory enzymes, which is quite useful for a deeper understanding of its biosynthetic mechanism. Watanabe et al reconstituted the activity of several PKSs from a filamentous fungi \textit{Chaetomium globosum} and were able to identify and characterize their respective products (Tsunematsu, Ishiuchi et al. 2013). A similar approach was performed by Wang et al. by performing a promoter-replacement strategy on every NRPKS gene in \textit{A. nidulans} (Ahuja, Chiang et al. 2012).

\subsection*{1.3.3 In vitro reconstitution of protein activity}

Among the techniques discussed here, in vitro reconstitution of enzyme activity provides the cleanest slate because the starting materials and the enzymes and their amounts could all be controlled. We gain knowledge not only on the mechanism but also on their kinetics. However, there are a lot of challenges to this method. Firstly, proteins, particularly the large ones like PKSs, do not always express solubly. In recent years, this has improved significantly with the development of yeast hosts that are better than bacteria-based systems in expressing megasynthases (Ma, Li et al. 2009; Zhou, Qiao et al. 2010; Zhou, Qiao et al. 2010). The yields are not optimal but are sufficient for the small-scale in vitro reactions. Secondly, the expression hosts are incapable of processing introns that are commonly found in the fungi. Previously, manual splicing of introns are performed based on annotations from prediction softwares; however, these are not always reliable. More recently, cDNA constructed from total or gene-specific mRNA have also been used as the intronless transcript for cloning (Tsunematsu, Ishiuchi et al. 2013). Lastly,
this method requires the precise starting materials, cofactors and conditions in order to work. This often entails optimization of the reaction conditions. Additionally, these reactions also require standards in order to characterize the compounds produced, especially since these reactions are often run in microliter-scales.

1.4 Engineering PKSs and tailoring enzymes for NP diversification


1.4.1 Engineering PKSs

Polyketide synthases (PKSs) are a family of enzymes responsible for the synthesis of a wide array of natural products, including the cholesterol-lowering lovastatin, the immunosuppressant rapamycin and the anti-bacterial tetracycline (Staunton and Weissman 2001; Hertweck 2009). The assembly of the polyketide backbone resembles that of fatty acid synthesis, with major differences in building block utilization, β-carbon reduction patterns and intramolecular cyclization (Staunton and Weissman 2001). The programming rules of the PKSs allow them to produce a far more diverse group of compounds relative to the fatty acid synthases (FASs). Understanding of rules governing the different types of PKSs has enabled the engineering of these enzymes towards generation of new compounds.

PKSs are often classified into three classes, loosely based on FAS nomenclature: a) Type I PKSs are multi-domain megasynthases that act either modularly (as in some bacterial PKSs) (Khosla, Kapur et al. 2009) or iteratively (as in most fungal PKSs) (Cox 2007) in order to catalyze successive polyketide chain extensions; b) Type II PKSs, on the other hand, consist of discrete
monofunctional proteins that are used repeatedly in producing the polyketide backbone (Hertweck, Luzhetskyy et al. 2007); and c) Type III PKSs, most commonly found in plants, are homodimeric enzymes characterized by directly using coenzyme-A (CoA) as acyl carrier instead of a cognate carrier domain/protein and catalyzing the series of condensation reactions in a single multifunctional active site (Austin and Noel 2003).

A minimum PKS unit is composed of a β-ketoacyl synthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP) (except in type III). The KS domain is responsible for the decarboxylative Claisen-like condensation between the growing polyketide and a carboxylated acyl building block to yield an elongated polyketide chain. In type II PKSs, the KSα/KSβ heterodimer also dictates the length of the polyketide product (Hertweck 2009). The AT domain selects and transfers activated acyl group, typically from CoA thioesters, to the phosphopantetheinyl arm of the ACP; and therefore safeguards the identity of the polyketide building blocks (Chan, Podevels et al. 2009). In addition, other PKS enzymes or domains such as β-ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) are recruited in different combinations after each chain extension step to modify the β-keto positions (Weissman and Leadlay 2005). In comparison to fatty acids, in which the β-positions are completely reduced to methylenes, polyketide backbones can be completely unreduced as in the case of aromatic polyketides (Das and Khosla 2009), or only partially reduced in different permutations such as in the macrolides. An important strategy employed by different PKSs to generate structural diversity is via the incorporation of different starter and extender acyl units (Moore and Hertweck 2002; Chan, Podevels et al. 2009). This is attractive in creating structural analogues that are not observed in nature. Additionally, mutational studies aimed at altering the specificity of the chain elongation
and reductive steps have also been successfully utilized to generate new derivatives of natural products, some of which have been found to have enhanced bioactivity.

1.4.1.1 Controlling Chain Length and Substrate Specificity

One of the key objectives of engineering PKS is to relax the machinery’s stringency towards starter unit selection and to alter the product size, which is reflected in the polyketide chain length. Towards this end, Abe et al. performed mutagenesis studies on the active site residues of the Type III PKS, aloeone synthase (ALS) from *Rheum palmatum*, based on homology modeling with the well-known chalcone synthase (CHS) (Abe, Watanabe et al. 2006). Ferrer and his colleagues have previously solved the *Medicago sativa* CHS crystal structure, which guided the selection of ALS residues for mutation (Ferrer, Jez et al. 1999). Despite high sequence similarity, these two enzymes are functionally distinct - ALS naturally selects an acetyl starter unit followed by six condensations with malonyl-CoA to produce the heptaketide aloeone, while CHS loads a bulkier 4-coumaroyl starter unit with three extensions to yield chalcone. Guided by the CHS model, some of the active-site residues in the ALS were mutated to the CHS counterparts, which resulted in production of new compounds with chalcone-like properties (Abe, Watanabe et al. 2006). This study also provided insights into the respective roles of gatekeeping residues - Gly256 is important in the selection of the starter unit, Thr197 is involved in controlling the product chain length, while Ser38 is needed to guide the growing chain into the cavity.

Similarly, Abe et al. also mutated the *Aloe arborescens* pentaketide chromone PKS to produce aromatic polyketides SEK4 and SEK4b, which are octaketide shunt products in actinorhodin biosynthesis (Abe, Utsumi et al. 2005). The crystal structures of the wild type chromone synthase and the M207G mutant that synthesized the octaketides were solved, which showed numerous residues that control product chain length based on stercics (Morita, Kondo et al. 2006).
Additional mutations of the bulky tyrosine and phenylalanine residues in the cavity yielded a longer nonaketide napthopyrone product (Figure 1.10) (Abe, Morita et al. 2007). Using a similar strategy, Klundt converted a benzophenone synthase (BHS) to a phenylpyrone synthase via the point mutation T135L. This mutation altered both the number of chain extension steps and the cyclization pattern of the resulting polyketide (Klundt, Bocola et al. 2009).

Among Type II PKSs, which are constituted of discrete monofunctional enzymes, Tang et al. studied the Chain Length Factor (CLF) subunit in the KS/CLF heterodimer (Tang, Tsai et al. 2003). The KS/CLF catalyzes the iterative condensation of malonyl building blocks to yield an unreduced polyketide backbone. By aligning sequences of CLF involved in the synthesis of polyketides of different lengths (from C_{16} to C_{24}) and building a homology structure of the actinorhodin KS-CLF, four residues at the heterodimer interface were proposed to be involved in determining the chain length. These residues were noted to become progressively smaller in CLFs that are involved in the synthesis of longer polyketides, and were seen in the structure to define the size of the polyketide cavity. By changing bulkier residues (e.g. phenylalanine) found in C_{16}-specific CLF into smaller amino acids, C_{20} polyketides were preferentially produced, validating the role of CLF in controlling the number of decarboxylative condensations steps in type II PKS. The roles of these residues were subsequently confirmed by the X-ray crystal structure (Keatinge-Clay, Maltby et al. 2004).
Figure 1.10. Increasing Type III PKS product chain length. (A) The active site cavity of the natural pentaketide chromone synthase (PDB 2D3M). It was shown that M207 dictates the cavity size; (B) The M207G mutation allowed entrance of the growing polyketide chain to the otherwise inaccessible cavity and led to the formation of the unnatural octaketide (PDB 2D52). Further mutations of the bulky Y82 and F80 residues found at the base of the cavity to alanine yielded the even larger nonaketide.

1.4.1.2 Incorporating Different Extender Unit

The AT domains in PKSs have been engineered toward the biosynthesis of polyketides that contain alternative extender units. For example, Reeves et al. succeeded in changing the specificity of the fourth AT (AT4) in the Type I modular PKS 6-deoxyerythronolide B synthase (DEBS) involved in erythromycin biosynthesis from its natural substrate methylmalonyl-CoA to malonyl-CoA (Reeves, Murli et al. 2001). By aligning sequences of ATs with different specificities and modeling to E. coli FabD, three key regions in the active site cleft were identified to discriminate the different extender units. Site-directed mutagenesis on eight different residues in these regions, both individually and in combination, were performed, each time changing the methylmalonyl-specific residues into the malonyl-specific ones. The mutants were introduced into the
heterologous production strain *Streptomyces lividans* and production of the expected 6-desmethyl analogue was observed. While AT swaps have been successfully employed on other DEBS modules to change the extender unit specificity (McDaniel, Thamchaipenet et al. 1999), this strategy failed when performed on AT4.

### 1.4.1.3 Altering PKS Reduction Specificity

The reductive domains of PKSs, such as the KR and ER domains, have different regio- and stereospecificities, thus further contributing to the structural diversity of partially-reduced polyketides. Engineering these domains through mutagenesis is therefore one strategy that can be utilized in the production of analogues with precise modifications in stereochemistry.

The KR domains of DEBS were modeled against the short chain dehydrogenase/reductase (SDR) superfamily by Reid et al. and a putative conserved triad of Tyr, Lys and Ser was identified (Reid, Piagentini et al. 2003). Mutagenesis of these residues in KR6 of the truncated DEBS Module6+TE (M6+TE), followed by *in vitro* assay using a synthetic diketide precursor showed that the KR was completely inactivated. Subsequently, the same mutations were introduced into KR6 of the full length DEBS expressed in *S. lividans*. Each of the mutations, which are K2664Q, S2686A and Y2699F, resulted in the production of the expected 3-deoxy-3-oxo-6-dEB, although only the Y2699F mutation fully inactivated the KR; both the K2664Q and S2686A mutations resulted in mixed products. KR deletions have been performed in other DEBS modules that yielded the expected inactivation product (McDaniel, Thamchaipenet et al. 1999); however, in the case of KR6, it also affected the substrate specificity of the neighboring AT domain leaving the protein less efficient in producing the desired derivative. This is in contrast to the Y2699 mutation that exclusively produced the analogue and had better catalytic properties when tested *in vitro*. 
Ding et al. used site-directed mutagenesis to inactivate the KR domain of a 6-methylsalicylic acid synthase (6-MSAS), converting it into a functional orsellinic acid synthase (OSAS) (Ding, Lei et al. 2010). 6-MSAS differs from OSAS in that a functional KR domain is present. By homology modeling, a conserved tyrosine residue that is part of the catalytic triad was mutated to phenylalanine in the 6-MSAS ChlB1 from the spirotetronate pathway. The host strain bearing the Y1540F ChlB1 mutant was found to accumulate orsellinic acid, demonstrating inactivation of KR did not affect the remaining activities of the PKS. The engineered ChlB1 was reintroduced into the spirotetronate biosynthetic pathway, and the downstream enzymes were able to incorporate the orsellinate into a new spirotetronate analogue with similar bioactivity as the original antibacterial agent.

In addition to inactivation studies, the KR domains have also been engineered to produce analogues with altered stereochemistry. This was demonstrated by Leadlay and coworkers in successfully switching the stereospecificity of DEBS KR1 (Baerga-Ortiz, Popovic et al. 2006). Using the synthetic substrate 2-methyl-3-oxopentanoyl-S-N-acetylcysteamine (SNAC), the wild type KR1 was shown to have strict stereospecificity in the reduction to afford only the (2S,3R)-2-methyl-3-hydroxypentanoyl product. Comparing KR1 to the downstream KR2 known to have the opposite stereospecificity, non-conserved residues that lie in the active sites were identified. In an accompanying study by O’Hare et al., saturation mutagenesis was performed on each of these residues and a library of KR1 mutants was constructed for product screening (O'Hare, Baerga-Ortiz et al. 2006). In order to select for active enzymes from this library and narrow down the number of mutants, a high-throughput spectrophotometric method that detected the reduction of trans-1 decalone was performed. The double mutant F141W/P144G was found to completely switch the stereospecificity and to produce the (2S,3S) product exclusively.
With regards to the stereospecificity of the ER domain, Kwan et al. identified a tyrosine residue that can possibly influence the stereochemistry of the α-methylated product, which originates from a methylmalonyl extender unit (Kwan, Sun et al. 2008). From surveying different ER domains with different stereospecificity, a correlation between the presence of a tyrosine residue in the active site and an S α-methyl-containing product was found. Substitution of the tyrosine with other residues, such as the prevalently found valine, resulted in a product with R configuration. To confirm the role of the tyrosine, module 4 ER of DEBS and module 13 ER of rapamycin synthase (RAPS), which have opposite stereospecificities, were studied. The tyrosine to valine mutation in DEBS ER4 switched the stereospecificity of enoylreduction and afforded the product in which the corresponding methyl is in the R configuration. However, the valine to tyrosine mutation in RAPS ER13 was not able to reverse the stereospecificity, suggesting additional residues may be involved. Probing deeper into this subject, Kwan performed mutagenesis on other active site residues in RAPS ER13 and examined individually and in combination the effect on the stereospecificity (Kwan and Leadlay 2010). When the combination of mutations V46L, V47I, V52Y and N52P were introduced to match those in DEBS ER4, the desired product with S stereochemistry was observed, albeit in low levels.

1.4.1.4 Producing Polyketide Analogues with Enhanced Bioactivity

KR inactivation proved useful in producing analogues of the antifungal polyketide amphotericin B, produced by a Type I modular PKS. Using site-directed mutagenesis, the KR domains of module 12 and module 16 were both inactivated by site-directed mutagenesis of the conserved tyrosine residues (Y720F and Y6165F, respectively) (Power, Dunne et al. 2008). The 16-descarboxyl-16-methyl-amphotericin produced by KR12 inactivation maintained the same antifungal activity as the amphotericin B, while the oxo-derivative from KR16-inactivated PKS
was shown to have better activity, enhanced solubility and reduced hemolytic properties. Brautaset et al. performed numerous site-directed mutagenesis studies to investigate the structure-activity-toxicity correlations of nystatin analogs generated from the inactivation of a) \textit{nys}J \text{ DH15} responsible for the dehydration of C-10 hydroxyl bond; b) \textit{nys}N, a P450 that putatively oxidizes the C16 methyl; c) \textit{nys}J \text{ KR16} and KR17 responsible for the reduction at C5 and C7, respectively; and d) some combinations of the above (Brautaset, Sletta et al. 2008). The nystatin analogue BSG020 (16-decarboxy-16-methyl-28,29-didehyronystatin) displayed enhanced activity over amphotericin B. More recently, the same group performed additional biosynthetic engineering of nystatin analogues to further elucidate structure-to-activity correlations, especially of the polyol region (Brautaset, Sletta et al. 2011). Double and triple inactivation experiments involving ER5, DH15, \textit{nys}N and the other P450 monooxygenase \textit{nsy}L were performed to generate four novel analogues. Notably, the authors found that the C16 methyl that resulted from \textit{nys}L inactivation was beneficial in improving the activity and in lowering the hemolytic activity.

The same technique was utilized by Kim et al. in producing analogues of geldanamycin, an Hsp90 inhibitor that is a potential chemotherapeutic agent (Kim, Lee et al. 2009). By inactivation of DH1 via site-directed mutagenesis at the His2041 residue in the seven-module Type I PKS, the expected C15-hydroxylated analogue was detected, along with other derivatives that resulted from an intramolecular reaction involving the C15 hydroxyl moiety. This demonstrated that the downstream modules and the tailoring enzymes were tolerant of this structural change originating from the first module. Further mutagenesis of tailoring enzymes, including the monooxygenase Gel7 and the \textit{O}-carbamoyl transferase Gel8, yielded more analogues of geldanamycin, among which one compound exhibited a near 5-fold improvement in Hsp90 ATPase activity.
1.4.2 Engineering tailoring enzymes

The diversity of natural products synthesized by the different classes of biosynthetic machineries is further increased by the action of tailoring enzymes. These enzymes decorate the natural product scaffold with various functional groups that are often essential for the observed bioactivity of the final compounds. Enzymes such as acyltransferases, glycosyltransferases and oxygenases are frequently found in various combinations in a biosynthetic pathway. Protein engineering approaches have been applied to these enzymes to either improve the catalytic activities or more importantly, broaden the substrate specificity towards unnatural natural products (Zhou, Xie et al. 2008).

1.4.2.1 Acyltransferases

Acyltransferases catalyze the transfer of an acyl group from a donor substrate to a nucleophilic site (for example, amine or hydroxyl) on an accepting substrate. Oftentimes, it is desirable to express and purify acyltransfearses from a heterologous host in order to investigate their functionality. One common problem hampering these studies is the lack of soluble and functional protein, which can result from unfavorable interactions between surface-exposed residues and water. In studying the specificity of the plant 2-O-benzoyltransferase from *Taxus cuspidate* towards different acyl substrates, Nawaranthe et al. first improved the solubility of the protein, which was initially expressed as non-functional inclusion bodies (Nawarathne and Walker 2010). By homology modeling to other functionally defined AT orthologues and to the solved crystal structure of a related viorine synthase, they identified two key residues, Glu19 and Asn23, that deviated from the consensus sequence and were also found to lie in the exposed regions of the protein. Using site-directed mutagenesis to incorporate the Q19P and N23K mutations, they
obtained a five-fold improvement in its solubility, which enabled them to further probe the specificity of this acyltransferase.

Besides solubility, protein engineering has also been performed to improve the catalytic activity of acyltransferases, as in the case of LovD, an enzyme involved in lovastatin biosynthesis (Gao, Xie et al. 2009). LovD was previously shown to mediate the transfer of the acyl group from the lovastatin diketide synthase LovF, as well as from the small molecule substrate α-dimethylbutyryl-S-methyl-mercaptopropionate, to monacolin J, and yielding either lovastatin or the semisynthetic simvastatin (Xie, Watanabe et al. 2006). Gao and coworkers used multiple rounds of directed evolution via error-prone PCR and saturation mutagenesis to enhance its catalytic efficiency, thermal stability and solubility. They coupled these methods to a screen based on an agar-diffusion assay, where simvastatin production inhibited the growth of embedded spores of Neurospora crassa. After seven rounds of directed evolution, a mutant that displayed 11-fold increase in whole-cell activity in simvastatin synthesis was recovered. A fifth-round mutant was crystallized and the X-ray structure was compared to that of the wild type LovD. It was proposed that mutations led to conformation changes of the enzyme and stabilized the active site in a configuration that is more favorable for catalysis.
1.4.2.2 Glycosyltransferase

The ligation of deoxysugars to the natural product aglycon is an essential step in introducing bioactivity to the final molecules. Therefore the corresponding glycosyltransferases

![Diagram](image)

**Figure 1.11.** Protein engineering of a “universal” glycosyltransferase. (A) The wild-type OleD catalyzes the transfer of glucose from UDP-Glu to oleandomycin. (B) A triple mutant P67T/S132F/A242V, OleD*, isolated via directed evolution using the fluorescent 4-methylumbelliferone as the aglycon acceptor was able to transfer a wide array of sugar substrates (denoted as U(T)DP-X).
are valuable in the synthesis of analogues containing altered glycosylation patterns. Most of the studies are focused on expanding the substrate specificities of the enzymes to be more promiscuous. A prime example is the expansion of both the donor and acceptor specificities of the oleandomycin glycosyltransferase, OleD, that naturally transfers uridine diphosphoglucose (UDP-Glc) to the oleandomycin aglycon to form the glucosylated product (Williams, Zhang et al. 2007). By performing several rounds of directed evolution, a “universal glycosyltransferase” that can transfer fifteen different deoxysugars onto eight different acceptor molecules, including the natural substrate oleandomycin, 4-methylumbelliferone and novobiocic acid, etc., was isolated (Figure 5). This paved the way for glycorandomization of several natural products via pairing of different deoxysugars and alycons. The various monosaccharide substrates that were accepted by the P67T/A242V/S132F triple mutant are shown in Figure 5B. Gantt and coworkers further probed the aglycon specificity of these OleD mutants by testing 137 different drug-like acceptors, more than half of which were glycosylated (Gantt, Goff et al. 2008).

In a more focused approach, Williams and his coworkers isolated an OleD mutant that specifically and efficiently glycosylates novobiocic acid to produce novobiocin (Williams, Goff et al. 2008). Using “hot spot” saturation mutagenesis on the sequence previously identified to be involved in the acceptor promiscuity, they obtained a glycosyltransferase mutant with several hundred-fold improvement in catalytic activity. In a similar manner, Ramos and coworkers mutated several residues in the elloramycin glycosyltransferase, a moderately promiscuous enzyme, in order to make it more specific towards glycosylation of the important antitumor drug, elloramycin (Ramos, Olano et al. 2009).
1.4.2.3 Oxygenases

Oxidative modification on natural products is another important tailoring step, and can lead to drastic changes in product polarity and structure, as well as installing sites for further enzymatic modifications such as acylation and glycosylation. Nature uses different types of oxygenases, including flavin-dependent, cytochrome P450 and α-ketoglutarate/iron dependent enzymes, to catalyze a variety of redox modifications, such as hydroxylation, epoxidation and Baeyer-Villiger oxidative cleavages, etc (Rix, Fischer et al. 2002). Several cytochrome P450 oxygenases have been engineered to date, the main goals of which are to process a wider array of substrates, to enhance the catalytic activity and to alter the enantioselectivity. Xiang et al. succeeded in relaxing the substrate specificity of the EryF P450 by mutation of Ala245 to the highly conserved threonine found among its homologs (Xiang, Tschirret-Guth et al. 2000). The conserved threonine residue among this family P450s is believed to participate in the O-O bond scission step during catalysis, and its absence in EryF is compensated by the 5-hydroxyl group in the substrate. With the A245T mutation, this strict substrate constraint is eliminated and conferred EryF promiscuity towards a much wider array of substrates. For example, the mutant catalyzed the efficient oxidation of testosterone in the 1-, 11α, 12- and 16α positions, activities that were not observed with the wild type.

Zocher and his coworkers solved the X-ray crystal structures of four isoforms of the P450 AurH involved in the synthesis of aureothin (Zocher, Richter et al. 2011). AurH catalyzes the two-step reaction including an initial hydroxylation at C7, followed by heterocyclization to yield the tetrahydrofuran ring. Structural insights guided mutational studies to characterize residues involved in the stereospecific oxidation reaction. By analyzing the “open” and “closed” (inhibitor-bound) conformations of the enzyme, as well as modeling substrate docking, Ser66 and Asn91
were predicted and confirmed to undergo conformational changes to accommodate the binding of both deoxy- and hydroxyaureothin substrates to the active site. Interestingly, in an effort to understand the roles of other residues lining the binding site, two mutants, F89W and T239F were found to have altered regiospecificity. The mutants selectively oxidized the C9a methyl instead of the natural C7 methyl into a carboxylic acid.

1.5 Motivation for study

Despite the numerous polyketides isolated and characterized, there is still a big reservoir of untapped potential revealed by the genome sequence of hundreds of fungal species (Machida, Asai et al. 2005; Pel, de Winde et al. 2007; van den Berg, Albang et al. 2008). In the era of genomic abundance and sufficient computing power, the method of tapping into this cryptic box is revolutionized. The significance of sequence to compound relationship becomes the focal point of these efforts.

Two approaches will be utilized to this end: a) the bottom-up approach will start with studying the primary sequence and extrapolating a possible product of structural and/or pharmacological merit; b) the top-down approach, on the other hand, will commence with an important molecule, from which the biological machinery responsible for its synthesis (and eventually the genomic code) could be deduced from. This bioinformatics approach will be coupled with genetic and molecular biology techniques that would validate the established association. This involves gene inactivation, monitoring gene expression, as well as heterologous protein expression for in vitro studies. These efforts could lead to discovery of novel compounds, as well as generation of compound analogues. More fundamentally, it will lead to an increased understanding of the biosynthesis of these polyketides and bring to light the cryptic programming rules.
2 Results

2.1 Discovery and Biosynthesis of Azanigerones from *Aspergillus niger* ATCC 1015

2.1.1 Genome mining in *Aspergillus niger* ATCC 1015

2.1.1.1 Industrial relevance of *A. niger*

*Aspergillus niger* is a member of the black aspergilli group (*Nigri*) of filamentous fungi. Taxonomically, they are characterized by brown to black conidia with the characteristic “aspergillum” shape, although there are inconsistencies on the classification of the sub-species. Other *Aspergillus* strains, such as *A. fumigatus* and *A. flavus* are known to produce mycotoxins like the carcinogen aflatoxin and cause invasive aspergillosis among immune-compromised targets, in addition to widespread agricultural damage to crops (Latge 2001; Hedayati, Pasqualotto et al. 2007). On the other hand, *A. niger* is given the GRAS status (generally regarded as safe) by the Unites States and Food and Drug Administration (Schuster, Dunn-Coleman et al. 2002).

