Natural product biosynthesis and discovery in actinomycetes

Low, Zhen Jie

2018

Low, Z. J. (2018). Natural product biosynthesis and discovery in actinomycetes. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/90198>

<https://doi.org/10.32657/10220/47317>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**Natural Product Biosynthesis and Discovery in
Actinomycetes**

Low Zhen Jie
SCHOOL OF BIOLOGICAL SCIENCES
2018

Natural Product Biosynthesis and Discovery in Actinomycetes

Low Zhen Jie

SCHOOL OF BIOLOGICAL SCIENCES

**A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for the
degree of Doctor of Philosophy**

2018

Acknowledgements

First and foremost, my postgraduate education in Nanyang Technological University wouldn't be possible without the funding and support of Ministry of Education Singapore and NTU. I am deeply grateful for their support and the promotive environment, which allowed me to focus fully on learning and building a scientific career.

Next, I would like to thank my supervisor, Associate Prof. Liang Zhao-Xun for his guidance throughout my time in the lab. During my PhD, we frequently held discussions that invoked me to think more critically as a scientist. My learning wouldn't have been complete without the freedom provided by him. His macro-management style inculcated an independent learning habit which drove me to develop problem-solving skills in the midst of exploring new ideas. Like most novice researchers who were often pre-occupied with the technical details, I was also taught by him to articulate my ideas and thoughts more impactfully. Till today, his mastery of scientific communication remains a skill I hope to demonstrate in future. Adding on, our communication went beyond science, as he frequently advised me on personal issues and my character development. Such is his friendliness that his door is always open (literally) for anyone who may want a word or two (science or life issues) with him. For all these, I am eternally grateful to him as my supervisor, mentor, senior and friend.

Given the cross-disciplinary nature of our work, it wouldn't have been possible for me to do everything within the confines of the lab. Thankfully I was rendered much help and had the wonderful opportunity to work with many collaborators outside our lab. In no sequential order, they are Dr. Juan Xiong (Fudan University), Assistant Prof. Fang Mingliang, Dr. Zhang Yingdan, Assoc. Prof. Yang Liang, Dr. Ding Yichen, July Fong, Assistant Prof. Miao Yansong, Xie Ying, Assoc. Prof. Liu Xuewei, He Jingxi, Dr. Kim Le Mai Hoang, Assoc. Prof. Liu Chuanfa, Dr. Takyi William John, Dr. Yoganathan Kanagasundaram (A*STAR).

Moving on, I would like to express my heartfelt gratitude to my team mate (and soul mate), Pang Li Mei. We started as project partners trying to establish something from scratch and continued as life partners when we wedded this year. We braved whatever challenges in our path and always had each other's back. Instead of generating conflicts, our contrasting characteristics complemented nicely to ensure that things ran

holistically in the lab. She was also my personal 'food panda', supplementing me with meals beyond my typical coffee and bread diet.

Continuing to the rest of the lab, I am deeply grateful to Hoa Tran. She was like my extra pair of hands, sharing some of my experimental workload with the cloning experiments. Her hardworking attitude and willingness to shoulder responsibility smoothened daily operations of the lab.

I am also grateful to Ye Ruijuan, Cheang Qing Wei, Hartono Candra, Xin Lingyi, Howard Saw, Daniel Wong, Melvin Lim Wei Sheng, Lee Jia Jun, Daryn Tan Fu Ern, Rachel Andrea Chea, Sheng Shuo and all other members/ former members of the lab (in no sequential order) for all their support, guidance and assistance in one way or another.

Away from science, I have my family and friends to thank for their love. My parents were always there for me and my brother always offered a listening ear whenever I needed. My close friends also offered great support and took my mind off work whenever I was slipping towards despondency.

Overall, this thesis is merely a summation of a piece of work, that is not just my achievement, but a dedication to those who had helped made this possible, including the helpful administrative staff of my school. Thank you everyone for being the wind beneath my wings as I seek to soar through the skies.

Table of Contents

Acknowledgements	III
Table of Contents	V
List of Figures	X
List of Tables	XIII
List of Supplementary Figures	XIV
List of Supplementary Tables	XVII
Abbreviations	XVIII
Abstract	XXI
Chapter 1: Introduction	1
1.1. Natural products and its importance	1
1.2. Actinomycetes as a source of natural products	3
1.2.1. Types of natural products from actinomycetes	4
1.2.1.1. Polyketide Synthase (PKS)-derived natural products	5
1.2.1.2. Nonribosomal Peptide Synthase (NRPS)-derived Natural Products	10
1.2.1.3. Terpenoid natural products	12
1.2.1.4. Ribosomally synthesized post-translationally modified peptides (RiPPs)	14
1.3. Renewal of interest in actinomycete natural product discovery in the Post-Genomic Era	16
1.4. Approaches to discovery of natural product in actinomycetes	17
1.4.1. Approaches to screen for natural products	17
1.4.2. Traditional approaches to activation of natural product biosynthesis in actinomycetes	20
1.4.2.1. Chemical elicitation	21
1.4.2.2. Co-cultivation	23
1.4.2.3. Culture condition screening	24
1.4.3. Genomics guided approaches to natural product biosynthesis and discovery in actinomycetes	25
1.4.3.1. Synthetic biology tools to study BGCs in vivo	28
1.5. Aims and Objectives	38
2. Chapter 2: Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a <i>Streptomyces</i> strain isolated from mangrove sediment	40
2.1. Introduction	40
2.2. Methods and materials	43
2.2.1. <i>Streptomyces</i> sp. SD85 isolation and strain maintenance	43

2.2.2. Genomic DNA isolation from <i>Streptomyces</i> sp. SD85	43
2.2.3. Genome sequencing of <i>Streptomyces</i> sp. SD85	44
2.2.4. Metabolite Profiling of <i>Streptomyces</i> sp. SD85.....	45
2.2.5. Large scale fermentation of <i>Streptomyces</i> sp. SD85 and isolation of sceliphrolactam	45
2.2.6. Construction of <i>Streptomyces</i> sp. SD85 Δ sceN mutants via CRISPR/Cas9 system.....	46
2.2.7. CRISPR/Cas9 mediated sceQ-R gene fusion	47
2.2.8. HPLC analysis of sceliphrolactam production for the <i>Streptomyces</i> sp. SD85 Δ sceN and sceQ-R fusion mutants	48
2.2.9. Primers used in this study.....	48
2.3. Results and discussion	48
2.3.1. <i>Streptomyces</i> sp. SD85 produces sceliphrolactam and other secondary metabolites under laboratory cultivation conditions.....	48
2.3.2. Complete genome sequencing suggests that <i>Streptomyces</i> sp. SD85 is a biosynthetically talented strain	50
2.3.3. Identification of sceliphrolactam BGC by gene inactivation	53
2.3.4. Annotation of non-PKS biosynthetic genes in the sce gene cluster	56
2.3.5. Analysis of the Sce Type I modular PKS	59
2.3.6. Evidence disfavours involvement of a <i>trans</i> -PKS module in sceliphrolactam biosynthesis	62
2.3.7. Sceliphrolactam biosynthesis potentially involves an iterative PKS module	63
2.4. Conclusion	64
Chapter 3: Discovery of meijiemycin, a novel linear polyene-polyol antifungal compound from <i>Streptomyces</i> sp. SD50.....	66
3.1. Introduction.....	66
3.2. Methods and materials	68
3.2.1. <i>Streptomyces</i> sp. SD50 isolation and strain maintenance	68
3.2.2. Genomic DNA isolation from <i>Streptomyces</i> sp. SD50	68
3.2.3. Metabolite profiling of <i>Streptomyces</i> sp. SD50.....	69
3.2.4. Genome sequencing of <i>Streptomyces</i> sp. SD50.....	70
3.2.5. Construction of <i>Streptomyces</i> sp. SD50 Δ nanA1 Δ ozmH double mutant, incapable of producing nanchangmycin and oxazolomyin to facilitate screening of novel compounds.....	70
3.2.6. Construction of <i>Streptomyces</i> sp. SD50 Δ mjmB and Δ mjmE mutants using <i>Streptomyces</i> sp. SD50 Δ nanA1 Δ ozmH double mutant to validate relationship between <i>mjm</i> BGC and meijiemycin	71

3.2.7. HPLC analysis of meijiemyacin production	71
3.2.8. Large scale fermentation of <i>Streptomyces</i> sp. SD50 for isolation and purification of meijiemyacin	72
3.2.9. Characterization of meijiemyacin	73
3.2.10. Determination of minimum inhibitory concentration (MIC) of meijiemyacin against fungal strains	74
3.2.11. Primers used in this study	74
3.3. Results and discussion	75
3.3.1. <i>Streptomyces</i> sp. SD50 is phylogenetically related to <i>Streptomyces bingchenggensis</i> and possesses immense biosynthetic potential for secondary metabolism.....	75
3.3.2. Metabolite Profiling of <i>Streptomyces</i> sp. SD50 mutants reveals peak containing hexaene UV fingerprint	77
3.3.3. Optimization of meijiemyacin yield and validation of its biosynthetic origin through CRISPR/Cas9 mediated gene inactivation.....	79
3.3.4. Structural elucidation of meijiemyacin, a novel linear polyene natural product.....	80
3.3.5. Analysis of meijiemyacin BGC	82
3.3.6. Preliminary fungal inhibition assay shows narrow-spectrum activity against <i>Candida albicans</i>	87
3.4. Conclusion	88
Chapter 4: Genomics-guided discovery of novel natural products from a <i>Micromonospora</i> strain.....	89
4.1. Introduction	89
4.2. Methods and materials	91
4.2.1. <i>Micromonospora</i> sp. MD118 isolation and strain maintenance	91
4.2.2. Bioactivity screening of <i>Micromonospora</i> sp. MD118 via agar overlay assay.....	91
4.2.3. Construction of pSK152 and pHZ-cod9.....	92
4.2.4. Genomic DNA isolation from <i>Micromonospora</i> sp. MD118	92
4.2.5. Genome sequencing of <i>Micromonospora</i> sp. MD118	92
4.2.6. Construction of gene overexpression mutants for cluster activation	93
4.2.6.1. Construction of <i>Micromonospora</i> sp. MD118A <i>mmd2050</i> gene overexpression mutants to validate functionality of pSK152	93
4.2.6.2. Construction of <i>Micromonospora</i> sp. MD118A7, a <i>lpmR1R2X</i> triple gene overexpression mutant, and <i>Micromonospora</i> sp. MD118AE, a <i>sgdR2R7</i> double gene overexpression mutant	93
4.2.7. Metabolite profiling of <i>Micromonospora</i> sp. MD118	94

4.2.8. Construction of <i>Micromonospora</i> sp. MD118 mutants using CRISPR/Cas9 system.....	95
4.2.8.1. Construction of <i>Micromonospora</i> sp. MD118A, a $\Delta mmd2050$ mutant to validate CRISPR/Cas9 functionality in the strain	96
4.2.8.2. Construction of $\Delta lpmB$ mutant to establish relationship between <i>lpm</i> BGC and lumpurmycin	96
4.2.8.3. Construction of $\Delta sgdX1$, $\Delta sgdX2$, $\Delta sgdX3$ and $\Delta sgdU15$ single mutants for <i>in vivo</i> study of sungeidine biosynthesis.....	96
4.2.9. HPLC analysis of lumpurmycin and sungeidine analogs.....	97
4.2.10. Large scale fermentation of <i>Micromonospora</i> sp. MD118A for compound isolation.....	97
4.2.10.1. Lumpurmycin	98
4.2.10.2. Sungeidine.....	98
4.2.10.3. Sungeidine intermediate (A1).....	98
4.2.11. Characterization of compounds.....	99
4.2.11.1. Lumpurmycin	99
4.2.11.2. Sungeidine.....	100
4.2.11.3. Sungeidine intermediate (A1).....	101
4.2.12. Plasmids constructed and primers used in this study.....	101
4.3. Results and discussion	102
4.3.1. <i>Micromonospora</i> sp. MD118 possesses a distinctive enediyne BGC and other uncharacterized BGCs.	102
4.3.2. Generic tools developed for <i>Streptomyces</i> are compatible in <i>Micromonospora</i> sp. MD118 for genetic manipulation	103
4.3.3. Modified vectors for efficient genetic manipulation of <i>Micromonospora</i> sp. MD118	106
4.3.3.1. pHZ-cod9, a modified pCRISPR-Cas9 plasmid with counterselectable marker facilitates selection of plasmid free mutants.....	106
4.3.3.2. pSK152, an integrative and conjugative plasmid with synthetic modular elements for gene overexpression.....	107
4.3.4. Discovery of lumpurmycin	110
4.3.4.1. Constitutive overexpression of three genes <i>lpmR1</i> , <i>R2</i> and <i>X</i> elicited the production of lumpurmycins.....	110
4.3.4.2. Structural elucidation of lumpurmycin, a novel benzoquinone ansamycin ..	113
4.3.4.3. Proposed biosynthetic mechanism for lumpurmycin	115
4.3.5. Discovery of sungeidine	121
4.3.5.1. Disruption of a <i>unbL</i> homologous gene, <i>sgdU15</i> , abolished production of sungeidine.....	121
4.3.5.2. Constitutive overexpression of <i>sgdR2</i> and <i>sgdR7</i> upregulated <i>sgd</i> BGC...	122

4.3.5.3. structural elucidation of sungeidine and A1	124
4.3.5.4. <i>sgd</i> BGC is unique from other identified anthraquinone-fused enediyne BGCs	125
4.3.5.5. Analysis of mutants revealed that SgdX2 is crucial for formation of sungeidine	128
4.4. Conclusion	131
Chapter 5: Future work.....	133
References.....	143
Appendix	167
Supplementary tables and figures for chapter 2	167
Supplementary tables and figures for chapter 3	176
Supplementary tables and figures for chapter 4	192
Publications.....	219

List of Figures

Figure 1: Sources of small molecule drugs from 1981 to 2010.	2
Figure 2: Domain architecture and mechanism of different Type I PKSs.	9
Figure 3: Domain architecture and mechanism of Type II and Type III PKSs.	10
Figure 4: NRPS mechanism and NRPS-derived natural products.	11
Figure 5: A schematic of the terpenoid biosynthetic pathway.....	13
Figure 6: A schematic of generic mechanism involved in RiPP biosynthesis.	15
Figure 7: Strategies to screen for novel compounds.	18
Figure 8: Traditional approaches to activation of natural product biosynthesis in actinomycetes.	21
Figure 9: Approaches to natural product discovery in actinomycetes.	25
Figure 10: A schematic of data generated by antiSMASH analysis of an actinomycetal genome.	27
Figure 11: Two approaches to genetic manipulation in actinomycetes.	31
Figure 12: Strategies for activating a target BGC <i>in vivo</i>	32
Figure 13: Reporter-based screening to identify conditions that activate targeted BGCs.....	35
Figure 14: Strategy for heterologous expression of BGCs.	37
Figure 15: Polyene macrolactams that contain a nitrogen-containing moiety derived from L-glutamate via a β -amino acid starter unit.	42
Figure 16: Growth of <i>Streptomyces</i> sp. SD85 on solid medium.	49
Figure 17: Complete genome of <i>Streptomyces</i> sp. SD85 and the PKS-based biosynthetic gene clusters (BGCs) with confirmed products.	54
Figure 18: Inactivation of sceliphrolactam biosynthetic gene <i>sceN</i> using CRISPR/Cas9-based method.	56
Figure 19: Proposed biosynthetic mechanism for sceliphrolactam.....	61
Figure 20: CRISPR/Cas9 mediated fusion of <i>sceQ</i> and <i>sceR</i>	63
Figure 21: Meijiemycin and other related polyene natural products.	67
Figure 22: Genome map of <i>Streptomyces</i> sp. SD50 annotated with secondary metabolite BGCs.	76
Figure 23: Metabolite profiling of <i>Streptomyces</i> sp. SD50 mutants.....	78
Figure 24: Meijiemycin biosynthetic origin and yield optimization.	80

Figure 25: Summary of ^1H - ^1H COSY and key HMBC NMR correlations observed for meijiemycin.	81
Figure 26: Summary of diagnostic proton-proton coupling constants (red) and NOE correlations (blue arrows).	82
Figure 27: Proposed biosynthetic mechanism of meijiemycin (Part 1)..	84
Figure 28: Proposed biosynthetic mechanism of meijiemycin (Part 2).	85
Figure 29: Genome map of <i>Micromonospora</i> sp. MD118 and generation of <i>Micromonospora</i> sp. MD118A, a $\Delta mmd2050$ mutant.	105
Figure 30: plasmid map of pHZ-cod9.	107
Figure 31: Validation of pSK152, an integrative and conjugative plasmid for gene overexpression.	109
Figure 32: Discovery of lumpurmycin through <i>lpmR1R2X</i> overexpression and culture condition screening	112
Figure 33: Lumpurmycin and other notable ansamycin natural products.	113
Figure 34: Diagnostic ^1H - ^1H COSY and HMBC correlations observed for lumpurmycin.	114
Figure 35: Summary of diagnostic proton-proton coupling constants and NOE correlations observed for lumpurmycin.	114
Figure 36: Proposed biosynthesis of lumpurmycin (part 1).	116
Figure 37: Proposed biosynthesis of lumpurmycin (part 2).	118
Figure 38: Sungeidine and other 10-membered enediyne natural products. Enediyne cores are depicted in blue.	121
Figure 39: Identification of sungeidine BGC and metabolite profiling of associated mutants.	124
Figure 40: 2D NMR correlations of sungeidine and A1.	125
Figure 41: Comparison of <i>sgd</i> BGC with other homologous enediyne BGCs.	126
Figure 42: Comparison of HPLC chromatogram of mutants and identification of <i>sgd</i> BGC associated peaks.	129
Figure 43: Proposed biosynthetic mechanism of sungeidine.	131
Figure 44: Construction of pBACS3, a <i>Streptomyces</i> - <i>E. coli</i> shuttle vector for genomic DNA library construction..	135
Figure 45: Construction of pSTAR3, a tripartite shuttle vector for TAR cloning and heterologous expression in <i>Streptomyces</i>	136

Figure 46: Examination of substrate specificity of AT domains in sce BGC through sequence alignment.	137
Figure 47: Possible scenarios resulting from altered substrate specificity of SceQ-AT and SceR-AT domains.	138
Figure 48: Organization of putative C40 BGC.	140
Figure 49: Organization of putative C45 BGC..	141

List of Tables

Table 1: Comparison of approaches for microbial natural product discovery	20
Table 2: AntiSMASH-predicted BGCs for <i>Streptomyces</i> sp. SD85	52
Table 3: Predicted function of the genes from the sceliphrolactam BGC.	57
Table 4: MIC values from preliminary fungal inhibition assay.....	87

List of Supplementary Figures

Figure S 1. Prediction of stereochemistry of β -hydroxyl group and olefin structures through KR fingerprint.	168
Figure S 2: Verification of sequence boundaries between sceP to sceS.	169
Figure S 3. ^1H NMR spectrum of sceliphrolactam in DMSO- d_6 , 500 MHz.....	170
Figure S 4. ^{13}C NMR spectrum of sceliphrolactam in DMSO- d_6 , 500 MHz.....	170
Figure S 5. COSY spectrum of sceliphrolactam in DMSO- d_6 , 500 MHz.	171
Figure S 6. TOCSY spectrum of sceliphrolactam in DMSO- d_6 , 500 MHz.	171
Figure S 7. HSQC spectrum of sceliphrolactam in DMSO- d_6 , 500 MHz.	172
Figure S 8. HMBC spectrum of sceliphrolactam in DMSO- d_6 , 500 MHz.....	172
Figure S 9. 1D ^1H NOESY spectrum of sceliphrolactam (DMSO- d_6 , 500 MHz) with selective excitation of H-15 to establish the geometry of olefin connecting H-15 and H-14.	173
Figure S 10. Mass spectrum of sceliphrolactam.....	173
Figure S 11. Analysis of docking domains and effect of sceR-Q gene fusion on sceliphrolactam biosynthesis.....	174
Figure S 12: Inactivation of <i>nanA1</i> gene.	178
Figure S 13: Inactivation of <i>ozmH</i> gene.	179
Figure S 14: Inactivation of <i>mjmB</i> gene.....	179
Figure S 15: Inactivation of <i>mjmE</i> gene.....	180
Figure S 16: ^1H NMR spectrum (400 MHz, in DMSO- d_6) of meijiemyacin.	182
Figure S 17: ^{13}C NMR spectrum (400 MHz, in DMSO- d_6) of meijiemyacin.	183
Figure S 18: ^1H - ^1H COSY NMR spectrum (400 MHz, in DMSO- d_6) of meijiemyacin.	183
Figure S 19: HSQC NMR spectrum (400 MHz, in DMSO- d_6) of meijiemyacin.	184
Figure S 20: HMBC NMR spectrum (500 MHz, in DMSO- d_6) of meijiemyacin.	185
Figure S 21: NOESY NMR spectrum (400 MHz, in DMSO- d_6) of meijiemyacin.	186
Figure S 22: ^1H NMR spectrum (400 MHz, in $\text{C}_5\text{D}_5\text{N}$) of meijiemyacin.....	187
Figure S 23: ^1H - ^1H COSY NMR spectrum (400 MHz, in $\text{C}_5\text{D}_5\text{N}$) of meijiemyacin. ...	187
Figure S 24: High resolution mass spectrum of meijiemyacin.	188
Figure S 25: AT domain analysis of Mjm PKS.....	189
Figure S 26: KR domain analysis of Mjm PKS.	190
Figure S 27: DH domain analysis of Mjm PKS.	191
Figure S 28: Agar overlay bioactivity assay of <i>Micromonospora</i> sp. MD118.....	192