Since 1919, *A. niger* has been used in industrial fermentation for the production of citric acid, an important acidifying agent used in food and beverage, pharmaceutical, cosmetics and toiletries, detergents, electroplating and other applications that has an estimated market of $9 \times 10^6$ tons per year (Roukas 2000; Karaffa and Kubicek 2003). It has also been used in the production of useful enzymes, including pectinase, protease and aminoglucosidase and hemicellulases (Schuster, Dunn-Coleman et al. 2002). Due to its classification as a safety strain, it is also often used in the heterologous expression of related enzymes and has been subject to genetic
modifications to make it more amenable to such changes (Olempska-Beer, Merker et al. 2006; Lubertozzi and Keasling 2009).

For these reasons, several species of A. niger have been fully sequenced. The first genome sequence was of A. niger CBS 513.88, which was publicly released in 2007 and was published on Nature Biotechnology (Pel, de Winde et al. 2007). This strain was the ancestor of the current strain for production of industrial enzymes, specifically of glucoamylase. The genome was found to contain numerous putative secondary metabolic gene clusters, including 17 NRPSs, 34 PKSs and 7 PKS-NRPS hybrids. A few years later, by the initiative of the Joint Genome Institute, the wild-type A. niger strain for citric acid production, ATCC 1015, was also sequenced and annotated. The comparative genomics between the two strains uncovered genome rearrangements, deletions, horizontal gene transfer (from A. oryzae RIB40 to CBS 513.88) and a high level of single nucleotide polymorphisms (Andersen, Salazar et al. 2011). Among the shared secondary metabolites between the two species are aurasperone B, fumonisin B2, pyranonigrin A and tensidol B (Figure 2.1). The mycotoxin ochratoxin was found in CBS 513.88, but not in ATCC 1015, and was supposed to be a result of gene deletion (Andersen, Salazar et al. 2011). In both cases, the chemical potential of each strain is high, as evidenced by the higher number of secondary metabolic clusters than the compounds isolated.

2.1.1.2 Genome mining for PKS targets in A. niger ATCC 1015

We chose to work with the A. niger ATCC 1015 strain, which we purchased from the ARS Fungal collection. Our interest was primarily on the polyketide secondary metabolic clusters so we mined the genome specifically for the different fungal PKS types, namely HR-PKS, NR-PKS, PR-PKS and PKS-NRPS hybrids. Using the protein sequence of a representative ketosynthase (KS) domain from the lovastatin nonaketide synthase, we searched the sequenced genome of A.
niger ATCC 1015 using the BLAST (Basic Local Alignment Search Tool) software (Altschul, Madden et al. 1997). From this, we uncovered 37 putative PKS clusters

![Chemical structures](image1.png)

**Citric acid**  **Aurasperone B**  **Tensidol B**  **Pyranonigrin A**

**Figure 2.1.** Secondary metabolites produced by *A. niger*.

(Figure 2.2), of which HR-PKSs are the most common (23/38), significantly outnumbering the other subclasses. We also annotated the genes upstream and downstream of the PKSs to predict the putative products of the clusters (Table 2.1). The PKS genes are named according to the ASPNIDRAFT annotation that is currently used in the NCBI website.

An important part of genome mining is determining which targets to investigate further. Two approaches are commonly observed in literature: a) the top-down approach starts with a molecular target, where we deduce the enzyme chemistries required to achieve the final molecular scaffold; b) on the other hand, the bottom up approach starts with looking at the genomic sequence and extrapolating a possible product of structural and/or pharmacological merit. The complete metabolome of ATCC 1015 has not been fully explored, although the secondary metabolites from other species of the black aspergilli *Nigri* have been reported previously (Nielsen, Mogensen et al. 2009). From comparison of the PKS clusters to the compounds, it is quite clear that there are...
clusters that are not productive in the culturing conditions that have been investigated. We therefore chose to study these so-called “orphan clusters” to elucidate their biosynthetic products.

**Figure 2.2.** Breakdown of PKSs found from genome mining of *A. niger* ATCC 1015.

There are different strategies in activating silent clusters. The OSMAC strategy (one strain, many compounds) has been a reliable method for decades now that requires growing the same strain in different culture conditions to induce production of new compounds (Bode, Bethe et al. 2002). Similarly, epigenetic methods using histone deacetylase and DNA methyltransferase inhibitors have also been productive in de-repressing silent genes to produce new metabolites (Williams, Henrikson et al. 2008). However, these methods are non-specific and do not require the genomic information that we have at hand. One commonly-utilized targeted approach is activation of the cluster by integration of a pathway-specific regulator (Hertweck 2009). From scanning our mined genome, we found that most clusters contain a fungal transcription factor. We therefore examined these clusters more closely.
Table 2.1. Detailed list of PKS gene clusters found in *A. niger* ATCC 1015.

<table>
<thead>
<tr>
<th>No.</th>
<th>Gene Name (ASPNIDRAFT_)</th>
<th>Gene Location</th>
<th>Domain Architecture</th>
<th>PKS Class</th>
<th>Closest characterized homolog, % ID</th>
<th>Neighboring genes</th>
<th>Biosynthetic product (Reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>128638</td>
<td>Contig 15: 603261-610244</td>
<td>KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td><em>Cochliobolus heterostrophus</em>, polypetide synthase 2, 32% (T-toxin, blastX)</td>
<td>Permease, Glycosyl hydrolases, Flavin containing amine oxidoreductase</td>
<td>None known at time of study (See 2.1.2, p 42)</td>
</tr>
<tr>
<td>3</td>
<td>118598</td>
<td>Contig 14: 500929-509350</td>
<td>KS-AT-DH-MT-ER-KR-ACP</td>
<td>HR-PKS</td>
<td><em>Cochliobolus heterostrophus</em>, AAB08104, 32% (T-toxin, blastX)</td>
<td>ATP-NAD kinase, Arginase, Amino acid permease, KR, Reductase, ER, Oxidoreductase</td>
<td>None known at time of study (See 2.1.2, p 42)</td>
</tr>
<tr>
<td>4</td>
<td>181803</td>
<td>Contig 6: 671718-680022</td>
<td>KS-AT-DH-MT-ER-KR-ACP</td>
<td>HR-PKS</td>
<td><em>Pyrenophora tritici-repentis</em> Pt-1C-BFP,ppsA, 35% (phenolpthiocerol)</td>
<td>Carboxylesterase, oxidoreductase, kinase, Tannase and feruloyl esterase, transcription factor</td>
<td>None known at time of study (See 2.1.2, p 42)</td>
</tr>
<tr>
<td>5</td>
<td>56946</td>
<td>Contig 11: 1291712-1299626</td>
<td>(SAT)-KS-AT-(PT)-ACP-MT-TE</td>
<td>NR-PKS</td>
<td><em>Monascus purpureus</em>, citrinin polypetide synthase, 47%</td>
<td>P450, dehydrogenase, CoA ligase, KR, Phospholipase/Carboxylesterase, NRPS, ER, HR-PKS, Zn:Cys, regulator</td>
<td>None known at time of study (See 2.1.2, p 42)</td>
</tr>
<tr>
<td>6</td>
<td>188817</td>
<td>Contig 11: 1271817-1280064</td>
<td>KS-AT-DH-MT-ER-KR-ACP</td>
<td>HR-PKS</td>
<td><em>Penicillium citrinum</em>, BAC20566, 52% (compactin)</td>
<td>ER, NR-PKS (56946), P450, Oxidoreductase, Aminotransferase, fatty acid synthase, Enoyl-CoA hydratase/isomerase, PKS?(ACP-MT-TE), oxidoreductase,</td>
<td>None known at time of study (See 2.1.2, p 42)</td>
</tr>
<tr>
<td>7</td>
<td>179462</td>
<td>Contig 4: 1912782-1915643</td>
<td>(SAT)-KS-AT</td>
<td>NR-PKS</td>
<td>AptA, 75% (blastX)</td>
<td>Beta-lactamase, monoxygenase, ACP, sugar transporter, dehydrogenase,</td>
<td>TAN-1612 (Li, Chooi et al. 2011)</td>
</tr>
<tr>
<td>Contig</td>
<td>ID</td>
<td>Description</td>
<td>Function</td>
<td>Organism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>118617</td>
<td>Contig 2: 159269-167810 KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>Pyrenophora tritici-repentis Pt1C-BFP, lovB, 30% (blastX) Oxidoreductase, oxygenase, Permease, Amidohydrolase,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>194381</td>
<td>Contig 19: 583914-590870 (SAT)-KS-AT-(PT)-ACP-ACP-MT</td>
<td>NR-PKS</td>
<td>Monascus purpureus, citrinin polyketide synthase, 35% dehydrogenase; P450, Acetyltransferase, oxygenase, Aminotransferase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>128601</td>
<td>Contig 15: 700665-712297 KS-AT-DH-KR-ACP-C-A-T-R</td>
<td>PKS-NRPS</td>
<td>Fusarium heterosporum, equisetin synthetase, 36% (blastX) ER, hydrolase, P450, NRPS, transporter,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>118629</td>
<td>Contig 4: 2490985-2498440 KS-AT-DH-MT-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>Aspergillus terreus LovB, 36% Serine hydrolase, CDP-alcohol phosphatidylintransferase, hydrolase, aldolase,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>39026</td>
<td>Contig 4: 1163904-1171308 KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>Microsporum canis CBS 113480, PKSN polyketide synthase for alternapyrone biosynthesis, 54% ER, P450, dehydrogenase, protein kinase,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>181201</td>
<td>Contig 5: 2229939-2231991 KS-AT-DH-?</td>
<td>HR-PKS</td>
<td>Aspergillus terreus LovF, 44% Sulfatase, Glycosyl hydrolases, Metallo-beta-lactamase,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>187099</td>
<td>Contig 10: 108868-115966 KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>Aspergillus carbonarius, CAQ16344.1, 90% (ochratoxin A ) P450, FAD binding Domain, Dienelactone hydrolase, sugar transporter oxidoreductase, O-MT,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>51499</td>
<td>Contig 22: 10677-16095</td>
<td>(SAT)-KS-AT-(PT)-ACP</td>
<td>NR-PKS</td>
<td>Bipolaris oryzae, BAD22832, 39%</td>
<td>O-MT, MT, dehydrogenase, P450</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>56896</td>
<td>Contig 11: 1475115-1482726</td>
<td>(SAT)-KS-AT-(PT)-ACP</td>
<td>NR-PKS</td>
<td>Emericella nidulans, WA, 70%</td>
<td>Zn$_2$Cys$_6$ regulator, Thymidylate kinase, Methylenetetrahydrofolate reductase, monooxygenase</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>191422</td>
<td>Contig 14: 570392-577055</td>
<td>(SAT)-KS-AT-(PT)-ACP-TE</td>
<td>NR-PKS</td>
<td>Emericella nidulans, WA, 45%</td>
<td>P450, O-MT, O-MT, Aminotransferase</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>225587</td>
<td>Contig 1: 1974582-1976177</td>
<td>KS-AT-?</td>
<td>HR-PKS</td>
<td>Microsporum canis CBS 113480, PKSN polyketide synthase for alternapyrone biosynthesis, 32%</td>
<td>Aldo/keto reductase, HR-PKS</td>
<td></td>
</tr>
</tbody>
</table>

Possible fumonisin cluster (Baker 2006)
<table>
<thead>
<tr>
<th>Contig ID</th>
<th>Accession</th>
<th>Gene Information</th>
<th>Gene Function</th>
<th>Protein ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>190014</td>
<td>Contig 12: 548769-556528</td>
<td>(SAT)-KS-AT-(PT)-ACP-MT-TE</td>
<td>NR-PKS</td>
</tr>
<tr>
<td>31</td>
<td>194290</td>
<td>Contig 19: 254910-256148</td>
<td>(SAT)-KS-AT-(PT)-ACP-TE</td>
<td>NR-PKS</td>
</tr>
<tr>
<td>33</td>
<td>179079</td>
<td>Contig 4: 1055222-1063904</td>
<td>KS-AT-DH-MT-ER-KR-ACP</td>
<td>HR-PKS</td>
</tr>
<tr>
<td>34</td>
<td>118666</td>
<td>Contig 9: 608011-614743</td>
<td>KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
</tr>
<tr>
<td>35</td>
<td>43495</td>
<td>Contig 11: 1420047-1428259</td>
<td>KS-AT-DH-MT-ER-KR-ACP</td>
<td>HR-PKS</td>
</tr>
<tr>
<td>37</td>
<td>211885</td>
<td>Contig 10: 575986-583981</td>
<td>KS-AT-DH-MT-(ER)-KR-ACP (?)</td>
<td>HR-PKS</td>
</tr>
</tbody>
</table>
2.1.2 Characterization of a Silent Azaphilone Gene Cluster from *Aspergillus niger* ATCC 1015 Reveals a Hydroxylation-Mediated Pyran-Ring Formation.”

This section was originally published with the same title in Chem. Biol.: Zabala, A.O., Xu W. Chooi, Y.H., Tang, Y. 2012, Volume 19, Issue 8, Pages 1049-1059, ISSN 1074-5521, except for the final paragraph in section 2.1.2.2.2. The compounds are independently numbered from the other sections.

2.1.2.1 Introduction

Filamentous fungi are known to be prolific producers of secondary metabolites, such as the penicillin, lovastatin and cyclosporine, and are an important resource for discovering small molecules of pharmaceutical and industrial value (Keller, Turner et al. 2005). In the last decade, whole genome sequencing of various fungi has revealed that these microorganisms have immense biosynthetic potential that far surpasses the chemical diversity that we observe in laboratory culture (Sanchez, Somoza et al. 2012). For example, the genome of many aspergilli are found to encode for a combined 30 to 80 polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs) and PKS-NRPS hybrids, which far exceeds the total number of known polyketides and nonribosomal peptides (Sanchez, Somoza et al. 2012). Of these, the fungal PKSs are of considerable interest due to their interesting enzymology and the polyketide structural diversity.

Fungal type I PKSs contain multiple catalytic domains and resemble the animal fatty acid synthases, where a single set of catalytic domains is used iteratively. The chain extension by decarboxylative condensation of malonyl-CoA units is catalyzed by the minimal PKS domains, including ketosynthase (KS), malonyl-CoA:ACP transacylase (AT) and acyl carrier protein (ACP)
Non-reducing PKSs (NR-PKSs) synthesize a poly-\(\beta\)-ketone backbone which is cyclized by a product template (PT) domain to yield aromatic compounds such as orsellinic acid and norsolorinic acid (Crawford, Korman et al. 2009). In contrast, highly-reducing PKSs (HR-PKSs) utilize different combinations of ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains following each chain extension to reduce the \(\beta\)-keto positions in different extent, and produces reduced polyketides such as lovastatin and fumonisin (Cox 2007). Together with tailoring enzymes that are typically clustered in a biosynthetic pathway at the genetic level, the different fungal PKSs produce a large array of polyketides (Keller, Turner et al. 2005).

Bioinformatic analyses of different fungal genomes have revealed that it is common for two PKSs to be located in the same gene cluster (Sanchez, Somoza et al. 2012). The polyketide products of several of these dual PKS-containing gene clusters are known, including hypothemycin (Reeves, Hu et al. 2008), asperfuranone (Chiang, Szewczyk et al. 2009) and lovastatin (Kennedy, Auclair et al. 1999; Ma, Li et al. 2009). The two PKSs can either work in sequence or in convergence to synthesize the polyketide product. When the two PKSs function sequentially, the polyketide chain formed by the first PKS is transferred to the second PKS to continue the chain extension process. This has been demonstrated in the biosynthesis of the resorcylic acid lactones and asperfuranone, in which the upstream HR-PKS produces a partially reduced polyketide chain that is transferred to the downstream NR-PKS to be further elongated (Chiang, Szewczyk et al. 2009; Zhou, Qiao et al. 2010). In the convergent model, the two PKSs can function independently in parallel, and the two polyketide products are ultimately connected via accessory enzymes. An example is the biosynthesis of the lovastatin, in which the nonaketide and a diketide chains produced by two different HR-PKSs are combined via the action of the acyltransferase LovD (Xie, Meehan et al. 2009). With a limited number of dual-PKS systems characterized so far, it is
currently not possible to predict which mode of crosstalk (sequential or convergent) between the two PKSs will take place through bioinformatic means alone. Therefore, characterization of additional dual PKS-containing pathways will facilitate our understanding of the molecular and genetic basis that underlie the differences between the PKS-PKS partnerships, and enable better prediction of the gene cluster products.

![Chemical structures](image)

**Figure 2.3.** Examples of azaphilone compounds isolated from fungi. The conserved bicyclic core (highlighted in blue) is the hallmark of these compounds.

*Aspergillus niger* and closely related black aspergilli are known to produce a large number of secondary metabolites, with up to 145 compounds catalogued (Nielsen, Mogensen et al. 2009). Annotation of the sequenced *A. niger* genomes unveiled an impressive number of PKS genes, including 34 PKSs and 7 PKS-NRPS hybrids (Pel, de Winde et al. 2007; Andersen, Salazar et al. 2011). In this work, we mined the genome of *A. niger* ATCC 1015 and identified a polyketide gene cluster that contains a pair of HR-PKS and NR-PKS. Overexpression of a pathway-specific transcriptional regulator found in the gene cluster led to the overproduction of a number of previously unknown azaphilone natural products. Azaphilones possess a signature bicyclic chromophore and is known to react readily with amines to produce the vinylogous γ-pyridones.
This family is structurally diverse and includes bioactive compounds such as the antifungal lunatoic acid (Nukina and Marumo 1977), the anti-inflammatory agent monascorubrin (Yasukawa, Takahashi et al. 1994), the lipoxygenase inhibitor sclerotiorin (Chidananda and Sattur 2007), the food dye ankaflavin (Manchand, Whalley et al. 1973) as well as the nephrotoxic mycotoxin citrinin (Endo and Kuroda 1976; Sakai, Kinoshita et al. 2008) (Figure 2.3). We show here that the two partnering PKSs function in a convergent manner toward azaphilone biosynthesis.

2.1.2.2 Results

2.1.2.2.1 Bioinformatic analysis of A. niger ATCC 1015 genome reveals a gene cluster with two PKS genes

We scanned the sequenced genome of A. niger ATCC 1015 for dual PKS gene clusters and found a single gene cluster containing an HR- and an NR-PKS. The NR-PKS gene ASPNIDRAFT_56946 and the HR-PKS gene ASPNIDRAFT_188817 (herein referred to as azaA and azaB, respectively) appear to be homologs of afoE and afoG (44% and 43% protein identity) in the A. nidulans afo cluster, which is responsible for the synthesis of asperfuranone (Chiang, Szewczyk et al. 2009) (Figure 2.4 and Table 2.2). The azaA gene encodes for a Clade III NR-PKS according to a previous fungal PKS phylogenetic classification scheme (Kroken, Glass et al. 2003). In addition to the minimal PKS domains, it also includes an N-terminus starter unit: acyl-CoA transacylase (SAT) and a C-terminus didomain of C-methyltransferase (CMeT) and reductase (R). On the other hand, AzaB is a typical HR-PKS consisting of the minimal PKS domains, the β-keto processing domains (DH, ER and KR) and CMeT domain. The domain organizations of AzaA and AzaB are parallel to that of AfoE and AfoG, respectively.
Figure 2.4. Comparison of the *A. niger* aza cluster with the *A. nidulans* afo cluster and their respective products.

Analysis of nearby genes in the dual-PKS cluster reveals the presence of ASPNIDRAFT_132962 (azaR) encoding for a Zn(II)$_2$Cys$_6$ zinc finger transcription factor, which could regulate the coordinated expression of the genes in a cluster (Keller, Turner et al. 2005). Upstream and downstream of the PKSs, we find several tailoring enzymes that may modify the polyketide product. Interestingly, most of these genes have a corresponding homolog in the asperfuranone gene cluster, including azaC (afoC homolog), azaG (afoF homolog), azaH (afoD homolog), azaJ (AN1030.3 homolog), azaL (afoF homolog), each with $>$ 30% identity to their afo counterparts (Table 2.2). However, four new genes are also found in the cluster including an acyltransferase (azaD), a ketoreductase (azaE), an AMP-dependent CoA ligase (azaF) and a P450 monooxygenase (azaI). The extensive overlap between the aza and afo gene clusters suggests that the aza cluster may encode for the production of an asperfuranone-like compound.
Table 2.2. Putative functions and homologs of the genes in the *aza* cluster.

<table>
<thead>
<tr>
<th>Locus Tag (ASPNIDRAFT_)</th>
<th>Gene Name</th>
<th>Putative function</th>
<th>Characterized Homolog</th>
<th>Protein Identity / Similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>54405</td>
<td>n/a</td>
<td>hydrolase</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>127101</td>
<td>n/a</td>
<td>thioesterase</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>50208</td>
<td><em>azaC</em></td>
<td>esterase/lipase</td>
<td>AN1032.3, <em>afaC</em></td>
<td>44/56</td>
</tr>
<tr>
<td>189181</td>
<td><em>azaD</em></td>
<td>acyltransferase</td>
<td><em>F. sporotrichioides</em> Tri101</td>
<td>33/49</td>
</tr>
<tr>
<td>212676</td>
<td><em>azaE</em></td>
<td>ketoreductase</td>
<td><em>S. cerevisiae</em> GRE2</td>
<td>35/50</td>
</tr>
<tr>
<td>188806</td>
<td><em>azaF</em></td>
<td>acyl:CoA ligase</td>
<td><em>Physcomitrella patens</em> 4CL2</td>
<td>35/53</td>
</tr>
<tr>
<td>189194</td>
<td><em>azaG</em></td>
<td>FAD-dependent oxygenase</td>
<td>AN1035.3, <em>afaF</em></td>
<td>37/64</td>
</tr>
<tr>
<td>188800</td>
<td><em>azaH</em></td>
<td>salicylate monoxygenase</td>
<td>AN1033.3, <em>afaD</em></td>
<td>41/58</td>
</tr>
<tr>
<td>43449</td>
<td><em>azaI</em></td>
<td>cytochrome P450</td>
<td><em>A. flavus</em> CypX</td>
<td>30/49</td>
</tr>
<tr>
<td>56946</td>
<td><em>azaA</em></td>
<td>NR-PKS</td>
<td>AN1034.3, <em>afaE</em></td>
<td>44/60</td>
</tr>
<tr>
<td>132962</td>
<td><em>azaR</em></td>
<td>Zn$_2$Cy$_6$ regulator</td>
<td>AN1029.3, <em>afaA</em></td>
<td>34/48</td>
</tr>
<tr>
<td>43447</td>
<td><em>azaJ</em></td>
<td>dehydrogenase</td>
<td>AN1030.3</td>
<td>40/53</td>
</tr>
<tr>
<td>188912</td>
<td><em>azaK</em></td>
<td>efflux transporter</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>132654</td>
<td><em>azaL</em></td>
<td>FAD-dependent oxygenase</td>
<td>AN1035.3, <em>afaF</em></td>
<td>33/55</td>
</tr>
<tr>
<td>188817</td>
<td><em>azaB</em></td>
<td>HR-PKS</td>
<td>AN1036.3, <em>afaG</em></td>
<td>43/60</td>
</tr>
<tr>
<td>188504</td>
<td>n/a</td>
<td>cytochrome P450</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>50199</td>
<td>n/a</td>
<td>oxidoreductase</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

2.1.2.2.2 Overexpression of *azaR* activates the *aza* pathway

Among the polyketide metabolites catalogued for *A. niger* (Nielsen, Mogensen et al. 2009), none appears to be a likely candidate of the *aza* cluster. Reverse transcription PCR (RT-PCR) analysis of *azaA* and *azaB* showed that both PKS genes were weakly transcribed under our laboratory culturing conditions (Figure 2.5). Ectopic integration and overexpression of pathway-specific transcriptional activator has been used as a strategy to activate the expression of silent gene cluster for production of cryptic metabolites (Bergmann, Schumann et al. 2007). In order to interrogate the *aza* cluster, we constructed the overexpression plasmid pAZ44 by cloning *azaR* into pBARGPE1 (obtained from FGSC), which contains a P$_{gpdA}$ for gene expression in fungi and
a *bar* gene that confers resistance to glufosinate. Upon PEG-mediated protoplast transformation and selection, several transformants emerged and produced a yellow pigmentation not observed in the wild type strain. The genomic DNA from one of the transformants (T1) was extracted and integration of the vector was confirmed via amplification of P<sub>gpdA</sub>::*azaR* present in the pAZ44 plasmid.

**Figure 2.5.** Transcriptional analysis by RT-PCR of the *aza* genes in *A. niger* WT (i) and the activated T1 strain (ii). PCR from genomic DNA (iii) are shown for comparison.<sup>1</sup>  

In order to determine which genes are transcriptionally activated by the regulator, RT-PCR analysis was performed on the wild type (WT) and the *A. niger* T1 strains under identical conditions. We amplified a ~600 bp segment in each gene, particularly in those regions with introns in order to distinguish between the products from spliced cDNA and potential genomic DNA carryover during RNA extraction. In addition to *azaA*, *azaB* and *azaR*, 10 other genes were

---

<sup>1</sup> # Denotes intronless genes; therefore no size difference is expected between lanes ii and iii.
significantly upregulated (Figure 2.5). The genes immediately upstream of \textit{azaC} and downstream of \textit{azaB} showed no increase in transcription levels compared to the WT. We therefore hypothesize that the 13 transcribed genes comprise the \textit{aza} cluster.

![Figure 2.6. Metabolic profiles of the activated \textit{A. niger} T1 strain on days 2, 4 and 7 and of the WT on day 2.](image)

LC-MS analysis of the metabolic profile of the activated \textit{A. niger} T1 reveals the presence of several new compounds not found in the WT strain (Figure 2.6). The culture was scaled up to 2 L to obtain sufficient compound for structure elucidation. In addition, a time course of the metabolites produced by T1 was performed in order to identify possible biosynthetic intermediates of the \textit{aza} pathway (Figure 2.6).

From the second and fourth day cultures of \textit{A. niger} T1, compound 1 (\(m/z\) 363 [M+H]⁺, \(\lambda_{\text{max}}\) 330 nm) was the most abundant product (Figure 2.6); it was purified as a yellow gum from the 4-day culture at an approximate yield of 60 mg/L. High-resolution mass of 1 is consistent with
the molecular formula of C$_{19}$H$_{22}$O$_7$ (observed m/z 363.1465 [M+H]$^+$, calculated 363.1444). The pure compound was characterized by $^1$H and $^{13}$C NMR spectroscopy, which showed all of the 19 carbon and 22 proton signals (Table 4.1 and Figure 4.5, Figure 4.6). From the 2D HSQC-INEPT135 and HMBC correlations (Figure 4.7), it is apparent that there are two substructures in the molecule – a fully-saturated aliphatic chain and a conjugated cyclic moiety.

From the 2D HSQC-INEPT135 and HMBC correlations (Figure 4.7), it is apparent that there are two substructures in the molecule – a fully-saturated aliphatic chain and a conjugated cyclic moiety.

![Structures of 2 - 7 isolated from the activated A. niger T1 strain.](image)

The highly-reduced chain is consistent with a 2,4-dimethyl hexanoyl structure characterized by 3 methyls, 2 methylenes and 2 methines. We did not find any HMBC correlation beyond the carbonyl carbon with a chemical shift of 176.86 ppm (Figure 4.8). Meanwhile, we assigned the rest of the signals to a bicyclic ring system consisting of three aromatic protons [$\delta_H$ 7.93 ppm (s), 7.31 ppm (s), 5.88 ppm (s)] with a carboxylic acid moiety, which corresponded to a low-field proton signal [$\delta_H$ 10.2 ppm (br s)]. We identified that the only available position on the bicyclic ring system for the linkage to the 2,4-dimethyl hexanoyl moiety is the oxygen atom bonded to the

\[ * \text{Denotes compounds elucidated by full NMR spectroscopy.}\]
methylated carbon (δC 84.06 ppm). We therefore propose the structure of 1 as shown in Figure 2.4, with the two substructures connected by an ester bond at C1’ and C4. MS/MS analysis supports this linkage, which showed a fragment consistent with the mass of the bicyclic core (m/z 237 [M+H]+) (Figure 4.2). Thus, 1 was determined to be a new member of azaphilones, which bears the typical pyrano-quinone bicyclic core, and was named azanigerone A. The configurations of the chiral centers in 1 are yet to be determined. The structure of 1 resembles lunatoic acid, a previously characterized azaphilone from the plant pathogen Cochliobolus lunatus (Nukina and Marumo 1977) (Figure 2.3). The aliphatic acyl chains of the two molecules are identical as evidenced by comparison to published values for lunatoic acid (Nukina and Marumo 1977). The main structural difference is the side chain on C9, where a longer acrylic acid is in place in lunatoic acid instead of the C1 carboxylic acid in 1.

In addition to 1, several other compounds were observed in the culture at different time points. Compound 2 (m/z 251 [M+H]+, λmax 300 nm) is produced on day 2 but disappeared at day 4 and beyond (Figure 2.6, Figure 2.7). This compound was purified and its structure was determined to be the tetra-substituted benzaldehyde FK17-P2a (Figure 2.7 and Table 4.1) that was previously isolated from Aspergillus versicolor and Pseudobotrytis terrestris (Arai T 1994; Yamaguchi, Masuma et al. 2004). FK17-P2a should be a polyketide that originates from a methylated hexaketide that can be synthesized by a NR-PKS, which undergoes C2-C7 intramolecular cyclization and reductive release to yield a 1,3-diketo benzaldehyde (Bailey, Cox et al. 2007). Further ketoreduction of the terminal ketone affords 2.

Compound 3 (m/z 377 [M+H]+, λmax 330 nm) was produced at a much lower yield during the earlier culture periods (Figure 2.6, Figure 2.7, Table 4.2). The bicyclic core structure was resolved to be similar to 1, but instead of the carboxylic acid moiety three new proton signals were
detected: a terminal methyl doublet ($\delta_H$ 1.32 ppm), a methylene doublet ($\delta_H$ 2.54 ppm) and a methane multiplet ($\delta_H$ 4.16 ppm) (Table 4.2, Figure 4.13). These signals are consistent with a 2-hydroxypropyl substitution at the C9, which is consistent with the $m/z$ 251 [M+H]$^+$ fragment from MS/MS analysis (Figure 4.3). The 2-hydroxypropyl moiety is reminiscent of the terminal portion of the 2-oxo-4-hydroxypentyl substituent in 2, which suggests that 2 may be an intermediate in the synthesis of 3 (Figure 2.7). Meanwhile, 4 ($m/z$ 393 [M+H]$^+$, $\lambda_{max}$ 330 nm) was produced at an even lower titer and its structure was not fully elucidated by NMR due to insufficient material obtained (Figure 2.6, Figure 2.7). Through comparison of MS/MS peaks of 3 and 4, as well as the molecular formula of 4 obtained through high resolution MS (C$_{21}$H$_{28}$O$_7$, observed $m/z$ 393.1908 [M+H]$^+$, calculated 393.1913) (Figure 4.4), we propose that 4 could be a 10-hydroxy derivative of 3 (Figure 2.7), and may also be an intermediate toward the biosynthesis of 1.

After seven days of culturing, a new compound 5 ($m/z$ 362 [M+H]$^+$, $\lambda_{max}$ 357 nm) replaces 1 as the dominant product in the A. niger T1 culture (Figure 2.6, Figure 2.7). 5 was purified from a 7-day culture and its structure was confirmed via NMR to be the vinylogous $\gamma$-pyridone derivative of 1 (Figure 2.6, Figure 2.7, Table 4.3). This compound is likely originated from 1 via a characteristic azaphilone reaction with amines present in the culture medium. Indeed, a compound with the identical retention time, mass and $\lambda_{max}$ was obtained by reacting purified 1 with NH$_4$OH (data not shown).

To determine the stereochemistry of the compounds, we used circular dichroism to compare the spectra of each compound with other known azaphilones for which the absolute stereochemistry of the core have been determined (Steyn and Vleggaar 1976). All the azanigerones displayed a positive Cotton effect at 350 nm and negative effects at 273 nm and 241, indicative of an R-configuration on C4.
2.1.2.2.3 Knockout of azaB demonstrates a convergent synthesis of 1 by AzaA and AzaB

Inspection of the structure of 1 suggests that the molecule is assembled convergently from two discrete polyketide chains. We propose that the dimethylhexanoate moiety is produced by the HR-PKS AzaB, while the bicyclic core is cyclized from an unreduced polyketide backbone synthesized by the NR-PKS AzaA. This is in stark contrast to the sequential mode of collaboration between AfoE and AfoG in the highly parallel asperfuranone gene cluster (Chiang, Szewczyk et al. 2009), and underscores the difficulties in accurately predicting product structures among fungal PKSs.

Therefore in the convergent synthesis model, disruption of azaB in T1 should not affect the function of AzaA, and should lead to accumulation of the pyrano-quinone. To test this model, genetic disruption of azaB in A. niger T1 was performed using a double-crossover recombination with the zeocin-resistant gene ble (Figure 2.8). Two desired ΔazaB transformants were identified via diagnostic PCR (Figure 2.8A). After culturing and metabolite analysis, the production of 1 was confirmed to be abolished in the ΔazaB strains. Instead, two new compounds (6 and 7) that have UV spectra characteristic of pyrano-quinone (identical to that of 1) emerged (Figure 2.8B). Compound 6 and 7 have molecular masses of 250 and 292, respectively (Figure 4.1). Both compounds were purified and their structures were elucidated by 1D and 2D NMR (Table 4.3). 6 contains the same bicyclic core and the 2-hydroxypropyl side chain as 3, but lacks the 2,4-dimethylhexanoyl ester as a result of azaB disruption (Figure 2.8C). Compound 7 is the O4 acetylated version of 6 (Figure 2.8C). In addition, the benzaldehyde 2 observed in the A. niger T1 strain was present in the early cultures of the ΔazaB strains (Figure 2.6). Upon closer examination, trace amounts of 6 and 7 were also found in A. niger T1 culture on Day 2 (Figure 2.6, Figure 2.7); the non-acetylated 6 is likely an intermediate in the biosynthesis of 1.
2.1.2.2.4 *In vitro* reaction of AzaH with the benzaldehyde intermediate confirms its roles in hydroxylation and pyran-ring formation

The formation of the bicyclic core of azaphilones has not been elucidated to date. It has been suggested that it proceeds from a benzaldehyde precursor in asperfuranone biosynthesis (Chiang, Szewczyk et al. 2009), which is structurally similar to 2 isolated from *A. niger* T1 in early culturing times. Since 2 is isolated in stable form, the route to the formation of the pyran ring in the biosynthesis of 1 must be enzyme-catalyzed.

---

**Figure 2.8.** (A) Knockout strategy employed for *azaB*. Correct integration is verified by using the three sets of primers shown. (B) HPLC traces demonstrating the abolishment of 1 production in the strain and the accumulation of 6 and 7.
We initially proposed that AzaC is involved in the pyran-ring cyclization. NCBI Conserved Domain Search results indicate that both AzaC and its homologs (CtnB from the citrinin pathway, and AfoD) belong to the esterase-lipase family of serine hydrolases and are not oxidoreductases as assigned previously for CtnB and AfoD (Sakai, Kinoshita et al. 2008; Chiang, Szewczyk et al. 2009). Since 1, citrinin and asperfuranone all contain a fused heterocyclic ring system, AzaC, CtnB and AfoB may be responsible for their corresponding heterocycle formation; a similar hypothesis has been proposed in a recent study as well (Davison, Al Fahad et al. 2012). Accordingly, the recombinant AzaC protein (30 kDa) was purified and assayed with 2. However, no conversion of the substrate 2 was observed in the in vitro reaction (Figure 2.9). We therefore reasoned that AzaC might act on a different substrate, possibly the C4-hydroxylated derivative of 2.

We then searched for genes in the aza cluster that may hydroxylate the C4 position in 2. AzaH, an FAD-dependent monooxygenase, was found to be homologous to a recently characterized monooxygenase TropB (43% identity, 61% similarity), which was recently shown

\[ \text{Figure 2.9. In vitro reaction with AzaH. (A) HPLC analysis of the in vitro assays of AzaH and/or AzaC (10 mM) with 2 (50 mM), and comparison with the standard 6. (B) Time course of AzaH-catalyzed conversion of 2 to 6.} \]
to hydroxylate the corresponding carbon of 3-methylorcinaldehyde during tropolone biosynthesis (Davison, Al Fahad et al. 2012). To examine its role in the synthesis of 1, AzaH was expressed and purified from E. coli BL21(DE3) and a series of in vitro assays including AzaH and AzaC/AzaH in the presence of 2 was performed. Surprisingly, incubation of both purified AzaH or AzaC/AzaH with 2 and NADPH yielded a new product that has an identical UV spectrum, retention time and mass as 6 (Figure 2.9A). This indicates that AzaH alone is capable of converting 2 to 6. However, in the absence of NADPH, which is needed for the regeneration of the reduced flavin cofactor, no substrate conversion by AzaH was observed. Therefore, the flavin-dependent hydroxylation of C-4 is likely a prerequisite for the subsequent pyran formation. Further LC/MS time-course analysis of the reaction of AzaH with 2 showed that the conversion proceeds rapidly with no detectable uncyclized C4-hydroxylated intermediate (Figure 2.9B). These results confirmed the critical role of AzaH in morphing the benzaldehyde intermediate into the bicyclic, pyran-containing azaphilone chromophore.

2.1.2.2.5 Proposed pathway for biosynthesis of the azanigerones

The RT-PCR transcriptional analysis indicates that there are a total of 12 genes in the aza cluster, which were activated upon overexpression of the pathway-specific regulator azaR (Figure 2.5). Combining the structural characterization of the compounds from the activated A. niger T1 culture and T1 ΔazaB, as well as in vitro confirmation of the role of AzaH, a biosynthetic pathway leading to the production of 1 can be proposed (Figure 2.10). Based on the time course studies of the metabolic profile of A. niger T1 (Figure 2.6), 1 is most likely to be the end product of the pathway, as it is the dominant compound in the culture after 4 days. We propose that the biosynthesis of 1 begins with the polyketide assembly by AzaA, which is a typical Clade III NR-PKS with a domain organization of SAT-KS-AT-PT-ACP-CMeT-R from the N- to C-terminal.
AzaA forms the hexaketide precursor from successive condensations of five malonyl-CoA units presumably with a simple acetyl-CoA starter unit. The reactive polyketide chain then undergoes a PT-mediated C2-C7 cyclization to afford the aromatic ring and is eventually released as an aldehyde through the R-domain, in a manner similar to the previously characterized 3-methylorcinaldehyde synthase in *Acremonium strictum* (Bailey, Cox et al. 2007). The putative ketoreductase, AzaE, a homolog of the methylglyoxal reductase in *Saccharomyces cerevisiae* (35% identity), is proposed to catalyze the reduction of the terminal ketone resulting in the early culture product 2.

The monooxygenase AzaH was demonstrated to be the only enzyme required to convert 2 to 6 *in vitro*. We propose that AzaH first hydroxylates the benzaldehyde intermediate 2 at C4, which triggers the formation of the pyran-ring to afford 6 (Figure 2.10). In parallel, the 2,4-dimethylhexanoyl chain is synthesized by the HR-PKS AzaB, and is proposed to be transferred to the C4-hydroxyl of 6 by the acyltransferase AzaD directly from the ACP domain of AzaB via a similar mechanism to the LovD acyltransferase (Xie, Meehan et al. 2009). Alternatively, the 2,4-dimethylhexanoyl chain may be offloaded from the HR-PKS as a carboxylic acid and converted to an acyl-CoA by AzaF, which shares 35% identity to the 4-coumarate:CoA ligase (Silber, Meimberg et al. 2008). The resulting acyl-CoA molecule could then be taken up as a
substrate by AzaD to form 3 (Figure 2.10). Like the homolog acetyltransferase Tri101 in trichothecene biosynthesis (33% identity to AzaD) (Garvey, McCormick et al. 2008), AzaD belongs to the BAHD family of acyltransferases that catalyze transfer of various acyl-CoA substrates (D’Auria 2006). It is interesting to note that the C4-O-acetylated 7 is also observed in the activated A. niger T1 and ΔazaB cultures (Figure 2.6). This suggests that AzaD may have a relaxed specificity toward acyl-CoA substrates, although it is entirely possible that other endogenous acyltransferases might be involved in this reaction.

In order to yield the carboxylic acid substituent in 1, the hydroxypropyl side chain of 3 would need to undergo a C-C oxidative cleavage. We propose that this C-C bond cleavage is
catalyzed by a cytochrome P450 Azal encoded in the cluster. Similar reactions have been observed in other P450s, with the CYP11A involved in the decomposition of cholesterol into pregnenolone and 4-methylpentanal as the archetypal enzyme (Burstein and Gut 1971). Such P450 is proposed to act on a vicinal diol that leads to a C-C bond scission either through an alkoxyradical intermediate or a peroxy complex (Ortiz de Montellano 2005). In the biosynthesis of 1, we propose that 3 first undergoes hydroxylation at C10, possibly catalyzed by one of the two FAD-dependent monooxygenases encoded in the cluster, AzaG or AzaL, resulting in the vicinal diol (Figure 2.10). A small amount of this putative intermediate 4, with mass and fragmentation pattern consistent with such a diol, was isolated from the 2-day culture of A. niger T1 (Figure 2.10). Oxidative cleavage 4 by Azal would yield the corresponding aldehyde derivative of 1. A homolog of Azal, CypX, also exists in the aflatoxin biosynthesis in A. parasiticus (35% identity) and is found to convert averufin to hydroxyversicolorone via a still unknown oxidative cleavage and ring rearrangement step (Wen, Hatabayashi et al. 2005). Finally, the dehydrogenase AzaJ (homologous to the AN1030.3, 40% identity) is proposed to convert the aldehyde functional group into the carboxylic acid, completing the conversion from 3 to 1. Alternatively, the oxidation of aldehyde to carboxylic acid may be catalyzed by the same P450 enzyme Azal via consecutive oxidation or by endogenous alcohol dehydrogenase. Interestingly, the proposed oxidative chain-shortening does not occur in the absence of the 2,4-dimethylhexanoate ester. Both 5 and 6 contain the hydroxypropyl chain and are not converted to the corresponding acids in ΔazaB culture.

2.1.2.3 Discussion

Genome mining, in particular activation of silent secondary metabolic pathways by overexpression of pathway-specific regulator, is becoming a useful strategy for natural product discovery in fungi. From overexpression of the cluster-specific regulator, we have uncovered novel
compounds that belong to the family of azaphilones. The hallmark feature of azaphilones is the presence of a highly oxygenated pyran-quinone bicyclic core of polyketide origin (Osmanova, Schultze et al. 2010). The name derives from their tendency to react with primary amines to yield γ-pyridones that result from the exchange of the pyrane oxygen with nitrogen. Over 170 different azaphilone compounds have been identified from hundreds of fungal species across 23 genera from 13 families, and many of the azaphilones exhibit diverse biological activities, including antibacterial, antifungal and antitumor (Osmanova, Schultze et al. 2010). Several common modifications that contribute to the structural diversification of the common bicyclic chromophore include but are not limited to a) aliphatic side chains at C9; b) aliphatic esters at O4; c) halogen substitution on the bicyclic core, as in sclerotiorin (Chidananda and Sattur 2007); d) additional fused rings, as in ankaflavin (Manchand, Whalley et al. 1973); e) aromatic esters at O4, as in mitorubrinol (Quang, Hashimoto et al. 2005); and f) molecular dimers, as in chaetoglobin A (Ge, Zhang et al. 2008) (Figure 1).

Despite their apparent ubiquity among fungi, the genetic and molecular basis for the biosynthesis of these azaphilones has not been thoroughly explored or understood. Besides labeling studies suggesting a polyketide origin for the bicyclic core (Ogihara, Kato et al. 2000), only the azaphilone-like compound asperfuranone and citrinin have been linked to their corresponding biosynthetic gene clusters (Sakai, Kinoshita et al. 2008; Chiang, Szewczyk et al. 2009). In this study, we have discovered a silent gene cluster in A. niger responsible for producing a series of azaphilones. Characterization of this aza gene cluster has provided important insights on the biosynthesis of this family of compounds.

In total, six new azaphilones, azanigerone A-F, were isolated and all share a substituted bicyclic core indicative of a common biosynthetic origin. We also observed their signature reaction
with amine groups in the conversion of the major product 1 into the vinylogous \( \gamma \)-pyridone 5. All six compounds contain a C4 hydroxyl group, which is a post-PKS tailoring modification observed in a multitude of other azaphilones. With the exception of 6, the tertiary alcohol at C4 of azanigerones is acylated with an aliphatic acyl or an acetyl group, which is another common feature among many azaphilones. The isolation of pyran intermediates at early stages of the \textit{A. niger} T1 culture underscores the convergent actions of the two PKSs in the production of these molecules (Figure 2.10).

A key discovery in this study is the enzyme responsible for the formation of the characteristic pyrano-quinone core common to all azaphilones. It has been proposed that the biosynthesis of azaphilones proceeds from a benzaldehyde intermediate (Chiang, Szewczyk et al. 2009), but the enzymology of the pyran-ring formation step has not been elucidated. We have discovered that the cyclization of the benzaldehyde is mediated by the C4-hydroxylation catalyzed by the FAD-dependent monooxygenase AzaH. We propose that the hydroxylation at C4 by AzaH dearomatizes the ring and causes keto-enol tautomerization at the C1 aldehyde leading to the condensation between the C1 enol and the C9 carbonyl to yield the pyran ring (Figure 6). Our finding that the C4-hydroxylation and pyran-ring cyclization occur very rapidly suggests that these two features are coupled. This is supported by the fact that with a few exceptions, majority of the azaphilones contain such geminal methyl, hydroxyl (or ester) substitution at the pyrano-quinone core (Figure 2.3) (Osmanova, Schultz et al. 2010). For the few azaphilones that do not have this feature, such as in the case of citrinin, a different mechanism of cyclization might take place. Coincidentally, the citrinin gene cluster lacks the \textit{azaH} homolog (Sakai, Kinoshita et al. 2008).

It is curious to note that with the exception of a few genes, the \textit{aza} and the \textit{afo} clusters are highly similar, with multiple gene homologs sharing >30\% protein sequence identity (Figure 2.4,
Table 2.2). However, the structures of the two compounds are considerably different. A striking disparity between the two biosynthetic pathways is on how the dual PKS systems partner to form their respective products. In the case of the *afo* cluster, the HR-PKS AfoG synthesizes the tetraketide intermediate, which is passed on to the SAT domain of the NR-PKS AfoE. Such sequential mode of collaboration occurs in other HR/NR-PKS tandem systems, such as in the biosynthesis of resorcylic acid lactones (Zhou et al. 2010). We were expecting the two PKSs in *aza* cluster, AzaA and AzaB, to follow the same tandem mode of PKS-PKS collaboration. To our surprise, the two PKSs acted independently of each other in the convergent synthesis of the two substructures of 1. Whereas the Δ*afoG* culture completely disrupted the production of asperfuranone or any appreciable intermediate, compounds 6 and 7 lacking the reduced triketide chain were isolated in the Δ*azaB* culture. Previous studies have pointed to the N-terminus SAT as a specificity-conferring domain that selects the starter unit (Crawford, Vagstad et al. 2008). By multiple protein sequence alignment, it was noticed that the SAT domains of AzaB and AfoG have a lower identity (35% vs. 43%) compared to the rest of the protein, although both contain the GXCXG motif found in most SATs. It is likely that the selectivity of the SAT and its ability to interact with HR-PKS hold the key for which mode of collaboration will take place between a pair of HR- and NR-PKSs. The *azaA-azaB* dual PKS system in *A. niger* is the first example of a convergent mode of collaboration between an HR and an NR-PKS; elucidation of the basis for the intriguing difference between *aza* and *afo* pathways could facilitate better prediction of polyketide products from such pathways.

Another notable difference between the structures of azanigerones and asperfuranone is the formation of a furan in the latter pathway instead of a pyran ring that is common in azaphilones. Both pathways are proposed to go through a benzaldehyde intermediate. However, in the case of
asperfuranone, it is likely that the benzylic carbon hydroxylation precedes the dearomatizing hydroxylation. It was previously proposed that AfoD catalyzes this benzylic carbon hydroxylation, which is the key branching point in the biosynthesis of asperfuranone and the azaphilone sclerotiorin (Davison, Al Fahad et al. 2012; Somoza, Lee et al. 2012). However, it is more likely that AfoD has a similar function as the homolog AzaH characterized in this study, i.e. the formation of the tertiary alcohol. Thus, AfoF, another FAD-dependent monooxygenase, may catalyze this benzylic carbon hydroxylation step in the asperfuranone pathway. Two AfoF homologs, AzaG and AzaL, are also found in the A. niger aza cluster (Table 2.2). We proposed that one of them may act later in the pathway in the hydroxylation of C10 to form the pre-cleavage vicinal diol.