Figure S 29: <i>Micromonospora</i> sp. MD118 grown on 6 different agar media.	199
Figure S 30: (-)-HRMS of lumpurmycin.	199
Figure S 31: ¹ H NMR spectrum (400 MHz, in C ₅ D ₅ N) of lumpurmycin.	201
Figure S 32: ¹³ C NMR spectrum (400 MHz, in C ₅ D ₅ N) of lumpurmycin.	201
Figure S 33: ¹³ C DEPT-135 NMR (400 MHz, in C ₅ D ₅ N) of lumpurmycin.	202
Figure S 34: HSQC NMR spectrum (400 MHz, in C ₅ D ₅ N) of lumpurmycin.	202
Figure S 35: COSY NMR spectrum (400 MHz, in C ₅ D ₅ N) of lumpurmycin.	203
Figure S 36: HMBC NMR spectrum (400 MHz, in C ₅ D ₅ N) of lumpurmycin.	203
Figure S 37: NOESY NMR spectrum (400 MHz, in C ₅ D ₅ N) of lumpurmycin.	204
Figure S 38: ROESY NMR spectrum (400 MHz, in C ₅ D ₅ N) of lumpurmycin.	204
Figure S 39: ¹ H NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of lumpurmycin.	205
Figure S 40: ¹³ C NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of lumpurmycin.	205
Figure S 41: HMBC NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of lumpurmycin.	206
Figure S 42: Inactivation of <i>lpmB</i> gene.	206
Figure S 43: AT domain analysis of Lpm PKS.	207
Figure S 44: KR domain analysis of Lpm PKS.	208
Figure S 45: DH domain analysis of Lpm PKS.	209
Figure S 46: Inactivation of <i>sgdU15</i> and <i>sgdX3</i> genes.	211
Figure S 47: Inactivation of <i>sgdX1</i> gene.	211
Figure S 48: Inactivation of <i>sgdX2</i> gene.	212
Figure S 49: ¹ H NMR spectrum (700 MHz, in DMSO- <i>d</i> ₆) of sungeidine.	213
Figure S 50: ¹³ C NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of sungeidine.	214
Figure S 51: HSQC NMR spectrum (700 MHz, in DMSO- <i>d</i> ₆) of sungeidine.	214
Figure S 52: TOCSY NMR spectrum (700 MHz, in DMSO- <i>d</i> ₆) of sungeidine.	215
Figure S 53: HMBC NMR spectrum (700 MHz, in DMSO- <i>d</i> ₆) of sungeidine.	215
Figure S 54: NOESY NMR spectrum (700 MHz, in DMSO- <i>d</i> ₆) of sungeidine.	216
Figure S 55: High resolution mass spectrum of sungeidine.	216
Figure S 56: ¹ H NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of A1.	217
Figure S 57: ¹³ C NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of A1.	218
Figure S 58: HSQC NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of A1.	218
Figure S 59: COSY NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of A1.	219
Figure S 60: HMBC NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of A1.	219
Figure S 61: ROESY NMR spectrum (400 MHz, in DMSO- <i>d</i> ₆) of A1.	220
Figure S 62: High resolution mass spectrum of A1.	220

Figure S 63: High resolution MS/MS of A1 with collision induced dissociation set at 40 eV.....	220
--	-----

List of Supplementary Tables

Table S 1: Primers and sequences used for study discussed in Chapter 2.....	167
Table S 2: ^1H and ^{13}C NMR data for sciliphrolactam (DMSO- d_6 , 500 MHz)	174
Table S 3: Primers and sequences used in chapter 3	176
Table S 4: Secondary metabolite gene clusters in <i>Streptomyces</i> sp. SD50 genome	177
Table S 5: ^1H and ^{13}C NMR spectroscopic data of meijiemycin.	180
Table S 6: Genes involved in meijiemycin biosynthesis and their putative functions	188
Table S 7: List of medium recipes for fermentation of <i>Micromonospora</i> sp. MD118	192
Table S 8: List of plasmids	193
Table S 9: List of primers used in this study	193
Table S 10: List of sequences used in this study.	195
Table S 11: List of putative BGCs identified from <i>Micromonospora</i> sp. MD118.	196
Table S 12: List of genes in <i>lpm</i> BGC and their proposed functions.	197
Table S 13: ^1H and ^{13}C NMR spectroscopic data of lumpurmycin (in $\text{C}_5\text{D}_5\text{N}$, 400 MHz) ^a	199
Table S 14: List of genes in <i>sgd</i> BGC and their proposed functions.	209
Table S 15: NMR data of sungeidine (in DMSO- d_6) ^a	212
Table S 16: NMR data of A1 (in DMSO- d_6 , 400 MHz) ^a	216

Abbreviations

3-meAsp	(2S, 3S)-3-methylaspartate
5' UTR	5' untranslated region
5FC	5-fluorocytosine
5FU	5-Fluorouracil
A	Adenylation
ACN	acetonitrile
ACP	Acyl Carrier Protein
AHBA	3-amino-5-hydroxybenzoic acid
aminoDAHP	3,4-dideoxy-4-amino-D-arabino-heptulosonic acid 7-phosphate
aminoDHQ	5-deoxy-5-aminodehydroquinic acid
aminoDHS	5-deoxy-5-aminodehydroshikimic acid
antiSMASH	antibiotics & Secondary Metabolite Analysis Shell
AT	Acyltransferase
BAC	Bacterial Artificial Chromosome
BGC	Biosynthetic Gene Cluster
BioMAP	Broad classification through Antibiotic mode of action profile
bp	base pair
C	Condensation
CATCH	Cas9-assisted targeting of chromosome
CFU	Colony Forming Unit
CLF	Chain Length Factor
CoA	Coenzyme A
CodA	Cytosine Deaminase
COSY	Correlation spectroscopy
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTAB	Cetyltrimethylammonium bromide
DEPT	distortionless enhancement of polarisation transfer
DH	Dehydratase
DMAPP	Dimethylallyl pyrophosphate
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
Dyn	dynemicin A
fas	fatty acid synthase
<i>g</i>	force of gravity
g	gram
GBL	γ -butyrolactone
gDNA	genomic DNA
GlcNAc	N-acetylglucosamine
HDR	homology directed repair
HGAP2	Hierarchical Genome Assembly Process 2
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High Performance Liquid Chromatography

hr	hour
HSQC	Heteronuclear Single Quantum Coherence
IE4P	imino-erythrose-4-phosphate
IPP	Isoprenyl pyrophosphate
ISP	International Streptomyces Project
KAIST	the Korea Advanced Institute of Science and Technology
KR	Ketoreductase
KS	β -Ketosynthase
LAP	Linear azol(in)e-containing peptide
LB	Luria-Bertani
LC-HRMS	Liquid chromatography coupled with high-resolution mass spectrometry
LLHR	Linear plus linear homologous recombination
Lpm	Lumpurmycin
M	Molar
m	metre
MeOH	Methanol
MEP	Non-mevalonate
MIBig	Minimum Information about a Biosynthetic Gene cluster
MIC	Minimum Inhibitory Concentration
min	minute
Mjm	Meijiemycin
MRSA	Methicillin-resistant <i>Staphylococcus aureus</i>
MVA	mevalonate
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NRPS	Nonribosomal Peptide Synthase
OD	Optical Density
OSMAC	One Strain Many Compounds
PAM	protospacer adjacent motif
PCP	Peptidyl carrier protein
PCR	Polymerase Chain Reaction
PKS	Polyketide Synthase
PKSE	enediyne PKS
PPTase	4' Phosphopantetheinyl Transferase
PVP	Polyvinylpyrrolidone
RBS	ribosomal binding site
RiPP	Ribosomally synthesized post-translationally modified peptides
ROI	Reads of inserts
RPM	Revolutions per minute
SDS	Sodium dodecyl sulfate
SMRT	Single Molecule Real Time
sp.	species
SPE	Solid phase extraction
Sgd	sungeidine

<i>t</i>	time
TAR	Transformation associated Recombination
TE	Thioesterase
Tnm	Tiancimycin A
TOCSY	Total Correlation Spectroscopy
TOMM	thiazole/oxazole-modified microcins
Ucm	uncialamycin
UDP	Uridine diphosphate
v/v	volume/volume
w/v	weight/volume
WT	Wild Type
YAC	Yeast Artificial Chromosome
Ypm	yangpumicin A
ZMW	zero-mode wavelength

Abstract

Natural products remain a valuable source of drug leads in the 21st Century. Revival of actinomycete natural product discovery in the Post Genomics Era is driven by realization of largely concealed biosynthetic potential revealed through genome mining efforts. Implementation of bioinformatics also prompted a paradigm shift in natural product discovery as a bottom up approach guided by genomics directs investigative efforts at specific biosynthetic gene clusters (BGCs). In this work, a genomics guided approach was adopted to study natural product discovery and biosynthesis in three actinomycete strains – 2 *Streptomyces* strains and 1 *Micromonospora* strain. Effectiveness of this approach was underlined by successful dereplication of known clusters and prioritization of novel gene clusters. Genome mining was combined with CRISPR/Cas9, a powerful tool for genetic engineering, to identify four novel BGCs through gene knockout studies *in vivo*. Development of other synthetic biology tools also facilitated the upregulation of cryptic BGCs, through constitutive overexpression of positive regulators, for compound discovery. Whilst one of the four identified BGCs encoded a known polyene macrolactam (sceliphrolactam), it piqued the possible involvement of an iterative, partially reducing polyketide synthase module, which is unprecedented. One of the three novel compounds discovered was meijiemyacin, a novel linear polyene that exhibited antifungal properties. The prospect of bioengineering its BGC holds promise in generating other structural analogs with improved therapeutic properties. The other two novel compounds were discovered from the *Micromonospora* strain, reflecting the promise of unlocking the immense biosynthetic potential displayed by rare actinomycetes using modern synthetic biology approaches. Discovery of lumpurmycin, the largest ansamycin natural product identified till date, adds onto a growing family of benzoquinone ansamycin natural products that are antagonistic against heat shock protein 90. The final compound, sungeidine, was derived from a novel 10-membered enediyne BGC, whose biosynthetic pathway is currently under investigation. Overall, this work demonstrated the potential of actinomycetes as a continued source of novel natural products and more importantly, established a methodical approach to uncover novel natural products in other actinomycete strains.

Chapter 1: Introduction

1.1. Natural products and its importance

Natural products typically refer to compounds synthesized by organisms in nature. They could be broadly divided into two categories – primary metabolites and secondary metabolites. Primary metabolites are vital compounds that are involved in the metabolic pathways of an organism, which are essential for sustaining life. These metabolic pathways can be categorized into anabolic or catabolic pathways that enable efficient physiological functioning of an organism through balance of energy and metabolic fluxes. One such example is fatty acid, which serves as precursors in both catabolic and anabolic pathways. It can be catabolized via β -oxidation¹ into eventual acetyl-CoA substrates that enter the citric acid cycle to generate energy in the form of ATP.² On the other hand, secondary metabolites refer to molecules that are not typically involved in sustaining proper physiological functions in an organism. Instead, these molecules participate in other cellular processes such as virulence,³ cell-cell communication⁴⁻⁵ and self-defence⁶ to render the host some competitive edge against other organisms in the natural environment.

Throughout the history of Mankind, secondary metabolites have always been relied upon as therapies against a range of diseases. Today, even with the knowledge and technological advances made in synthetic chemistry, the significance of natural products in drug development is still unmatched by synthetic compounds (**FIGURE 1**) This is despite commercially available chemical space being 2 orders of magnitude larger than natural product chemical space.⁷ The biological relevance of natural product chemical space is unsurprising given nature's capability in synthesizing molecules with novel chemical scaffolds that are hard to conceive by combinatorial chemistry. It is also worth noting that these remarkable chemical scaffolds from Nature are derived from years of evolution that have enhanced their functionality as bioactive compounds, an intrinsic process that is very much akin to the application of combinatorial chemistry in drug discovery.⁸ Even so, these natural products were evolved to suit the host's requirements as opposed to Mankind's needs. Hence, structural modifications are necessary to improve their pharmacological properties and repurpose them as clinical drugs.⁹

Figure 1: Sources of small molecule drugs from 1981 to 2010. Apart from 36 % of the drugs that are conceived without natural product conception, all the remaining drugs are either natural products or synthesized with natural product conception. N = natural product, ND = natural product derived, S* = synthetic compound with natural product pharmacophore, S*/NM = Mimic of natural product which is synthesized with natural product pharmacophore, S/NM = Mimic of natural product which is synthesized totally, S = totally synthetic compound with no natural product conception.

Despite substantial evidence demonstrating the effectiveness of natural products as leads in drug discovery, the fraction of FDA-approved new molecular entities being presented by natural products has been on a decline since 1930s.¹⁰ This is partly due to a lack of emphasis on natural product screening, which requires tedious sample preparation and dedicated procedures for different samples. Despite the increased throughput workflow adopted by pharmaceutical companies, the Golden Era of Antibiotic discovery (1950s to 1970s) ended as constant rediscovery of known compounds and absence of new natural products prompted them to abandon natural product screening in pursuit of chemical synthesis for generation of drug leads. However, in recent years, success of natural products in drug development has persuaded pharmaceutical companies to revert their attention back to natural products for drug leads. The identification of new natural products and subsequent construction of these natural product libraries will significantly expand the natural product chemical space for drug discovery.¹¹⁻¹² Though all organisms can produce natural products,

their ability to biosynthesize secondary metabolites differs at taxonomic level. Consequently, the influence they possess on discovery of new drug leads is dissimilar. A study in 2008 revealed that nearly half of the natural products in drug development pipeline were of plant origin, whereas microbial natural products accounted for merely 14%.¹³ While the role of microbial natural products in drug development appeared less impressive, it should be noted that their role is downplayed by the fact that chemical space in microbes is vastly hidden as most biosynthetic gene clusters (BGCs) encoding for secondary metabolites are cryptic in laboratory conditions. Furthermore, nearly 99% of microbes are uncultivable under normal laboratory conditions,¹⁴ which restricts the supply of microbial natural products that are truly accessible at the moment. Exploring unculturable microbes is a way of potentially uncovering new natural products with therapeutic values, as shown in a recent study where a new antibiotic called teixobactin was discovered.¹⁵ Likewise, activation of these cryptic BGCs can also significantly expand the natural product chemical space. This revival of interest in natural products is poised to reinforce its importance in the field of drug discovery.

1.2. Actinomycetes as a source of natural products

Actinomycetales, commonly known as actinomycetes, are an order of *Actinobacteria* and possess DNA with high GC content. These Gram Positive bacteria are prolific producers of bioactive secondary metabolites that are widely recognized for their contributions in the field of antibiotics.¹⁶ During the Golden age of Antibiotic discovery, nearly 60% of the new antibiotics were discovered from *Streptomyces*, a genus of the actinomycetes.¹⁷ For example, tetracycline, an important drug against microbial infections like MRSA, was derived from streptomycetes. Another notable antibiotic which is used to treat *Clostridium difficile* and other multiple drug resistant Gram-Positive bacterial infections is daptomycin. Antifungal polyene drugs like amphotericin, candicidin and nystatin were also discovered from actinomycetes. Apart from antibiotics, secondary metabolites from Actinomycetes also contributed to other fields of Medicine. Ansamycin compounds like geldanamycin had shown promise as potential cancer therapeutics while polyether compounds like salinomycin had shown activity against insect-borne *Flaviviridae* viruses like Zika and Dengue. Even as research in actinomycetes decreased in recent years, natural products derived from

actinomycetes still account for nearly one-third of the entire microbial metabolites with antibiotic properties. As mentioned previously, the decline in bioprospecting of actinomycetes by pharmaceutical companies is partly due to the repeated discovery of known natural products from isolated actinomycetes. Far from being exhausted, the vast number of hidden BGCs and unidentified actinomycetes necessitates a paradigm shift in the approach undertaken for bioprospecting actinomycetes. Strategies for natural product discovery in actinomycetes can be revamped on two levels – strain level and BGC level. On the strain level, despite having an estimate of 10^{26} actinomycete strains in the top 10 cm of soil covering earth, only about 10^7 of actinomycete strains had been surveyed for natural products.¹⁸ This statistic suggests that a significant fraction of actinomycetes that can potentially contribute to natural product discovery remained unknown. Given that evolution diversifies secondary metabolism,¹⁹ isolation of novel actinomycetes is equally significant in facilitating the discovery of new natural products. Instead of systematic isolation of actinomycetes from soil samples using conventional media recipes, researchers started seeking new isolation strategies that could preferentially recover rare and novel actinomycete strains. Pre-treatment methods such as microwave irradiation²⁰ was developed to reduce the recovery of non-actinomycete microbes while enhancing the recovery of slow-growing actinomycetes. Development of specific isolation culture media also facilitated the isolation of rare actinomycetes that might require unique trace chemicals to grow.²¹⁻²² Exploration of new habitats like marine sediments and caves also uncovered rare actinomycetes that had evolved to adapt to their natural environments.²³⁻²⁴ *In situ* trapping and cultivation of filamentous actinomycetes also eased the adaptation of the isolated strains to laboratory cultivation conditions, hence enabling isolation of actinomycetes that were otherwise, neglected by typical direct plating strategy.²⁵

1.2.1. Types of natural products from actinomycetes

Actinomycetes produce diverse natural products, which are synthesized by enzymes encoded by genes that are typically clustered together as a BGC. Hence, a typical BGC is responsible for the production of a single secondary metabolite, although there are exceptions whereby aberrant enzymatic reactions resulted in biosynthesis of by-products.²⁶ BGCs also contain genes for export of synthesized natural products and

immunity to the host organism. Aided by development in genomics and molecular biology, research in natural product biosynthesis in actinomycetes has accelerated throughout the past two decades.

Like other natural products, the biosynthesis of natural products in actinomycetes is a multi-step process which involves the assembly of basic substrate units into a chemical scaffold that undergoes further modification to form the eventual product. Though diverse in nature, these chemical compounds are synthesized from a small set of homologous proteins that are involved in secondary metabolism. Enzymes involved in structural backbone biosynthesis are classified with reference to their functions. Other proteins that are involved in structural backbone modification are known as tailoring enzymes.

1.2.1.1. Polyketide Synthase (PKS)-derived natural products

PKS enzymes catalyze the biosynthesis of polyketide natural products through step-wise incorporation of extender units to a nascent polyketide intermediate via decarboxylative Claisen condensation.²⁷ They belong to a family of large proteins which function as a single, multidomain protein or in the form of a multi-enzyme complex. Each domain of a PKS catalyzes a specific reaction in a chronological fashion to yield the eventual final polyketide compound. For each round of chain elongation, the nascent polyketide chain is first tethered to β -Ketosynthase (KS). Acyltransferase (AT) domain transfers a specific extender unit, typically malonyl-Coenzyme A (malonyl-CoA), to the Acyl Carrier Protein (ACP) domain of the same PKS module, which is previously activated through phosphopantetheinylation by a 4'-Phosphopantetheinyl Transferase (PPTase).²⁸ This extender unit is subsequently incorporated to the polyketide chain via decarboxylative Claisen condensation, which is performed by KS. Processing domains like Ketoreductase (KR) and Dehydratase (DH) might be present, which catalyzes reduction of the β -keto group to a β -hydroxyl group and dehydration of β -hydroxyl with an α -hydrogen to form a α,β -unsaturated moiety. The final polyketide product is released from the ACP via Thioesterase (TE) mediated hydrolysis of the thioester bond.

In general, PKSs can be divided into three classes – Type I PKSs, Type II PKSs, Type III PKSs. Type I PKSs can be subdivided into Type I modular PKSs and Type I iterative PKSs, *Trans*-AT modular PKSs and enediyne PKS (PKSE).

Type I modular PKSs are large multidomain proteins that function in a linear, processive manner to yield a final polyketide product (**FIGURE 2a**). Each module is typically responsible for a single round of chain elongation, although there are exceptions whereby some modules may be iterative, either in a programmed fashion²⁹⁻³⁰ or as a result of aberrant iteration.³¹ Intermolecular PKS modules possess N and C terminal docking domains that allow them to recognize and interact with successive PKS modules to facilitate transfer of nascent polyketide intermediate.³² After every round of chain elongation, the polyketide intermediate is shuttled from the ACP of the preceding module to the KS domain of the succeeding module. The procedure continues until the polyketide reaches the terminal end of the entire assembly line.

Type I iterative PKSs consists of a single PKS module, which is responsible for multiple rounds of chain elongation, in a programmed, iterative fashion, to yield a final polyketide product (**FIGURE 2b**). They can be non-reducing, partially reducing or even fully reducing. Both non-reducing and partially reducing Type I iterative PKSs can be found in actinomycetes, whereas fully reducing Type I iterative PKSs are mainly found in Fungi.³³ Non-reducing Type I iterative PKS do not contain other processing domains apart from KS, AT, ACP, and is typically restricted to generation of aromatic polyketides.³⁴ Whereas partially reducing Type I iterative PKS contains processing domains like KR and DH, which are programmed to process the nascent polyketide chain at specific rounds of elongation.^{35-36b}). They can be non-reducing, partially reducing or even fully reducing. Both non-reducing and partially reducing Type I iterative PKSs can be found in actinomycetes, whereas fully reducing Type I iterative PKSs are mainly found in Fungi.³³ Non-reducing Type I iterative PKS do not contain other processing domains apart from KS, AT, ACP, and is typically restricted to generation of aromatic polyketides.³⁴ Whereas partially reducing Type I iterative PKS contains processing domains like KR and DH, which are programmed to process the nascent polyketide chain at specific rounds of elongation.³⁵⁻³⁶

Trans-AT modular PKSs are modular PKSs that act in similar chemical fashion as modular Type I PKSs, however, their PKS modules do not contain AT domains

(**FIGURE 2c**).³⁷ The *trans*-acting AT domains, which are structurally closer to Type II AT domains than Type I AT domains,³⁸ fulfil the role of *cis*-AT domains in an otherwise modular Type I PKS setup. In some cases, these *trans*-acting AT domains serve as checkpoints in a polyketide chain assembly line by proofreading the extender units loaded onto the ACPs.^{39c}).³⁷ The *trans*-acting AT domains, which are structurally closer to Type II AT domains than Type I AT domains,³⁸ fulfil the role of *cis*-AT domains in an otherwise modular Type I PKS setup. In some cases, these *trans*-acting AT domains serve as checkpoints in a polyketide chain assembly line by proofreading the extender units loaded onto the ACPs.³⁹

Enediyne PKSs are unique fully reducing iterative type I PKS that are found in BGCs that produce natural products with 9 or 10-membered enediyne core moiety (**FIGURE 2d**). They have a characteristic PKS architecture that is distinctive from other iterative Type I PKSs.⁴⁰ PKSEs produce a heptaene metabolite⁴¹⁻⁴² which is subsequently converted by other tailoring enzymes within the BGC to form the eventual enediyne natural product. d). They have a characteristic PKS architecture that is distinctive from other iterative Type I PKSs.⁴⁰ PKSEs produce a heptaene metabolite⁴¹⁻⁴² which is subsequently converted by other tailoring enzymes within the BGC to form the eventual enediyne natural product.