Among the aza genes that share a homolog to the afo cluster, the function of AzaC remains enigmatic. The possibility of its involvement in the heterocycle formation has been excluded by the results in this study. In line with the NCBI Conserved Domain Search results that suggest an esterase-lipase function for this enzyme, we propose that AzaC might participate in the offloading of the 2,4-dimethylhexanoyl chain from the ACP of AzaB (Figure 2.10). The role of the homologs AfoC and CtnB in the biosynthesis of asperfuranone and citrinin respectively is also unclear and should be the subject of further investigation.

2.1.2.4 Significance

Activation of the silent aza cluster has led to the discovery of new compounds named azanigerones, which have not been isolated from A. niger before. More importantly, characterization of the pathway has shed light to the biosynthesis of the azaphilone group of compounds, which are structurally diverse and exhibit a wide range of bioactivities. This study provides an important basis for further molecular genetics and biochemical investigations of azaphilone biosynthesis. The sequence information will aid in identification of the gene clusters
encode for known bioactive azaphilones and discovery of new azaphilones by genome mining. The many structural diversifications on a common bicyclic core observed among this family of fungal natural products also offer a unique opportunity for combinatorial biosynthesis to expand the library of azaphilones and screen for enhanced bioactivities.

2.1.2.5 Experimental Procedures

2.1.2.5.1 Strain and culture conditions

* A. niger* ATCC 1015 was obtained from Agricultural Research Service (ARS) culture collection (as NRRL 328) and cultured at 28 °C in potato dextrose agar medium. The activated *A. niger* T1 strain and the ΔazaB mutant strains were maintained in glucose minimal medium (GMM) containing 10 mM ammonium tartrate as nitrogen source (Chooi, Cacho et al. 2010) or yeast glucose medium (YG) (Szewczyk, Nayak et al. 2006). *E. coli* strains XL1 (Stratagene) and TOPO10 (Invitrogen) were used for routine cloning. *Saccharomyces cerevisiae* BJ5464 was used for in vivo yeast DNA recombination cloning. *E. coli* BL21(DE3) was used for protein expression.

2.1.2.5.2 Molecular genetic manipulations

Polymerase chain reactions were performed using Phusion high-fidelity DNA polymerase (New England Biolabs), Platinum Pfx DNA polymerase (Invitrogen) and GoTaq Green Master Mix (Promega). PCR products were cloned into a PCR-Blunt vector (Invitrogen) for DNA sequencing and subcloning. Restriction enzymes (New England Biolabs) and T4 ligase (Invitrogen) were used respectively for the digestion and ligation of DNA fragments. Primers used for amplification were synthesized by Integrated DNA Technologies and are listed in Table S4.

Genomic DNA of *A. niger* ATCC 1015 was extracted using the ZYMO ZR fungal/bacterial DNA kit according to manufacturer’s protocols. The azaR gene (including introns) was amplified
from the genomic DNA and inserted into the fungal shuttle vector pBARGPE1 (digested with 
*Bgl*II and *Eco*RI) to yield pAZ44. The pBARGPE1 vector was obtained from Fungal Genetic Stock 
Center (FGSC) (Pall and Brunelli 1993). The intronless *azaH* and *azaC* were also amplified from 
genomic DNA and inserted into the p8HIS expression vector (Jez, Ferrer et al. 2000) digested with 
*Eco*RI and *Not*I to yield pAZ81 and pAZ83, respectively.

The knockout cassette for *azaB* was assembled in yeast using the yeast recombination 
method (Colot, Park et al. 2006). The two homologous regions were amplified from *A. niger* 
genomic DNA using primers containing overlapping regions with the yeast vector and the zeocin- 
resistance cassette. The zeocin-resistance cassette containing the *ble* gene under the *gpdA* promoter 
was amplified from pAN8-1 (Punt, Oliver et al. 1987) using *gpdA*for and blerev. The three DNA 
fragments were co-transformed with the linearized vector backbone derived from YEplac195 
(Gietz and Sugino 1988) into the *S. cerevisiae* BJ5464 (Jones 1991) using *S. c.* EasyComp™ 
Transformation Kit (Invitrogen) and selected on uracil-dropout semisynthetic media. The resulting 
transformants were screened by colony-direct PCR and the plasmid in the correct transformant 
was rescued using the Zymoprep Yeast Plasmid Miniprep Kit (Zymo Research) and transformed 
into *E. coli* for propagation and sequencing verification. The resulting plasmid was designated 
AZ61. This plasmid was used as template to obtain the linear *azaB* knockout cassette for 
transformation of *A. niger* T1.

2.1.2.5.3 *Fungal transformation*

The preparation of *A. niger* protoplasts and the PEG-mediated transformations were carried 
out as previously described (Li, Chooi et al. 2011). The *A. niger* transformed with pAZ44 for *azaR* 
overexpression was selected on approximately 8 mg/mL of glufosinate prepared as described 
previously (Chooi, Cacho et al. 2010). For genetic manipulation in the activated *A. niger* strain T1,
a modified protocol was used derived from the one developed for *Aspergillus nidulans* (Szewczyk, Nayak et al. 2006). The following modifications were implemented: a) the 2x protoplasting solution was prepared with 3g of VinoTaste Pro; b) the PEG solution used comprised of 25% PEG 4000, 10mM CaCl\(_2\) and 10mM Tris-HCl; c) the transforming protoplasts were first plated on non-selective media for five hours before overlaying with soft agar (8g/L) with 200 µg/mL of Zeocin; and d) the protoplasts were stabilized with 1.2M sorbitol.

2.1.2.5.4 *Expression analysis by reverse transcription polymerase chain reaction (RT-PCR)*

The total RNA of *A. niger* mutant and wild-type strains were extracted using the Ambion RNA extraction kit. The first strand cDNA was synthesized using the Oligo-dT primer and Improm-II reverse transcription system (Promega) according to the manufacturer’s instructions. cDNA was then amplified with GoTaq Green Master Mix using gene-specific primers (Table 4.4).

2.1.2.5.5 *Expression and purification of AzaC and AzaH*

pAZ81 and pAZ83 were used to transform the BL21(DE3) cells for expression of AzaC and AzaH, respectively. The cells were cultured at 37 °C and 250 rpm in 500 mL of LB medium supplemented with 35 µg/mL kanamycin to a final OD\(_{600}\) of 0.4-0.6 followed by addition of 0.1 M isopropylthio-β-D-galactoside (IPTG) to induce protein expression. The proteins were purified by Ni\(_2+\) affinity chromatography as described previously (Li, Chooi et al. 2011). Purified AzaC and AzaH was concentrated and exchanged into PBS (50mM, pH 7.4) with the centriprep filters (Amicon). Protein concentration was determined with the Bradford assay using bovine serum albumin as a standard.
2.1.2.5.6 *In vitro assays with AzaC and AzaH*

The reactions were performed at 50 μL scale containing 100 mM phosphate buffer (pH 7.4) in the presence of 50 μM of 2, 0.5 mM NADPH and 10 μM of AzaH and/or AzaC. After 2 hours, the reactions were quenched by precipitating the enzyme/s with 100 μL of acetonitrile. The acetonitrile/aqueous solution was concentrated 5-fold and directly analyzed on the LC-MS (Figure 5A). The time course analysis was performed with the same assay condition except that 0.5 μM AzaH was used and the reactions were quenched at different time points (Figure 5B).

2.1.2.5.7 *LC/MS analysis*

For small-scale analysis, the *A. niger* wild-type and transformants were grown in 10 mL GMMT under shaking conditions at 28 °C. The cultures were extracted with equal volume of ethyl acetate (EA) with 0.1% trifluoroacetic acid (TFA) and evaporated to dryness. The dried extracts were dissolved in methanol for LC/MS analysis using a Shimadzu 2010 EV Liquid Chromatography Mass Spectrometer with positive and negative electrospray ionization and Phenomenex Luna 5µL 2.0x10mm C18 reverse phase column. The samples were resolved on a linear gradient from 5 to 95% with CH$_3$CN/H$_2$O + 0.05% formic acid solvent system.

2.1.2.5.8 *Isolation of compounds*

Most compounds were purified from shake flask cultures, except for 2, which was obtained from 2-day static culturing. 1, 3 and 4 were extracted with EA after four days of culturing, 5 after seven days, and the azaB knockout compounds, 6 and 7, after two days. The following scheme was generally used to purify each compound from the crude extract: a) chloroform-water partitioning; b) hexane - 9:1 methanol:water partitioning; c) Sephadex LH-20 (GE) chromatography; and d) preparative high pressure liquid chromatography with Phenomenex Luna 5µL 250x1000mm C18 reverse phase column using CH$_3$CN/H$_2$O + 0.1% TFA or +0.2% formic
acid solvent system. DRX500 or ARX500 instruments were used to perform the NMR spectroscopy of the different compounds.

2.1.3 Heterologous expression of *aza* proteins

2.1.3.1 Cloning and in vivo reconstitution of *AzaA* activity in yeast

The pyrano-quinone core of the azanigerones is proposed to be synthesized by the NR-PKS AzaA. This was based on the bioinformatic and transcriptional analysis performed on the activated *A. niger* T1 strain. However, in order to interrogate the mechanism of the enzyme, it is important to reconstitute its activity in vitro. To do this, we cloned out the uninterrupted coding region of the 7.9-kb gene by performing RT-PCR on mRNA under producing conditions. In order to ensure the fidelity of the transcription, we divided the gene into six equal-sized fragments; each fragment was amplified from the cDNA and was subcloned and sequence-verified to confirm correct splicing of the introns (PCRB constructs of each fragment were designated as pAZ97 to pAZ102). After confirmation, the respective pieces were re-amplified from the PCRB constructs with primers that contain overlapping regions to the next segment of the gene. The pieces were transformed together with the linearized vector backbone derived from YEplac195 containing the ADH2 promoter and terminator with N-terminus FLAG-tag and C-terminus hexahistidine tag (Gietz and Sugino 1988) into *S. cerevisiae* BJ5464-npgA (Jones 1991) using S. c. EasyComp™ Transformation Kit (Invitrogen) and selected on uracil-dropout semisynthetic media. The resulting transformants were screened by colony-PCR and the plasmid in the correct transformant was rescued using the Zymoprep Yeast Plasmid Miniprep Kit (Zymo Research) and transformed into *E. coli* for propagation and sequencing verification. The verified plasmid was designated pAZ103.
The resulting yeast strain containing the pAZ103 plasmid was then cultured to observe for production of new metabolites from AzaA.

![Diagram of cloning strategy](image)

**Figure 2.11.** Cloning strategy to construct the *azaA*-harboring plasmid, pAZ103.

We first inoculated the seed culture using minimal SDCT(A,T) media and grew it for three days before inoculating 1:1000 into a 25-ml sterile culture of YPD. As a control, the untransformed yeast was grown at the same time using the same procedure. To monitor for the production of new metabolites, we extracted a 500-μL sample of each culture with an equal volume of ethyl acetate with 1% acetic acid and dried the extract in the SpeedVac. Following this, the sample was redissolved in methanol and injected in the LC-MS for analysis using the water/acetonitrile solvent system with a linear gradient from 5-95%. On Day 3, we observed the production of a new compound 8, with an *m/z* 233 [M+H]^+ and λ\(_{\text{max}}\) of 340 nm, that was not found in the culture of the control strain (Figure 2.12). We proceeded to purify it in a large-scale culture in order to elucidate its structure. Additionally, we verified the expression of the protein via SDS gel electrophoresis.
We cultured the strain in 2L of YPD and after three days, extracted it with an equal amount of ethyl acetate, followed by chloroform/water and 90% methanol/hexane partitioning. Unfortunately, 8 was unstable and was easily converted to a new compound 9 (m/z 247 [M+H]^+, \( \lambda_{\text{max}} \) 340 nm), whose mass was +14 relative to 8. We purified 7 to homogeneity (5 mg) using HPLC and subjected it to full NMR spectroscopy to elucidate the structure. The mass of 7 was consistent with the molecular formula of C_{13}H_{10}O_{5}. We found all 13 carbon and 10 proton signals from the corresponding 1D spectra. From the \(^1\)H spectrum, 7 was found to contain two aromatic proton (\( \delta_H \) 8.20, 7.43), two methyl groups (\( \delta_H \) 2.64, 1.90) and two hydroxyl protons (\( \delta_H \) 10.90 and 12.00). Meanwhile, the \(^{13}\)C spectrum revealed the presence of eight olefinic carbons (\( \delta_C \) 162.32, 155.60, 134.46, 129.11, 126.41, 123.45, 120.84) and three carbonyls (\( \delta_C \) 201.19, 183.36, 180.00). Based on the UV-Vis spectra and a detailed analysis of the HMBC correlations, we determined 7 to have the naphthaquinone chromophore. Furthermore, the NMR values for this compound match a previously reported metabolite (6-acetyl-2,7-dihydroxy-3-methylnaphthalene-1,4-dione) produced by an NR-PKS from \emph{A. terreus} when it was heterologously expressed in \emph{A. nidulans} (Chiang, Oakley et al. 2013). The oxidation at the C5 position was not expected based on the
biosynthetic logic of polyketides and is most likely a result of aerial oxidation of the actual product 8. While 8 could not be isolated, based on the mass and its relation to 9, we can deduce its structure as shown in Figure 2.13.

(A)

Day 1

Day 2

Day 7

(B)

Figure 2.13. (A) Compound 8 is easily converted to 9. (B) The structure of 9 was elucidated from the NMR while 8 is deduced to be its unoxidized form.

Compound 8 is the product of the spontaneous cyclization of the unreduced hexaketide from AzaA. After the C2-C7 cyclization and reductive release by the R domain to an aldehyde, the acidic C10 methylene flanked by the two ketone groups becomes deprotonated and performs a nucleophilic attack on the C1 carbonyl. Subsequently, following dehydration and rearrangement, the 1,3 naphthoquinone 8 is formed. This spontaneous reaction proceeds quickly, which explains why we were not able to detect the putative uncyclized intermediate, as previously proposed. Instead, the tetrasubstituted benzaldehyde 2 was isolated from the early cultures of the activated
strain. This result suggests that the ketoreduction at the terminal carbonyl is essential to stop the spontaneous formation of 8.

2.1.3.2 Role of AzaE in controlling the spontaneous cyclization

As we have initially proposed in the biosynthesis of azanigerones on Figure 2.10, we investigated the role of AzaE in controlling the spontaneous second-ring cyclization of the unreduced chain. The 1.2-kb gene was predicted to have four intron regions and was similarly amplified from the cDNA synthesized from the producing culture’s mRNA. In order to be used for the co-transformation with the previously constructed pAZ103, we cloned out the gene into the TRP3 marker-vector digested with Smal and NdeI using conventional cloning methods. The sequenced plasmid was designated as pAZ123. Subsequently, pAZ103 and pAZ123 were cotransformed in the yeast-NpgA strain and selected on uracil and tryptophan dropout plates. The presence of both plasmids on the single colonies was confirmed via amplification of a short region of each gene. The correct strains were then inoculated on a seed culture similarly deficient in uracil and tryptophan, followed by 1:1000 inoculation into sterile 25-ml YPD. As controls, the AzaA-expressing strain and the untransformed yeast strain were cultured alongside.

Figure 2.14. HPLC traces of the extract from yeast expressing AzaA and AzaE, AzaA only and negative control. Notice the production of 2 and loss of 6 when the ketoreductase was present.
After three days of culturing, the previously characterized 2 was observed, while 8 and 9 were not found (Figure 2.14). While this confirms the involvement of the ketoreductase AzaE, the timing of events is not very clear. A likely possibility is that the ketoreduction occurs while the growing chain is still tethered to the ACP, before the reductive release. This way, the reactivity of C10 is reduced and would not spontaneously cyclize (Figure 2.15).

![Diagram showing AzaE activity and spontaneou cyclization](image)

**Figure 2.15.** Possible timing of AzaE activity in controlling the spontaneous cyclization into the naphthoquinone 6.

### 2.1.4 Bioactivity assays on azanigerone A

For novel compounds, it is important to test their bioactivity in order to establish whether they have potential as drug leads. Majority of the compounds from the broad family of fungal azaphilones are known to have a variety of bioactivity including antibacterial, antifungal, and anticancer, among many others (Osmanova, Schultze et al. 2010). In this study, we subjected the major compound Azanigerone A for antibacterial and anti-yeast testing.

Five different bacteria were used in the 96-well plate based assay namely *Escherichia coli* DH10B, *Bacillus sp.*, *Pseudomonas sp.*, *Salmonella sp.* and *Staphylococcus sp.* The cultures were grown overnight in LB. Subsequently, the culture was diluted 800-fold to an estimated OD of 0.0005, while a 200-fold dilution was plated to determine the starting CFU. To each well, 50μL of
the diluted culture was added. Meanwhile, azanigerone A was dissolved first in DMSO and added to blank sterile LB to a final concentration of 128μg/ml. This was serially diluted 10 times. To each well, an equal volume of the compound+LB was added. Additionally, a no-drug control and a no-cell control were also set up. Each was done in triplicates. The plates were incubated overnight at 37°C and were visually inspected and checked on the plate reader (set at 600 nm to determine cell density) after 24 hours. All plates had OD\textsubscript{600} similar to the no-drug control, indicating that the compound did not inhibit the growth of any of the bacterial strains tested. The IC\textsubscript{50} of Azanigerone A is >128 μg/ml.

A similar procedure was performed against \textit{Saccharomyces cerevisiae} BJ5464-NpgA and \textit{Candida albicans}, with the exception of the incubation conditions. The plates were incubated at 30°C for 72 hours. Azanigerone A also did not exhibit any inhibition on their growths at the concentration tested (<256 μg/ml).

Bioactivity testing against the aforementioned bacteria and yeast with the other azanigerone analogs have not been performed. Testing on mammalian cell lines have also not been carried out. This could be the subject of further investigations since most azaphilones have been demonstrated to have bioactive functions.

\section*{2.2 Biosynthesis of Brefeldin}

\subsection*{2.2.1 Significance and activity of Brefeldin A}

Brefeldin A was initially discovered as an antiviral agent (Tamura, Ando et al. 1968), but its repertoire of bioactivity has quickly grown to include antifungal (Crabbe and Betina 1994), antitumor and cytotoxic against several cancer cell lines (Shao, Shimizu et al. 1996). While its poor pharmacokinetic properties limit its use as a drug, it is frequently utilized in biological labs
as a broad spectrum protein-transport inhibitor to study protein transport. Its mode of action is characterized by non-competitive binding to the Arf-GDP/Sec7 complex, which prevents the formation of the Golgi apparatus and forces the redistribution of the Golgi proteins into the endoplasmic reticulum (Zeghouf, Guibert et al. 2005). BFA acts on all eukaryotes, from fungi to plants to mammals.

2.2.2 Genome mining for BFA-producing cluster in Eupenicillium brefeldianum ATCC 58665

Brefeldin A is known to be produced by various species of fungi, including Eupenicillium brefeldianum (Scott and Stolk 1967), Penicillium decumbens (Singleton, Bohonos et al. 1958), Penicillium cyaneum (Betina, Barath et al. 1962), Aschocyta imperfectabeti (Suzuki, Tanaka et al. 1970) and Cylindrocarpon obtusisporum (Yu, Zhu et al. 2010). The biosynthetic origin of this compound has been studied extensively in the past using labeled substrates. Due to its similarity to the prostaglandins, it was previously hypothesized to be derived from fatty acids. A report by Bu’Lock and Clay described an intact incorporation of labeled [9-14C] palmitic acid, but has since been retracted due to its non-reproducibility in other labs (Bu'Lock and Clay 1969). In the late 70s and early 80s, Prof. Hutchinson pioneered the studies on BFA biosynthesis using labeled substrates including [18O2, 2-3H]-, [14C, 3H]-, [2-13C, 2H3]-, and [2H3] acetates that established its polyketide origin (Mabuni, Garlaschelli et al. 1977; Hutchinson, Kurobane et al. 1981; Yamamoto, Hori et al. 1985). Beyond these, there are no information identifying the cluster or the genes involved in the synthesis of this important protein-transport inhibitor.

We obtained the E. brefeldianum ATCC 58665 (also called P. brefeldianum Dodge) from the Fungal Genetics stock. This strain has been reported to produce BFA and has been the same
strain used in the previous labeling studies. It has also been the subject of optimization studies to improve production (McCloud, Burns et al. 1995). Our strategy was to sequence the strain and perform a detailed bioinformatic analysis to identify or narrow down the cluster candidates.

The genomic DNA of the BFA-producing strain *E. brefeldianum* ATCC 58665 was sequenced using Roche (454) GS FLX Titanium series and Illumina HiSeq 2000. The resulting GS FLX Titanium reads were first assembled using GS de novo assembler; the output contigs in FASTA format were then combined with the supplementary HiSeq 2000 reads in a hybrid assembly using the Geneious Assembler embedded in the Geneious software suite (Drummond, Ashton et al. 2011). The hybrid assembly generated 708 scaffolds consisting of nearly 36 Mbases of non-redundant reads that roughly reflects the *E. brefeldianum* genome size. A local BLAST database was created using the scaffolds.

<table>
<thead>
<tr>
<th>Number of Contigs</th>
<th>Unused Reads</th>
<th>All contigs</th>
<th>&gt;100</th>
<th>&gt;1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>189</td>
<td>708</td>
<td>708</td>
<td>568</td>
</tr>
<tr>
<td>Median</td>
<td>88</td>
<td>170</td>
<td>170</td>
<td>1,007</td>
</tr>
<tr>
<td>Mean</td>
<td>662</td>
<td>5,464</td>
<td>5,464</td>
<td>10,948</td>
</tr>
<tr>
<td>Max</td>
<td>8,734</td>
<td>1,793,048</td>
<td>1,793,048</td>
<td>1,793,048</td>
</tr>
<tr>
<td>N50</td>
<td>218,001</td>
<td>218,001</td>
<td>218,001</td>
<td>218,001</td>
</tr>
<tr>
<td>Non-redundant reads</td>
<td>125,284</td>
<td>36,118,073</td>
<td>36,118,073</td>
<td>36,093,928</td>
</tr>
</tbody>
</table>

In order to identify the PKS gene clusters, we used the protein sequence for the well-studied PKS, LovB, which is the nonaketide synthase (LNKS) involved in the production of the cholesterol-lowering drug lovastatin (Ma, Li et al. 2009). We reasoned that the target PKS should be significantly similar to LovB since the product brefeldin A is an octaketide molecule. This generated a complete list of contigs that contain a PKS gene that fall into the following categories:
HRPKS, NRPKS, PRPKS or PKS-NRPS. A total of 23 genes were located, with majority belonging to the HRPKS group (Figure 2.16).

Figure 2.16. Breakdown of PKS genes found in E. brefeldianum ATCC 58665.

Subsequently, the genes upstream and downstream of the PKSs were also annotated in order to deduce the possible transformations on the polyketide scaffold. Besides the target cluster, several other interesting clusters were found, including a likely azaphilone cluster (19 and 20 on Table 2.4 below), as well as possible clusters for other known metabolites, including eupenifeldin (Mayerl, Gao et al. 1993; Hsiao, Hsiao et al. 2011), palitantin (Demetriadou, Laue et al. 1988) and fulvic acid (Kurobane, Hutchinson et al. 1981). In addition, several clusters contain some fungal transcription factors that could be utilized in order to selectively activate these orphan clusters.

A more careful analysis of the domains of the HRPKS shows that only five of the 11 do not have the C-methylation domain that is involved in the S-Adenosylmethionine-dependent methylation at the α-position in the polyketide backbone. Due to the lack of methylation on BFA, we looked at these clusters in more detail (cluster 1, 4, 5, 6 and 9 on Table 2.4 below).
Table 2.4. Detailed list of PKS gene clusters found in *E. brefeldianum* ATCC 58665.