Type II PKSs function as multi-enzyme complex, in an iterative fashion, to generate a polyketide product (**FIGURE 3a**). Due to the absence of processing domains like KR and DH, polyketides generated by Type II PKSs are typically unstable and undergo spontaneous cyclization to yield aromatic moieties, such as angucyclines and anthraquinones.⁴³ KS domains of Type II PKSs are also unique from Type I PKS in which they exist as a α/β -heterodimer complex with the α -subunit catalysing decarboxylative Claisen condensation during chain elongation, and the β -subunit serving as a chain length factor (CLF) to control the polyketide chain length.⁴⁴⁻⁴⁶

Type III PKSs are stand-alone KS domains that function in the absence of ACP (**FIGURE 3b**). Although Type III PKSs are long associated with plants,⁴⁷ advances in genome mining have revealed putative Type III PKS BGCs in microbes like actinomycetes.⁴⁸ Though rare, such PKSs found in actinomycetes can produce interesting molecules like germicidin⁴⁹ and violapyrones⁵⁰.

PKS is capable of constructing unique chemical scaffolds that present a significant source of leads in drug discovery.⁵¹ Some notable examples of clinical drugs derived from polyketide natural products include avermectin,⁵² a compound that is used to treat parasitic infections and amphotericin B,⁵³ which has been used as antifungal drug for more than half a century.

(a) Type I PKS (modular)

(b) Type I PKS (iterative)

(c) *Trans* AT PKS

(d) Eneidyne PKS

Figure 2: Domain architecture and mechanism of different Type I PKSs. (a) Type I modular PKS generates a polyketide product in a linear processive manner. (b) Type I iterative PKS generates a polyketide product through iterative rounds of chain elongation by a single module. (c) *Trans*-AT PKS architecture lacks *cis*-AT domains, whose role is fulfilled by *trans*-acting AT domain. (d) Eneidyne PKS generates a heptaene moiety that is subsequently converted into 9 or 10-membered enediynes core. A refers to adenylation domain, ACP refers to acyl carrier protein domain, AT refers to acyltransferase, DH refers to dehydratase, KR refers to β -ketoreductase, KS refers to β -ketosynthase, PPT refers to phosphopantetheinyl transferase, TE refers to thioesterase.

(a) Type II PKS (iterative)

(b) Type III PKS (ACP-independent, iterative)

Figure 3: Domain architecture and mechanism of Type II and Type III PKSs. (a) Type II PKS functions iteratively as a multi-enzyme complex to generate a polyketide through multiple rounds of chain elongation. (b) Type III PKS functions independently without ACP and generates a polyketide product through rounds of chain elongation. ACP refers to Acyl carrier protein, KS α refers to α -subunit of β -Ketosynthase, KS β refers to β -subunit of β -Ketosynthase.

1.2.1.2. Nonribosomal Peptide Synthase (NRPS)-derived Natural Products

NRPS are a family of large, multidomain proteins that are responsible for synthesis of peptides through a mechanism that is independent of messenger RNA and typical cellular translational machinery.⁵⁴ Most NRPSs function in a modular assembly fashion,⁵⁴ synthesizing nonribosomal peptides from a diverse range of amino acids, which includes the twenty canonical amino acids found in typical ribosomal peptide synthesis. The adenylation (A) domain of a NRPS selects a specific amino acid and activates it as an aminoacyl-AMP, which is subsequently transferred to a peptidyl carrier protein (PCP) as an aminoacyl thioester (**FIGURE 4a**). Like ACPs of PKS, PCPs of NRPS must be phosphopantetheinylated by a PPTase before they are able to tether to any substrates. Condensation (C) domain of the succeeding module catalyzes the formation of an amide bond between the amino group of a preceding aminoacyl thioester with the thioester group of another downstream aminoacyl thioester. Bioinformatic analysis of several taxonomic families of actinomycetes revealed that a wealth of potential NRPS BGCs in actinomycetes, with *Pseudonocardaceae* and *Streptomyetaceae* containing an average of around seven such clusters.⁵⁵

(a) Type I NRPS (modular)

(b) Examples of NRP natural products

Figure 4: NRPS mechanism and NRPS-derived natural products. (a) NRPS domain architecture and mechanism of nonribosomal peptide synthesis. (b) Examples of prominent antibiotics that are biosynthesized by NRPSs.

Non-modular NRPS do exist and has been reported in a previous study on bleomycin biosynthesis, which identified a stand-alone PCP protein (BlmI).⁵⁶ Non-modular NRPSs are also prevalent in the biosynthesis of siderophores like vibriobactin⁵⁷ and enterobactin.⁵⁸ However, identification and characterization of non-modular NRPS BGCs still lag behind typical modular NRPS BGCs. Recently, a genome mining study of 2699 genomes has shed light on the common occurrence of non-modular NRPS,⁵⁹ with more than half the NRPS BGCs identified being classified as non-modular. In that study, although actinomycetes represented <20% of the strains, it accounted for nearly 1/3 of all BGCs identified. This finding provides optimism that apart from nonribosomal peptides, NRPSs also play a significant role in biosynthesis of other alkaloid natural products and actinomycetes could provide a source of alkaloid natural products.

NRPS provides interesting peptide scaffold (**FIGURE 4b**) that include the incorporation of non-canonical amino acids. Some of these peptides exhibit potent antimicrobial activities and have been utilized as drugs or are currently in late stage clinical trial as treatment against bacterial infections.⁶⁰⁻⁶²

1.2.1.3. Terpenoid natural products

Terpenoids are a large class of natural products that are synthesized by terpene synthases or terpene cyclases.⁶³ Depending on the mechanism for formation of initial carbocation, a terpene cyclase can be classified as either Class I or Class II.⁶⁴ In Class I terpene cyclase, a trinuclear metal cluster ionizes an isoprenoid pyrophosphate substrate to generate a reactive allylic cation and its pyrophosphate leaving group.⁶⁵ On the other hand, Class II terpene cyclase utilizes an acid to protonate the pyrophosphate group, which departs from the isoprenoid substrate to yield a tertiary carbocation.⁶⁶

While terpenoids have long been considered as secondary metabolites from plants and fungi, recent genome mining has identified numerous terpene synthases in bacteria.⁶⁷ The discovery of these bacterial terpene synthases certainly explains the widespread nature of terpenoids in actinomycetes.⁶⁸ Even though different terpene synthases require different isoprenoid precursors, these precursors are all generated from the most basic C₅ isoprenoid units, dimethylallyl pyrophosphate (DMAPP) and isoprenyl pyrophosphate (IPP). As such, all isoprenoid precursors contain carbons in multiples of five. DMAPP and IPP can be synthesized by two different metabolic pathways – mevalonate (MVA) pathway and non-mevalonate (MEP) pathway (**FIGURE 5**). Like most bacteria, actinomycetes generally synthesize DMAPP and IPP through the MEP pathway as they lack enzymes from the MVA pathway. However, some terpenoid-producing actinomycetes do contain a cluster of genes which encode proteins involved in MVA pathway.⁶⁹⁻⁷¹

Examples of terpenoids produced by actinomycetes include platencin, platensimycin⁷² and carotenoids,⁷³ which are produced by many strains from the family *Micromonosporaceae*.

Terpenoid biosynthetic pathway

Figure 5: A schematic of the terpenoid biosynthetic pathway. The MVA pathway is depicted in blue while MEP pathway is depicted in black.

1.2.1.4. Ribosomally synthesized post-translationally modified peptides (RiPPs)

Another class of natural products that is commonly found in actinomycetes is ribosomally synthesized and post translationally modified peptides (RiPPs), which has been attracting attention throughout the past decade. Since the advent of genome mining, knowledge regarding RiPP has expanded rapidly. Prior to the availability of genome sequencing, different classes of RiPPs were regarded as distinctive classes of natural products. However, the accumulation of knowledge on vast number of RiPP BGCs in the post-genomic era has led to suggestion that RiPPs share a common biosynthetic origin. In 2013, a universal nomenclature for RiPPs was finally created.⁷⁴ All these RiPPs identified thus far can be classified into more than 20 classes, with some classes of RiPPs being restricted to certain taxonomic groups – E.g. cyanobactins are only produced by Cyanobacteria.⁷⁵

All RiPP BGCs possess a small gene that encodes the precursor peptide, which is ribosomally synthesized. This precursor peptide consists of a leader peptide attached to the N terminus of a core peptide (**FIGURE 6a**). While the core peptide undergoes post translational modifications to form the eventual RiPP, the leader peptide facilitates recognition of the precursor peptide by tailoring enzymes for post translational modifications and export.⁷⁶ Unusual cases exist whereby the precursor peptide itself either lacks a leader peptide for recognition⁷⁷ or is reliant on a trans-acting leader peptide.⁷⁸ All RiPPs share this generic biosynthetic mechanism but are further classified into different categories based on the biosynthetic mechanism of the RiPP.⁷⁴

(a) RiPP Biosynthesis

(b) Examples of RiPP natural products

Figure 6: A schematic of generic mechanism involved in RiPP biosynthesis. (a) The precursor peptide which is synthesized by the cell's translational machinery is being modified and hydrolysed to form the eventual mature RiPP. (b) Examples of notable RiPP natural products from actinomycetes. Thioestrepton is a thiopeptide produced by several *Streptomyces* strains while bottromycin is produced by *Streptomyces bottropensis*.

Till date, six classes of RiPPs have been discovered in actinomycetes - Lasso peptide,⁷⁹⁻⁸⁰ Lanthipeptide,⁸¹⁻⁸² Linaridin,⁸³ Bottromycin,⁸⁴ Linear azol(in)e-containing peptide (LAP),⁸⁵ and thiopeptides.⁸⁶⁻⁸⁷ Lasso peptides are RiPPs that contain a signature knotted structure which resembles a lasso fold.⁸⁸ With a length of 16 - 21 amino acids, they possess an N-terminal macrolactam that is generated by an amide linkage between the N-terminal amino group with the carboxylate side chain of an aspartic acid or glutamic acid at position 8 or 9. Lanthipeptides are characterized by the presence of two alanine residues bonded by a thioether linkage. This linkage was catalyzed by a Michael addition reaction between a Cys residue and a previously dehydrated Ser or Thr residue. Linaridins possess thioether linkages and are structurally similar to lanthipeptides. However, they are classified differently from lanthipeptides due to the absence of lanthipeptide dehydratases in their BGCs. Bottromycins (**FIGURE 6b**) are a unique class of RiPPs that are discovered in *Streptomyces bottropensis* only. They are characterized by the presence of a macrocyclic amidine moiety with a C-terminal thiazole ring.⁸⁴ LAPs, also known as thiazole/oxazole-modified microcins (TOMMs) are characterised by the presence of thiazol(in)e and oxazol(in)e heterocycles in their structures. The formation of these heterocycles is catalysed by a heterotrimeric protein complex containing a cyclodehydratase and dehydrogenase. Cyclodehydration is achieved by a nucleophilic

attack of the thiol or hydroxyl group of the side chain on the carbonyl group of a preceding amide bond to form a thiazoline or oxazoline structure, which is subsequently reduced into a thiazole or oxazole heterocyclic structure. Thiopeptides are RiPPs that share similar structural features as LAPs and lanthipeptides due to the presence of similar enzymes in their BGCs. The major difference between thiopeptides and other RiPPs is the signature six-membered nitrogen-containing heterocycle which bonds the peptide in a cyclic form.⁸⁹ Its oxidation state is dependent on its substitution pattern, which varies between different thiopeptides. As such, the heterocyclic ring can exist as either pyridine, piperidine or dehydropiperidine.

1.3. Renewal of interest in actinomycete natural product discovery in the Post-Genomic Era

Since the arrival of the Millennium, the Genomic Era has overseen huge leaps in genome sequencing technologies, which led to expansion of DNA sequencing capacity and reduction in incurring costs.⁹⁰ The affordability of DNA sequencing has rendered the application of genomics in scientific research, including natural product discovery and biosynthesis in actinomycetes. Surveying of natural environment for actinomycete biodiversity is also a viable option with the availability of single cell sequencing technology.⁹¹ Unsurprisingly, these advancements have led to a sharp increase in genomic data and development of bioinformatic tools for analysis of the burgeoning knowledge generated. Presumably exhausted after rediscovery of known natural products, genomic data from actinomycetes has revealed many potential BGCs with no accountable natural products yet. Given that analyses performed on known BGCs revealed that they are cryptic in at least 70% of the strains surveyed,⁵⁵ it is likely that a significant portion of entire natural product repertoire remains hidden. For pharmaceutical companies that embarked on natural product screening in the Golden Age of Antibiotics, diminished success in the later years of such programs could be attributed to two main factors. Firstly, their pursuit for potential antimicrobials against intended targets skewed the objective of their search towards bioactivity based screening. However, compounds with low yield tend to be neglected in these routine screens, which typically have a higher detection limit than chemical instrumentation. Secondly, the throughput approach which deployed systematic fermentation methodologies also failed to activate many BGCs, which typically require specific

culture conditions to be active. Therefore, rather than displacing the traditional approach, genomics guided approach offers an alternate path to facilitate natural product discovery. Instead of deploying non-pleiotropic methods to discover novel compounds, genome mining offers a more targeted approach at prioritizing and targeting microbes for intended discovery of novel compounds.⁹²⁻⁹³ This targeted approach has generated optimism that genome mining of actinomycetes for novel natural products will trigger a revival in natural product discovery in the current post Genomic era.⁹⁴⁻⁹⁵

1.4. Approaches to discovery of natural product in actinomycetes

1.4.1. Approaches to screen for natural products

Since the inception of natural product research, interest in this topic has grown globally. Despite the acknowledged withdrawal of natural product screening programs in pharmaceutical companies, the rate of novel natural product discovery has been on the uptrend since 1940s.⁹⁶ Screening outcomes were either judged by metabolite profiling or detection of bio-based screening (**FIGURE 7**). Despite showing initial successes, the lack of a robust dereplication strategy grew in significance as a key limitation in such procedures. Although this issue was partially rectified with the gradual establishment of databases for chemical dereplication,⁹⁷⁻⁹⁸ bioactivity screening outcomes remained difficult to derePLICATE apart from broad classification through Antibiotic mode of action profile (BioMAP).⁹⁹

Figure 7: Strategies to screen for novel compounds. Screening of novel compound (depicted as green circle) from a crude mixture can be performed via chemical profiling or bio-based screening. Detection of positive hit necessitates further compound isolation and structural elucidation to validate the identity of the compound.

Even without a robust dereplication system in place, bioactivity screening remains favourable due to the ease of data analysis and targeted outcomes generated from such assays. Pharmaceutical companies that continued natural product screening programs were constantly developing novel bioactivity screening methods to identify hits against intended organisms or receptors. As such, bioactivity screening has evolved widely from basic whole cell based screening to more targeted approaches. These alterations address some of the underlying limitations and support the continued usage of bioactivity screening in identifying potential hits. One notable method, which led to the discovery of platensimycin and platencin, used a target based whole cell screening assay which involves the detection of bacterial sensitivity via differential antisense transcription.¹⁰⁰⁻¹⁰¹ Such strategy enables identification of hits which are easily neglected by conventional whole cell based screening as it might require higher yields to register a hit. Another similar strategy which targets the central metabolism of an organism also addresses the limitation created by usage of rich media used in typical whole cell screening.¹⁰² It allows for identification of inhibitors against specific enzymes which are essential for survival of the microbe in the environment of its infected host.¹⁰³

Bioactivity screening can also be performed via *in vitro* assays to identify hits against target receptor of interest. Although such strategy discounts the influence of other factors such as membrane permeability and other mechanisms of resistance, it offers a preliminary hit for design of leads with better pharmacological properties. Receptor-ligand binding assays have been used extensively in drug discovery and development as they are easily scalable to facilitate high throughput screening.¹⁰⁴ However, they require special instrumentation and expensive components like purified recombinant proteins, which restricts their application to pharmaceutical companies. Recently, an *in vitro* RNA polymerase inhibition assay has enabled the identification of a novel nucleoside-analog compound, pseudouridimycin, from crude extract screening as a potential lead against drug-resistant pathogens.¹⁰⁵ This demonstrates the continued effectiveness of screening as a strategy for screening new compounds when applied appropriately. Apart from binding assays, receptor-ligand binding based screening is also a promising method for identifying novel compounds with therapeutic potential.¹⁰⁶ This method, also known as affinity-based screening, hinges on the receptor-ligand binding affinity to screen for potential ligands with significant binding to the target receptor of interest. One recent application identified potential inhibitors towards bacterial transglycosylase by screening crude mixtures for potential ligands which bind to binding domains of bacterial transglycosylase.¹⁰⁷ The chromatographic platform utilized for the screening was established by binding recombinant His-tagged fused protein binding domains to nickel chelation beads. The reusability and flexibility of this system makes it an appealing option for screening for leads against other target receptors of interest.

A bioactivity-based screening is not without its drawbacks. The specific aim of each bioactivity assay neglects novel compounds that harbour therapeutic properties which are irrelevant to the biological target in question. This limitation is remarkable as a growing number of 'old' natural products have been identified as new drug leads in recent times.⁹ A noteworthy example is salinomycin, which was discovered as an antibiotic in 1974,¹⁰⁸ but was later on found to possess selective killing effect against cancer stem cells in 2009.¹⁰⁹ Also, the limit of detection and sensitivity of biological based assays are usually inferior when compared with chemical instrumentation. Rather than displacing bioactivity screening, metabolite profiling serves as an indiscriminate approach to identify all novel compounds regardless of their therapeutic

properties. In this aspect, improvements in chemical instrumentation has resulted in lowered detection limits for identification of compounds produced in minute yields. Due to its unparalleled limit of detection, use of LC-MS technology soon became a prevalent tool for this purpose. However, the increased sensitivity also posed a challenge for deciphering positive hits from an immense data background. Hence, data mining strategies like molecular networking was developed to facilitate identification of positive hits through visualization of big data.¹¹⁰⁻¹¹¹

Both bio-based and chemical based screening approaches ensure emerging applications continue to get developed from an ever-expanding natural product repertoire.

Table 1: Comparison of approaches for microbial natural product discovery

Conventional Approach	Genomics guided approach
Unable to identify strains with novel BGCs	Enables prioritization of strains with novel BGCs
Screening can proceed without genomics	Requires genomics, bioinformatics and optimization of tools for genetic manipulation
Randomized screening for cluster activation	Targeted activation of cryptic clusters
High rate of rediscovery of known compounds	Avoidance of known compounds through identification of their corresponding BGCs eliminates risk of rediscovery
Laborious effort required to mine for novel compound	Workflow is streamlined to focus on targeted BGCs
Tedious compound isolation and structural elucidation without information regarding the compounds	Guidance from genomics provide plausible biosynthetic logic to facilitate structure elucidation

1.4.2. Traditional approaches to activation of natural product biosynthesis in actinomycetes

Before the conception of genomics in natural product research, little information was available on the biosynthetic potential of all the secondary metabolite producing microbes. Crude extracts generated from fermented cultures were analysed *per se*, without any pointers other than the identity of the microbe. The fact that these microbes ferment differently under different growth and media conditions is acknowledged and thus, utilized in hope of discovering novel natural products. However, the lack of understanding of every microbe's biosynthetic potential also meant that these approaches are reliant on blind chance generated through pleiotropic methodologies.

Generally, the main limitation of these pleiotropic approaches lies in the unpredictable nature of their outcomes, hence, experimentation with a variety of conditions might be necessary to achieve the desired outcome.

Figure 8: Traditional approaches to activation of natural product biosynthesis in actinomycetes. These non-pleiotropic approaches include the use of chemical elicitors, co-cultivation with other microbes or screening of different fermentation conditions to activate natural product biosynthesis in actinomycetes.

1.4.2.1. Chemical elicitation

Supplement of chemicals in culture media is a useful technique to trigger activation of cryptic BGCs in actinomycetes (**FIGURE 8**). These chemicals mimic molecules in the natural environment, which affect certain signalling pathways in the affected actinomycete's metabolism. Such chemicals act by influencing proteins involved in global regulatory network, triggering stress in the bacteria, or increase precursor supply for efficient yield production. The complexity of the entire metabolome and influence from associated pathways mean that the outcome is usually random and unpredictable. However, ongoing efforts do exist to understand such regulatory pathways and their influence in secondary metabolism. For example, N-

acetylglucosamine (GlcNAc), which is a monomer of chitin, is notable for eliciting the production of cryptic secondary metabolites when used appropriately.¹¹²⁻¹¹⁴ It influences secondary metabolism in actinomycetes by serving as a signaling molecule that conveys nutritional status information to the cell and mediates the activity of DasR, a GntR family regulator protein.¹¹⁵ DasR binds to operator sequences of several genes, including those involved in growth and secondary metabolism,¹¹⁶ hence inhibiting transcription of these genes. Binding to glucosamine-6-phosphate, a metabolic product of GlcNAc, inactivates DasR, which consequently leads to transcriptional activation of the affected genes. The effect GlcNAc has on secondary metabolism is dependent on the medium composition that the actinomycete is cultured in. Under rich media conditions, GlcNAc represses antibiotic production. Conversely, in poor media conditions, it activates antibiotic production. It has been hypothesized that the opposing effects elicited by GlcNAc under rich or poor media conditions is a result of the cell being programmed in a feasting or famine state, respectively.¹¹³ As GlcNAc is also a component of bacterial cell wall, its presence in a poor media could relate to bacterial cell death, which is inferred as a state of famine. While in a rich media, it could mean a feasting state with an abundant supply of chitin, an important source of carbon and nitrogen for actinomycetes.

Another chemical elicitor is γ -butyrolactone (GBL), which activates secondary metabolite production by promoting aerial hyphae growth.¹¹⁷⁻¹¹⁹ Being a quorum sensing molecule, GBL mediates key processes linking metabolism and cell development.¹²⁰

Interestingly, supplementation of media with sub-inhibitory concentrations of antibiotics had also been reported to elicit secondary metabolism in actinomycetes.¹²¹⁻¹²² In high throughput screening of chemical elicitors, antibiotics showed stronger eliciting effect on secondary metabolism than other general chemicals,¹²³ which may suggest that modulation of secondary metabolism by supplemented antibiotics was a defence mechanism by the host actinomycete.¹²⁴ Other chemical elicitors which triggers activation of secondary metabolism are compounds like DMSO,¹²⁵ ethanol, heavy metals, etc,¹²⁶ whose underlying mechanisms of activation remained unknown.

Apart from activation of secondary metabolism, chemicals could elicit production of certain secondary metabolite simply by increasing the precursor supply available for

biosynthesis. In the production of polyketide natural products, supplementation of sodium acetate or sodium propionate in millimolar quantities had been reported to boost yields by several folds.¹²⁷⁻¹²⁸

With deeper understanding of the regulatory mechanisms governing secondary metabolism and identification of potential genes influenced by global regulator proteins,¹²⁹ chemical elicitation holds potential to be a useful tool for activation of secondary metabolism in actinomycetes.

1.4.2.2. Co-cultivation

Co-cultivation is another non-pleiotropic approach that has been explored in recent times for activation of cryptic clusters (**FIGURE 8**). In general, a binary combination of strains is employed as cultivation of multiple strains concurrently complicates the growth monitoring of each strain. The target actinomycete strain could be cultivated together with either a helper strain¹³⁰ or a competitor strain¹³¹⁻¹³². Helper strains might facilitate activation of BGCs through quorum sensing, while competing strains usually produce compounds that act as chemical elicitors to trigger activation of cryptic BGCs. For example, mycolic acid, produced by bacteria like *Mycobacteria*, is known to elicit natural product production in *Streptomyces*.¹³³ Hence, co-cultivation with such bacteria facilitates the activation of cryptic BGCs in *Streptomyces*.¹³⁴⁻¹³⁶

Establishing a Co-cultivation system requires detailed optimization of culture conditions of individual strains to maintain proper growth of the strains. Also, stringent control of experimental conditions is necessary to obtain reproducible experimental outcomes. Co-cultivation can be performed on plate cultures¹³⁷ whereby growth development of target actinomycete strain and competing strain can be observed to understand the effects of combined cultures on the growth of each strain¹³⁸. Observation of growth inhibition of competing strain also serves as a visual cue for detection of potential antimicrobial compounds. However, the barrier to scalability limits this cultivation system to low yield compound production. Liquid fermentation is typically preferred. In order to isolate compounds of higher yields from co-culture systems.¹³⁶ To achieve co-cultivation without actual contact of the two microbes, unique culture flasks which are separated from each other by 0.2 μm filter will be

required.¹³⁹ This allows exchange of small molecules between both cultures while allowing cultures of individual strain to be monitored for growth development.

Due to the valuable knowledge that can be obtained from microbial interspecies communication, growing interest in microbial co-cultures has caused a paradigm shift in microbiology.¹⁴⁰ As co-cultivation becomes an increasingly important tool for natural product activation in actinomycetes,¹⁴¹⁻¹⁴³ their application together with other modern technologies can help to uncover more novel compounds from cryptic sources.¹³⁹

1.4.2.3. Culture condition screening

One of the most conventional way of natural product activation in actinomycete is simply by altering their fermentation conditions (**FIGURE 8**). Although manipulation of culture conditions has long been known to influence microbial growth and secondary metabolism, the term 'One Strain Many Compounds' (OSMAC) was only conceived in 2002.¹⁴⁴ As the name suggests, each microbial strain has the capacity to provide different compounds under different fermentation conditions, hence, leading to a diverse range of compounds that can potentially be uncovered from a single microbe. Conditions like temperature,¹⁴⁵ pH,¹⁴⁶ phosphate level,¹⁴⁷ carbon source,¹¹⁶ nitrogen source,¹⁴⁸ salt level,¹⁴⁹ trace elements level,¹⁵⁰ etc can be altered to explore the biosynthetic potential of individual microbe. The extent of components and range that can be tested render the number of possible fermentation conditions seemingly infinite. As the success rate of OSMAC is highly variable across different actinomycete strains, testing numerous conditions in a high throughput manner is advisable for generating possible hits. Apart from activation of cryptic clusters, culture condition screening, particularly in a statistical manner with a chemically defined basal medium,¹⁵¹ is also a good way to optimize yield of a desirable natural product.¹⁵²⁻¹⁵⁴ Given the ease of culture condition screening, it remains a mainstay in natural product activation, which is generally applied as an initial step to interrogate the biosynthetic capacity of actinomycetes before other approaches are considered.

1.4.3. Genomics guided approaches to natural product biosynthesis and discovery in actinomycetes

In the Post Genomic Era, blooming Genomic data has led to a paradigm shift in the approach taken towards natural product discovery in actinomycetes. Instead of relying on conventional screening to detect novel compounds, genome mining offers a powerful tool for predicting the biosynthetic potential of individual strains prior to extensive wet-lab experimentation. This allows natural product discovery in actinomycetes to be oriented towards systematically analyzed putative BGCs.^{95, 155} Hence, allowing actinomycete strains to be prioritized for further pursuit of targeted BGC through pleiotropic or non-pleiotropic approaches. Predictions on molecular framework encoded by putative BGCs also facilitate a targeted approach to be deployed for compound isolation and structural elucidation. The effectiveness of genome mining has led to its extension to a new approach termed ‘molecules-to-genes-to-molecules’, where a BGC encoding for a newly elucidated compound was used to query for other similar BGCs, which led to the discovery of structurally similar analogs.¹⁵⁶

Figure 9: Approaches to natural product discovery in actinomycetes. Traditional approach relies on screening to identify hits before proceeding to compound isolation and structural elucidation. The Post Genomic Era has seen a paradigm shift towards a genomics guided approach,¹⁵⁷ which facilitates identification of clusters for activation. The structures of isolated compounds from successfully activated BGCs are subsequently elucidated.

To facilitate genome mining efforts, researchers have developed a repertoire of computational tools for related purposes such as identification of potential BGCs, prediction of substrate specificity and connection between genomic data and mass-spectral data.¹⁵⁸ Depending on the algorithms implemented, these tools can either provide a fast yet reliable overview of potential BGCs present or detect novel BGCs with low similarity, which can potentially encode for novel chemical structures. Due to the limitations of each tool, users are required to select the ideal tool suited for their bioinformatic analysis.

One particular bioinformatics tool that is widely used for genome mining is antiSMASH,¹⁵⁹ which was first conceived as a concerted effort to develop genomic mining as a robust tool in natural product discovery. AntiSMASH uses prediction algorithms from other standalone prediction tools, which facilitates an untargeted, comprehensive identification of all putative BGCs available in a given DNA sequence (**FIGURE 10**). Key functions include gene annotations, cluster identification, cluster comparison, substrate prediction, etc. With such comprehensive analysis offered by antiSMASH, it is unsurprising that the tool has garnered significant increase in user traffic since its first inception. Apart from facilitating the identification of orphan BGCs and other clusters for further pursuit, results from antiSMASH aid dereplication efforts through homology comparison with known BGCs from MIBiG¹⁶⁰, a database of all known BGCs. With consistent upgrades and maintenance, antiSMASH has reached the 4th version whereby improvements are made to chemistry prediction and cluster boundary prediction.¹⁶¹

Identification of putative BGCs

BGC boundary and gene annotation

c00001_scf7180... - Cluster 9 - T1pks-nrps

Gene cluster description

c00001_scf7180... - Gene Cluster 9. Type = t1pks-nrps. Location: 5013021 - 5080660 nt. Click on genes for more information.
 Show PHMM detection rules used

Legend:

■ core biosynthetic genes ■ additional biosynthetic genes ■ transport-related genes ■ regulatory genes ■ other genes ■ TTA codon

Gene prediction

Comparison with known BGCs from MIBiG

Comparison with known sequences from NCBI

Figure 10: A schematic of data generated by antiSMASH analysis of an actinomycetal genome. First page of analysis reveals putative BGCs identified by antiSMASH. Clicking onto each cluster will reveal more details regarding the cluster boundary and gene annotations. Gene prediction data is also provided alongside with homology comparison with known BGCs from MIBiG. Enabling ClusterFinder algorithm will reveal comparison with known sequences from NCBI.

Apart from genome mining tools, databases pertaining to gene clusters have also been established and evolved through the years to accommodate the surge in information. These databases facilitate bioinformatic exploration of natural products by offering analysis tools and knowledge on biosynthesis of known natural products. IMG-ABC, the largest gene cluster database till date, provides information on experimentally validated and computationally predicted gene clusters in genomes and metagenomes, which can be analyzed in depth with integrated tools found within the database.¹⁶² antiSMASH has also implemented its personal database, which includes novel and

known BGCs from common microbial strains whose genomic information is available in NCBI.¹⁶³ Another aforementioned database, MIBiG, works as a standalone database or in association with antiSMASH. It provides comprehensive gene annotation and classification on experimentally validated BGCs.¹⁶⁰

Chemical compound databases have also expanded greatly due to the continuous discovery of novel natural products. Dictionary of Natural Products is the most comprehensive natural product database available, providing information on basic properties and biological origin of the natural products. The significance of actinomycetes as a source of natural product is highlighted by the construction of a database that is solely dedicated to natural products derived from *Streptomyces*.¹⁶⁴

1.4.3.1. Synthetic biology tools to study BGCs in vivo.

Till date, the translation of genome mining knowledge into natural product has been greatly hindered by the cryptic nature of the BGCs identified. Despite harboring numerous BGCs in their genomes, actinomycetes only express a fraction of these BGCs. The cryptic nature of most BGCs under typical laboratory conditions necessitates additional approaches for activation. Despite the success of conventional approaches in activation of some clusters, the unpredictability of those approaches demands experimenting numerous conditions to achieve favorable outcomes. Such workflow is laborious and faces a high failure rate. Development of tools in synthetic biology addresses the limitations of conventional approaches by enabling targeted activation of BGCs either through cluster manipulation or screening of targeted BGCs.

Years of research on actinomycetes has generated a wealth of plasmids for overexpression of genes. These shuttle plasmids are capable of replication in *E. coli* for ease of cloning and maintenance. As conjugation is still the method of choice for transforming actinomycetes with plasmids, these shuttle plasmids usually contain like OriT to facilitate conjugal transfer of plasmid from *E. coli* to actinomycetes. Conjugal transfer of plasmid from *E. coli* to actinomycetes necessitates the formation of pilus between the donor *E. coli* and actinomycete strain and RP4 mediated transfer of the OriT containing plasmid in the form of single stranded DNA. The required machinery is encoded by genes present in a broad host range RK2 plasmid.¹⁶⁵ Hence, the donor *E. coli* must possess both the shuttle plasmid and a broad range RK2 plasmid

(pUZ8002) for successful conjugation. Certain actinomycete strains possess strict methylated DNA restriction system that hinders the receipt of methylated DNA from donor *E. coli*.¹⁶⁶ In these cases, a non-methylating donor *E. coli* is used instead.¹⁶⁷ Overexpression plasmids can be categorized into replicative and integrative versions. Replicative plasmids possess origin of replication with different copy numbers.¹⁶⁸⁻¹⁷⁰ They are useful for maintaining small sized inserts. but are generally not suitable for stable maintenance of large inserts (>40 kb). Heterologous expression of such inserts requires the use of integrative plasmids, which integrate the inserts into specific sites in the chromosomes of actinomycetes.¹⁷¹⁻¹⁷³ Integrated inserts tend to maintain in a stable manner and do not require selective pressure from antibiotics. Integrative plasmids harness the serine integrases from phages to integrate DNA to specific attachment sites in actinomycetal chromosomes.¹⁷⁴ Till date, three attachment sites – *attB*^{BT1}, *attB*^{C31}, *attB*^{VWB}, associated with their corresponding phage integrases have been discovered.¹⁷⁵⁻¹⁷⁷ They offer possibility of integrating up to three inserts into different regions of an actinomycetal genome.

While useful, overexpression plasmids are incapable of manipulating the genome, which is important for purposes like gene functionality studies and engineered biosynthesis. Before the discovery of Clustered Regularly Interspaced Short Palindromic Repeats/Cas9 (CRISPR/Cas9) as a genome editing tool, PCR targeting is routinely used for performing genetic manipulation of actinomycetal genomes.¹⁷⁸ It involves a two-step cloning procedure. First, a gene knock-out mutant, containing apramycin resistance gene, is generated via DNA homologous recombination mediated double-crossover event and selected in the presence of apramycin. Then, the same antibiotic selection marker is removed via DNA homologous recombination mediated double crossover event from the mutant, leaving behind a 81 bp scar (**FIGURE 11A**). The entire procedure is laborious as it requires several rounds of recombination and cloning in *E. coli* to generate plasmids with homologous sequences for intended homologous recombination. A cosmid library is also preferred to provide long homologous sequences necessary for homologous recombination event to take place in the absence of any DNA strand breaks.

Other techniques like Cre/loxP recombination¹⁷⁹ and use of Meganuclease I-SceI¹⁸⁰ were being developed as alternatives to PCR targeting. However, both applications

were largely impeded by limitations such as lengthy experimental time requirement and limited sites for genome editing, respectively.

In contrast, the development of CRISPR/Cas9 as a tool for genetic manipulation in actinomycetes has greatly streamlined the entire workflow.¹⁸¹⁻¹⁸⁴ CRISPR/Cas9 allows cleavage of any specific DNA sequences that contain a NGG protospacer adjacent motif (PAM). The flexibility of CRISPR/Cas9, coupled with the fact that Actinomycetal DNA possesses a high GC content, makes it an ideal tool for targeting most DNA sequences available in the actinomycetal genomes. Double stranded DNA breaks generated by Cas9-sgRNA complex mediates homology-directed repair by actinomycetes (**FIGURE 11B**), hence, increasing the efficiency of DNA homologous recombination double crossover event. As such, this approach discounts the requirement of a cosmid genomic DNA library. Furthermore, genome editing can be achieved in a single cloning step without the generation of any DNA scars at the target site. This reduces the time required for a single round of genome editing to two-third (17 days) of that required for conventional PCR targeting (25 days).¹⁸⁵

Figure 11: Two approaches to genetic manipulation in actinomycetes. (A) Workflow of PCR targeting which illustrates homologous recombination between cosmid and genomic DNA to disrupt a target gene in the first round of cloning. Second round of homologous recombination is performed with cosmid that lacks apramycin resistance gene. Eventual deleted gene has a 81 bp scar at the target site. Apramycin resistance gene (depicted in yellow) serves as a selection marker. FRT sequences (depicted in green) facilitates excision of apramycin resistance gene via recombination in *E. coli*. (B) Workflow of CRISPR/Cas9 which illustrates Cas9-sgRNA complex mediated double stranded DNA break and subsequent homology directed repair. A scarless deleted gene is eventually obtained after a single round of cloning.

Limitations exist which hinder the application of synthetic biology tools to natural product studies in actinomycetes. One limitation is the lack of availability of DNA sequence data for some strains or clusters. Despite the reduced cost of DNA sequencing, obtaining genomic sequencing data for all actinomycete strains is still impractical due to lower throughput as compared to conventional approaches. Another limitation is the genetic intractability of some actinomycete strains, which requires in-depth knowledge and elaborate experimental workflows for genetic manipulation of clusters to become a reality. Therefore, synthetic biology serves as a targeted approach to complement existing conventional approaches in natural product activation and engineered biosynthesis in actinomycetes.

1.4.3.1.1. Strategies to activate BGCs via genetic manipulation

Master or global regulator proteins which control secondary metabolism and growth development of actinomycetes can be manipulated to activate cryptic BGCs.¹⁸⁶⁻¹⁸⁹ However, they will not be elaborated in detail as the focus of this introduction is solely on activation of specific BGCs.

Figure 12: Strategies for activating a target BGC *in vivo*. A cryptic cluster can be activated by knocking out its repressor protein gene, refactoring the cluster via replacement of native promoters with strong constitutive ones, or overexpression of its activator protein gene in *cis* or *trans*.

In general, there are many manipulation approaches to activate a target BGC (**FIGURE 12**). Certain approaches are synergistic to each other, such as the tandem knock out of a repressor and overexpression of a positive regulator.¹⁹⁰ *In silico* analysis is typically performed to identify regulator proteins, which are compared with characterized homologues to predict their roles in the regulation of BGCs. However, predictions are usually putative as scenarios exist whereby the actual function of regulators opposes their canonical relatives.¹⁹¹⁻¹⁹² Knocking out repressor proteins facilitates access of promoter sequences by transcriptional factors, hence activating transcription of affected genes.¹⁹³⁻¹⁹⁴ Overexpression of positive regulators can be performed either in *trans* by transforming the strain with a plasmid that harbors the regulator under control of a strong, constitutive promoter,¹⁹⁵⁻¹⁹⁶ or by replacing the

native promoter with a strong, constitutive promoter in cis.¹⁹⁷ Exceptions exist whereby cluster activation was only achieved when a positive pathway-specific regulator was driven by a temporal promoter instead of a strong, constitutive promoter.¹⁹⁸

Another strategy involves refactoring the entire BGC by replacing the native promoters of all operons with strong, constitutive promoters. Suitability of refactoring approach is dependent on the number of operons in the BGC. Some BGCs consist of multiple (>3) operons that make refactoring an impractical option. For clusters with less operons, refactoring has proven to be a viable option for activation of BGCs in vivo.¹⁹⁹⁻²⁰¹

1.4.3.1.2. Genetic manipulation in vivo for engineered biosynthesis of novel natural product analogues

In actinomycetes, biosynthesis of a natural product usually involves several rounds of chemical transformation of associated substrates, which are catalyzed by various enzymes. By understanding the roles of these individual enzymes in the biosynthetic pathway, opportunities for engineering biosynthesis of novel analogues can be explored for natural product diversification. Rather than simply expanding the natural product repertoire, such chemical diversification is particularly useful for improving pharmacological properties of an original compound which has been identified as a clinically relevant drug lead.²⁰²

Engineering strategies usually target tailoring enzymes, which function at latter stages of the biosynthetic pathway. As main biosynthetic core enzymes typically function upstream of the pathway, the risk of abrupt collapse of entire pathway is much higher if the altered intermediates are unable to serve as substrates for downstream enzymes.

Engineered biosynthesis has been applied successfully to generate novel compounds with better therapeutic properties. For example, polynik A and polyoxin N are two compounds with more potent antifungal properties, which are produced via engineered combinatorial biosynthesis of the nucleoside moiety of nikkomycin with peptidyl moieties of polyoxin.²⁰³

Another class of clinically relevant compounds being targeted for engineered biosynthesis is polyene macrolides. Polyene macrolides possess potent antifungal properties and amphotericin, a polyene macrolide compound, has been in clinical use

against systematic antifungal infections for more than half a century.²⁰⁴ To tackle the rising rate of fungal infections²⁰⁵ coupled with problems of resistance against existing antifungal treatments,²⁰⁶ engineering the biosynthesis of novel amphotericin and other polyene macrolide analogues can provide leads for development of new antifungal drugs.²⁰⁷⁻²¹⁰ Given the limited number of tailoring enzymes available for manipulation, the range of generated analogues of polyene macrolides is typically restricted to modifications like the addition of a sugar moiety, deletion of a carboxyl or hydroxyl group.

Overall, though useful, engineered biosynthesis is mainly restricted to the functionality of the enzymes in question. At times, regions of the chemical scaffold that are untargeted by enzymes in a BGC must be altered via chemical synthetic strategies instead. Such chemical syntheses have also proven successful as shown by the non-toxic amphotericin B analogues being created recently.²¹¹ This implies that rather than relying on either strategy alone, complementing both chemical synthesis with engineered biosynthesis is more realistic for generating clinically relevant natural product diversification. A similar concept called GenoChemetics has been introduced recently.²¹² This unique approach hyphenates synthetic biology with chemical synthesis by introducing chemically synthesized molecules as precursors for an otherwise biosynthetic pathway. The result is generation of novel analogues with moieties that are not conceived naturally.

1.4.3.1.3. Reporter-based screening of targeted BGCs

Unlike other strategies which involves direct manipulation of the target BGCs for activation, reporter-based screening facilitates the identification of targeted BGCs that are activated via conventional, pleiotropic approaches. The concept is to correlate transcriptional activation of the targeted BGC with a reporter gene. For such correlation to be implemented, the reporter gene, typically a fluorescence protein^{123, 213} or an enzyme that synthesizes a colored pigment²¹⁴⁻²¹⁵ for ease of detection, is placed under the control of a native promoter obtained from a crucial operon or gene within the targeted BGC (**FIGURE 13**). Attempts to activate the cluster are typically pleiotropic in nature, but the simplified readout minimizes tedious inspection of screening outcomes, making the method scalable to high throughput capacity.

Reporter-based Screening

Figure 13: Reporter-based screening to identify conditions that activate targeted BGCs.

Targeted activation of BGCs generally requires genetic manipulation of host genome, which is intrinsically low throughput. With the adaptability of putting a reporter gene under different native promoters, reporter-based screening enables high throughput screening of multiple BGC targets for identification of activated clusters.