<table>
<thead>
<tr>
<th>No.</th>
<th>Contig number</th>
<th>Gene Location</th>
<th>Domain Architecture</th>
<th>PKS Class</th>
<th>Closest characterized homolog, % ID</th>
<th>Neighboring genes</th>
<th>Biosynthetic product (Reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72</td>
<td>31214-41766</td>
<td>KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>ChPKS7 (CladeI HR, LDKS) 37%</td>
<td>dehydrogenase, MFS, oxygenase, Cinamoyl alcohol dehydrogenase, DUF1212 membrane protein, MFS</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>72</td>
<td>239484-246451</td>
<td>KS-AT-DH-Met-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>polyketide synthase [Metarhizium anisopliae ARSEF 23] 42%</td>
<td>Prohibitin, N-methyltransferase, DUF221 transmembrane, Reductase, Myb-like DNA-binding domain</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>130</td>
<td>84533-91430</td>
<td>KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>BfPKS8 (HR Clade 3, T-toxin nonaketide) 45%,</td>
<td>sulfotransferase, MFS, oxidoreductase, thioesterase, kinase</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>286</td>
<td>16191-23507</td>
<td>KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>BfPKS10 (fumonisin) 51%</td>
<td>4 P450s, alpha-beta hydrolase, WD40, kinesin</td>
<td>brefeldin (2.2.3)</td>
</tr>
<tr>
<td>6</td>
<td>189</td>
<td>11767-20477</td>
<td>KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>Gz PKS4 30%</td>
<td>P450, ER, 2 activators, oxidase, dehydratase, dehydrogenase</td>
<td>palitantin?</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>89039-99199</td>
<td>KS-AT-DH-Met-ER-KR-ACP-CarnAT</td>
<td>HR-PKS</td>
<td>BfPKS2 (CladeI HR, LDKS) 43%</td>
<td>proteinase inhibitor, tRNA hydrolase, aminotransferase, epimerase, MFS, regulatory, ICL-PEPM, C2H2 Zn finger, hydrolase</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>261</td>
<td>1745-9680</td>
<td>KS-AT-DH-Met-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>PDE 03455 [Penicillium oxalicum 114-2], 66%</td>
<td>esterase/lipase, MFS, 1,3, betaglucanosyltransferase</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>63</td>
<td>124574-131879</td>
<td>KS-AT-DH-ER-KR-ACP</td>
<td>HR-PKS</td>
<td>Hpm8, 32% LNKS, 73%</td>
<td>membrane protein, PKS-NRPS, 2 P450's, dioxygenase, Zn2Cys6, dehydrogenase, esterase lipase, reductase</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>192</td>
<td>15917-22559</td>
<td>KS-AT-ACP-Met-TE</td>
<td>NR-PKS</td>
<td>Bf PKS17 (NR Clade 3) 35%, 52% citrinin polyketide synthase [Monascus purpureus] 31%, 49%</td>
<td>FAD-dependent oxygenase, geranyl-geranyl pyrophosphate synthetase, P450, prenyl transferase, membrane protein, MFS, dehydrogenase/oxy</td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Start</td>
<td>End</td>
<td>Gene</td>
<td>Type</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>28</td>
<td>427292-434840</td>
<td>KS-AT-ACP-TE</td>
<td>NR-PKS</td>
<td>Naphthopyrone WA1 PKS 43, 60%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>125</td>
<td>61192 - 68031</td>
<td>KS-AT-ACP-TE</td>
<td>NR-PKS</td>
<td>conidial pigment polyketide synthase PksP/Alb1 [Arthroderma benhamiae CBS 112371] 54, 70%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>311</td>
<td>23074-29285</td>
<td>KS-AT-ACP-Met-R</td>
<td>NR-PKS</td>
<td>methylorcinaldehyde synthase [Acremonium strictum], 53%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>52</td>
<td>123846 - 130847</td>
<td>KS-AT-ACP-TE-ER-KR-ACP</td>
<td>NR-PKS</td>
<td>PKS3 [Fusarium pseudograminearum CS3096] 64/77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- doreductase, ABC exporter,
- membrane protein, laccase, Cu-oxidase, ACP, Ado-Met, C6 transcription factor
- leucine-rich, Zn2cys6 activator
- monooxygenase, S-transferase, activator
- Cu-oxidase (ascorbase), cytokinesis protein, pyrophosphatase
- helicase, MFS, p450, activator, acetyl hydrolase
- activator, RTA, MFS, maltase, hydrolase
<table>
<thead>
<tr>
<th>No.</th>
<th>Gene</th>
<th>Accession</th>
<th>Description</th>
<th>Protein</th>
<th>Function</th>
</tr>
</thead>
</table>
2.2.3 Fungal Polyketide Synthase Product Chain-Length Control by Partnering Thiohydrolase

This section was originally published with the same title in ACS Chem. Biol.: Zabala, A.O., Chooi, Y.H., Choi, S.M., Lin, H-C., Tang, Y. 2014, ACS Chem Biol, in press.

2.2.3.1 Abstract

Fungal highly-reducing polyketide synthases (HRPKSs) are an enigmatic group of multidomain enzymes that catalyze the biosynthesis of structurally diverse compounds. This variety stems from their intrinsic programming rules, which permutate the use of tailoring domains and determine the overall number of iterative cycles. From genome sequencing and mining of the producing strain Eupenicillium brefeldianum ATCC 58665, we identified an HRPKS involved in the biosynthesis of an important protein transport-inhibitor Brefeldin A (BFA), followed by reconstitution of its activity in Saccharomyces cerevisiae and in vitro. Bref-PKS demonstrated an NADPH-dependent reductive tailoring specificity that led to the synthesis of four different octaketide products with varying degrees of reduction. Furthermore, contrary to what is expected from the structure of BFA, Bref-PKS is found to be a nonaketide synthase in the absence of an associated thiohydrolase Bref-TH. Such chain-length control by the partner thiohydrolase was found to be present in other HRPKS systems, and highlights the importance of including tailoring enzyme activities in predicting fungal HRPKS functions and their products.

2.2.3.2 Introduction

Fungal polyketides constitute an important group of natural products that includes statins, antibiotics and anti-cancer agents (Marinelli 2009; Chooi and Tang 2012). The biosynthesis of polyketide in fungi is performed by Type I Iterative polyketide synthases (IPKSs), which are
multidomain megasynthases (Cox 2007). The highly reducing IPKS (HRPKS) is a large subgroup of the IPKSs that are associated with the biosynthesis of highly reduced compounds, such as lovastatin and fumonisin. The minimal PKS domains of HRPKSs, which consist of ketosynthase (KS), malonyl-CoA:acyl carrier protein transacylase (MAT) and acyl carrier protein (ACP), catalyze the selection of malonyl building blocks and the repeated Claisen-like chain extension steps. HRPKS also consists of a set of tailoring domains, including ketoreductase (KR), dehydratase (DH), enoylreductase (ER); and the Cα-methyltransferase (MT). Although the domain organization of HRPKSs bears strong resemblance to that of mammalian fatty acid synthases (FASs), the HRPKSs operate in a much more sophisticated fashion. (Smith and Tsai 2007) Most notably, the HRPKSs use the single set of tailoring domains in different permutations during every extension cycle, which results in the high degree of variability at individual α- and β- positions in the products. Product chain-length control also varies between different HRPKSs, which results in polyketides that have a wide range of sizes. Additionally, HRPKS differs from FAS in that there is no dedicated and fused thioesterase (TE) domain at the C-terminus of the megasynthase, and instead rely on the in trans interaction with discrete TE or acyltransferase-like enzymes for product release (Chooi and Tang 2012). Our current understanding of these unique features of HRPKSs has remained at an early stage, thereby limiting our ability to link reduced polyketides to corresponding HRPKSs, and to predict product structures from the vast number of HRPKSs identified from fungal genome sequencing efforts.

In order to better understand these enigmatic features of HRPKSs, it is important to work with a suitable model system for biochemical analysis. First, the model HRPKS should be a standalone enzyme that generates a product of substantial chain length that could be detected and analyzed (i.e. UV active). This circumvents the dependence on downstream enzymes (e.g. other
PKS) for further modifications that convolutes product analysis. Second, the HRPKS should be programmed to synthesize a product with variable degrees of β-reduction within each extension cycle to allow investigation of the permutative tailoring rules. Lastly, it is ideal to work with an HRPKS that is involved in the biosynthesis of a bioactive polyketide product to aid the re-engineering of the HRPKS for analogue generation.

Several HRPKSs involved in the biosynthesis of pharmaceutically important polyketides have been investigated, such as LovB and Hpm8 that are responsible for the biosynthesis of lovastatin and hypothemycin, respectively. In the LovB system, however, a yet unresolved Diels-Alder cyclization step embedded among the chain extension steps has complicated analysis of the HR-PKS alone (Auclair, Sutherland et al. 2000; Ma, Li et al. 2009). On the other hand, HRPKSs from resorcylic acid lactone (RAL) pathways (such as Hpm8) and from HRPKS-nonribosomal peptide synthetase (NRPS) hybrids require downstream enzymes/domains for product transfer and further modification, which represents an added level of complexity in deconvoluting HRPKS functions and products (Xu, Cai et al. 2010; Zhou, Qiao et al. 2010; Zhou, Qiao et al. 2010).

![Figure 2.17. Putative biosynthethic pathway for BFA. The HRPKS is proposed to synthesize the precisely reduced octaketide precursor, which could then be directly offloaded by the thiohydrolase enzyme followed by a P450-mediated formation of the cyclopentane ring and macrocyclization to afford the 7-deoxy BFA 11 (top scheme). Alternatively, the first ring annulation can also occur on the ACP-tethered intermediate before the thiohydrolase release and lactonization (bottom scheme). The C7-hydroxylation is believed to be the final step in the process to obtain the final structure of 10, BFA.](image-url)
In this study, we chose the HRPKS responsible for brefeldin A (BFA) as a model system (Figure 2.17). BFA is a protein-transport inhibitor isolated from several species of filamentous fungi (Singleton, Bohonos et al. 1958; Betina, Barath et al. 1962; Scott and Stolk 1967; Suzuki, Tanaka et al. 1970; Yu, Zhu et al. 2010). It is used to study protein transport among eukaryotes, but has also been found to have antiviral, antifungal and antitumor properties (Betina 1992). The polyketide origin of this 16-membered macrolactone was previously established through feeding studies with labeled acetate (Coombe, Foss et al. 1969; Mabuni, Garlaschelli et al. 1977; Gonzalez de la Parra and Hutchinson 1987). The proposed biosynthesis of BFA involves formation of an acyclic polyketide chain that is differentially tailored throughout the backbone (Figure 2.17). The presence of the terminal hydroxyl group, along with the strategically positioned double bonds, is proposed to enable cyclization of the acyclic precursor into the fused, bicyclic structure seen in BFA. We hypothesized that a single HRPKS should be sufficient to generate the entire carbon backbone without the need for an additional PKS. If the acyclic product indeed contains conjugated double bonds as proposed, the biosynthetic product should be readily identifiable and isolated. Collectively, the BFA HRPKS appears to fit the mold as a good model HR-PKS for biochemical analysis. Here, we first identified the HRPKS most likely responsible for BFA biosynthesis from *Eupenicillium brefeldianum* ATCC 58665. We further demonstrate that using a product based assay, important programming rules of HRPKSs were elucidated, including NADPH concentration-dependent extent of reduction, and the control of HRPKS product chain length by an associated discrete thiohydrolase (TH). These insights further improve our understanding of HRPKSs functions.
2.2.3.3 Results

2.2.3.3.1 Identification of the brefeldin A PKS by bioinformatic and transcription analysis

The genomic DNA of the BFA-producing strain *E. brefeldianum* ATCC 58665 was sequenced using Roche (454) GS FLX Titanium series and Illumina HiSeq 2000. The resulting GS FLX Titanium reads were first assembled using GS de novo assembler; the output contigs in FASTA format were then combined with the supplementary HiSeq 2000 reads in a hybrid assembly using the Geneious Assembler embedded in the Geneious software suite (Drummond,
Ashton et al. 2011. The hybrid assembly generated 708 scaffolds consisting of nearly 36

\[
\text{Figure 2.18} \quad \text{Production of BFA correlated with transcription of Contig 286 HRPKS. (A) LC-MS profile of } E. \text{ brefeldianum} \ \text{ATCC58665 culture grown in MEM media after three days. This media was used to optimize production of BFA and reduce production of other metabolites. The cells for mRNA extraction were collected from the same culture. (B) Gene organization of the HR-PKS-containing contigs (Domain organization: KS-AT-DH-ER-KR-ACP) in } E. \text{ brefeldianum} \ \text{and transcriptional analyses on the PKS genes from each contig.}
\]
Mbases of non-redundant reads that roughly reflects the *E. brefeldianum* genome size. A local BLAST database was created using the scaffolds. Using the KS domain of the nonaketide synthase LovB as a query sequence, 24 putative PKSs were identified: 11 HRPKSs; eight non-reducing PKSs (NRPKSs); two partially-reducing PKSs (PRPKS); and two HRPKS-NRPS hybrids (Figure 2.16). The lack of Cα-methylation in BFA excluded the MT-containing HRPKSs, narrowing down the search to five HRPKS-containing gene clusters. Subsequently, RT-PCR was performed on the mRNA of the BFA-producing culture to determine the transcription of the HRPKS genes, of which only Contig_286 PKS was highly transcribed at the time point that coincided with BFA production (Figure 2.18), indicating the high likelihood of this HRPKS being involved in BFA biosynthesis.

The HRPKS of Contig_286 (*orf7*) contains the following domains linearly juxtaposed from N- to C- terminus: KS, MAT, DH, ER, KR and ACP. Using the maximum likelihood statistical method on MEGA Version 5, a phylogenetic tree with HRPKSs of known natural products was constructed (Tamura, Peterson et al. 2011). Contig_286 HRPKS claded with those that biosynthesize relatively longer (> C10), acyclic polyketides such as the polyene portion of fumagillin and fumonisin (Figure 2.19). The closest homolog is an uncharacterized HRPKS from *Trichoderma virens* Gv29-8, TRVIDRAFT_151590 (59% identity, 73% similarity). Within the 37,663 bp of Contig_286, nine other open-reading frames (ORFs) were identified using Softberry prediction (Table 2.5, Figure 2.20). These include an α–β hydrolase (*orf2*), which interestingly has the closest homolog, TRVIDRAFT_53350 (60, 77%), encoded immediately upstream of the TRVIDRAFT_151590 PKS gene in *T. virens* Gv29-8. Hence, this may be the partnering enzyme involved in the release of the ACP-tethered polyketide thioester product via either hydrolysis to yield an acyclic product (as a thiohydrolase) or macrocyclization to yield a cyclized product (Figure 2.17). The neighboring genes are consistent with that of a possible BFA biosynthetic gene
cluster, including four genes encoding P450 monooxygenases (orfs 3-6). Previous feeding studies have shown that the C4 and C7 hydroxylation in BFA resulted from oxidative tailoring, while the cyclopentane ring formation was similarly proposed to be P450-mediated (Mabuni, Garlaschelli et al. 1977; Yamamoto, Hori et al. 1985). Hence the collection of P450s here may be responsible for these transformations on the polyketide product. To analyze if these genes are co-transcribed with the HRPKS gene in this cluster, a transcriptional analysis was performed using gene-specific primers (Figure 2.20). All P450s and TH encoding genes were highly transcribed in the producing culture, along with orf8 of unknown function. These genes were putatively assigned to comprise the bref cluster, with the HRPKS and TH renamed to Bref-PKS and Bref-TH, respectively.

**Figure 2.19.** Phylogenetic tree containing BrefPKS and HRPKSs of known products. The alignment was performed on Mega 5.05.
Table 2.5. Genes annotated in Contig_286.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Proposed Function (Domain Organization)</th>
<th>Sequence Homologs (Identities/Positives)</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf1</td>
<td>F-box domain protein</td>
<td><em>Aspergillus oryzae</em> RIB40 BAE55018 (26/42)</td>
</tr>
<tr>
<td>orf2 / bref-TH</td>
<td>Alpha-beta hydrolase</td>
<td><em>Trichoderma virens</em> Gv29-8 TRIVIDRAFT_53350 (60/77)</td>
</tr>
<tr>
<td>orf3</td>
<td>Cytochrome p450</td>
<td><em>Bipolaris maydis</em> C5 COCHEDRAFT_1151230 (42/60)</td>
</tr>
<tr>
<td>orf4</td>
<td>Cytochrome p450</td>
<td><em>Macrophomina phaseolina</em> MS6 MPH_06433 (46/61)</td>
</tr>
<tr>
<td>orf5</td>
<td>Cytochrome p450</td>
<td><em>Aspergillus oryzae</em> CYP628C1 (52/70)</td>
</tr>
<tr>
<td>orf6</td>
<td>Cytochrome p450</td>
<td><em>Macrophomina phaseolina</em> MPH_06428 (46/64)</td>
</tr>
<tr>
<td>orf8</td>
<td>large tegument protein UL36</td>
<td><em>Penicillium chrysogenum</em> Wisconsin 54-1255 Pc12g10520 (51/65)</td>
</tr>
<tr>
<td>orf9</td>
<td>WD-40</td>
<td><em>Aspergillus terreus</em> NIH2624 ATEG 09923 (52/67)</td>
</tr>
<tr>
<td>orf10</td>
<td>Kinesin</td>
<td><em>Bipolaris victoriae</em> FI3 COCVIDRAFT_89854 (71/80)</td>
</tr>
</tbody>
</table>

Figure 2.20. Transcriptional analysis of genes in Contig_286 determines the putative boundary of the bref cluster. (A) Arrangement of genes in Contig_286. (B) RT-PCR analysis on the annotated genes within the contig. The template mRNA was extracted from a Day2 BFA-producing culture of *E. brefeldianum* in the optimized production media, MEM.
2.2.3.3.2 Cloning and expression of Bref-PKS and Bref-TH

The producing strain *E. brefeldianum* ATCC 58665 is a non-sporulating filamentous fungi and attempts at genetic manipulation were futile due to the multinucleated protoplasts. To examine the activity and product of the Bref-PKS, the 7.1-kb intron-less gene was assembled from five cDNA fragments and placed under control of the ADH2 promoter by yeast recombination in *S. cerevisiae* BJ5464-NpgA () (Ma, Li et al. 2009). Subsequently, BJ5464-NpgA harboring the Bref-PKS expression plasmid was cultured and grown to stationary phase for protein expression. The hexahistidine-tagged Bref-PKS was solubly expressed and was purified via nickel affinity chromatography at a yield of 2 mg/L (Figure 4.36). The intron-less gene encoding Bref-TH was similarly constructed from cDNA and was subsequently expressed in *E. coli* BL21(DE3) via an IPTG-inducible-T7 promoter and purified using nickel affinity chromatography (Figure 4.36).

We first assayed the activities of the minimal PKS domains (KS, MAT and ACP) by incubating Bref-PKS with 2 mM malonyl-CoA in PBS buffer, pH 7.4. In the absence of the reducing cofactor NADPH, the reductive domains are expected to be inactive and should yield an unreduced polyketide product. After 16 hours, the reaction was either extracted directly or first treated with 1 M NaOH followed by extraction for product analysis. In the absence of base hydrolysis, no product was recovered. With NaOH treatment, we observed the production of 4-hydroxy-6-methyl-2H-pyran-2-one (TKL) (Figure 2.21), which forms through the spontaneous cyclization and release of the unreduced triketide (Winter, Sato et al. 2012). This result indicates that the minimal PKS components were active and product release from the PKSs requires additional factors.
2.2.3.3.3 Reconstitution of Bref-PKS activities

Most HRPKSs do not have a dedicated domain for product offloading. They often rely on an \textit{in trans} acyltransferase or TH to release the polyketide from the ACP-tether (Chooi and Tang 2012). Such partnering enzyme plays an important role in terminating the chain elongation and consequently in determining the length of the final product. To investigate the influence of the \textit{in trans} Bref-TH on HRPKS function, the purified enzyme was added to the Bref-PKS in vitro reaction at an equimolar ratio (20 μM) with 2 mM malonyl-CoA and 10 mM NADPH. The reaction was left at room temperature for 16 hr and was extracted with ethyl acetate and analyzed by LC-MS. Four relatively hydrophobic compounds 3-6 emerged at the end of gradient (5%-95% acetonitrile in H$_2$O, 30 min), with 3 being the dominant product (Figure 2.22). The corresponding masses for compounds 12, 13, 14 and 15 are 264, 268, 264 and 266, respectively (Figure 4.37). Compounds 12 and 14 displayed $\lambda_{\text{max}}$ of 260 nm, indicating the presence of a slightly conjugated...
structure, while compounds 13 and 15 did not have significant absorption above 220 nm. The masses of these compounds are consistent with that of an octaketide that has undergone several reductive modifications (for reference, the molecular weight of palmitic acid is 256).

![Figure 2.22](image)

**Figure 2.22.** In vitro products of Bref-PKS and Bref-TH are acyclic octaketides with variable degrees of β-reduction. (A) HPLC and EIC trace of the in vitro reaction between Bref-PKS and Bref-TH. (B) Production of compounds 3 to 6 from the *S. cerevisiae*-NpgA strain co-expressing Bref-PKS and Bref-TH. Notice the change in the production profile between days 1 and 3. The compounds were purified according to their peak production period. (C) Elucidated structures of compounds 3 to 6 from the corresponding NMR spectra.

In order to isolate sufficient amounts of 12-15 for structure elucidation, Bref-PKS and Bref-TH were co-expressed under the ADH2 promoter in the yeast host. Compounds 12 and 14 were harvested after 36 hours of inoculation (1 mg/L and 0.5 mg/L final yield, respectively), while 13 and 15 were extracted after 72 hours (2 mg/L and 4 mg/L final yield, respectively), depending on
their highest production period (Figure 2.22). Both 12 and 14 disappeared at the 72 hour time point, suggesting the possible conversion to 13 and 15, respectively. Each compound was purified to homogeneity and subjected to full NMR spectroscopy to elucidate their structures.