1.4.3.1.4. Strategies for heterologous expression of BGCs

Despite widespread interest in natural product discovery in actinomycetes, most genetic tools and protocols have been developed for *Streptomyces*, the largest and most well studied genus of Actinomycetales. Given that non-streptomycete actinomycetes are also rich sources of natural products and that secondary metabolism diversity is correlated to taxonomic diversity,²¹⁶⁻²¹⁸ efforts to explore the 'dark matter' encrypted by silent BGCs in these actinomycetes cannot be forsaken. While there has been successful attempts to develop synthetic biology tools for manipulating non-streptomycete actinomycetes, such as *Saccharopolyspora erythraea*²⁰¹ and other *Micronomospira* strains²¹⁹, majority of the wild type actinomycetes remain genetically intractable. Time-consuming optimization of protocols for genetic manipulation of individual actinomycete strains makes heterologous expression of BGCs from these strains a more feasible option.

Heterologous expression of targeted BGCs requires two essential tools – well-characterized actinomycete hosts and robust cloning systems for cloning BGCs. Currently, there are several *Streptomyces* hosts for heterologous expression of actinomycete-derived BGCs. Their selection is merited by possession of one or more favourable characteristics like fast growth, small genome size, facile genetic tractability, etc. These strains are typically well-characterized, with consistent growth and

metabolic profiles that aid downstream diagnostics, thus rendering them reliable chassis for heterologous expression. To suit their purpose for secondary metabolite production, redundant genes may be deleted from their genomes, alongside with the addition of extra copies of useful genes (**FIGURE 14A**).

One of those well characterized actinomycete hosts is *Streptomyces albus* J1047, which grows fast, possesses a small genome size of ~6.8 Mb and is genetically tractable.²²⁰ It has been successfully applied for heterologous production of compounds like napyradiomycin,²²¹ cyclooctatin,²²² iso-migrastatin,²²³ thiocoraline,²²⁴ etc. Recently, mutations had been introduced to the strain to regulate its metabolic flux for increased secondary metabolism capacity.²²⁵ *Streptomyces avermitilis*, notable for producing avermectin,²²⁶ is another host that had undergone deletion of more than 1.4 Mb of non-essential genes in order to increase its finesse for heterologous expression of BGCs.²²⁷ One of the genome minimized mutants, *Streptomyces avermitilis* SUKA17, was subsequently used as host for heterologous expression of several BGCs.²²⁸⁻²²⁹ *Streptomyces coelicolor*, which is a model organism for *Streptomyces*, has also been developed as a heterologous expression host. Several genes were deleted from the model strain to eliminate the production of known secondary metabolites such as prodiginine and actinorhodin, while additional mutations were made to its 30S ribosomal subunit protein gene (RpsL) and β subunit of its RNA polymerase gene (RpoB) to enhance its heterologous expression capacity.²³⁰

The aforementioned strains remain a fraction of existing heterologous actinomycete hosts, with more efforts placed on developing more strains for such intention.¹⁷² Since successful heterologous expression is dependent on factors like the hosts' metabolic flux, protein stability, precursor availability, etc, experimenting with several host strains provides insurance against failure due to host-associated elements.

Figure 14: Strategy for heterologous expression of BGCs. (A) Optimization of actinomycete host through deletion of redundant genes and overexpression of useful genes to increase finesse for secondary metabolism. (B) Constructed gDNA library, digested gDNA and chemically synthesized PCR fragments are cloned into linearized vector for subsequent expression in a surrogate host.

Targeted BGCs can be cloned using several strategies (**FIGURE 14B**). However, the cloning method used is typically dependent on the size of the cluster. The most conventional way of cloning involves the construction of genomic DNA (gDNA) library of an actinomycete using cosmid²³¹ or bacterial artificial chromosome vectors²³². These vectors can accommodate insert size of around 40 kb (cosmid) and 100 kb Bacterial Artificial Chromosome (BAC), hence, facilitating studies of actinomycete derived BGCs, which typically range from 20 to 200 kb in size. Despite larger average size of DNA fragments warranted by BAC library,²³³⁻²³⁴ ease of cosmid DNA library preparation versus the meticulous experimental procedure of preparing and handling large DNA fragments makes cosmid library preparation a more common option. Its application has also been expanded to the construction of environmental DNA library. Construction of gDNA library is advantageous if a strain contains several BGCs of interest. Its application has also been expanded to the construction of environmental DNA library. Otherwise, a directed cloning approach like Transformation associated Recombination (TAR) cloning or Gibson assembly can be applied.

TAR cloning is a cloning technique that was initially developed in 1996 to facilitate cloning of large DNA fragments while reducing DNA handling which could result in DNA artefacts.²³⁵ It is performed *in vivo* by utilizing the effective homologous recombination system in *S. cerevisiae* to clone large fragments of DNA into a Yeast

Artificial Chromosome (YAC) vector via homologous recombination. Since then, several optimization studies have been performed to improve this technique.²³⁶⁻²³⁹ The efficiency and versatility of TAR cloning resulted in its wide application in the field of molecular biology; studying genetic variation in microbial gene locus to cloning an entire bacterial genome.²⁴⁰⁻²⁴³ Despite the low GC content of *S. cerevisiae* genomic DNA, TAR cloning has been successfully applied to high GC,²⁴⁴ proving that the homologous recombination repair system in yeast was not biased against high GC DNA. In recent years, TAR cloning has been developed as a tool to clone BGCs for biosynthetic studies using heterologous expression,^{191, 243, 245} thus effectively sidestepping the requirement of a gDNA library. Its functionality can also be applied to assembly of large BGCs from chemically synthesized DNA fragments²⁴⁶⁻²⁴⁷ or truncated DNA inserts,²⁴³ rendering it a useful tool for natural product biosynthesis. While TAR cloning is performed in yeast, Gibson assembly is an *in vitro* cloning technique that exploits homology between DNA fragments for seamless assembly into a single fragment.²⁴⁸⁻²⁴⁹ While a 72 kb cluster encoding pristinamycin had been successfully cloned in a stepwise manner using Gibson Assembly for subsequent heterologous expression,²⁵⁰ its usefulness was further enhanced by the development of a powerful technique called Cas9-assisted targeting of chromosome (CATCH).²⁵¹ CATCH harnesses the flexibility of CRISPR/Cas9 in cleaving genomic DNA at any desired regions, containing the essential recognition sequences, to generate homologous ends that enable a single step assembly with a linearized vector via Gibson assembly.

As natural product discovery and biosynthesis progressively shifts towards a genomics based approach, other cloning techniques like Linear plus linear homologous recombination (LLHR)²⁵² are continually being experimented and developed for heterologous expression. The progressive nature of heterologous expression is significant as more than 90% of all bacteria are unculturable, making this approach a prospective and possibly, only option for accessing the concealed supply of compounds reserved by these bacteria.

1.5. Aims and Objectives

This project has two aims. The first and primary aim is the discovery of novel natural products from the lab's actinomycete culture collection. The secondary aim involves

studying their biosynthesis through mechanistic studies guided by bioinformatic analysis of genomic data.

To achieve these aims, both chemical profiling and genomics guided approaches were used in tandem to prioritize strains for novel compounds discovery. The chemical profiling approach was implemented with a robust conventional chemical profiling platform using HPLC instrumentation. Genomic mining of actinomycete strains via bioinformatic analyses guided dereplication efforts and identification of putative BGCs of interest. Instrumental modifications and microbial workflows were developed for subsequent purification of target molecules from crude extracts of fermentation cultures, while spectral data was complemented with bioinformatic predictions for structural elucidation of natural products. For biosynthetic studies and work on silent BGCs, existing synthetic biology tools were used to genetically manipulate host actinomycete strains.

2. Chapter 2: Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a *Streptomyces* strain isolated from mangrove sediment

2.1. Introduction

The quest for discovery of novel natural products from actinomycetes is impossible without actinomycete strains. While culture collections promise an ample supply of actinomycete strains for immediate investigation, the restrictive conditions bound by agreements means that useful applications derived from any research work on those strains will be hindered by potential contractual implications. This compelled efforts to survey the vastly untapped natural habitats of Singapore for rare actinomycetes. A combination of pre-treatment and isolation strategies were applied for selective isolation of actinomycetes,^{20, 22} while *in situ* trapping was also attempted to isolate rare, dormant actinomycetes.²⁵ Overall, a collection of nearly 200 actinomycetes was eventually isolated from three locations - Sungei buloh Mangrove reserve, Pulau Ubin quarry lake and secondary forest within the vicinity of Nanyang Technological University. One of those initially isolated actinomycete strains is *Streptomyces* sp. SD85, which was isolated from Sungei buloh Mangrove reserve. *Streptomyces* are the most well studied genus of actinomycetes and possess more BGCs on average as compared to other actinomycetes.²⁵³ The prospect of manipulating *Streptomyces* sp. SD85's genome with readily available tools developed for *Streptomyces*, coupled with its fast growing characteristic and huge collection of BGCs, encouraged its prioritization for further investigation.

Preliminary metabolite profiling of *Streptomyces* sp. SD85 revealed a putative polyene compound which corresponded to an uncharacterized polyene macrolactam BGC identified from genome mining. Actinobacteria produce a superfamily of macrolactams using polyketide biosynthetic pathways. Within the macrolactam superfamily, there is a structurally distinct group of macrolactams containing a polyene skeleton and a nitrogen-containing moiety derived from L-glutamate via a β -amino acid starter unit (**FIGURE 15**). This group of polyene macrolactams can be further divided into two subgroups according to the different β -amino acid starter units. Members of the first subgroup are synthesized using a (2S, 3S)-3-methylaspartate (3-meAsp) as starter unit²⁵⁴⁻²⁵⁷; whereas members of the second subgroup are synthesized using a 3-

aminobutyrate as starter unit ²⁵⁸⁻²⁶⁰. Both 3-meAsp and 3-aminobutyrate are synthesized from L-glutamate by dedicated pathways. The biosynthetic gene clusters (BGC) for vicanistatins, micromonosporin A, lobosamide A and mirlactam A have been identified and the biosynthetic mechanisms have been proposed ^{255, 259-262}. The proposed biosynthetic mechanisms share some common features, with the macrolactam skeleton assembled by a Type I modular PKS system from the L-glutamate-derived starter unit and malonyl-CoA or methylmalonyl-CoA extender units. The PKS product undergoes cyclization and tailoring to yield the final macrolactam. The so-called collinearity rule²⁶³⁻²⁶⁴ for polyketide biosynthesis is strictly followed in the reported biosynthetic mechanisms. Structural diversity in the polyene macrolactams is generated via the utilization of different starter units (i.e. 3-meAsp or 3-aminobutyrate), varying composition of PKS modules and different tailoring enzymes.

This chapter describes a genomics-guided approach to study natural product discovery in *Streptomyces* sp. SD85, which successfully demonstrated the effectiveness of genome mining as a tool for dereplication of known compounds and identification of interesting BGCs. Notably, sceliphrolactam, a polyene macrolactam that shares structural similarity and potentially similar biosynthetic mechanism with vicanistatin and other known macrolactams, was isolated. Subsequent genetic manipulation of the putative sceliphrolactam BGC with a powerful genome editing tool, CRISPR/Cas9, validated the relationship between the cluster and molecule.

Figure 15: Polyene macrolactams that contain a nitrogen-containing moiety derived from L-glutamate via a β -amino acid starter unit. All the representative polyene macrolactams shown here are produced by members of the *Streptomyces* genus with the exception of micromonosporin and macrotermycin A, which are produced by *Micromonospora* and *Amiclatopsis* strains respectively.

2.2. Methods and materials

2.2.1. *Streptomyces* sp. SD85 isolation and strain maintenance

Streptomyces sp. SD85 was isolated from marine sediment samples collected from the Sungei Buloh Wetland Reserve, Singapore. The sediment samples were collected from the sediment (1–2 meter depth) in the mangrove swamps to harvest the microbial communities from the submerged sediment. The sediment samples were rinsed and filtered after being air dried and treated with microwave irradiation. With the filtrate, a series of 10-fold dilutions were performed and a total of 0.1 mL from each dilution was plated onto the International *Streptomyces* Project 3 (ISP3) medium plate. 30 µg/mL Cycloheximide and 25 µg/mL nalidixic acid were used to inhibit fungal and fast growing Gram-negative bacteria. All culture plates were incubated for up to three weeks. Individual colonies were obtained after a few rounds of streaking and the strain was maintained in GYM medium agar (4 g/L glucose, 4 g/L yeast extract, 10 g/L malt extract, 20 g/L agar, pH 7.2) at 28 °C.

2.2.2. Genomic DNA isolation from *Streptomyces* sp. SD85

High quality genomic DNA was extracted using a modified genomic DNA extraction protocol for *Streptomyces* ²⁶⁵. Frozen spores of *Streptomyces* sp. SD85 was inoculated in a 50 mL liquid CRM medium (10 g/L glucose, 103 g/L sucrose, 15 g/L tryptic soy broth powder, 5 g/L yeast extract, 0.12 g/L MgCl₂·6H₂O) and grown at 30 °C, 180 RPM for 2-4 days until it reached late log phase of growth. The mycelium pellets were harvested from liquid culture by centrifuging at 2000 *g* for 5 min and subsequently washed with 15 mL of Lysis buffer (25 mM Tris-Cl pH 8, 25 mM EDTA pH 8, 0.3 M sucrose). Washed mycelia was centrifuged again and resuspended in 8 mL of Lysis buffer. The mycelia suspension was homogenized with a 15 mL handheld Dounce tissue grinder until finely dispersed. Lysozyme solution (50 mg/mL) was added to the cell suspension until an eventual concentration of 2 mg/mL and incubated at 37 °C for 1-2 hr. The lysozyme-treated cells were lysed by adding 10% Sodium dodecyl sulfate (SDS) to a final concentration of 0.5% with gentle mixing. Proteinase K solution (20 mg/mL) was subsequently added and mixed to the lysed culture until a final concentration of 0.18 mg/mL and incubated at 55 °C for 1-2 hr for complete protein digestion. Upon the completion of protein digestion, 1.18 mL of 5 M NaCl solution was

added to the solution with gentle mixing before 0.68 mL of CTAB/NaCl solution (10% cetyltrimethylammonium bromide (CTAB) in 0.7 M NaCl solution) and 0.3 mL 10% Polyvinylpyrrolidone (PVP) solution were subsequently added with incubation at 55 °C for 15 min. The mixture was cooled to 37 °C and mixed gently with equal volume of (25:24:1) Phenol-chloroform-isoamyl alcohol solution until homogenous. Phases were separated by centrifuging at a slow speed of 3000 *g* for 10 min and the aqueous phase, containing genomic DNA, was collected with wide bore pipette. Phenol-chloroform-isoamyl alcohol treatment was repeated again and the aqueous phase was washed twice with equal volume of chloroform to remove residual phenol. Aqueous phase was supplemented with RNase A to a final concentration 100 µg/mL and incubated at 37 °C for 1-2 hr for RNA digestion. High molecular weight genomic DNA was finally spooled from the aqueous mixture by adding 0.6 volume of isopropanol. Spooled DNA was washed twice with 5 mL of 70% aqueous ethanol and air dried under ambient conditions. Genomic DNA was subsequently re-dissolved in 10 mM Tris buffer, pH 8.0 and aliquoted into small volumes for storage in -20 °C freezer.

2.2.3. Genome sequencing of *Streptomyces* sp. SD85

The genomic DNA was first sequenced using Illumina Miseq technology to obtain a draft genome which contains large gaps. Subsequent complete genome sequence of *Streptomyces* sp. SD85 was obtained using Single Molecule Real Time (SMRT) sequencing technology (Pacific Biosciences, California, USA) offered by PacBio RS II platform technology. A single SMRT cell was used and the sequencing run was performed 3 times. The PacBio long sequencing reads have reached an approximate 114 X coverage of the genome and were successfully assembled into a linear chromosome with a size of 8,625,764 bp by using the Hierarchical Genome Assembly Process 2 (HGAP2) protocol from SMRT Analysis version 2.0. The DNA sequences for the scilipholactam biosynthetic gene cluster obtained from Illumina and PacBio sequencing are identical and have been deposited in GenBank (Accession number: KX230849). For Illumina MiSeq sequencing, a total of 3,906,106 reads were obtained with an average read length of 145.17 bases. 3,896,491 reads were matched with an average length of 145.34 bases. Sequencing reads were trimmed and *De novo* assembly was performed using CLC Genomics Workbench (CLC bio, Denmark). A total of 112 contigs were obtained with an average length of 76,406 bases and a total

length of 8,557,547 bases. For the long-read sequencing performed the PacBio technology, a single SMRT cell was used with a three sequencing runs. A read quality value of 84 was achieved with 86,939 zero-mode-waveguides (ZMWs), with average coverage value of 114.0. Mean polymerase read length was 15,610 bp with mean reads of inserts (ROI) readlength value of 13,164 bp. A single contig of length 8,625,764 bases was obtained after assembly.

2.2.4. Metabolite Profiling of *Streptomyces* sp. SD85

Streptomyces sp. 85 was cultured in GYM liquid medium (100 mL in 500 mL baffled Erlenmeyer flask) for 96 hr at 28 °C. Mycelia were separated from culture broth by centrifugation. Mycelia and culture broth were extracted separately with methanol and ethyl acetate, respectively. The organic layers were dried under vacuum and re-dissolved in methanol. HPLC analysis of the crude extract was performed with Grace® VisionHT™ C18-HL (4.6 mm x 250 mm, 5 µm) using an Agilent 1200 HPLC system equipped with DAD detector. A 60 min gradient elution at 1 mL/min was employed with 10% aqueous acetonitrile to 100% acetonitrile. Mobile phase solvents were supplemented with 0.1% formic acid. Analytes were monitored at λ = 220 nm, 280 nm, 330 nm, 400 nm. LC-MS was performed using Thermo Scientific Accela LC system coupled with an LTQ XL™ Linear Ion Trap Mass Spectrometer. The conditions for LC-MS were similar to those employed for HPLC-DAD analysis. Mass spectra were measured using positive and negative ion mode of ESI (capillary voltage 3.5 kV; cone voltages 30V/50V; source temperature 400 °C; cone gas flow 60 L/hr; desolvation gas flow 450 L/hr) with a mass range of 100 – 2000 Da and a scan time of 0.5 s.

2.2.5. Large scale fermentation of *Streptomyces* sp. SD85 and isolation of sceliphrolactam

Streptomyces sp. SD85 was cultured in GYM liquid medium (800 mL in 2 L baffled Erlenmeyer flask) for 7 days at 29 °C, in the dark. Combined mycelia from four culture flasks were extracted with one liter of acetone twice. The organic layer was dried under vacuum to afford a crude extract (~4 g). The extract was re-dissolved in 90% methanol and partitioned four times with equal volume of hexane to remove unwanted lipid material. The washed extract was subjected to semi-preparative reversed phase HPLC using Shimadzu Prominence Preparative HPLC system equipped with Zorbax

Eclipse XDB C18 semi-prep column (9.4 mm x 250 mm, 5 μ m). An isocratic solvent system of H₂O:CH₃OH (45:55) with a flow rate of 4.7 mL/min was used for compound isolation. Elution of the compounds was monitored at 333 nm and a peak corresponding to sceliphrolactam was observed at t = 22 min. Pure sceliphrolactam (2 mg/L) was obtained under dark conditions to minimize degradation of the polyene compound.

2.2.6. Construction of *Streptomyces* sp. SD85 Δ sceN mutants via CRISPR/Cas9 system

Construction of the plasmid pCRISPR-Cas9-SceN for CRISPR/Cas9-mediated *sceN* gene mutation in *Streptomyces* sp. SD85 was performed using pCRISPR-Cas9, which was a kind gift from Sang Yup Lee's lab in the Korea Advanced Institute of Science and Technology (KAIST). All cloning steps were carried out using *E. coli* TOP10 (Invitrogen, US). Identification of protospacer sequences in *sceN* was performed using the Benchling Server. Selected *sceN* DNA spacer was introduced into pCRISPR-Cas9 by ligating a PCR-generated sgRNA sequence into *Sna*BI and *Nco*I linearized pCRISPR-Cas9 vector. To generate an 883 bp deletion via homology directed repair, two DNA templates that are homologous to the 5'- and 3'- regions of the *sceN* gene and are 883 bp apart from each other were amplified from genomic DNA using Phusion® high fidelity DNA polymerase. The *sceN* spacer-containing pCRISPR-Cas9 was digested using *Stu*I restriction enzyme and jointed with the template DNA fragments using Gibson Assembly kit (New England BioLabs, US). The sgRNA and template-containing pCRISPR-Cas9 plasmid was introduced into the *Streptomyces* sp. SD85 parental strain by conjugation method using *E. coli* ET12567/pUZ8002 as described ²⁶⁶. Briefly, *Streptomyces* spores were collected from seven-day old culture grown on GYM agar containing 10 mM MgCl₂ and 10 mM CaCl₂, and subjected to heat shock at 45 °C for 5 min. They were allowed to germinate at 30 °C for 6 hr and mixed with an exponential phase culture of *E. coli* ET12567/pUZ8002 containing the modified pCRISPR-Cas9 plasmid at a donor to recipient ratio of 3:1. The resultant mixture was plated onto MS agar supplemented with 40 mM MgCl₂ and 20 mM CaCl₂, and incubated for 22.5 hr at 30 °C. One millilitre of sterile water containing 1 mg of apramycin and 500 μ g of nalidixic acid was overlaid onto the plates for selection of successful exconjugants. Plates were further incubated at 30 °C for 4-5 days before

exconjugants were picked onto fresh MS agar plates containing 40 µg/mL apramycin and 25 µg/mL nalidixic acid.

Exconjugants harbouring the modified pCRISPR-Cas9 plasmid were determined by colony PCR. True exconjugants were patched onto MS agar that contain 40 µg/mL apramycin and 10 µg/mL thiostrepton for induction of Cas9 expression. Clearance of plasmids was achieved by patching resultant colonies onto MS agar and incubating them at 37 °C for 4 days. Apramycin sensitivity was confirmed by replica plating onto selective and non-selective plates.