From the NMR spectra (Figure 4.42, Figure 4.46, Table 2.6), 12 is determined to be a linear 16-carbon carboxylic acid ($\delta_{C1}$ 171.30) (Figure 2.22). The $^1$H NMR spectrum revealed three olefinic groups, one oxygenated methine, seven methylene groups and one methyl group. Based on COSY correlations, two of the double bonds (C2-C3 and C4-C5) form a diene that by HMBC correlations, was further conjugated to carbonyl C1. This conjugated dienoic acid accounts for the higher $\lambda_{\text{max}}$ 260 nm of 3 compared to BFA ($\lambda_{\text{max}}$ = 220 nm). Additionally, the large coupling constants of the connected protons ($J_{HH}$ 15.3) indicate trans configurations of the double bonds. The remaining olefinic carbons are found to be at the C10-C11 position from detailed analysis of the COSY and HMBC correlations. The position of the C-OH group at C15 was confirmed by the direct correlation of the doublet C16 methyl protons to the oxygenated methine group. To determine whether this compound is a macrolactone or an acyclic molecule, we compared the carbon chemical shift of C15 to the corresponding carbons in BFA and in an uncyclized BFA analog. The C15 chemical shift of 3 ($\delta_{C15}$ 68.5) accorded with the uncyclized form ($\delta_{C}$ 69.3) (Yu, Zhu et al. 2010) as opposed to the slightly downfield shift in the lactone ($\delta_{C}$ 71.8) (Glaser, Shifman et al. 2000).
Table 2.6. 1H NMR (500 MHz, CD$_3$OH) and 13C NMR (125 MHz, CD$_3$OH) data for 12, 13, 14 and 15

<table>
<thead>
<tr>
<th>No</th>
<th>12 $^{13}$C NMR δ [ppm]</th>
<th>$^{1}$H NMR δ [ppm] (area, m, $J_{HH}$ [Hz])</th>
<th>13 $^{13}$C NMR δ [ppm]</th>
<th>$^{1}$H NMR δ [ppm] (area, m, $J_{HH}$ [Hz])</th>
<th>14 $^{13}$C NMR δ [ppm]</th>
<th>$^{1}$H NMR δ [ppm] (area, m, $J_{HH}$ [Hz])</th>
<th>15 $^{13}$C NMR δ [ppm]</th>
<th>$^{1}$H NMR δ [ppm] (area, m, $J_{HH}$ [Hz])</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>171.3</td>
<td>-</td>
<td>177.8</td>
<td>-</td>
<td>171.2</td>
<td>-</td>
<td>177.1</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>121.1</td>
<td>5.78 (CH, d, 15.3)</td>
<td>35.6</td>
<td>2.31 (CH$_2$, d, 6.0)</td>
<td>122.4</td>
<td>5.82 (CH, d, 15.3)</td>
<td>35.2</td>
<td>2.32 (CH$_2$, d, 6.1)</td>
</tr>
<tr>
<td>3</td>
<td>146.5</td>
<td>7.22 (CH, dd, 15.3, 10.6)</td>
<td>29.2</td>
<td>2.27 (CH$_2$, m)</td>
<td>144.4</td>
<td>7.13 (CH, dd, 14.9, 10.9)</td>
<td>29.0</td>
<td>2.27 (CH$_2$, m)</td>
</tr>
<tr>
<td>4</td>
<td>129.8</td>
<td>6.23 (CH, dd, 15.1, 10.6)</td>
<td>129.7</td>
<td>5.44 (CH, dt, 3.8, 1.9)</td>
<td>129.3</td>
<td>6.22 (CH, dd, 15.1, 10.6)</td>
<td>129.6</td>
<td>5.44 (CH, dt, 3.8, 1.9)</td>
</tr>
<tr>
<td>5</td>
<td>145.4</td>
<td>6.16 (CH, m)</td>
<td>132.5</td>
<td>5.45 (CH, m)</td>
<td>143.6</td>
<td>6.10 (CH, m)</td>
<td>132.6</td>
<td>5.38 (CH, m)</td>
</tr>
<tr>
<td>6</td>
<td>33.8</td>
<td>2.19 (CH$_2$, m)</td>
<td>33.6</td>
<td>1.99 (CH$_2$, m)</td>
<td>33.4</td>
<td>2.19 (CH$_2$, m)</td>
<td>33.4</td>
<td>1.99 (CH$_2$, m)</td>
</tr>
<tr>
<td>7</td>
<td>29.3</td>
<td>1.45 (CH$_2$, m)</td>
<td>30.2</td>
<td>1.35 (CH$_2$, m)</td>
<td>29.9</td>
<td>1.41 (CH$_2$, m)</td>
<td>30.1</td>
<td>1.36 (CH$_2$, m)</td>
</tr>
<tr>
<td>8</td>
<td>30.2</td>
<td>1.38 (CH$_2$, m)</td>
<td>30.1</td>
<td>1.35 (CH$_2$, m)</td>
<td>30.2</td>
<td>1.41 (CH$_2$, m)</td>
<td>30.0</td>
<td>1.36 (CH$_2$, m)</td>
</tr>
<tr>
<td>9</td>
<td>33.6</td>
<td>2.01 (CH$_2$, m)</td>
<td>33.4</td>
<td>1.99 (CH$_2$, m)</td>
<td>33.00</td>
<td>1.99 (CH$_2$, m)</td>
<td>33.4</td>
<td>1.99 (CH$_2$, m)</td>
</tr>
<tr>
<td>10</td>
<td>131.4</td>
<td>5.41 (CH, m)</td>
<td>131.6</td>
<td>5.40 (CH, m)</td>
<td>130.6</td>
<td>5.42 (CH, m)</td>
<td>132.3</td>
<td>5.40 (CH, m)</td>
</tr>
<tr>
<td>11</td>
<td>131.6</td>
<td>5.41 (CH, m)</td>
<td>131.4</td>
<td>5.40 (CH, m)</td>
<td>128.6</td>
<td>5.46 (CH, m)</td>
<td>130.7</td>
<td>5.40 (CH, m)</td>
</tr>
<tr>
<td>12</td>
<td>33.4</td>
<td>2.01 (CH$_2$, m)</td>
<td>33.5</td>
<td>1.99 (CH$_2$, m)</td>
<td>33.8</td>
<td>2.14 (CH$_2$, m)</td>
<td>32.9</td>
<td>1.99 (CH$_2$, m)</td>
</tr>
<tr>
<td>13</td>
<td>26.9</td>
<td>1.42 (CH$_2$, m)</td>
<td>26.9</td>
<td>1.46 (CH$_2$, m)</td>
<td>24.7</td>
<td>1.60 (CH$_2$, m)</td>
<td>24.2</td>
<td>1.60 (CH$_2$, m)</td>
</tr>
<tr>
<td>14</td>
<td>39.7</td>
<td>1.44, 1.39 (CH$_2$, m)</td>
<td>39.7</td>
<td>1.50, 1.42 (CH$_2$, m)</td>
<td>43.5</td>
<td>2.48 (CH$_2$, t, 7.3)</td>
<td>43.5</td>
<td>2.46 (CH$_2$, t, 7.3)</td>
</tr>
<tr>
<td>15</td>
<td>68.5</td>
<td>3.71 (CH, dd, 11.6, 6.0)</td>
<td>68.5</td>
<td>3.71 (CH, dd, 12.0, 6.2)</td>
<td>212.3</td>
<td>-</td>
<td>212.0</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>23.5</td>
<td>1.14 (CH$_3$, d, 6.2)</td>
<td>23.5</td>
<td>1.14 (CH$_3$, d, 6.2)</td>
<td>29.5</td>
<td>2.11 (CH$_3$, s)</td>
<td>29.1</td>
<td>2.09 (CH$_3$, s)</td>
</tr>
</tbody>
</table>
The structure of 13 ($m/z = 267 \text{ [M+H]}^+$) was similarly elucidated from the 1D- and 2D-NMR spectra and from comparison to 12. Instead of the olefinic protons ($\delta_H 5.78, 7.22$) at the C2-C3 position, the C2-C3 carbons in 13 were found to have methylene protons ($\delta_H 2.27, 2.31$) that were connected to the C4-C5 double bond in COSY (Table 2.6, Figure 4.47, Figure 4.50). This is in accordance with the loss of conjugation in the molecule ($\lambda_{\text{max}}$ 220 nm) relative to 12. The rest of the NMR signals of 13 are consistent with 12. Interestingly, 14 ($m/z = 265 \text{ [M+H]}^+$, $\lambda_{\text{max}}$ 260 nm) and 15 ($m/z = 267 \text{ [M+H]}^+$, $\lambda_{\text{max}}$ 220 nm) were found to be structurally related to 12 and 13, respectively, differing only by the absence of the C15 hydroxyl group ($\delta_{C15} 68.5$, $\delta_{H15} 3.71$). Instead, an aliphatic carbonyl signal ($\delta_{C15} 212$) was observed in their respective $^{13}$C NMR spectra, which was validated by HMBC correlations and by the appearance of C16 methyl group as a singlet (Figure 4.51-Figure 4.58, Table 2.6).

It is noteworthy that several important structural features of 12 are consistent with that of BFA including the C16-backbone. Additionally, the C2-C3 and C10-C11 olefinic groups in 12 are retained in BFA. The terminal hydroxyl group that is important for macrocyclization to form BFA is also present. On the other hand, while not observed in BFA, the C4-C5 double bond in 12 appears to be well-positioned for formation of the cyclopentane ring in BFA, presumably mediated by a P450 oxygenase encoded in the bref cluster (Figure 2.17). The structural parallels, both in size and sites of unsaturation/hydroxylation, between 3 and BFA therefore strongly indicate Bref-PKS is indeed the HRPKS involved in BFA biosynthesis.

The discovery of compounds 12-15 containing varying degrees of $\beta$-reduction by Bref-PKS is unexpected. In generation of the ketones 14 and 15, it appears the KR domain is prone to skipping ketoreduction of the diketide intermediate which results in the C15 hydroxyl group. To
further investigate the effect of NADPH concentration on the reductive programming rules of Bref-PKS, we performed in vitro assays by varying the NADPH concentrations from 0.5 mM to 10 mM. While 12 remained as the dominant product at NADPH concentrations higher than 2 mM, compound 14 was found to be the dominant product at lower, but still physiologically relevant concentration (< 2 mM) (Figure 2.24). Hence, the KR domain is highly sensitive to availability of reducing cofactors, albeit only at the first ketoreduction step during Bref-PKS function. Interestingly, the C15 ketone observed in 14 is also found in the 7-dehydrobrefeldin A acid analog that contains the cyclopentane ring but is not macrocyclized (Yu, Zhu et al. 2010). Therefore, the KR domain may indeed be imprecisely programmed to act at the C15 ketone in the native producer as well. Such NADPH-dependent tailoring can add another degree of complexity to HRPKSs in generating diversity among fungal polyketides.

The other octaketide products of Bref-PKS, 13 and 15, lack the α-β double bond (C2-C3) observed in BFA and in 12 and 14. Both were minimally produced in vitro but were the end products in the yeast in vivo culture after more extended fermentation (Figure 2.22). The accumulation of 13 and 15 in yeast coincided with the disappearance of 12 and 14, which strongly suggests the enoylreduction of 12 and 14 by endogenous yeast enzymes. This was further confirmed by the bioconversion experiments with purified 12 and 14, in which both compounds were fully converted into 13 and 15, respectively, by untransformed S. cerevisiae BJ5464-NpgA within 24 hours of addition to the culture (Figure 2.23). However, the production of 13 and 15 in the in vitro assay also suggests that the ER domains of Bref-PKS can partially reduce the enoyl during the last iteration. No aberrant enoylreduction of the C4-C5 and C10-C11 positions were observed among the products, therefore pointing to a “specific” overreduction of the ER domain at the last iteration.
Figure 2.23. Yeast biotransformation of compounds 12 and 14 to 13 and 15, respectively. In order to confirm that 12 and 14 were indeed converted to 13 and 15 by yeast endogenous enzymes, we grew two 2-ml culture of *S. cerevisiae* BJ5464-NpgA (empty host) on YPD for 24 hours. Purified 12 and 14 were then added to each culture and incubated O/N with shaking at 28°C. After 24 hours, 500μL of each culture was extracted with EA+acid and dried subsequently. The extract was analyzed on LC/MS, showing that 12 and 14 were indeed nearly fully converted to 13 and 15, respectively. This explains why only 13 and 16 were left in the yeast culture harboring *bref-PKS* and *bref-TH* after more than three days of culturing.
2.2.3.3.4 **Bref-TH controls the programmed release of the octaketide product from Bref-PKS**

Having reconstituted the activities of the Bref-PKS, we then investigated the role of Bref-TH in chain termination and product release. The Bref-PKS assay was repeated without Bref-TH for 16 hrs, followed either by direct extraction with ethyl acetate or pre-treatment with 1M NaOH.
at 65°C (base hydrolysis) before extraction. The amount of product turnover decreased significantly (10-fold) in the absence of Bref-TH, consistent with the proposed role of the hydrolase in facilitating chain release. When subjected to base hydrolysis or upon subsequent addition of equimolar amount of Bref-TH after 16 hours, a higher level of product release was observed (Figure 4.39). The octaketides 3-6 were recovered as previously, with the level of 4 now similar to that of 3. The increased amount of ER over-reduction in the absence of Bref-TH may be due to stalling of the polyketide products on the Bref-PKS in the absence of TH-mediated release. More importantly, two new products 16 (m/z 295 [M+H]+) and 17 (m/z 297 [M+H]+) were isolated from the above Bref-PKS assay, with 16 now being the dominant product of all polyketide products (Figure 2.25). Both compounds displayed \( \lambda_{\text{max}} \) of 215 nm, which corresponds to absence of the conjugated dienoic acid moiety. The +28 mu increase in masses of 16 and 17 compared to that of 12 and 14 suggests the incorporation of a completely reduce ketide unit (-CH\(_2\)-CH\(_2\)-) as a result of an additional round of chain elongation and reduction by the Bref PKS (Figure 2.25). Due to the single-turnover nature of the assay in the absence of the Bref-TH, compounds 16 and 17 cannot be sufficiently obtained for structural elucidation. Expression of Bref-PKS alone in yeast did not lead to detectable amounts of 16 and 17 either. Therefore to confirm compound 16 is indeed a nonaketide instead of octaketide, we performed the in vitro assay using 2-\(^{13}\)C-malonate and the MatB system, which generates the 2-\(^{13}\)C-malonyl-CoA in situ (An and Kim 1998). As expected an increase of 9 mu in molecular weights was observed for both labeled 16 and 17, confirming the incorporation of nine ketide units derived from malonate into the backbone (Figure 4.38). Combining the UV and mass data (Figure 4.37, Figure 4.38), we propose the structure of 16 and 17 as shown in Figure 2.25, derived from an additional round of chain elongation from 12 and 14, respectively, followed by full \( \beta \)-reduction. Selected ion-monitoring of the mass of 16 and
17 in the Bref-PKS assay that contained Bref-TH yielded no trace of these two compounds. These results demonstrate that in the absence of the Bref-TH, the Bref-PKS functions primarily as a nonaketide synthase.

**Figure 2.25.** In vitro reactions with Bref-PKS demonstrate the TH-controlled PKS chain length release. (A) EIC spectra of the in vitro reactions showed the variation in the product profiles of: Bref-PKS with Bref-TH; Bref-PKS with base hydrolysis; Bref-PKS with Bref-TH H276A; and Bref-PKS with other *in trans* releasing enzymes CazE and Fma-AT. The reactions consist of 20μM Bref-PKS, 2mM mCoA and 10mM NADPH with either 20μM of the releasing enzyme or base hydrolysis (1M NaOH at 65°C for 10 min). (B) Proposed structures of 16 and 17.
A protein family database search of Bref-TH indicated that the enzyme belongs to the Abhydrolase_6 family (E-value of 3.4e-14). This diverse family is characterized by an alpha/beta hydrolase fold and function via a proposed catalytic triad (Ollis, Cheah et al. 1992). Using Phyre2, we modeled the enzyme and identified the putative catalytic triad to be S116, D247 and H276 that lie within catalytic distances from each other (Figure 2.26) (Kelley and Sternberg 2009). Each residue was mutated to alanine and the resulting recombinant proteins were expressed and purified from E. coli at comparable yields as wild-type for the in vitro reaction with Bref-PKS (Erijman, Dantes et al. 2011). Surprisingly, adding either S116A or D247A mutant to Bref-PKS resulted in a similar product profile as those seen with the wild type Bref-TH. Only the H276A mutant failed to catalyze release of polyketide products and yielded 16 as the major product upon base hydrolysis (Figure 2.25). This suggests that only H276 is essential for the chain length-specific hydrolysis of the PKS product, possibly serving as the general base to facilitate the thiohydrolysis. We also probed the specificity of the hydrolase-PKS interactions by incubating Bref-PKS with CazE from chaetoviridin pathway (Winter, Sato et al. 2012) and Fma- AT from the fumagillin pathway (Lin, Chooi et al. 2013), both known to participate in release of reduced polyketide products from the respective HPPKS partners. In each case, the noncognate releasing enzymes did not lead to turnover of the octaketides, and the nonaketide products were instead detected using base hydrolysis (Figure 2.25).
Figure 2.26. Modeling of Bref-TH to known hydrolases allows identification of its active site residues. (A) Sequence alignment of Bref-TE to other structurally-related hydrolases. The catalytic triad of Ser-116, Asp-247 and His-276 was conserved. (B) The model shows how the three residues are within catalytic distances of each other.
2.2.3.3.5 Chain-length control by the releasing enzyme is also observed in other HRPKS systems

Figure 2.27. Fma-PKS produces longer polyenes in the absence of cognate Fma-AT. (A) HPLC profiles of Fma-PKS with: Fma-AT; base hydrolysis; or Bref-TH. Fma-PKS produces a hexaketide polyene in the presence of the partner Fma-AT. In its absence or in the presence of the non-cognate Bref-TH, the PKS catalyzes 1 to 2 more extension steps to yield the heptaketide 10 and octaketide 11. The reactions consist of 20μM Fma-PKS, 2mM mCoA and 2mM NADPH with either 20μM of the releasing enzyme or base hydrolysis (1M NaOH at 65°C for 10 min). (B) Proposed structures of the polyene compounds produced in the in vitro assay.

To examine whether chain-length regulation by the releasing enzyme is also observed in other HRPKSs, we assayed the Fma-PKS involved in the biosynthesis of the polyene portion of the meroterpenoid fumagillin. Fma-PKS was previously shown to produce the highly conjugated hexaketide pentaenoic acid 18 ([M+H+] 191, λ_{max} 358 nm) in the presence of its releasing acyltransferase partner, Fma-AT (Lin, Chooi et al. 2013). When the Fma-PKS assay was
performed in the absence of Fma-AT and subjected to base-hydrolysis followed by extraction, two new products were observed, 19 ([M+H]$^+$ 217) and 20 ([M+H]$^+$ 243), with $\lambda_{\text{max}}$ of 378 and 398 nm, respectively. The periodic increases in both mass (+26 mu) and $\lambda_{\text{max}}$ compared to 18 therefore strongly indicate 19 and 20 are heptaketide and octaketide polyenes, respectively, as shown in Figure 2.27. This is also verified by the in vitro labeling studies with 2-$^{13}$C-malonate that showed the corresponding mass shift of +7 and +8 for 19 and 20, respectively (Figure 4.40, Figure 4.41).

When Bref-TH is used in the reaction with Fma-PKS, 19 and 20 were also dominantly produced relative to 18 (Figure 2.27).

### 2.2.3.4 Discussion

In this work, we identified a gene cluster in *E. brefeldianum* that is most likely to be involved in the biosynthesis of the protein transporter inhibitor BFA. BFA is a fungal polyketide that is derived from a highly reduced polyketide synthesized by a HRPKS. The Bref-PKS along with a partnering Bref-TH were reconstituted in *S. cerevisiae* and in vitro. The reconstitution experiments showed that the dominant product of Bref-PKS in the presence of Bref-TH is an acyclic polyketide 3 that is of the same length as BFA and exhibited the expected $\beta$-reduction patterns for downstream conversion into BFA (Figure 2.17). Unexpectedly, Bref-PKS synthesized longer polyketide products in the absence of Bref-TH, implicating an important role of Bref-TH in controlling the chain length of the HRPKS. This phenomenon was also observed in the Fma-PKS and Fma-AT pair involved in fumagillin biosynthesis. The in vitro reconstitution studies reported here were crucial in enabling single turnover experiments (chemical hydrolysis) that were not observable under in vivo conditions.

Reduced polyketides synthesized by HRPKSs requires accessory enzymes for product release. In most cases, this is completed by an assortment of enzymes belonging to the alpha-beta
hydrolase family. The protein-protein interaction between the HRPKS and the releasing enzyme, which facilitates product turnover and chain length control, is therefore an intricate part of the overall programming rule of these enzymes. When the releasing enzyme is an acyltransferase, the polyketide is either transferred to the hydroxyl group of a small molecule acceptor such as catalyzed by LovD in lovastatin biosynthesis (Xie, Meehan et al. 2009); or the free thiol of the ACP domain of a partnering PKS catalyzed by the Starter-Unit:ACP acyltransferase (SAT) in dual PKS systems (Crawford, Dancy et al. 2006; Zhou, Qiao et al. 2010; Zhou, Qiao et al. 2010). Thiohydrolases such as Bref-TH hydrolyzes the polyketide thioester to release the product. Other examples of TH include LovG that hydrolyses dihydromonacolin L from LovB (Xu, Chooi et al. 2013). Interestingly, the releasing alpha-beta hydrolases are widely varied in sequences and also in catalytic mechanisms. Whereas LovD, LovG and SAT domain utilize covalent catalysis via active site nucleophiles, enzymes such as Bref-TH and CazE apparently operate via noncovalent, general base catalysis similar to the trichothecene acyltransferase (Garvey, McCormick et al. 2008). Each releasing enzyme is also highly specific for the partnering HRPKS ACP domain, as shown in our results. This exclusive protein-protein interaction likely arose during evolution to ensure minimal crosstalk between HRPKS gene clusters, as well as triggering allosteric structural changes required for catalysis (Jimenez-Oses, Osuna et al. 2014).

Using the Bref-PKS and Bref-TH pair (as well as the Fma pair), we showed that the releasing enzyme is important in ensuring the proper chain length control of HRPKSs. In the absence of the TH enzymes, both Bref-PKS and Fma-PKS synthesized longer products than what is reflected in the final product. This result can be rationalized in a competition model in which the polyketide chain can be either offloaded by the TH/AT at the correct size, or can be recaptured by the KS domain for another round of elongation (Figure 2.28). In the absence of the TH, the
polyketide chain remains attached to the ACP domain, which allows reentry into the KS domain. If the KS can accommodate a product of longer size, an additional extension step can take place as observed in products 16 and 19. Hence the THs must be precisely programmed to only recognize acyl products of the correct length. We did not detect any shorter polyketides in the in vitro assays, indicating the high substrate specificities of the TH towards the correct acyl group. However, when equimolar amount of Bref-TH was added 16 hours after initiation of the in vitro Bref-PKS reaction, the longer nonaketide products (16 and 17) were again observed at similar levels as the base-hydrolyzed reaction (Figure 4.39). This indicates that Bref-TH is capable of hydrolyzing longer chain length but preferably hydrolyze the correct octaketide chain in a timely manner when co-incubated with Bref-PKS. Interestingly, in both Bref-PKS and Fma-PKS, the extra ketide(s) that form as a result of excluding the TH in the reaction were completely processed by the available reduction domains. The recognition of the longer (and unnatural) substrates by these domains may similarly be due to stalling of the polyketide on the ACP domain, which led to the observed modification. Detailed kinetics studies using model substrates of varying length will provide insights into the substrate specificities of these tailoring domains.

These results support that KS domain remains an important element in determining polyketide chain length. This is also evident in the phylogeny based classification of HRPKSs using KS domain sequences, which led to clading of the HRPKSs based on chain size (short, medium and long, as shown in Figure 2.19). Additionally, some HRPKS systems appear to maintain a high fidelity in producing the correct product chain length; for instance, both lovastatin diketide and nonaketide synthases always produce the correct chain length (Ma, Li et al. 2009; Xie, Meehan et al. 2009), as well as the solanapyrone synthase that releases the product via pyrone formation (Kasahara, Miyamoto et al. 2010). However, as demonstrated here in Bref-PKS and
Fma-PKS, these domains appear to also have some flexibility on the chain length programming and are insufficient to terminate at the correct chain length alone. In other words, KS domains are capable of synthesizing products of longer size without downstream enzyme control. This may rationalize an in vivo study involving the biosynthesis of fumonisins. Zhu et al reported a successful complementation of the \( C_{16} \)-PKS ALT1 for the \( C_{18} \)-PKS FUM1-disruption strain to yield the \( C_{18} \) fumonisins (Zhu, Vogeler et al. 2008). In this case, offloading of the polyketide via a decarboxylative condensation with an alanine by Fum8 dictates the ultimate chain length in the final product (Gerber, Lou et al. 2009). Releasing enzyme control of chain length was also observed in the HRPKS-NRPS involved in the biosynthesis of preaspyridone, a precursor of the natural product aspyridone. In that example, when the fused NRPS module was excised from the megasynthetases, the PKS module was able to synthesize a polyketide product that is longer than that present in preaspyridone (pentaketide instead of tetraketide) (Xu, Cai et al. 2010). That is again most likely due to stalling of the polyketide chain on the PKS in the absence of a downstream domain (C domain in the NRPS) that offloads the correctly sized polyketide for subsequent modification. Additional factors that affect chain length control outside of KS were previously reported, such as the KR-dependent product size observed in PKS-NRPS domain swapping experiments (Fisch, Bakeer et al. 2011).

While it may appear that some HRPKSs become aberrant in their chain-length control in the absence of cognate releasing partners, it could be argued that these HRPKSs might have originated from ancestral PKS clusters that synthesize final products of longer chain length. In other words, this is Nature’s alternative and quick way of generating chain length diversity without modifying the KS or other PKS components, which could be more difficult. In this way, the same HRPKS can be adapted to produce polyketide product of different chain lengths by partnering with
releasing enzymes of different chain length specificities. Thus, Bref-PKS may have originated from an ancestral HRPKS cluster that produces a nonaketide product, but was adapted to produce octaketide by coupling with the Bref-TH with shorter chain length specificity. This lesson from Nature has important implications for engineering of polyketide biosynthesis, as it suggests that we can attempt to alter the chain length specificity of the releasing enzyme instead of focusing on the HRPKS components to manipulate the chain length of the final product.

**Figure 2.28.** Summary of the programming rules exhibited by Bref-PKS. From the in vitro studies, we were able to fully reconstitute the complex programming of this model HRPKS. Bref-PKS uses different permutations of the reductive domains at each extension cycle and selectively offloads the correct octaketide products with the partner Bref-TH or the longer nonaketide products with base hydrolysis. Compounds 13 and 15 that resulted from additional enoyl reduction at the final extension are italicized.
2.2.3.5 Implications of in vitro Bref-PKS data on the biosynthesis of Brefeldin A

In light of the results of the reconstitution studies on Bref-PKS, we can improve our current understanding of the biosynthesis of BFA. From our data, we can conclude that Bref-PKS and Bref-TH are minimally required to produce the octaketide backbone of BFA. Additionally, Bref-PKS is capable of producing the correctly tailored intermediate for further modification to achieve the final structure of BFA. The C2-C3 and C10-C11 double bonds are identical to those in BFA, while the C4-C5 double bond is strategically placed in 12 to facilitate the oxidative steps leading to the intramolecular annulation that installs the cyclopentane ring. The actual mechanism would require additional studies on the enzymes encoded by the genes in the bref cluster. We have also demonstrated that Bref-PKS synthesizes products with variable degrees of reduction, which explains the isolation of some BFA analogs, such as the un lactonized 7-dehydrobrefeldin A acid (Yu, Zhu et al. 2010).