To screen for successful mutants, individual clones were cultured in liquid GYM at 30 °C for 3 days for genomic DNA isolation using a previously described protocol.²⁶⁷ Briefly, mycelia was harvested and resuspended in 0.5 mL Lysis buffer (400 mM Tris-HCl pH 8.0, 60mM EDTA pH 8.0, 150mM NaCl and 1% SDS) and incubated at 55 oC for 10 min. 0.15 mL of potassium acetate solution (60% v/v 5 M potassium acetate, 11.5% v/v glacial acetic acid, 28.5% v/v ultrapure water, pH 4.8) was added and the mixture was vortexed vigorously for 2 min to ensure thorough mixing. The resultant suspension was centrifuged at 10,000 g for 5 min and the supernatant was mixed with equal volume of isopropanol. Upon thorough mixing, crude nucleic acids from the mixture was precipitated and washed twice with 70% aqueous ethanol. The DNA pellets were dried at 55 oC for 5 min and re-dissolved in sterile water for downstream analysis. Diagnostic PCR was performed to establish a deletion of 883 bp in the *sceN* gene.

2.2.7. CRISPR/Cas9 mediated *sceQ-R* gene fusion

A dual sgRNA-containing pCRISPR-Cas9 plasmid was constructed by designing a dual sgRNA synthetic construct with a configuration of (*Nco*I site – spacer - sgRNA1tracr – terminator -*gapdhp* (EL) - spacer - sgRNAtracr – *Sna*BI site). This construct was cloned into *Sna*BI and *Nco*I linearized pCRISPR-Cas9 vector by Gibson Assembly. The resulted plasmid was linearized using the *Stu*I restriction enzyme and ligated with the HDR template to yield the plasmid pCRISPR-cas9-dual-sceQR. Introduction of this plasmid into *Streptomyces sp. SD85* and subsequent induction of Cas9 expression was performed as described above. Successful mutant strains were

validated by diagnostic PCR and sequencing of the fragments amplified from genomic DNA of individual clones.

2.2.8. HPLC analysis of sciliphrolactam production for the *Streptomyces* sp. SD85 Δ sceN and sceQ-R fusion mutants

The mycelia of 7 days old wild-type (WT) *Streptomyces* sp. SD85 and the mutant strains were extracted twice with equal volume of acetone and dried under vacuum. The crude extracts were re-dissolved in methanol and subjected to HPLC analysis with an Agilent 1200 HPLC system equipped with DAD for UV detection. A Gradient elution program was employed, starting with 40% aqueous methanol at 0 min to 70% aqueous methanol at 50 min and 100% methanol at 60 min. The analytes were monitored at $\lambda = 220$ nm, 280 nm, 330 nm, 420 nm. Experiments were repeated twice to ensure reproducibility.

2.2.9. Primers used in this study

Primers used for this study were synthesized by Integrated DNA Technologies, US. Sequences of all primers in this study are available in **TABLE S 1**.

2.3. Results and discussion

2.3.1. *Streptomyces* sp. SD85 produces sciliphrolactam and other secondary metabolites under laboratory cultivation conditions

Streptomyces sp. SD85 was initially identified as an actinomycete through 16S rRNA gene sequencing. It exhibited fast growth, producing reddish colonies with thick aerial mycelia on GYM agar (**FIGURE 16**) After the cultivation in GYM *Streptomyces* medium, the secondary metabolites produced by *Streptomyces* sp. SD85 were extracted from the culture broth and mycelium using ethyl acetate and methanol respectively. The secondary metabolites were subsequently profiled using liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS). As genome mining of the strain identified three known BGCs (**FIGURE 17**) encoding factumycin, filipin and streptorubin biosynthesis, these compounds were highlighted for chemical dereplication. With the aid of an in-house natural product library, these compounds were subsequently identified by chemical profiling (a member of prodiginine family,

$m/z = 392.2694$ $[M+H]^+$, calculated $m/z = 392.2701$ $[M+H]^+$, m/z error: -1.784485 ppm), 6,8-O-dimethylreticulol ($m/z = 251.0916$ $[M+H]^+$, calculated $m/z = 251.0919$ $[M+H]^+$, m/z error: -1.194782 ppm), 6-O-dimethylreticulol ($m/z = 237.0760$ $[M+H]^+$, calculated $m/z = 237.0763$ $[M+H]^+$, m/z error: -1.265415 ppm), filipin III ($m/z = 655.4048$ $[M+H]^+$, calculated $m/z = 655.4057$ $[M+H]^+$, m/z error: -1.373195 ppm) and factumycin ($m/z = 801.4301$ $[M+Na]^+$, calculated $m/z = 801.4302$ $[M+Na]^+$, m/z error: -0.124777 ppm). Apart from the exception of dimethylreticulol derivatives, the agreement of chemical profiling outcome with genome mining predictions demonstrated the reliability of genome mining as part of dereplication efforts. The red pigment streptorubin B was produced in abundance and accounted for the red/orange appearance of the colonies (FIGURE 16).

Figure 16: Growth of *Streptomyces* sp. SD85 on solid medium. *Streptomyces* sp. SD85 was inoculated on a modified ISP 2 medium (4 g/L glucose, 4 g/L yeast extract, 10 g/L malt extract, 2 g/L calcium carbonate and 12 g/L of bacto agar) and incubated at 28 °C for 4 days.

In addition, *Streptomyces* sp. SD85 produces several potentially new compounds whose spectral signatures were not found in the database. The structure for one of the “unknown” compounds was established using HRMS and NMR spectroscopy (FIGURE S 3 TO FIGURE S 9, TABLE S 2). This compound, which was isolated from the mycelium, turned out to be sceliphrolactam ($m/z = 482.2532$ $[M+H]^+$, calculated for $C_{28}H_{35}NO_6$ 482.2542 $[M+H]^+$, m/z error: -2.073595 ppm), which is a 3-amino-2-methylpropionate-containing polyene macrolactam that was isolated recently from a wasp-associated *Streptomyces* strain²⁶².

2.3.2. Complete genome sequencing suggests that *Streptomyces* sp. SD85 is a biosynthetically talented strain

Genome of *Streptomyces* sp. SD85 was initially sequenced via Miseq from Illumina sequencing technology. The identification of a putative polyene macrolactam BGC from Miseq sequencing which disagreed with the structure of sceliphrolactam (two carbons more than prediction), led to suspicion of errors in the draft genome sequences generated from Miseq sequencing. Considering that the short reads from Illumina sequencing may cause mis-assembly, the genome was re-sequenced for the second time using SMRT technology (PacBio) to obtain the complete genome. Complete genome from PacBio sequencing revealed the same DNA sequence for the putative polyene macrolactam BGC, reassuring that its integrity was not compromised by any sequencing errors. Additionally, full genome sequence, which is highly similar to the draft genome except for the gap regions, revealed that the chromosome consists of 8,625,724 base pairs with an average GC content of 72.3%. Phylogenetic analysis based on 16S rDNA sequences suggests *S. hiroshimensis* NBRC 3839(T) as the closest relative with a shared nucleotide identity of 99.52%.

AntiSMASH²⁶⁸ analysis using antiSMASH 3.0 suggested that the genome contains at least 52 BGCs for the biosynthesis of different classes of secondary metabolites. The 52 BGCs occupy a total of 1.82 Mb and constitute 21.2% of the genome, which is higher than the 10.8% and 16.6% occupancy for *S. coelicolor* and *S. avermitilis*²⁶⁹⁻²⁷⁰. Like *S. coelicolor* and some other *Streptomyces* strains, most of the BGCs reside in the two subtelomeric regions of the genome (**FIGURE 17A**). More than half of the BGCs are predicted to produce polyketide and ribosome or non-ribosome peptide-derived secondary metabolites (**TABLE 2**). Seventeen BGCs have genes that encode type I, II or III PKS systems, including three type I modular PKS systems. Ten BGCs possess genes predicted to produce lantipeptide, thiopeptide or lasso-peptides; and eleven BGCs contain NRPS proteins. A few terpene BGCs were identified as evidenced by the presence of terpene synthase genes. Several BGCs are predicted to be hybrid BGCs containing genes that code for more than one type of scaffold-synthesizing enzymes (**TABLE 2**). Considering that fewer than ten secondary metabolites were detected in our LC/MS-based metabolite profiling, some of the BGCs could be suppressed due to transcriptional or translational regulation.

For PKS and NRPS proteins to be enzymatically active, the carrier protein domains of the PKS and NRPS proteins must undergo phosphopantetheinylation, a post-translational modification catalyzed by Sfp or AcpS-type 4'-phosphopantetheinyl transferases (PPTase) ²⁷¹⁻²⁷². The *Streptomyces sp.* SD85 genome harbors at least three genes encoding Sfp-type PPTases and one gene-encoding AcpS-type PPTase. All four PPTases are predicted to be discrete proteins with none of them integrated into the PKS or NRPS proteins as observed for some PKSs²⁷³.

Table 2: AntiSMASH-predicted BGCs for *Streptomyces* sp. SD85

BGC	Position		Product/type
	From	To	
Cluster 1	128373	191397	NRPS
Cluster 2	191639	228433	Lantipeptide
Cluster 3	254893	292628	T3PKS
Cluster 4	338768	374816	Thiopeptide
Cluster 5	449647	472959	Lantipeptide
Cluster 6	544798	631900	Filipins
Cluster 7	632222	683087	NRPS
Cluster 8	752764	792909	Phosphonate
Cluster 9	814767	848264	NRPS
Cluster 10	1079383	1094356	Terpene
Cluster 11	1237686	1312223	Sceliphrolactam
Cluster 12	1368633	1383668	NRPS
Cluster 13	1460470	1474512	NRPS-T1PKS
Cluster 14	1544288	1554571	Bacteriocin
Cluster 15	1588034	1600491	Terpene
Cluster 16	1646863	1698126	NRPS-T1PKS
Cluster 17	1714938	1736137	Terpene
Cluster 18	2135476	2146348	Bacteriocin
Cluster 19	2170308	2196218	Siderophore
Cluster 20	2196485	2241910	T3 PKS-Terpene
Cluster 21	2365889	2376287	Ectoine
Cluster 22	2829683	2850624	Lantipeptide
Cluster 23	3125189	3133324	Melanin
Cluster 24	4226586	4269086	T2 PKS
Cluster 25	4476131	4497579	Lasso peptide
Cluster 26	4624864	4667988	T3 PKS
Cluster 27	4985954	5012231	Lantipeptide
Cluster 28	5634528	5639110	Melanin
Cluster 29	5999342	6017297	Siderophore
Cluster 30	6504488	6559786	NRPS
Cluster 31	6670408	6706365	Prodiginine
Cluster 32	6734497	6758278	Terpene
Cluster 33	6779444	6838577	T1 PKS-Terpene
Cluster 34	6838626	6874184	NRPS
Cluster 35	6948157	6973315	Lantipeptide
Cluster 36	7169469	7244251	PKS/FAS
Cluster 37	7248671	7339459	Factumycin
Cluster 38	7363591	7384450	Lasso peptide
Cluster 39	7568556	7585720	Terpene
Cluster 40	7611812	7663138	NRPS
Cluster 41	7695774	7723406	Lantipeptide
Cluster 42	7805055	7854201	T1 PKS-Terpene
Cluster 43	7997743	8031379	NRPS-T1 PKS
Cluster 44	8083462	8121926	NRPS
Cluster 45	8125632	8127719	γ -Butyrolactone
Cluster 46	8128024	8185901	PKS/FAS
Cluster 47	8251315	8271035	Terpene
Cluster 48	8274362	8305247	Terpene
Cluster 49	8312476	8367627	PKS/FAS
Cluster 50	8409881	8446192	Unknown
Cluster 51	8465511	8538072	PKS/FAS
Cluster 52	8549285	8564374	Lantipeptide

2.3.3. Identification of sceliphrolactam BGC by gene inactivation

With the availability of complete genome sequence, the sceliphrolactam-producing *Streptomyces sp.* SD85 strain provided an opportunity to identify the sceliphrolactam BGC and decipher the biosynthetic pathway. As the macrolactam skeleton of sceliphrolactam is most likely generated by a type I modular PKS system, we first searched BGCs that contained type I modular *pks* genes. The genome of *Streptomyces sp.* SD85 harbors several BGCs that are predicted to contain type I *pks* or *fas* (fatty acid synthase) genes, but only three (BGC6, 11 and 37 (**TABLE 2**) encode type I modular PKS systems. Based on the high similarity shared with the BGCs from the MIBiG database ²⁶⁸, BGC6 and BGC37 are predicted to produce filipins and kirromycin respectively (**FIGURE 17B**) ²⁷⁴⁻²⁷⁶. Assignment of the two BGCs is consistent with the detection of filipin III and factumycin (a close structural analogue of kirromycin) ²⁷⁷ in the culture broth. BGC11 encodes six PKS proteins and a dozen of other enzymes, transporters and regulatory proteins (**FIGURE 17C**). Three uncharacterized BGCs that share high similarity with BGC11 were also found in the genomes of *Streptomyces sp.* SirexAA-E, *Streptomyces sp.* PAMC26508 and *S. flavogriseus* ATCC3331 (**FIGURE 17D**).

Figure 17: Complete genome of *Streptomyces* sp. SD85 and the PKS-based biosynthetic gene clusters (BGCs) with confirmed products. A. Circular representation of *Streptomyces* sp. SD85 chromosome. The inner ring shows a normalized GC skew plot and the center ring depicts a normalized GC content plot. The outer ring shows the distribution of secondary biosynthetic gene clusters

(represented by the bars). B. Comparison of the putative filipin, Streptorubin B (a prodiginine) and factumycin gene clusters from *Streptomyces* sp. SD85 with characterized homologous gene clusters from other *Streptomyces* strains. C. Organization of the putative sceliphrolactam gene cluster (See **TABLE 3** for detailed gene annotation). D. Homologous gene clusters identified in three other *Streptomyces* strains.

To find out whether BGC11 is responsible for sceliphrolactam biosynthesis, a CRISPR/Cas9 strategy was used to delete an internal fragment of the *sceN* gene, which encodes the first of the six PKS modules. The pCRISPR-Cas9 plasmid was employed to express Cas9, sgRNA and a DNA template required for homology directed repair (HDR) ²⁷⁸. After cloning the sgRNA and homologous DNA sequences into the pCRISPR-Cas9 plasmid, we transformed *Streptomyces* sp. SD85 with the modified plasmid by the *E. coli* - *Streptomyces* conjugation method ²⁷⁸. After Cas9-mediated DNA cleavage and repair, DNA gel analysis and DNA sequencing confirmed that a portion of *sceN* was successfully deleted (**FIGURE 18A, B**). It should be noted that five sgRNAs were employed to target different protospacer adjacent motif (PAM) sites within the *sceN* gene and only one of the sgRNAs (sgRNA2) led to DNA cleavage and repair (**FIGURE 18B**). As already documented by other researchers, the efficiency of Cas9/sgRNA-mediated gene deletion in bacteria can vary greatly. ²⁷⁹ Subsequent metabolite analysis by HPLC confirmed the absence of sceliphrolactam in the culture broth of the Δ *sceN* mutant strain (**FIGURE 18C**). The gene inactivation experiment supports that the *sce* BGC encodes the enzymes and other proteins required for sceliphrolactam biosynthesis. This conclusion is further reinforced by the sequence similarity shared by the biosynthetic genes between *sce* and the vicenistatin BGC as discussed in the following subchapter 2.3.4.

Figure 18: Inactivation of sceliphrolactam biosynthetic gene *sceN* using CRISPR/Cas9-based method. A. Schematic illustration of the CRISPR/Cas9-mediated cleavage of genomic DNA and homology directed repair (HDR) to delete part of *sceN*. B. PCR results confirmed the deletion of 883 base pairs of *sceN* using sgRNA2 as guide. A full-sized image of the DNA gel is included in the supporting information. C. HPLC analysis of the Δ *sceN* mutant strain to show the abolishment of sceliphrolactam production. The sceliphrolactam peak is indicated by the arrow. The wavelength (λ) was set at 330 nm for the HPLC detector (Inset: on-line absorption spectrum of sceliphrolactam).

2.3.4. Annotation of non-PKS biosynthetic genes in the *sce* gene cluster

In addition to the five PKS-encoding genes, the *sce* BGC also contains other genes predicted to code for non-PKS enzymes, membrane-embedded ABC transporters and two LuxR-type transcriptional regulators. The predicted functions of the biosynthetic

enzymes, transporters and regulatory proteins are summarized in **TABLE 3**. Annotation of the *sce* genes provides further support for the involvement of the *sce* BGC in sceliphrolactam biosynthesis.

Table 3: Predicted function of the genes from the sceliphrolactam BGC.

Gene	Length of encoded protein	Closest homolog [source]	Identity /Similarity (%)	Proposed function	UniProt ID
<i>sceA</i>	456	Integral membrane transporter [<i>Streptomyces</i> sp. 769]	79/88	MFS transporter	A0A0A8EUD2
<i>sceB</i>	384	Thioredoxin [<i>Burkholderia cenocepacia</i> KC-01]	45/58	Electron transfer protein	V4ZEC6
<i>sceC</i>	73	Ferredoxin [<i>Saccharothrix</i> sp. ST-888]	61/82	Electron transfer protein	A0A0F0HKK3
<i>sceD</i>	406	Cytochrome P450 [<i>Streptomyces</i> sp. <i>SirexAA-E</i> / <i>ActE</i>]	81/87	Cytochrome P450 monooxygenase	G2NJJ2
<i>sceE</i>	404	Cytochrome P450 hydroxylase [<i>Streptomyces hygrosopicus</i> subsp. <i>jinggangensis</i> 5008]	81/88	Cytochrome P450 monooxygenase	H2JPP4
<i>sceF</i>	298	Amino acid amidase [<i>Saccharothrix</i> sp. ST-888]	81/89	Peptidase	A0A0F0HK57
<i>sceG</i>	314	ACP S-malonyltransferase [<i>Saccharothrix</i> sp. ST-888]	79/86	Acyl carrier protein aminoacyltransferase	A0A0F0HLD6
<i>sceH</i>	87	Phosphopantetheine-binding protein [<i>Saccharothrix</i> sp. ST-888]	69/77	Acyl carrier protein	A0A0F0HNA7
<i>sceI</i>	508	AMP-dependent synthetase [<i>Saccharothrix</i> sp. ST-888]	80/86	ATP-dependent aminoacyl-ACP	A0A0F0HHV6
<i>sceJ</i>	512	Peptide synthetase [<i>Saccharothrix</i> sp. ST-888]	80/87	ATP-dependent aminoacyl-ACP	A0A0F0HNB1
<i>sceK</i>	413	Decarboxylase [<i>Saccharothrix</i> sp. ST-888]	80/89	Decarboxylase	A0A0F0HI66
<i>sceL</i>	163	Methylmalonyl-CoA mutase [<i>Saccharothrix</i> sp. ST-888]	75/86	Glutamate mutase S subunit	A0A0F0HK60
<i>sceM</i>	432	Glutamate mutase E-chain [<i>Streptomyces hygrosopicus</i> subsp. <i>jinggangensis</i> 5008]	75/83	Glutamate mutase E subunit	H2JPP6
<i>sceN</i>	3626	Modular polyketide synthase [<i>Streptomyces hygrosopicus</i> subsp. <i>jinggangensis</i> 5008]	69/76	Polyketide synthase	H2JPP5
<i>sceO</i>	3508	Modular polyketide synthase [<i>Streptomyces hygrosopicus</i> subsp. <i>jinggangensis</i> 5008]	74/81	Polyketide synthase	H2JPP4
<i>sceP</i>	926	LuxR family transcriptional regulator [<i>Streptomyces roseochromogenus</i> subsp. <i>oscitans</i> DS 12.976]	57/70	LuxR family transcriptional regulator	V6KW85
<i>sceQ</i>	1824	Polyketide synthase [<i>Saccharothrix</i> sp. ST-888]	79/85	Polyketide synthase	A0A0F0HIL7
<i>sceR</i>	1582	Beta-ketoacyl synthase [<i>Saccharothrix</i> sp. ST-888]	77/85	Polyketide synthase	A0A0F0HP09
<i>sceS</i>	5150	Modular polyketide synthase [<i>Streptomyces hygrosopicus</i> subsp. <i>jinggangensis</i> 5008]	73/80	Polyketide synthase	H2JPR4

<i>sceT</i>	1294	Modular polyketide synthase [<i>Streptomyces hygroscopicus</i> <i>subsp. jinggangensis</i> 5008]	73/81	Polyketide synthase	H2JPR3
<i>sceU</i>	81	Uncharacterized protein [<i>Streptomyces</i> <i>roseochromogenus subsp.</i> <i>oscitans</i> DS 12.976]	55/69	Unknown	V6KW03
<i>sceV</i>	266	ABC transporter [<i>Saccharothrix</i> <i>sp.</i> ST-888]	87/92	ABC transporter	A0A0F0H J95
<i>sceW</i>	840	ABC transporter permease [<i>Streptomyces sp.</i> PBH53]	68/80	ABC transporter permease	A0A0H4C 931
<i>sceX</i>	433	Histidine kinase [<i>Saccharothrix</i> <i>sp.</i> ST-888]	77/83	Histidine kinase	A0A0F0H HC4
<i>sceY</i>	214	LuxR family transcriptional regulator [<i>Streptomyces</i> <i>roseochromogenus subsp.</i> <i>oscitans</i> DS 12.976]	87/94	LuxR-family transcriptional regulator	V6KY34

The *sce* gene cluster has several genes (*sceG-M*) that share homology with a set of genes found in *S. halstedii* HC-34 for vicenistatin biosynthesis (**FIGURE 19A**). This set of genes (*vinH-O*) is responsible for the synthesis of the 3-amino-2-methylpropionate moiety of vicenistatin from L-glutamate via 3-MeAsp^{254, 280}. Similar to vicenistatin, sceliphrolactam also contains a 3-amino-2-methylpropionate moiety. Considering the high sequence homology shared by the two sets of genes, the 3-amino-2-methylpropionate moiety of sceliphrolactam is likely to be synthesized by a similar pathway. Based on the function of starter-synthesizing enzymes in vicenistatin biosynthesis²⁸¹, SceL and SceM, which are the homologs of VinH (Identity/Similarity: 57%/74%) and VinI (Identity/Similarity: 62%/70%), are likely to form a coenzyme B₁₂-dependent glutamate mutase that catalyzes the transmutation of glutamate (**FIGURE 19B**). SceI, a homolog of VinN (Identity/Similarity: 61%/75%), is predicted to transfer 3-methylaspartate onto the carrier protein SceH (homolog of VinL: Identity/Similarity: 64%/75%). SceK is the homolog of VinO (Identity/Similarity: 67%/79%) and is likely to catalyze the decarboxylation and epimerization of SceH-tethered 3-methylaspartate to 3-amino-2-methyl propionate; whereas SceJ (VinM homolog, Identity/Similarity: 58%/67%) is likely to catalyze the formation of an amide to protect the reactive amino group. SceG shares high sequence homology with the aminoacyltransferase VinK (Identity/Similarity: 69%/86%), with all the catalytic residues of VinK conserved²⁸². Accordingly, SceG is likely responsible for transferring the 3-amino-2-methylpropionate group from SceH onto the ACP of the loading module of SceN^{280, 283}.

Apart from the starter biosynthetic genes, *sceF* encodes a protein (SceF) that shares high similarity with VinJ (Identity/Similarity: 71%/84%), an amidohydrolase or peptidase in vicanistatin biosynthesis^{280, 284}. VinJ, which is distinct from other more common serine peptidases adopting an α/β hydrolase fold, plays a role by catalyzing the amide hydrolysis to generate the free amino group, which must be protected during the polyketide chain extension to prevent premature cyclization of intermediates.²⁸³ Hence, the protection de-protection strategy for the reactive aminoacyl-ACP intermediates found in the biosynthesis of vicanistatin, amino marginolactone²⁸⁵ and other natural products is most likely to be shared by sceliphrolactam biosynthesis. In addition, the presence of two cytochrome P450 genes (*sceD* & *secE*) in the BGC is also consistent with the two hydroxylated carbon atoms (C-10 and C-12) in sceliphrolactam (**FIGURE 19B**). The genes *sceA* and *sceB* encode a pair of thioredoxin and ferredoxin, which are likely to be the electron transfer proteins that supply electrons to the cytochrome P450 hydroxylases.

2.3.5. Analysis of the Sce Type I modular PKS

Detailed sequence analysis was performed on the acyltransferase (AT), ketosynthase (KS), ketoreductase (KR) and dehydratase (DH) domains of all ten PKS modules encoded by the *sce* gene cluster. All AT, KS, KR and DH domains contain the essential catalytic residues and are predicted to be catalytically active. The substrate preference of AT domains of type-I PKS towards malonyl-CoA or methylmalonyl-CoA is defined by several key residues in the substrate binding pocket²⁸⁶⁻²⁸⁸. Sequence comparison suggests that the two AT domains from module-3 and module-8 are likely to use methylmalonyl-CoA as substrate. This is in accordance with the two methyl groups (C-26, C-27) in sceliphrolactam (**FIGURE 19B**). Meanwhile, protein sequence analysis suggests that the KR domains belong to the “B-type” KR that contains a conserved (L/V)D(D/N) motif, which is crucial for guiding polyketide intermediates into the active site of KR domains to produce “R”-configured β -hydroxyl groups (**FIGURE S 1**).

To assemble the macrolactam skeleton of sceliphrolactam, the modular PKS system is expected to have eleven modules to join the 3-amino-2-methylpropionate starter unit with nine acetyl and two propionyl extender units. Surprisingly, the *sce* BGC only encodes ten PKS modules found in SceN, SceO, SceQ, SceR and SceS. According to the collinearity rule for PKS²⁶³⁻²⁶⁴, the order of PKS modules in the biosynthetic

complex should be reflected in the chromosomal order of the *pks* genes. Considering the collinearity rule is strictly followed in the biosynthesis of other known macrolactam polyenes, the Sce PKS system could represent a deviation from the orthodox systems and may feature an iterative module (**FIGURE 19C, D**). To confirm that the shortage of a module is not because of erroneous assembly of the genome sequence, DNA segments at the boundaries of *sceO*, *sceQ*, *sceR* and *sceS* genes were amplified by PCR (**FIGURE S 2**). The result confirms that the four genes form a continuous DNA fragment, and thus, rules out the existence of a gene encoding a “missing” module. Meanwhile, we did not detect other macrolactams in the biomass or culture broth of *Streptomyces* sp. SD85, which makes it tenuous to argue that sceliphrolactam is a minor product generated by the aberrant stuttering of a PKS module.

Figure 19: Proposed biosynthetic mechanism for sceliphrolactam. A. BGCs of sceliphrolactam and vicenistatin share a set of genes (*sceG-M* vs. *vinH-O*) for β -amino acid starter unit biosynthesis. B. Overall biosynthetic mechanism for sceliphrolactam. The presence of a *cis*-double bond between the protons of C14 and C15 was confirmed by ^1H NOESY correlation (Fig. S10). The stereochemistry of C-11 was assigned based on the observation that the KR_6 domain is a B-type KR domain²⁸⁹⁻²⁹¹. C. A possible mechanism with SceQ containing the iterative module. D. A possible mechanism with SceR containing the iterative module/ (ACP: acyl carrier protein; AT: acyltransferase; DH, dehydratase; KR; ketoreductase; KS, ketosynthase; TE, thioesterase).

2.3.6. Evidence disfavouring involvement of a *trans*-PKS module in sceliphrolactam biosynthesis

There is a plausible explanation for the shortage of a PKS module in the Sce system without invoking the breakdown of the collinearity rule. This would involve the use of a *trans*-PKS module encoded by another gene located somewhere else in the genome. This *trans*-PKS module must be integrate into the mega-PKS protein complex by docking specifically between SceQ and SceR. Considering that the β -keto group remains unreduced for the chain extension catalyzed by the mysterious module, a potential *trans*-PKS module is most likely to have a KS-AT-ACP domain composition and contain docking domains that are compatible with the docking domains of SceR and SceQ. However, an inspection of the complete genome of *Streptomyces* sp. SD85 revealed that all the *pks* genes in the genome are accounted for by the 52 BGCs. We could not find any orphan PKS module with docking domains, let alone an orphan module with the KS-AT-ACP domain organization. We also searched the genome of the three *Streptomyces* strains that contain the homologous *sce* BGCs (Fig. 2C) and did not find such orphan PKS module. Moreover, we identified the N- and C-terminal portions of SceN, SceO, SceQ, SceR and SceS that are predicted to fold into docking domains crucial for holding the PKS complex together (**FIGURE S 11**)²⁹². The lack of a gene that encodes a potential *trans*-PKS module and the presence of compatible docking domains for SceR and SceQ seem to argue against the participation of a *trans*-PKS module.

To seek experimental evidence for or against the involvement of a *trans* PKS module, we constructed a mutant strain by fusing the *sceQ* and *sceR* genes on the chromosome. The rationale behind the experiment is that when the *sceQ* and *sceR* genes are fused, SceQ and SceR will be expressed as a single protein to prevent the insertion of any potential *trans*-PKS module. Production of sceliphrolactam is expected to be abolished for the fusion mutant because the *trans*-PKS module can no longer dock between the closely linked SceQ and SceR. Similar gene fusion experiment has been used to support an iterative mechanism in the biosynthesis of borrelin²⁹³. To create the *sceQ-R* fusion mutant, a pCRISPR-cas9 plasmid that harbors two sgRNA cassettes was constructed to target the *sceQ/sceR* intergenic region for the deletion of 14 base pairs (**FIGURE 20A**). Deletion of the 14 base pairs was expected to remove

the stop and start codons and lead to the fusion of *sceQ* and *sceR* genes. The sgRNA and homology template-containing pCRISPR-cas9 plasmid was transformed into *Streptomyces* sp. SD85 strain using the same *E. coli* - *Streptomyces* conjugation method as the gene inactivation experiment described earlier. Deletion of the 14 base pairs was confirmed by gene sequencing for the four colonies we tested (**FIGURE 20B**). After cultivating the *sceQ-R* fusion mutant strains inoculated using four positive colonies, we found that sceliphrolactam was still produced by all four mutant strains, despite at a lower yield than the wild type strain (**FIGURE 20C**). The results suggest that the polyketide intermediate is passed from SceQ directly to SceR, and thus, it is tenuous to argue that a *trans*-PKS module is involved in the biosynthesis of sceliphrolactam.

Figure 20: CRISPR/Cas9 mediated fusion of *sceQ* and *sceR*. A. Schematic illustration of the CRISPR/Cas9-mediated and dual sgRNA-guided DNA cleavage to fuse the *sceQ* and *sceR* genes. B. DNA sequencing result confirmed the fusion of the two genes upon the deletion of 14 base pairs (sequence in red). C. HPLC analysis comparing the production of sceliphrolactam between the *sceQ-R* fusion mutant and the wild type *Streptomyces* sp. SD85. The sceliphrolactam peak is indicated by the arrow. The wavelength (λ) was set at 330 nm for the HPLC detector.

2.3.7. Sceliphrolactam biosynthesis potentially involves an iterative PKS module

As the experimental evidence argue against the involvement of a *trans*-PKS module, the apparent lack of a PKS module in the biosynthesis of sceliphrolactam remains to be fully explained. One of the conceivable mechanisms involves the iterative use of a

PKS module. The iterative use of modules in PKS systems is increasingly known as a common phenomenon and has been reported in the biosynthesis of lankacidin ²⁹⁴, borrelidin ²⁹⁵, aureothin ²⁹, stigmatellin ²⁹⁶ and other polyketide secondary metabolites ²⁹⁷. There are two possibilities, with either SceQ or SceR acting iteratively (**FIGURE 19CD**). If SceQ acts iteratively, the β -keto group is reduced in the first chain-extension cycle to generate a hydroxyl group that is subsequently dehydrated, but remains unreduced in the second cycle. On the other hand, if SceR acts iteratively, the β -keto group is not reduced in the first chain-extension cycle; but reduced to yield hydroxyl group in the second cycle. Regardless of which module acts iteratively, the iterative module seems to be “programmed” with the activity of the KR domain switching on and off in two successive extension cycles. Hence, in comparison with most known iterative mechanisms that involve modules acting iteratively with the processing domains (e.g. KR, DH, ER or MT) generating the same chemical outcome in each extension cycle, the putative iterative module in scilipholactam biosynthesis is able to generate different outcome during the two successive extension cycles. Such “programmed iteration” is rare but not unprecedented. It was recently found that the biosynthesis of azalomycin involves a similar iterative module. This iterative module contains an ER domain whose activity can be “toggled” off and on to ensure that the ER domain only functions in the second extension cycle with ^{285, 298}. The selective keto-reduction on different polyketide intermediates by the SceQ or SceR module is also reminiscent of the programmed keto-reduction mechanism observed in partially reducing iterative PKSs ^{286, 299-301}.

2.4. Conclusion

In summary, a biosynthetically talented *Streptomyces* strain that harbors a large number of uncharacterized biosynthetic gene clusters was isolated. As an effort to explore the biosynthetic potential of the strain, the strain was sequenced to identify all putative clusters present in its genome. Analytical data obtained from genome mining effort proved to be an effective aid for chemical dereplication and further investigation of interesting BGCs. Although most of these interesting BGCs were silent under tested laboratory conditions, the prospect of uncovering novel compounds through activation of these novel BGCs remains bright. Realization of CRISPR/Cas9 as a genetic

manipulation tool also set up a platform for eventual genetic manipulation of other actinomycetes in the lab.

For the polyene macrolactam sceliphrolactam, identification of the *sce* BGC suggests a potentially unique modular PKS system for macrolactam biosynthesis. Experimental data suggest that the biosynthesis of sceliphrolactam potentially involves a “programmed” iterative PKS module. Among all the known polyene macrolactams, sceliphrolactam is the only one that features a biosynthetic mechanism that does not seem to follow the co-linearity rule by involving an iterative PKS module. Future mechanistic studies are required to identify the iterative module and elucidate its pathway of programmed keto-reduction.

Chapter 3: Discovery of meijiemycin, a novel linear polyene-polyol antifungal compound from *Streptomyces* sp. SD50

3.1. Introduction

The Post Genomic era has seen a renaissance in natural product discovery with the realization that most of the 'dark chemical matter' remain hidden under typical laboratory conditions.³⁰² Concurrently, awareness of the relevance of natural products in Medicine has fuelled efforts in expanding the natural product repertoire for therapeutic purposes.³⁰³ Actinomycetales, which is notable for its remarkable contribution to the field of antibiotics, particularly in the Golden Age of antibiotic discovery (1940s – 1970s),³⁰⁴ is one taxonomic order of prokaryotes that has garnered attention in natural product discovery. These high GC content bacteria possess many uncharacterized BGCs²⁵³, making them prospective sources of novel natural products.

As part of the lab's efforts to discover novel natural products from actinomycetes, sediments from Sungei Buloh Wetland reserve were collected and surveyed for interesting actinomycetes. These coastal mangrove trees grow in a brackish water environment that is supplied by the Straits of Malacca and Buloh Besar river. Such unique ecological environment provides a rich biodiversity which encourages microbial interactions and extensive horizontal gene transfer³⁰⁵. Given the favourable conditions for taxonomic and molecular diversification, it facilitates the isolation of interesting actinomycetes with unique BGCs for natural product discovery. Previously, as discussed in chapter 2, a scilphrolactam-producing *Streptomyces*, which was isolated from Sungei Buloh Wetland reserve had been studied.³⁰⁶ This chapter focuses on another *Streptomyces* strain isolated from the same location - *Streptomyces* sp. SD50.

Using a genomics guided approach, *Streptomyces* sp. SD50 was initially sequenced and putative BGCs were identified for dereplication and targeting efforts. Apart from BGCs encoding for known polyketide natural products like oxazolomycin, nanchangmycin and meilingmycin, a particularly interesting BGC encoding a polyene natural product was identified. Analysis of this BGC predicted a linear hexaene-polyol molecule which sparked further pursuit due to two factors – scarcity of linear polyene natural products and one that possesses a hexaene moiety. Identification of a

molecule containing a UV fingerprint which corresponds to a hexaene moiety prompted yield optimization efforts through culture condition screening. Eventual isolation and structural elucidation studies of the compound led to discovery of meijiemycin (**FIGURE 21**), a novel linear polyene-polyol molecule. Association of meijiemycin with its corresponding BGC was established through knock-out studies achieved using CRISPR/Cas9 mediated genetic manipulation.

Figure 21: Meijiemycin and other related polyene natural products.

Preliminary biological assay against several fungi demonstrated selective inhibition against a *Candida albicans* clinical isolate. In the face of growing public health concern³⁰⁷ generated from increasing incidence rate of fungal infections and emerging resistance of fungal pathogens towards existing limited range of antifungals agents available²⁰⁶, meijiemycin and potential manipulation of its BGC offers opportunities for creating prospective drug leads to boost the antifungal drug pipeline.

3.2. Methods and materials

3.2.1. *Streptomyces* sp. SD50 isolation and strain maintenance

Streptomyces sp. SD50 was isolated in a similar manner as described for *Streptomyces* sp. SD85 in chapter 2. Pure colony isolate was spread onto MS agar (20 g/L soy flour, 20 g/L mannitol, 20 g/L Bacto agar) and cultured at 30 °C for 2-3 weeks until sporulation. Spores were harvested by gentle scrapping of aerial mycelia from plates flooded with sterile water using 1 mL pipette tip, followed by centrifugation at 8000 *g* for 5 min. Pelleted spores were resuspended in 20% glycerol solution, aliquoted and kept frozen at 80 °C until required.

3.2.2. Genomic DNA isolation from *Streptomyces* sp. SD50

Genomic DNA of *Streptomyces* sp. SD50 was extracted using a modified DNA salting out procedure for isolation of genomic DNA from Gram Positive bacteria.³⁰⁸ *Streptomyces* sp. SD50 was cultured in liquid CRM culture for 2-4 days until it reached late log phase of growth. Mycelium pellets were harvested by centrifuging the culture at 6000 *g* for 15 min and resuspended in 5 mL of SET buffer (75 mM NaCl, 25 mM EDTA pH 8.0, 20 mM Tris-HCl pH 7.5). The mycelium suspension was homogenized using a 15 mL handheld Dounce tissue grinder until finely dispersed. Lysozyme was added to the solution until a final concentration of 1 mg/mL and the mixture was incubated at 37 °C for 1-2 hr. The eventual protoplasts were lysed by addition of 600 µL of 10% SDS solution. 140 µL of proteinase K (20 mg/mL in water) was added to the lysis mixture with gentle mixing, followed by a 2 hr incubation at 55 °C. Upon completion of protein digestion, 2 mL of 5 M NaCl solution was added to the mixture with gentle mixing. Upon cooling to 37 °C, equal volume of chloroform was added to the lysed cell solution and mixed by repeated gentle inversion for 5 min at ambient temperature. The emulsion was centrifuged at 4500 *g* for 15 min and the aqueous layer was collected. Chloroform step was repeated until a clear aqueous layer was obtained. Aqueous phase was supplemented with RNase A to a final concentration 100 µg/mL and incubated at 37 °C for 1-2 hr for RNA digestion. High molecular weight genomic DNA was finally spooled from the aqueous mixture by adding 0.6 volume of isopropanol. Spooled DNA was washed twice with 5 mL of 70% aqueous ethanol and air dried under ambient conditions. Genomic DNA was subsequently re-dissolved in

10 mM Tris buffer, pH 8.0 and aliquoted into small volumes for storage in -20 °C freezer.

3.2.3. Metabolite profiling of *Streptomyces* sp. SD50

For agar plate fermentation, frozen spores of *Streptomyces* sp. SD50 were spread evenly onto MS agar plates and incubated at 30 °C for 10 days. Agar containing mycelia was collected and blended with acetone for extraction of compounds from the agar bits and mycelia. The organic layers were dried under vacuum and re-dissolved in methanol. For shake flask liquid fermentation, a log phase culture of *Streptomyces* sp. SD50 was inoculated into 50 mL of fresh medium at 4% inoculum. Liquid cultures were grown at 30 °C at 180 RPM for 5-6 days until late stationary phase of growth. Mycelium pellets were separated from culture broth by centrifugation. Mycelia and culture broth were extracted separately with methanol and ethyl acetate, respectively. The organic layers were dried under vacuum and re-dissolved in methanol.

HPLC analysis of the crude extract was performed with Grace® VisionHT™ C18-HL (4.6 mm x 250 mm, 5 µm) using an Agilent 1200 HPLC system equipped with DAD detector. A gradient elution with a binary solvent system at 1 mL/min was employed using water as mobile phase A and methanol as mobile phase B. LC program was set at 0 min 10% B -> 5 min 20% B -> 35 min 70% B -> 50 min 90% B -> 60 min 100% B. Analytes were monitored at λ = 220 nm, 264 nm, 280 nm, 360 nm, 420 nm.

LC-MS was performed using Agilent 1260 LC system coupled with an Agilent 6550 Accurate-Mass QTOF spectrometer equipped with a Zorbax® Eclipse plus C18 (2.1 mm x 50 mm, 5 µm). Water with 0.1% formic acid was set as mobile phase A while acetonitrile with 0.1% formic acid was set as mobile phase B. A flow rate of 0.3 mL/min was used and the gradient elution program was similar to that employed for HPLC-DAD analysis. Mass spectra were measured using positive and negative ion modes of ESI and conditions of MS were set accordingly (capillary voltage 3 kV; nebulizer 45 psig; source temperature 300 °C; gas flow 11 L/min; MS scan rate 1/s, MS/MS scan rate 5/s, collision energy of 10, 20 and 40 eV) with a mass range of 100 – 2000 Da. Autotuning of MS was performed prior to every sample analysis.

3.2.4. Genome sequencing of *Streptomyces* sp. SD50

Initial genome sequencing was performed using Illumina Miseq sequencing technology. A total of 214 contigs with an average size of 53,914 bp was obtained after *De novo* assembly of trimmed sequencing reads was performed using CLC Genomics Workbench (CLC bio, Denmark). Analysis of draft contigs revealed many truncated PKS genes in the genome, which complicated genome mining efforts due to difficulty in ascertaining relationship between those truncated genes. Hence, another round of genome sequencing was attempted with PacBio to obtain longer DNA contigs. Single Molecule Real Time (SMRT) sequencing technology (Pacific Biosciences, California, USA) offered by PacBio RS II platform technology was used for obtaining long sequencing reads. A single SMRT cell was used and the sequencing run was performed 3 times. The long sequencing reads generated a 37 X coverage of the genome. A read quality value of 84 was achieved with 102,653 zero-mode-waveguides (ZMWs). Mean polymerase read length was 12,011 bp with mean reads of inserts (ROI) readlength value of 10,013 bp. A single contig of length 11,607,137 bases was obtained after assembly. The PacBio reads were assembled into two contigs by Canu v1.6.³⁰⁹ The Illumina reads were mapped to the a single contig by CLC Genomics Workbench v10.0, and the .bam files for the mapped reads were extracted and used to polish the assembled contigs by Pilon v1.22³¹⁰ to further minimise sequencing errors.

3.2.5. Construction of *Streptomyces* sp. SD50 $\Delta nanA1\Delta ozmH$ double mutant, incapable of producing nanchangmycin and oxazolomyin to facilitate screening of novel compounds

All cloning steps were carried out using *E. coli* TOP10 (Invitrogen, US). Experimental procedures for construction of plasmid for CRISPR/Cas9 mediated gene mutation, conjugal transfer of plasmid from donor *E. coli* to *Streptomyces* sp. SD50, identification of true exconjugants, induction of *cas9* gene expression and plasmid curing are similar to those described in Chapter 2.