The mechanism for macrolactonization of BFA is an intriguing aspect of the biosynthesis. Unlike the RAL systems, Bref-PKS does not partner with an NR-PKS that uses an in cis thioesterase for macrolactonization. However, we propose that the Bref-TH could similarly perform both offloading and lactonization, as it is the only candidate gene transcribed in the bref cluster. It is foreseeable that the mechanism would involve the C15-hydroxyl that has been observed in in vitro product 12, which is the proper position for an intramolecular attack on the thioester bond. The catalytic His276 could deprotonate the C15-hydroxyl group and make it a suitable nucleophile for lactonization. Our observation that all the products were linear, however, indicate the necessity of having the cyclopentane ring form first to “bend” the molecule and bring the nucleophile closer to the ACP to direct lactonization (bottom scheme in Figure 2.17). This is in contrast to what was previously hypothesized, where the macrolactonization precedes the
cyclopentane ring formation. (Mabuni, Garlaschelli et al. 1977) Mabuni et al. originally proposed a mechanism involving a C4-C5 epoxide intermediate. However, this epoxide-opening mechanism requires an alkene at C9-C10, which is an unusual position in a polyketide chain. (17) Recently, Zhang et al. proposed an NADPH-dependent reductive cyclization catalyzed by the alcohol dehydrogenase IkaC for the formation of a five-membered ring within the macrolactam ikarugomycin. (Zhang, Zhang et al. 2014) However, we did not find a homolog of this enzyme in the bref cluster. In light of the structure of 3 produced by Bref-PKS, we propose a mechanism via a C9 radical intermediate catalyzed by a P450 enzyme (Figure 2.29). Addition of the C9 radical to the C4-C5 alkene followed by an oxygen rebound will result in the cyclopentane ring and C4-hydroxyl of 1. The actual cyclization mechanism will be the subject of further investigations, including determining the roles of the P450s in the cluster. Additionally, the role of Bref-TH in the lactonization of BFA should also be probed in the future.

Figure 2.29. Proposed mechanism for the formation of the cyclopentane ring. We propose that a C9 radical initiated by a P450 enzyme could perform the alkene addition to the C4-C5 double bond to form the cyclopentane moiety. This is followed by an oxygen rebound which hydroxylated C4 and resulted in the final structure observed in BFA.
The mechanism for macrolactonization of BFA is an intriguing aspect of the biosynthesis. Unlike the RAL systems, Bref-PKS does not partner with an NR-PKS that uses an *in cis* thioesterase for macrolactonization. However, we propose that the Bref-TH could similarly perform both offloading and lactonization, as it is the only candidate gene transcribed in the *bref* cluster. It is foreseeable that the mechanism would involve the C15-hydroxyl that has been observed in in vitro product 12, which is the proper position for an intramolecular attack on the thioester bond. The catalytic His276 could deprotonate the C15-hydroxyl group and make it a suitable nucleophile for lactonization. Our observation that all the products were linear, however, indicate the necessity of having the cyclopentane ring form first to “bend” the molecule and bring the nucleophile closer to the ACP to direct lactonization (bottom scheme in Figure 2.17). This is in contrast to what was previously hypothesized that the macrolactonization precedes the cyclopentane ring formation (Mabuni, Garlaschelli et al. 1977). Further studies are required to confirm the role of Bref-TH in the lactonization of BFA.

As noted earlier, Bref-TH homolog is found in the vicinity of a Bref-PKS homolog in *Trichoderma virens*. Such HRPKS/*in trans*-TH pairs are also found in several other fungal genomes, including *Aspergillus nidulans* (AN7084.2/7083.2), *Botryotinia fuckeliana* (BeDW1_1087/1086), *Neofusicoccum parvum* (UCNR2P-2180/2181) and *Macrophomina phaeoeolina* (MPH_06436/06434). These HRPKS/ TH pairs may be responsible for production of similar linear reduced polyketide or macrolide compounds. Several non-RAL macrolides have been previously isolated from fungi such as putaminoxin (Evidente, Lanzetta et al. 1995), pinolidoxin (Evidente, Lanzetta et al. 1993), and balticolid (Shushni, Singh et al. 2011); the discovery and characterization of the Bref-PKS and Bref-TH therefore provide leads to the biosynthetic gene clusters that are responsible for production of such fungal macrolide scaffolds.
2.2.3.6 Conclusion

HRPKSs are still poorly understood enzymes that catalyze the synthesis of a wide array of compounds. In order to harness their biosynthetic potential, it is important to first improve our understanding of the underlying mechanism behind their activities by finding a model PKS to work with. In this paper, we report the heterologous expression and reconstitution of Bref-PKS involved in the biosynthesis of BFA to characterize the HRPKS programming rules in vitro. This system has the advantage of being a standalone enzyme with a sizeable product and different reduction patterns at each extension. Using this HRPKS, we uncovered an NADPH-dependent reductive tailoring by the PKS in synthesizing products of mixed reduction. We also demonstrated that chain-length determination is not dictated by the KS domain alone, but can be altered by the releasing enzyme. Such strategy could be Nature’s way of generating chain length diversity without having to evolve the large multidomain megasynthases. The TH-mediated control is highly specific with respect to protein-protein interactions, as well as on acyl substrate recognition. This study further underscores the importance of including tailoring enzyme functions in relating and predicting fungal HRPKS and their products.

2.2.3.7 Acknowledgment

This work was supported by the US NIH (1R01GM085128 and 1DP1GM106413) to Y.T.; A.Z. is supported by NIH Biotechnology Training Grant T32GM067555; Y.H.C. is supported by an Australian Research Council (ARC) Discovery Early Career Researcher Award fellowship; H-C.L. is supported by National Science Council of Taiwan (102-2917-I-564-008). We thank Dr. Jaclyn Winter for providing the purified CazE.
2.2.3.8 Accession codes

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries with an accession code KJ728786.

2.2.3.9 Experimental Procedures

2.2.3.9.1 Strain and Culture Conditions

*E. brefeldianum* ATCC 58665 was obtained from ATCC and maintained in GMM or SMM agar at 28°C. For BFA production and for mRNA extraction, the strain was grown in liquid MEM media at room temperature with 250 rpm shaking (McCloud, Burns et al. 1995). *Saccharomyces cerevisiae* BJ5464 was used for protein expression of Bref-PKS and subsequent in vivo production of 12 to 15. This strain was maintained and cultured in YPD, while the transformants carrying the recombinant plasmids were grown in synthetic defined dropout media with appropriate supplements. *E. coli* TOPO 10 was used for the subcloning steps, while *E. coli* BL21(DE3) was used for expressing Bref-TH.

2.2.3.9.2 Sequencing and Bioinformatic Analysis

The genomic DNA of *E. brefeldianum* was prepared from mycelium grown in stationary liquid culture (Karp, Isaac et al. 1998). Shotgun sequencing was performed at GenoSeq (UCLA Genotyping and Sequencing Core) with the GS FLX Titanium system (Roche) and at Ambry Genetics (Aliso Viejo, CA) using Illumina Hiseq 2000. The reads were assembled into contigs using SOAPdeNOVO.(Li, Zhu et al. 2010) The contigs were formatted to BLAST database format for local BLAST search using standalone BLAST software (v. 2.2.18). Gene predictions were performed using the FGENESH program (Softberry) and manually checked by comparing with homologous gene/proteins in the GenBank database. Functional domains in the translated protein...
sequences were predicted using Conserved Domain Search (NCBI). Modeling and alignment of the Bref-hydrolase was performed using Phyre2 server (Kelley and Sternberg 2009) and the image was generated using Boxshade 3.21.

2.2.3.9.3 Expression analysis by reverse transcription polymerase chain reaction (RT-PCR)

The total RNA of *E. brefeldianum* ATCC 58665 was extracted from the culture grown in MEM media using the Ambion RNA extraction kit. The first strand cDNA was synthesized using the Oligo-dT primer and Improm-II reverse transcription system (Promega) according to the manufacturer’s instructions. Desired cDNA was then amplified with GoTaq Green Master Mix (Promega) using gene-specific primers synthesized by Integrated DNA Technologies (Table 4.5). gDNA template (previously prepared) was used for comparison using the same primer mix as the cDNA reaction.

2.2.3.9.4 Molecular Genetic Manipulation

Polymerase chain reactions for cloning were performed using Phusion high-fidelity DNA polymerase (New England Biolabs) or Platinum Pfx DNA polymerase (Invitrogen). PCR products were cloned into a PCR-Blunt vector (Invitrogen) for DNA sequencing and subcloning. Restriction enzymes (New England Biolabs) and T4 ligase (Invitrogen) were used respectively for the digestion and ligation of DNA fragments. All primers were ordered from IDT (Table 4.5).

The intron-less transcript of *bref*-TH was obtained by amplification from the cDNA (as prepared previously) using primers containing the *EcoRI* and *NotI* restriction sites. The PCR product was digested with the corresponding enzymes and ligated to the linearized pHis8 vector.(Jez, Ferrer et al. 2000) The correct construct (designated pAZ93) was PCR-verified and
sequenced to ensure intact ORF. pAZ93 was also used as the template for constructing the yeast plasmid (Trp3 marker) for the double transformation experiments with bref-PKS (pAZ112), as well as for constructing the bref-TH single and triple mutants using transfer PCR (Erijman, Dantes et al. 2011).

To construct the intron-less transcript of bref-PKS, the 7.3-kb gene was divided into 5 equal-sized fragments; each fragment was amplified from the cDNA (as prepared previously) and was subcloned into PCR-blunt vector and sequence-verified with M13 primers to confirm absence of introns. After confirmation, the respective pieces were re-amplified from the PCR-blunt subcloning constructs with primers that contain overlapping regions to the next segment of the gene. The pieces were transformed together with the linearized vector backbone derived from YEplac195 containing the ADH2 promoter and terminator with N-terminus FLAG-tag and C-terminus hexahistidine tag (Gietz and Sugino 1988) into S. cerevisiae BJ5464-NpgA (Jones 1991) using S. c. EasyComp™ Transformation Kit (Invitrogen) and selected on uracil-dropout semisynthetic media. The resulting transformants were screened by colony-PCR and the plasmid in the correct transformant was rescued using the Zymoprep Yeast Plasmid Miniprep Kit (Zymo Research) and transformed into E. coli for propagation and sequencing verification. The resulting plasmid was designated pAZ94.

2.2.3.9.5 Protein expression and in vitro reactions

The bref-TH-pHis8 construct (pAZ93) was transformed into BL21(DE3) via electroporation and cultured and induced with IPTG for expression. The His-tagged protein was purified using nickel-affinity chromatography. The mutant hydrolases were expressed and purified the same way as the wild type. Similarly, the bref-PKS construct (pAZ94) was re-transformed into the yeast host for protein expression, as described elsewhere,(Lee, Da Silva et al. 2009) and
subsequently purified using nickel-affinity chromatography and eluted using 250 mM of imidazole.

Unless otherwise stated, the reactions involving Bref-PKS were set up on ice containing 100 mM phosphate buffer, 0.5 to 10 mM NADPH, 2 mM malonyl-CoA and Bref-PKS and Bref-TH enzymes (typically 20 μM) to a final volume of 100 μL. The reactions were left at room temperature overnight, extracted with 2x200 μL ethyl acetate with 0.1% acetic acid and dried completely for analysis. In cases where base hydrolysis was performed, 10 μL of 1M NaOH was added first and the reaction was heated to 65°C before extraction with the organic solvent.

For the $^{13}$C labeling experiments, Bref-PKS was incubated with 20 μM of MatB, 10 μM of ATP, 10 mM 10 mM NADPH, 100 mM [2-$^{13}$C]-malonate, 10 mM MgCl₂, 25 mM ATP, 10 mM CoA, 25 μM MatB, 1 mM NADPH in 100 mM PBS buffer, pH 7.4 in 100 μL total volume. The reactions were performed at room temperature overnight and either directly extracted with ethyl acetate + 0.1% acetic acid, or base-hydrolyzed using 1M of NaOH.

2.2.3.9.6 In vivo culturing and extraction of compounds

pAZ94 (bref-PKS, ura3 marker) and pAZ112 (bref-TH, trp3 marker) were co-trasformed into S. cerevisiae BJ5464-NpgA (Jones 1991) using S. c. EasyComp™ Transformation Kit (Invitrogen) and selected on uracil- and tryptophan-dropout semisynthetic media. The transformants were PCR-screened to confirm the presence of both plasmids and the correct colony was grown on minimal media as seed culture for three days. Subsequently, the seed culture is diluted 1000x onto 1L of YPD and grown for at most three days at 28°C with shaking at 250 rpm. The metabolites were monitored daily by sampling 500 μL of the culture and extracting with equal volume of ethyl acetate with 0.1% acetic acid for LC-MS analysis. For purification of compounds,
12 and 14 were harvested at the optimal production of 36 hrs, while 13 and 15 were extracted after 72 hours. 20 g of Amberlite XAD02 polymeric adsorbent resin (Supelco) was added to the media after centrifugation of the cells to extract the organic compounds and was left overnight with shaking. The resin was collected by filtration and the compounds were eluted using acetone. The solvent was dried using and the residue was partitioned with chloroform/water. The chloroform fraction was then dried completely and flash separated using the CombiFlash system. A C18 column was used with water and acetonitrile solvent system. The fractions were analyzed and those containing the desired compounds were pooled together for the final purification using semi-preparative HPLC with Phenomenex Luna 5µL 250x1000mm C18 reverse phase column using acetonitrile/water + 0.1% formic acid solvent system.

2.2.3.9.7 LC/MS Analysis

The dried samples were first dissolved in methanol and centrifuged for 8 minutes before injecting to the Shimadzu 2010 EV liquid chromatography mass spectrometer with positive and negative electrospray ionization and Phenomenex Luna 5 µl 2.0 × 10 mm C18 reverse-phase column. The samples were resolved on a linear gradient from 5% to 95% with CH₃CN/H₂O + 0.05% formic acid solvent system for 30 minutes.
3 Conclusion

Type I iterative PKSs still remain as important enzymes in the biosynthesis of natural products. Their products continue to inspire the scaffolds for new therapeutics and drugs in treatment of diseases afflicting mankind. Because of this, we need to gain better understanding for their mechanisms of action in order to develop better predictive rules in genome mining, as well as to be able to better engineer them in creating new product analogs.

To this end, this work has contributed to the study of polyketide synthases using two approaches: the bottom-up approach, where we used the sequence information to deduce its product; and the top down approach, where we used a known product to find the cluster responsible for its biosynthesis. Using these methods, we have successfully established two new polyketide synthase systems that is now added to the roster of biochemically characterized PKSs from the myriads of PKSs exposed by the genome sequencing efforts in recent years.

From the bottom up-approach, we investigated a silent cluster in the industrial fungi *A. niger* ATCC 1015 that contains a dual-PKS system (HRPKS and NRPKS). Through overexpression of a cluster-specific Zn$_2$Cys$_6$ regulator, we were able to induce the transcription of the genes belonging to the cluster. In turn, the strain started to produce several new compounds, which we isolated, characterized and identified to belong to the azaphilone group of natural products. These compounds, which we termed azanigerones, contained the characteristic pyranoquinone core and are found to readily react with amines. By knocking out the HRPKS in the cluster, we have demonstrated a convergent model for PKS collaborations; instead of the HRPKS producing a reduced precursor that is passed on to the downstream NRPKS, as in the case of previously characterized tandem systems of RALs and asperfuranone, the *aza* PKS enzymes performs two parallel biosynthesis. The NRPKS produces the unreduced hexaketide that becomes
regioselectively cyclized by the PT domain of the PKS and offloaded by the R-domain. From the in vivo analysis of the NRPKS in yeast, we found that the polyketide product readily cyclizes into a naphthoquinone compound that is susceptible to aerial oxidation, if the terminal ketone remains unreduced. Furthermore, we have also demonstrated that the ketoreductase AzaE in the cluster controls this spontaneous event by reducing the ketone to an alcohol. This was observed in the intermediate that was detected from the culture of the activated strain (previously identified as FK17-p2A). Meanwhile, the HRPKS independently synthesized a fully-reduced triketide that is offloaded and subsequently transferred to the azaphilone core possibly by an acyltransferase in the cluster.

One of the significant contributions of this work is in identifying the biosynthetic basis of the pyrano-quinone core shared by the almost 700 azaphilone compounds produced by fungi. We expressed the FAD-dependent monooxygenase AzaH in E. coli and found that one enzyme is sufficient to perform both the hydroxylation at the C5 position of the benzaldehyde intermediate FK17-p2A, as well as the cyclization of the second pyrone ring. This has implications on the genome mining of azaphilone compounds, since we can now easily identify the azaphilone clusters simply by looking for this gene fingerprint in the dual-PKS clusters. This has been successfully demonstrated by the proper identification of azaphilone clusters in other fungi, as reported in literature.

We utilized the top-down approach in order to elucidate the biosynthetic basis of the potent protein-trasnport inhibitor Brefeldin A. Due to the structural features of this classic reduced polyketide molecule, it is easy to envision its HRPKS enzyme as a good platform for mechanistic studies. The quest to find its cluster was simplified by the availability of affordable sequencing technology that gave a relatively complete account of the total metabolic potential of the organism.
Successful expression of Bref-PKS and its partner thiohydrolase provided us with some important insights on the programming of HRPKSs. In vitro, it demonstrated a cofactor-dependent use of the \( \beta \)-tailoring domains that contributed to product diversity. More importantly, we found that its cognate thiohydrolase regulates the chain length of the product, instead of it being inherently programmed in the PKS. Bref-PKS was capable of catalyzing another round of extension, as the chain stalls in the ACP in the absence of a release mechanism. This has similarly been demonstrated in other HRPKSs and highlights the significance of these accessory releasing enzymes in deciphering the product of the PKS.

There are still a lot of questions in polyketide biosynthesis and while it may be difficult to find general rules that would correlate the sequence to its product, progress is being made by the biochemical characterization of individual systems.
4 Appendix

4.1 Supplementary information for Section 2.1.2

The information presented here is also available on [http://www.sciencedirect.com/science/article/pii/S1074552112002268](http://www.sciencedirect.com/science/article/pii/S1074552112002268). The atom labeling is the same as it appears in the main text.

Table 4.1. NMR data for Azanigerone A (left) and FK17-p2A (right).

<table>
<thead>
<tr>
<th>C No.</th>
<th>( \delta^{13} \text{C} ) (ppm)</th>
<th>( \delta^{1} \text{H} ) (ppm) (area,m, ( J_{HH} ) (Hz))</th>
<th>( ^{13} \text{C} - ^{1} \text{H} ) HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>154.05</td>
<td>7.93 (1H, s)</td>
<td>C2, C7, C9</td>
</tr>
<tr>
<td>2</td>
<td>141.33</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>192.12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>84.06</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>194.41</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>111.45</td>
<td>5.88 (1H, s)</td>
<td>C4, C7, C8</td>
</tr>
<tr>
<td>7</td>
<td>115.55</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>118.76</td>
<td>7.31 (1H, s)</td>
<td>C6, C7, C9, C10</td>
</tr>
<tr>
<td>9</td>
<td>146.33</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>161.17</td>
<td>10.97 (1H, s) OH</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>21.9</td>
<td>1.54 (3H, s)</td>
<td>C3, C4, C5</td>
</tr>
<tr>
<td>1'</td>
<td>176.86</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2'</td>
<td>36.35</td>
<td>2.68 (1H, m)</td>
<td>C1’, C3’, C7’</td>
</tr>
<tr>
<td>3a’</td>
<td>40.85</td>
<td>1.74 (1H, m)</td>
<td>C2’, C4’, C7’, C8’</td>
</tr>
<tr>
<td>3b’</td>
<td></td>
<td>1.12 (1H, m)</td>
<td></td>
</tr>
<tr>
<td>4’</td>
<td>31.94</td>
<td>1.46 (1H, m)</td>
<td>C2’, C3’, C5’, C6’, C8’</td>
</tr>
<tr>
<td>5a’</td>
<td>29.57</td>
<td>1.31 (1H, m)</td>
<td>C4’, C6’, C8’</td>
</tr>
<tr>
<td>5b’</td>
<td></td>
<td>1.12 (1H, m)</td>
<td>C5’</td>
</tr>
<tr>
<td>6’</td>
<td>11.2</td>
<td>0.86 (3H, t, 7.4)</td>
<td>C5’</td>
</tr>
<tr>
<td>7’</td>
<td>17.68</td>
<td>1.17 (3H, d, 6.9)</td>
<td>C1’, C2’, C3’</td>
</tr>
<tr>
<td>8’</td>
<td>19.15</td>
<td>0.90 (3H, d, 6.6)</td>
<td>C3’, C4’, C5’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C No.</th>
<th>( \delta^{13} \text{C} ) (ppm)</th>
<th>( \delta^{1} \text{H} ) (ppm) (area,m, ( J_{HH} ) (Hz))</th>
<th>( ^{13} \text{C} - ^{1} \text{H} ) HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>195.89</td>
<td>9.8 (1H, s)</td>
<td>C2, C4</td>
</tr>
<tr>
<td>2</td>
<td>114.64</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>112.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>165.99</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>165.36</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>112.27</td>
<td>6.27 (1H, s)</td>
<td>C3, C8</td>
</tr>
<tr>
<td>7</td>
<td>139.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>48.44</td>
<td>4.09 (2H, s)</td>
<td>C2, C6, C7, C9</td>
</tr>
<tr>
<td>9</td>
<td>209.49</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10a</td>
<td></td>
<td>2.74 (1H, dd, 15.8 &amp; 8.1)</td>
<td>C9, C11</td>
</tr>
<tr>
<td>10b</td>
<td>52.88</td>
<td>2.65 (1H, dd, 15.8 &amp; 4.6)</td>
<td>C10, C12</td>
</tr>
<tr>
<td>11</td>
<td>66.07</td>
<td>4.29 (1H, m)</td>
<td>C10, C12</td>
</tr>
<tr>
<td>12</td>
<td>24.5</td>
<td>1.22 (3H, d, 5.9)</td>
<td>C11</td>
</tr>
<tr>
<td>13</td>
<td>8.05</td>
<td>2.04 (3H, s)</td>
<td>C2, C3, C4, C5, C6</td>
</tr>
</tbody>
</table>
Table 4.2. NMR data for Azanigerone B (left) and Azanigerone D (right).

<table>
<thead>
<tr>
<th>C No.</th>
<th>$\delta^{13}C$ (ppm)</th>
<th>$\delta^1H$ (ppm) (area,m, $J_{HH}$ (Hz))</th>
<th>$^{13}C - ^1H$ HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>153.61</td>
<td>7.86 (1H,s)</td>
<td>C2, C7</td>
</tr>
<tr>
<td>2</td>
<td>142.04</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>192.77</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>83.67</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>193.09</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>107.27</td>
<td>5.53 (1H, s)</td>
<td>C4, C7, C8</td>
</tr>
<tr>
<td>7</td>
<td>115.16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>110.5</td>
<td>6.19 (1H, s)</td>
<td>C6, C7, C9, C10</td>
</tr>
<tr>
<td>9</td>
<td>158.85</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10a</td>
<td>42.45</td>
<td>2.54 (1H, dd, 15 &amp; 5)</td>
<td>C8, C9, C11, C12</td>
</tr>
<tr>
<td>10b</td>
<td>65.33</td>
<td>4.16 (1H, m)</td>
<td>C10, C12</td>
</tr>
<tr>
<td>11</td>
<td>23.42</td>
<td>1.32 (3H, d, 6.2)</td>
<td>C10, C11</td>
</tr>
<tr>
<td>12</td>
<td>21.89</td>
<td>1.52 (3H, s)</td>
<td>C3, C4, C5</td>
</tr>
<tr>
<td>1’</td>
<td>176.26</td>
<td>1.34 (1H, m)</td>
<td>C1’, C3’, C7’</td>
</tr>
<tr>
<td>2’</td>
<td>36.08</td>
<td>2.70 (1H, m)</td>
<td>C1’, C3’, C7’</td>
</tr>
<tr>
<td>3a’</td>
<td>40.71</td>
<td>1.77 (1H, m)</td>
<td>C2’, C4’, C7’, C8’</td>
</tr>
<tr>
<td>3b’</td>
<td>31.69</td>
<td>1.51 (1H, m)</td>
<td>C2’, C3’, C5’, C6’, C8’</td>
</tr>
<tr>
<td>4’</td>
<td>31.69</td>
<td>1.34 (1H, m)</td>
<td>C2’, C3’, C5’, C6’, C8’</td>
</tr>
<tr>
<td>5a’</td>
<td>29.36</td>
<td>1.12 (1H, m)</td>
<td>C4’, C6’, C8’</td>
</tr>
<tr>
<td>6’</td>
<td>10.99</td>
<td>0.90 (3H, t, 7.4)</td>
<td>C4’, C5’</td>
</tr>
<tr>
<td>7’</td>
<td>17.51</td>
<td>1.21 (3H, d, 6.9)</td>
<td>C1’, C2’, C3’</td>
</tr>
<tr>
<td>8’</td>
<td>18.98</td>
<td>0.92 (3H, d, 6.6)</td>
<td>C3’, C4’, C5’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C No.</th>
<th>$\delta^{13}C$ (ppm)</th>
<th>$\delta^1H$ (ppm) (area,m, $J_{HH}$ (Hz))</th>
<th>$^{13}C - ^1H$ HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>141.93</td>
<td>8.08 (1H, s)</td>
<td>C2, C7, C9</td>
</tr>
<tr>
<td>2</td>
<td>139.17</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>192.31</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>85.84</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>196.71</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>101.61</td>
<td>5.57 (1H, s)*</td>
<td>C4, C7, C8</td>
</tr>
<tr>
<td>7</td>
<td>118.16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>122.48</td>
<td>7.45 (1H, s)</td>
<td>C6, C7, C9, C10</td>
</tr>
<tr>
<td>9</td>
<td>153.65</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>163.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>24.21</td>
<td>1.54 (3H, s)</td>
<td>C3, C4, C5</td>
</tr>
<tr>
<td>1’</td>
<td>178.42</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2’</td>
<td>38.55</td>
<td>2.7 (1H, m)</td>
<td>C1’, C3’, C7’</td>
</tr>
<tr>
<td>3a’</td>
<td>65.33</td>
<td>1.76 (1H, m)</td>
<td>C2’, C4’, C7’, C8’</td>
</tr>
<tr>
<td>3b’</td>
<td>43.19</td>
<td>1.13 (1H, m)</td>
<td>C2’, C4’, C7’, C8’</td>
</tr>
<tr>
<td>4’</td>
<td>33.95</td>
<td>1.57 (1H, m)</td>
<td>C2’, C3’, C5’, C6’, C8’</td>
</tr>
<tr>
<td>5a’</td>
<td>31.71</td>
<td>0.93 (1H, m)</td>
<td>C4’, C6’, C8’</td>
</tr>
<tr>
<td>5b’</td>
<td>12.35</td>
<td>0.97 (3H, t, 7.4)</td>
<td>C4’, C5’</td>
</tr>
<tr>
<td>6’</td>
<td>19.07</td>
<td>1.14 (3H, d, 7.0)</td>
<td>C1’, C2’, C3’</td>
</tr>
<tr>
<td>8’</td>
<td>20.24</td>
<td>0.96 (3H, d, 6.6)</td>
<td>C3’, C4’, C5’</td>
</tr>
</tbody>
</table>
Table 4.3. NMR data for Azanigerone E (left) and Azanigerone F (right).

<table>
<thead>
<tr>
<th>C No.</th>
<th>δ (^{13}\text{C}) (ppm)</th>
<th>δ (^{1}\text{H}) (ppm) (area,m, (J_{HH})(Hz))</th>
<th>(^{13}\text{C} - ^{1}\text{H}) HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>152.79</td>
<td>7.94 (1H, s)</td>
<td>C2, C7, C9</td>
</tr>
<tr>
<td>2</td>
<td>145.50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>196.05</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>83.16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>195.69</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>104.99</td>
<td>5.43 (1H,s)</td>
<td>C4, C7,C8</td>
</tr>
<tr>
<td>7</td>
<td>116.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>109.56</td>
<td>6.39 (1H,s)</td>
<td>C6, C7, C9, C10</td>
</tr>
<tr>
<td>9</td>
<td>160.29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10a</td>
<td>42.62</td>
<td>2.57 (1H, dd, 15 &amp; 5)</td>
<td>C11</td>
</tr>
<tr>
<td>10b</td>
<td>42.62</td>
<td>2.51 (1H, dd 15 &amp; 7)</td>
<td>C11</td>
</tr>
<tr>
<td>11</td>
<td>64.03</td>
<td>4.11 (1H,m)</td>
<td>C10, C12</td>
</tr>
<tr>
<td>12</td>
<td>22.87</td>
<td>1.22 (3H, d, 6.2)</td>
<td>C10, C11</td>
</tr>
<tr>
<td>13</td>
<td>27.72</td>
<td>1.43 (3H,s)</td>
<td>C3, C4, C5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C No.</th>
<th>δ (^{13}\text{C}) (ppm)</th>
<th>δ (^{1}\text{H}) (ppm) (area,m, (J_{HH})(Hz))</th>
<th>(^{13}\text{C} - ^{1}\text{H}) HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>153.73</td>
<td>7.89 (1H, s)</td>
<td>C2, C7, C9</td>
</tr>
<tr>
<td>2</td>
<td>142.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>193.05</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>84.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>192.93</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>108</td>
<td>5.55 (1H,s)</td>
<td>C4, C7,C8</td>
</tr>
<tr>
<td>7</td>
<td>115.16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>110.4</td>
<td>6.23 (1H,s)</td>
<td>C6, C7, C9, C10</td>
</tr>
<tr>
<td>9</td>
<td>159.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10a</td>
<td>43.5</td>
<td>2.56 (1H, dd, 15 &amp; 4.5)</td>
<td>C8, C9, C11, C12</td>
</tr>
<tr>
<td>10b</td>
<td>43.5</td>
<td>2.52 (1H, dd 15 &amp; 7.5)</td>
<td>C8, C9, C11, C12</td>
</tr>
<tr>
<td>11</td>
<td>65.7</td>
<td>4.18 (1H,m)</td>
<td>C10, C12</td>
</tr>
<tr>
<td>12</td>
<td>23.8</td>
<td>1.30 (3H, d, 6.2)</td>
<td>C10, C11</td>
</tr>
<tr>
<td>13</td>
<td>22.6</td>
<td>1.55 (3H,s)</td>
<td>C3, C4, C5</td>
</tr>
<tr>
<td>14</td>
<td>170</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>20.5</td>
<td>2.14 (3H, s)</td>
<td>C14</td>
</tr>
</tbody>
</table>
### Table 4.4. Primers used in this study.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’→ 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gw1_11.669BglII for</td>
<td>aaagatctatgtcagactccgaaccacaa</td>
</tr>
<tr>
<td>gw1_11.669Mfel rev</td>
<td>ttcaattgtcatgaccgcggatgtcctga</td>
</tr>
<tr>
<td>A.niger NR-PKS RT for</td>
<td>agcgcgtggaagggcaagcagt</td>
</tr>
<tr>
<td>A.niger NR-PKS RT rev</td>
<td>cgcgeaataacacttctgca</td>
</tr>
<tr>
<td>A.niger HR-PKS RT for</td>
<td>gcggccttcctcaggagaatggc</td>
</tr>
<tr>
<td>A.niger HR-PKS RT rev</td>
<td>tgcctattttgtaaagtgtgac</td>
</tr>
<tr>
<td>Niger AcT-RT2 for</td>
<td>ggaatgcggaggaaatagg</td>
</tr>
<tr>
<td>Niger AcT-RT2 rev</td>
<td>ggaacatctacccgctattg</td>
</tr>
<tr>
<td>Niger P4503-RT2 for</td>
<td>tcgttcaagcctattctcaang</td>
</tr>
<tr>
<td>Niger P4503-RT2 rev</td>
<td>ataccctttcctcaggagaag</td>
</tr>
<tr>
<td>Niger oxidoreductase-RT2 for</td>
<td>taggcacacagggcctttgaa</td>
</tr>
<tr>
<td>Niger oxidoreductase-RT2 rev</td>
<td>tctccgtctgatacttctgc</td>
</tr>
<tr>
<td>Niger FAD1-RT for</td>
<td>gttcagcccgatgctttgata</td>
</tr>
<tr>
<td>Niger FAD1-RT rev</td>
<td>gcgtagggagttgtgggcatg</td>
</tr>
<tr>
<td>Niger FAD2-RT for</td>
<td>cgcgttcaagctttctcaang</td>
</tr>
<tr>
<td>Niger FAD2-RT rev</td>
<td>aggcagcctacggtttttc</td>
</tr>
<tr>
<td>Niger FAD3 RT for</td>
<td>gactttcaatcaggagaagg</td>
</tr>
<tr>
<td>Niger FAD3 RT rev</td>
<td>taatgatggtgacgctttgccc</td>
</tr>
<tr>
<td>Niger MFS1-RT for</td>
<td>atgacactgtcctattctcgc</td>
</tr>
<tr>
<td>Niger MFS1-RT rev</td>
<td>tctccgagactgtagaactg</td>
</tr>
<tr>
<td>Niger MFS2-RT for</td>
<td>atgactgtcactttctcgc</td>
</tr>
<tr>
<td>Niger MFS2-RT rev</td>
<td>acaccaccaggagcacaag</td>
</tr>
<tr>
<td>Niger activator-RT for</td>
<td>gcgtagggtagttgtggtttag</td>
</tr>
<tr>
<td>Niger activator-RT rev</td>
<td>agtagacccctatgctggtagan</td>
</tr>
<tr>
<td>Niger TE-RT for</td>
<td>aaccgtgccgtaaagtttgtt</td>
</tr>
<tr>
<td>Niger TE-RT rev</td>
<td>cgtgacacgagaagccaaa</td>
</tr>
<tr>
<td>Niger epox-RT for</td>
<td>tcatgttccacagcttactgc</td>
</tr>
<tr>
<td>Niger epox-RT rev</td>
<td>caaactgcacatcggacccctc</td>
</tr>
<tr>
<td>Niger Btubulin-RT for</td>
<td>acaactgcacatagtaagggg</td>
</tr>
<tr>
<td>Niger Btubulin-RT rev</td>
<td>tctaacttgccagccagctgtcgg</td>
</tr>
<tr>
<td>Niger HR gpdA P1 for</td>
<td>ggaacaaagctggagctggagctgggtcatttagccatctcaactcaggcttcggttaatggttcggttggtagcactgc</td>
</tr>
<tr>
<td>Niger HR gpdA P1 rev</td>
<td>actgtcgtgacttctgctggcaggctggcagactctgtgaatgtaaagtggttaaatcttcgc</td>
</tr>
<tr>
<td>Niger HR gpdA P3 for</td>
<td>gcccctctctctcctcttcgcctccgcttctccttcgcttctctcctccggttagcgtaatg</td>
</tr>
<tr>
<td>Niger HR gpdA P3 rev</td>
<td>acgttggtaaaccaggcagcgttggatcagttggttggctcggcttcgcttcgcttcgcc</td>
</tr>
<tr>
<td>Zeo cas for</td>
<td>tctagccgctgtccctgcggcag</td>
</tr>
<tr>
<td>Zeo cas rev</td>
<td>catgcggagagaagccaggaagcgg</td>
</tr>
<tr>
<td>Niger AzaH EcoRI for</td>
<td>ggaatctgtcagatcagactcgcggagcaggttggtaggttaaatcttcgccagcttcgc</td>
</tr>
<tr>
<td>Niger AzaH NotI rev</td>
<td>gggcgcgcgattttcctccaccccagccta</td>
</tr>
<tr>
<td>An oxidoreductase Pet23a for</td>
<td>aagaattcatgcgcgccaaataccaga</td>
</tr>
<tr>
<td>An oxidoreductase Pet23a rev</td>
<td>ttgcgccgcctaacgatccagaagaaggg</td>
</tr>
</tbody>
</table>
Figure 4.1. Mass and UV spectra of compounds 1 to 7 (A - F).
Figure 4.2. Tandem MS spectra of 1.
Figure 4.3. Tandem MS spectra of 3.
Figure 4.4. Tandem MS spectra of 4.
Azanigerone A
default proton parameters
Solvent: Chloroform-d

Figure 4.5. 1H NMR Spectrum for 1. Measured in CDCl3 at 500 MHz.
Figure 4.6. 13C NMR Spectrum for 1. Measured in CDCl3 at 125 MHz.
Figure 4.7. Figure S5C. 1H-13C HSQC Spectrum for 1. Measured in CDCl3 at 125MHz.
Figure 4.8. 1H-13C HMBC Spectrum for 1. Measured in CDCl₃ at 125MHz.
Figure 4.9. 1H NMR Spectrum for 2. Measured in methanol-d4 at 500 MHz.
Figure 4.10. 13C NMR Spectrum for 2. Measured in methanol-d4 at 125 MHz.
Figure 4.11. 1H-13C HSQC Spectrum for 2. Measured in methanol-d4 at 125MHz.
Figure 4.12. 1H-13C HMBC Spectrum for 2. Measured in methanol-d4 at 125MHz.
Figure 4.13. 1H NMR Spectrum for 3. Measured in CDCl3 at 500 MHz.
Figure 4.14. 13C NMR Spectrum for 3. Measured in CDCl3 at 125 MHz.
Figure 4.15. 1H-13C HSQC Spectrum for 3. Measured in CDCl3 at 125MHz.
Figure 4.16. 1H-13C HMBC Spectrum for 3. Measured in CDCl3 at 125MHz.
Figure 4.17. 1H NMR Spectrum for 5. Measured in methanol-d4 at 500 MHz.
Figure 4.18. 1H NMR Spectrum for 5. Measured in acetone-d6 at 500 MHz.
Figure 4.19. 13C NMR Spectrum for 5. Measured in methanol-d4 at 125 MHz.
Figure 4.20. $^{13}$C NMR Spectrum for 5. Measured in acetone-$d_6$ at 125 MHz.
Figure 4.21. 1H-13C HSQC Spectrum for 5. Measured in acetone-d6 at 125MHz.
Figure 4.22. 1H-13C HMBC Spectrum for 5. Measured in acetone-d6 at 125MHz.
Figure 4.23. $^{1}H$ NMR Spectrum for 6. Measured in acetone-d6 at 500 MHz.
Figure 4.24. 13C NMR Spectrum for 6. Measured in acetone-d6 at 125 MHz.
Figure 4.25. 1H-13C HSQC Spectrum for 6. Measured in acetone-d6 at 125MHz.
Figure 4.26. 1H-13C HMBC Spectrum for 6. Measured in acetone-d6 at 125MHz.
Figure 4.27. 1H NMR Spectrum for 7. Measured in CDCl3 at 500 MHz.
Figure 4.28. 13C NMR Spectrum for 7. Measured in CDCl3 at 125 MHz.
Figure 4.29. 1H-13C HSQC Spectrum for 7. Measured in CDCl₃ at 125MHz.
Figure 4.30. 1H-13C HMBC Spectrum for 7. Measured in CDCl3 at 125MHz.
**Figure 4.31.** Circular dichroism spectra for azanigerones A-F (1,3-7).
4.2 Supplementary information for Section 2.1.3.

Figure 4.32. 1H NMR Spectrum for 9. Measured in CDCl3 at 500 MHz.
Figure 4.33. $^{13}$C NMR Spectrum for 9. Measured in CDCl3 at 125 MHz.
Figure 4.34. 1H-13C HSQC Spectrum for 9. Measured in CDCl3 at 125MHz.
Figure 4.35. 1H-13C HMBC Spectrum for 9. Measured in CDCl3 at 125MHz.
4.3 Supplementary information for Section 2.2.3.

The information presented here is also part of the supplementary information for the manuscript entitled “Fungal Polyketide Synthase Chain-Length Control by Partnering Hydrolase.”

(A)

(B)

(C)

Figure 4.36. Expression of Bref-PKS and Bref-TH and its mutants. Construction of expression plasmids (A) pAZ94 (Bref-PKS) for *S. cerevisiae* expression and (B) pAZ93 (Bref-TH) for *E. coli* expression. All fragments were amplified from cDNA. (C) The proteins were expressed and purified and the elution samples were run on SDS gel. Bref-PKS was run on 6% acrylamide, while Bref-TH and its mutants were run on 12% acrylamide.
Figure 4.37. MS and UV spectra of compounds 12 to 17 (A-F).
Figure 4.38. C13-labeling of the product polyketides using the MatB regeneration system. (A) HPLC chromatogram of the in vitro reaction with Bref-hydrolase and with base hydrolysis. (B) MS spectra for compounds 12 to 17. Comparison of each MS spectrum to Figure 4.37 shows the +8 mu increase for compounds 12 to 15, and +9mu increase for 16 and 17.
**Figure 4.39.** Effect of base hydrolysis on product release and post-addition of Bref-TH on product release. EIC spectra for specific product masses in the non-hydrolase in vitro reactions with and without base hydrolysis. Base hydrolysis is found to facilitate the release of the polyketides. Addition of equimolar concentration of Bref-TH after 16 hours of in vitro reaction yielded the same product profile. Therefore, in the absence of an immediate releasing partner, the polyketide chain stalled on the ACP, leading to a longer chained product.
Figure 4.40. MS and UV spectra of compounds 18 to 20 (A-C).
Figure 4.41. C13-labeling of the Fma-PKS heptaketide product. (A) HPLC trace and (B) positive ion mass spectra for the in vitro reaction of Fma-PKS with MatB regeneration system for C13-labeling of the heptaketide product, 10. A mass shift of +7 μu was observed.
Figure 4.42. $^1$H spectrum for compound 12. Measured in CD$_3$OD at 500 MHz.
Figure 4.43. $^{13}$C spectrum for compound 12. Measured in CD$_3$OD at 125 MHz.
Figure 4.44. HSQC-INEPT135 spectrum for compound 12.
Figure 4.45. HMBC spectrum for compound 12.
Figure 4.46. COSY spectrum for compound 12.
Figure 4.47. $^1$H spectrum for compound 13. Measured in CD$_3$OD at 500 MHz.
Figure 4.48. $^{13}$C spectrum for compound 13. Measured in CD$_3$OD at 125 MHz.
Figure 4.49. HSQC-INEPT135 spectrum for compound 13.
Figure 4.50. HMBC spectrum for compound 13.
Figure 4.51. $^1$H spectrum for compound 14. Measured in CD$_3$OD at 500 MHz.
Figure 4.52. $^{13}$C spectrum for compound 14. Measured in CD$_3$OD at 125 MHz.
Figure 4.53. HSQC-INEPT135 spectrum for compound 14.
Figure 4.54. HMBC spectrum for compound 14.
Figure 4.55. $^1$H spectrum for compound 15. Measured in CD$_3$OD at 500 MHz.
Figure 4.56. $^{13}$C spectrum for compound 15. Measured in CD$_3$OD at 125 MHz.
Figure 4.57. HSQC-INEPT135 spectrum for compound 15.
Figure 4.58. HMBC spectrum for compound 15.
Table 4.5. Primers used in this study.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5' --&gt; 3')</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primers for cloning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bref hydrointron for EcoRI</td>
<td>aagaattcatgctggtcgagaacgtgc</td>
<td>Bf286 hydrolase for pHis8 EcoRI for</td>
</tr>
<tr>
<td>Bref hydrointron rev 1 NotI</td>
<td>aagcgccgaacctgctggtcgggaacttgc</td>
<td>Bf286 hydrolase for pHis8 NotI rev</td>
</tr>
<tr>
<td>AZ3</td>
<td>ctagecgattataagaatagtatgtaaagtacttagtatgeggcctcaactctct</td>
<td>Bf286 HR-PKS for XW55 P1for</td>
</tr>
<tr>
<td>AZ4</td>
<td>tactcttgacagccaaatatactgagctgcctgtacagcactgtaatctaatgttc</td>
<td>Bf286 HR-PKS for XW55 P1rev</td>
</tr>
<tr>
<td>AZ5</td>
<td>ggttcgcggctagtgtcagctgcgatgacgctgaagcagccagcactgtatatgg</td>
<td>Bf286 HR-PKS for XW55 P2for</td>
</tr>
<tr>
<td>AZ6</td>
<td>cttggccctgtagccgctcactgtgttcgcttggaacactggcagatccacctttccaccc</td>
<td>Bf286 HR-PKS for XW55 P2rev</td>
</tr>
<tr>
<td>AZ7</td>
<td>atcttttaaagtccgctggtcgggaacttgc</td>
<td>Bf286 HR-PKS for XW55 P3for</td>
</tr>
<tr>
<td>AZ8</td>
<td>ggccagatgcggtcgggaacttgc</td>
<td>Bf286 HR-PKS for XW55 P3rev</td>
</tr>
<tr>
<td>AZ9</td>
<td>tggagactgcggtcgggaacttgc</td>
<td>Bf286 HR-PKS for XW55 P4rev</td>
</tr>
<tr>
<td>AZ10</td>
<td>gtaacgcactgctggtcgggaacttgc</td>
<td>Bf286 HR-PKS for XW55 P4rev</td>
</tr>
<tr>
<td>AZ11</td>
<td>tggcactggcactgcgtgcggcctgcctgtcactgcggtctgtaaagtccagcactgtatcctaatgacgcatctggctggggaagctgc</td>
<td>Bf286 HR-PKS for XW55 P5for</td>
</tr>
<tr>
<td>AZ12</td>
<td>atttaataaggtgtgtgtgtgtgtgcgctgtgtgactgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcgtgcggtgcg</td>
<td></td>
</tr>
<tr>
<td>Bf72 HR-75 RTr</td>
<td>cagacggagatctgcacgag</td>
<td>Bref 72 PKS, 600 bp region with intron rev</td>
</tr>
<tr>
<td>Bf130 HR RTf</td>
<td>ctgacgggtctagcatgac</td>
<td>Bref 130 PKS, 600 bp region with intron for</td>
</tr>
<tr>
<td>Bf130 HR RTr</td>
<td>ggctctttaaatctggcat</td>
<td>Bref 130 PKS, 600 bp region with intron rev</td>
</tr>
<tr>
<td>Bf63for RTPCR</td>
<td>agataagcagcagccatcg</td>
<td>Bref 63 PKS, 600 bp region with intron for</td>
</tr>
<tr>
<td>Bf63rev RTPCR</td>
<td>ttgcatagcggcttgaca</td>
<td>Bref 63 PKS, 600 bp region with intron rev</td>
</tr>
<tr>
<td>Bf189 HR RTf</td>
<td>cagacgctagcatctggc</td>
<td>Bref 189 PKS, 600 bp region with intron for</td>
</tr>
<tr>
<td>Bf189 HR RTr</td>
<td>ggattttctccttgatctggaaaccc</td>
<td>Bref 189 PKS, 600 bp region with intron rev</td>
</tr>
<tr>
<td>AZ188</td>
<td>atgaggattctctgcatttg</td>
<td>RT-PCR check on Contig 63 esterase for</td>
</tr>
<tr>
<td>AZ189</td>
<td>ctatgcctgagctatagcccag</td>
<td>RT-PCR check on Contig 63 esterase rev</td>
</tr>
<tr>
<td>AZ190</td>
<td>atgcgcaaccagcttcc</td>
<td>RT-PCR check on Contig 130 thioesterase for</td>
</tr>
<tr>
<td>AZ191</td>
<td>tcaagcttctgaccccttc</td>
<td>RT-PCR check on Contig 130 thioesterase rev</td>
</tr>
</tbody>
</table>

**Bref_286 genes**

| Bref Hydrolase RTf | atgcctgctggaacttgc | RT-PCR Bref_286 hydrolase, 600bp region for |
| Bref Hydrolase RTr | cagcactcggtgtcttc | RT-PCR Bref_286 hydrolase, 600bp region rev |
| AZ30 | atgtgccacatttaagactacactc | E. brefeldianum, 1.7kb P450-3a for |
| AZ31 | tcagcattcctgactacactc | E. brefeldianum, 1.7kb P450-3a rev |
| AZ32 | atgtatcctatcctgattccc | E. brefeldianum, 1.6kb P450-3b for |
| AZ33 | ctatgacccatattctggaacgc | E. brefeldianum, 1.6kb P450-3b rev |
| AZ34 | atggatcagctgactgaaga | E. brefeldianum, 1.5kb P450-4 for |
| AZ35 | tattggggaattggattctgtgc | E. brefeldianum, 1.5kb P450-4 rev |
| AZ36 | atgtgacatttatttattgacttg | E. brefeldianum, 1.5kb P450-5 for |
| AZ37 | cttaactgaacacagcacaacc | E. brefeldianum, 1.5kb P450-5 rev |
| AZ89 | atgcggccgacaacagcct | WD for RT-PCR for |
| AZ90 | ctggtcttgagttgcctgattgg | WD for RT-PCR rev |
| AZ202 | atgtctgactgtgtgtgatagcag | Gene 1 RT in Contig 286 for |
| AZ203 | gtctttaaccagtctgcaagac | Gene 1 RT in Contig 286 rev |
| AZ204 | atgcagggctctggactgacct | Gene 7 RT in Contig 286 for |
| AZ205 | tcaagcacacagcctgcttagc | Gene 7 RT in Contig 286 rev |
| AZ206 | atgcgcacacacccagcttc | Gene 9 RT in Contig 286 for |
5 References