To abolish production of nanchangmycin, a $\Delta nanA1$ knockout mutant was generated by a 751 bp frameshift deletion of *nanA1* gene in the genome of *Streptomyces* sp. SD50. Likewise, oxazolomycin production was abolished by generating a 692 bp frameshift deletion of $\Delta ozmH$ gene. Double $\Delta nanA1\Delta ozmH$ knockout mutant of

Streptomyces sp. SD50 was achieved by performing a subsequent *ozmH* gene deletion on a *Streptomyces* sp. SD50 Δnan mutant.

To screen for successful mutants, individual clones were cultured in liquid GYM at 30 °C for 3 days for genomic DNA isolation.²⁶⁷ Briefly, mycelia were harvested and resuspended in 0.5 mL Lysis buffer (20 g CTAB, 100 mL Tris-HCl pH 8.0, 40 mL 0.5 M EDTA pH 8.0, 280 mL 5 M NaCl, 10 g PVP, top up to 1 L with ultrapure water and final pH of 5.0). The mixture was homogenized using a micro-pestle and incubated at 55 °C for 15 min for lysis. An equal volume of chloroform was added to the mixture and vortexed vigorously for 1 min to ensure even mixing. The resultant emulsion was centrifuged at 10,000 *g* for 5 min. Aqueous layer containing genomic DNA was collected and mixed with equal volume of isopropanol. The mixture was centrifuged at 10,000 *g* for 5 min to pellet crude DNA precipitates. DNA was washed twice with 70% aqueous ethanol. The resultant pellets were dried at 55 °C for 5 min and re-dissolved in sterile water for downstream analysis. Diagnostic PCR was performed to establish deletion of 751 bp and 692 bp in *nanA1* and *ozmH*, respectively (**FIGURE S 12** and **FIGURE S 13**).

3.2.6. Construction of *Streptomyces* sp. SD50 $\Delta mjmB$ and $\Delta mjmE$ mutants using *Streptomyces* sp. SD50 $\Delta nanA1\Delta ozmH$ double mutant to validate relationship between *mjm* BGC and meijiemyacin

Experimental procedures for CRISPR/Cas9 mediated genetic manipulation are as described above. *mjmB* gene was deleted by introducing a 926 bp frameshift deletion to *Streptomyces* sp. SD50 $\Delta nanA1\Delta ozmH$ double mutant, resulting in the creation of a *Streptomyces* sp. SD50 $\Delta nanA1n\Delta ozmH\Delta mjmB$ triple mutant. Similarly, a frameshift deletion of 772 bp was used to delete *mjmE*, which generated a *Streptomyces* sp. SD50 $\Delta nanA1\Delta ozmH\Delta mjmE$ triple mutant.

Diagnostic PCR was performed to establish deletion of 926 bp and 772 bp in *mjmB* and *mjmE*, respectively (**FIGURE S 14** and **FIGURE S 15**).

3.2.7. HPLC analysis of meijiemyacin production

To optimize the production yield of meijiemyacin, *Streptomyces* sp. SD50 $\Delta nanA1\Delta ozmH$ double mutant was cultured in various media conditions like GYM agar,

MS agar, liquid GYM, liquid Gly-YM (4 g/L glycine, 4 g/L yeast extract, 10 g/L malt extract), liquid MYM (20 g/L mannitol, 4 g/L yeast extract, 10 g/L malt extract) for 7 days at 30 °C. Extraction of compounds from both liquid and solid cultures was performed as described above for metabolite profiling.

HPLC analysis of crude extracts was performed with Agilent Zorbax Eclipse XDB-C18 (4.6 mm x 250 mm, 5 µm) using an Agilent 1200 HPLC system equipped with DAD detector. A gradient elution with a binary solvent system at 1 mL/min was employed using water as mobile phase A and acetonitrile as mobile phase B. Both mobile phases were supplemented with 0.1% formic acid. LC program was set at 0 min 10% B -> 5 min 20% B -> 35 min 70% B -> 50 min 90% B -> 60 min 100% B. Analytes were monitored at λ = 220 nm, 264 nm, 280 nm, 360 nm, 420 nm.

3.2.8. Large scale fermentation of *Streptomyces* sp. SD50 for isolation and purification of meijiemycin

Streptomyces sp. SD50 was cultured in MYM (20 g/L mannitol, 4 g/L yeast extract, 10 g/L malt extract) liquid medium (5 X 800 mL in 2 L baffled Erlenmeyer flask) for 7 days at 30 °C, 220 RPM. Mycelia was harvested by centrifugation and extracted by stirring with 4 L of methanol for 1 hr at ambient temperature. The methanolic extract was filtered using a Buchner funnel and dried using a rotary evaporator to afford a crude extract (~ 3 g). The crude extract was redissolved in 500 mL of methanol and extracted thrice with equal volumes of hexane to remove fatty material. The extract was dried completely and washed with 100 mL of ultrapure water to get rid of salts and other water miscible materials. Residual crude was soaked in minimal amount of dimethylformamide (DMF) and sonicated in a water bath to promote compound solubility. Methanol and ultrapure water were subsequently added to the crude solution at a final concentration of 1% (v/v) DMF, 60% methanol, 39% water. Meijiemycin was purified from the crude using Shimadzu Prominence Preparative HPLC system equipped with an ACE C18 semi-prep column (10 mm X 250 mm, 5 µm). A gradient elution program was set up with the following conditions: 0 min 75% aqueous methanol -> 40 min 90% aqueous methanol -> 50 min 100% aqueous methanol. A flow rate of 4 mL/min was employed for the reversed phase column chromatography procedure. Fractions containing pure meijiemycin, which eluted at t = 36 min, were pooled and dried using a rotary evaporator. Final yield of NMR-pure meijiemycin was 25 mg.

3.2.9. Characterization of meijiemycin

Meijiemycin was isolated as a dark yellow solid; UV (MeOH) $\lambda_{\max}$ (log ϵ) 277 (4.48), 324 (4.38), 340 (4.71), 356 (4.90), 378 (4.89) nm; ^1H NMR (400 MHz, in $\text{C}_5\text{D}_5\text{N}$) δ 7.51 (1H, ddd, J = 15.6, 7.0, 6.7 Hz, H-3), 6.91 (1H, dd, J = 15.4, 11.6 Hz, H-10), 6.74 (1H, dd, J = 15.0, 10.7 Hz, H-43), 6.62 (1H, br d, J = 11.0 Hz, H-7), 6.43 (1H, dd, J = 11.0, 11.0 Hz, H-8), 6.36 (1H, d, J = 15.6 Hz, H-2), 6.25 (dd, m, H-44), 6.21-6.40 (7H, m, H-45 to H-50), 6.14 (1H, m, H-52), 6.12 (1H, dd, J = 15.4, 7.0 Hz, H-11), 6.10 (1H, dt, J = 15.0, 6.3 Hz, H-42), 6.08 (1H, dd, J = 11.6, 11.0 Hz, H-9), 5.71 (1H, dt, J = 15.0, 7.5 Hz, H-27), 5.69 (1H, m, H-53), 5.58 (1H, dt, J = 15.0, 7.5 Hz, H-26), 5.23 (1H, m, H-37), 5.23 (1H, dd, J = 7.5, 4.5 Hz, H-5), 5.16 (1H, m, H-33), 5.13 (1H, m, H-41), 4.95 (1H, br s, H-1'), 4.93 (1H, m, H-39), 4.75 (1H, m, H-15), 4.52 (1H, m, H-31), 4.41 (1H, m, H-13), 4.39 (1H, br s, H-2'), 4.21 (1H, m, H-29), 3.94 (1H, br d, J = 9.2 Hz, H-19), 3.87 (1H, br d, J = 7.9 Hz, H-17), 3.81 (1H, m, H-3'), 3.59 (1H, m, H-5'), 3.30 (1H, dd, J = 10.2, 10.2 Hz, H-4'), 2.98 (1H, dd, J = 10.5, 10.5 Hz, H-38), 2.82 (1H, ddd, J = 15.8, 7.5, 7.0 Hz, H-4a), 2.73 (1H, m, H-36a), 2.65 (1H, m, H-12), 2.56 (1H, ddd, J = 15.8, 6.7, 4.5 Hz, H-4b), 2.46 (1H, m, H-40a), 2.45 (1H, m, H-28a), 2.41 (1H, m, H-28b), 2.40 (1H, m, H-16), 2.39 (1H, m, H-40b), 2.17 (1H, dd, J = 11.0, 4.9 Hz, H-34a), 2.07 (1H, m, H-18), 2.07 (1H, m, H-14a), 2.07 (1H, m, H-34b), 2.06 (1H, m, H-21a), 2.05 (3H, br s, H₃-55), 2.02 (2H, m, H₂-25), 1.96 (1H, m, H-36b), 1.96 (2H, m, H₂-30), 1.90 (2H, m, H₂-32), 1.87 (1H, m, H-14b), 1.76 (1H, m, H-20), 1.65 (1H, d, J = 7.0 Hz, H-54), 1.46 (1H, d, J = 5.4 Hz, H-6'), 1.36 (2H, m, H₂-24), 1.36 (2H, m, H₂-23), 1.27 (3H, d, J = 6.7 Hz, H₃-56), 1.26 (3H, d, J = 6.7 Hz, H₃-58), 1.25 (1H, m, H-21b), 1.22-1.35 (2H, m, H₂-22), 0.98 (3H, d, J = 6.6 Hz, H₃-57), 0.85 (3H, d, J = 6.3 Hz, H₃-59); ^{13}C NMR (400 MHz, in $\text{DMSO}-d_6$) δ 174.7 (C-60), 168.0 (C-1), 145.2 (C-3), 141.3 (C-6), 139.2 (C-11), 134.2 (C-42), 133.9 (C-51), 133.1 (C-43), 132.4 (C-26), 131.5 (C-52), 131.1-133.9 (C-44 to C-50), 130.7 (C-53), 129.1 (C-9), 127.3 (C-27), 125.3 (C-10), 124.7 (C-2), 122.3 (C-8), 121.6 (C-7), 98.9 (C-1'), 97.6 (C-35), 78.8 (C-17), 76.1 (C-41), 74.1 (C-19), 73.3 (C-3'), 72.2 (C-5'), 70.8 (C-13), 70.6 (C-2'), 69.4 (C-29), 68.9 (C-15), 67.6 (C-31), 67.4 (C-5), 66.5 (C-39), 66.3 (C-33), 65.9 (C-37), 58.1 (C-38), 54.7 (C-4'), 46.6 (C-34), 45.8 (C-32), 44.6 (C-30), 44.5 (C-36), 43.7 (C-12), 42.2 (C-16), 41.3 (C-28), 38.9 (C-40), 38.5 (C-4), 37.1 (C-14), 36.3 (C-20), 34.5 (C-18), 32.9 (C-21), 32.7 (C-25), 29.8 (C-24), 29.8 (C-23), 26.5 (C-22), 18.8 (C-55), 18.7 (C-54),

18.6 (C-6'), 17.2 (C-56), 15.8 (C-59), 12.1 (C-57), 11.4 (C-58). HRMS m/z 1216.7357 [M+H]⁺ (calculated for C₆₆H₁₀₅NO₁₉H⁺, 1216.7359).

3.2.10. Determination of minimum inhibitory concentration (MIC) of meijiemycin against fungal strains

MIC was determined via a modified broth microdilution assay protocol adopted for aerobic bacteria.³¹¹ Lyophilized meijiemycin and nystatin powder were initially dissolved in DMSO and diluted with sterile water to form 2% DMSO aliquots for assays. 50 μ L of fungal culture, containing 1 million CFU in RPMI 1640 medium, was spiked with equal volume of compound solution and incubated at 30 °C with 180 RPM for 40 hr. Antifungal susceptibility was measured using a range of compound concentration from 0.25 μ g/mL to 128 μ g/mL via serial two-fold dilution. MIC values were determined from growth information obtained via turbidometry.³¹²

3.2.11. Primers used in this study

Primers used for this study were synthesized by Integrated DNA Technologies, US. Sequences of all primers in this study are available in **TABLE S 3**.

3.3. Results and discussion

3.3.1. *Streptomyces* sp. SD50 is phylogenetically related to *Streptomyces bingchenggensis* and possesses immense biosynthetic potential for secondary metabolism

Streptomyces sp. SD50 is a close relative of *Streptomyces bingchenggensis* BCW-1³¹³, with whom it shares a 16S rRNA gene sequence identity of 99.72%. Comparison of the genome of both strains using OrthoANI³¹⁴ revealed an average nucleotide identity value of 93.70 %, and this was demonstrated further in the number of similar BGCs³¹⁵ shared between both strains as identified by antiSMASH¹⁶¹ analysis. Similarly, both strains have a genome size of >11 Mb, which is higher than the average genome size of 8-10 Mb in most *Streptomyces*.

The larger genome size in *Streptomyces* sp. SD50 could be attributed to the large number of secondary metabolite encoding BGCs present in its genome. This immense potential that *Streptomyces* sp. SD50 possesses for producing secondary metabolites was demonstrated by the identification of 50 putative BGCs (**TABLE S 4**), which are distributed rather evenly throughout the entire genome (**FIGURE 22A**). Genome mining efforts confirmed at least three known PKS gene clusters (**FIGURE 22B**) that encode for oxazolomycin, meilingmycin and nanchangmycin for dereplication. Notably, a huge PKS cluster, which shares some degree of similarity to amphotericin BGC was putatively identified by antiSMASH analysis. Initial analysis of this BGC, which was validated later by further experiments to encode for meijiemyacin, suggested the formation of a linear polyene molecule as the 1st PKS module lacks a KR domain for reduction of the preceding β -keto group. Polyene macrolides are common in nature, with linear polyene molecules like ECO-02301³¹⁶ and mycangimycin³¹⁷ being atypical examples. Further analysis of the PKS system predicted the presence of a triene and hexaene moiety. Such hexaene containing natural products are rare, with only bahamaolides³¹⁸ and dermostatin³¹⁹ being notable examples. Encouraged by genetic analysis of meijiemyacin BGC, which hypothesized the biosynthesis of a novel compound, efforts were focused on identifying this polyene molecule.

Figure 22: Genome map of *Streptomyces* sp. SD50 annotated with secondary metabolite BGCs. (A) Circular representation of *Streptomyces* sp. SD50 chromosome. Normalized GC skew plot is depicted by inner ring with positive values in green and negative values in magenta. Center ring shows normalized GC content of the DNA sequence with positive values in yellow and negative values in black. Secondary metabolite BGCs are annotated as bars in outer ring with colors representing different classes of BGCs – PKS (red), NRPS (yellow), hybrid PKS/NRPS (orange), RiPP (green), Terpene (blue), Others (black). (B) Identification of known BGCs through comparison with characterized homologous clusters from other *Streptomyces* strains.

3.3.2. Metabolite Profiling of *Streptomyces* sp. SD50 mutants reveals peak containing hexaene UV fingerprint

Initial metabolite profiling of wild-type *Streptomyces* sp. SD50 revealed a small peak with a UV spectrum that contains two fingerprint signatures indicative of the presence of a triene and a hexaene moiety.³²⁰ The presence of only one polyene BGC in its genome suggested the likelihood that the identified peak corresponded to meijiemycin. However, its low yield necessitated optimization to generate enough material for structural studies and practical applications. Apart from meijiemycin BGC, *Streptomyces* sp. SD50 possesses other PKS BGCs encoding unidentified secondary metabolites. One such cluster is a T2 PKS cluster that encodes an angucycline type compound, which was also found in the HPLC analysis. However, two other interesting PKS BGCs encoding for polyether (Cluster 40) and spirotetronate (Cluster 46) natural products, which are valued for their therapeutical properties,³²¹⁻³²² were likely silent as reflected by their absence in the crude extract.

Observation of peaks in crude extracts of *Streptomyces* sp. SD50 with UV spectra that correspond to oxazolomycin and nanchangmycin, as reported in literature,³²³⁻³²⁴ prompted the disruption of oxazolomycin and nanchangmycin BGCs (Figure 23A) to free up precursor supply for biosynthesis of other polyketide natural products, particularly meijiemycin. Production of both oxazolomycin and nanchangmycin were abolished (Figure 23B) respectively by CRISPR/Cas9 mediated frameshift deletion of $\Delta ozmH$ ³²⁵ (FIGURE S 13) and $\Delta nanA1$ ³²⁶ (FIGURE S 12), which effectively disrupted the entire PKS biosynthetic pathway of both clusters. However, successful disruption of both clusters in *Streptomyces* sp. SD50 $\Delta ozmH\Delta nanA1$ double mutant did not result in significant increase in production of meijiemycin. While comparison of HPLC analysis between the mutants and wild type strain exhibited noticeable differences other than the targeted abolishment of oxazolomycin and nanchangmycin, such changes in metabolic profile were likely a reflection of pleiotropic effects generated by non-identical growth profiles. Nonetheless, an alternative approach through culture condition screening was attempted to increase the yield of meijiemycin.

Figure 23: Metabolite profiling of *Streptomyces* sp. SD50 mutants. (A) Disruption of oxazolomycin BGC and nanchangmycin BGC via CRISPR/Cas9 mediated genetic manipulation. (B) Comparison of HPLC chromatograms of wild-type strain and mutants. Chromatogram of wavelength (λ) = 220 nm is selected for comparison. Peaks corresponding to different compounds are numbered in red font with their on-line UV absorption spectra shown on the right.

3.3.3. Optimization of meijiemyacin yield and validation of its biosynthetic origin through CRISPR/Cas9 mediated gene inactivation

Meijiemyacin was originally detected in the crude extracts obtained from solid state fermentation on MS agar. After unsuccessful attempts to increase its yield through disruption of nanchangmycin and oxazolomyacin BGCs, *Streptomyces* sp. SD50 $\Delta\text{ozmH}\Delta\text{nanA1}$ was cultured in different culture conditions to elicit the production of meijiemyacin to a higher level. A switch from solid state to liquid fermentation appeared to repress secondary metabolism on a global level, resulting in lesser metabolites detected through HPLC analysis. However, fermentation of the double mutant in liquid MYM culture fortuitously increased meijiemyacin yield by at least 50 folds as compared to MS agar fermentation (**FIGURE 24B**).

Concurrently, to validate the hypothesis that *mjm* BGC is responsible for the biosynthesis of meijiemyacin, two genes (*mjmB* and *mjmE*) from *mjm* BGC were individually inactivated through CRISPR/Cas9 mediated frameshift deletion of the genes (**FIGURE 24C**). The abolishment of meijiemyacin production in the resultant *Streptomyces* sp. SD50 $\Delta\text{ozmH}\Delta\text{nanA1}\Delta\text{mjmB}$ and *Streptomyces* sp. SD50 $\Delta\text{ozmH}\Delta\text{nanA1}\Delta\text{mjmE}$ triple mutants validated the biosynthetic relationship between meijiemyacin and *mjm* BGC (**FIGURE 24D**).

Figure 24: Meijiemycin biosynthetic origin and yield optimization. (A) Organization of genes in *mjm* biosynthetic gene cluster. (B) HPLC chromatograms of *Streptomyces* sp. SD50 double mutant fermentation under different culture media. Peak corresponding to meijiemycin is depicted by blue arrow. Wavelength (λ) = 360 nm is shown for comparison. (C) Inactivation of *mjmB* and *mjmE* of *mjm* gene cluster via CRISPR/Cas9 mediated genetic manipulation. Gene is colored black to illustrate its inactivation. (D) HPLC chromatograms of *Streptomyces* sp. SD50 mutants fermented using liquid MYM. Peak corresponding to meijiemycin is depicted by blue arrow. Wavelength (λ) = 360 nm is shown for comparison.

3.3.4. Structural elucidation of meijiemycin, a novel linear polyene natural product

(+)-HRMS analysis of meijiemycin revealed a m/z value of 1216.7357 $[M+H]^+$ (**FIGURE S 24**), which matches its chemical formula of $C_{66}H_{105}NO_{19}$ as predicted by preliminary NMR spectral analysis (**TABLE S 5**, **FIGURE S 16 TO FIGURE S 23**). The prediction of a hexaene and triene moieties based on UV spectrum of meijiemycin was confirmed by two dimensional NMR spectra (COSY, HSQC, HMBC, NOESY), which also revealed the presence of a perosamine sugar, four aliphatic methyl groups, four aliphatic

methines, thirteen methylene protons, two vinyl methyl groups, eleven oxymethines and four sets of conjugated vinyl protons. The presence of two carboxyl groups at C-1 and C-60 were determined by their δ_c values of 168.0 and 174.7 ppm, respectively. The four separate spin systems were connected through HMBC correlations (**FIGURE 25**), which completes the structure of meijiemyacin.

Figure 25: Summary of ^1H - ^1H COSY and key HMBC NMR correlations observed for meijiemyacin.

Like most other polyene natural products, determination of stereochemistry of all chiral centers in the molecule is often an arduous task that is impeded by factors like low yield, compound stability or molecular complexity.^{127, 327} Apart from complete structural elucidation through X-ray crystal analysis,³²⁸ such absolute configuration is typically achievable only through chemical derivation of the compounds to ascertain their relative configurations with each other.^{317, 329-330}

Geometry of the lactone ring within meijiemyacin was deduced by NOE correlations between $\text{H}_{\text{a}}\text{-36}$ with $\text{H}_{\text{ab}}\text{-34}$ and H-38 (**FIGURE 26**). Presence of NOE generated by 1,3-diaxial interactions between $1'\text{-H}$ and both $3'\text{-H}$ and $5'\text{-H}$ also elucidated the configuration of perosamine, which has an axial $4'\text{-H}$ without any NOE correlation with any protons (**FIGURE 26**). Stereochemistry of all olefins could be assigned based on NOE and coupling constants, except for the hexaene conjugated system, whereby overlap in proton signals impeded the absolute assignment of vinyl protons from C-44 to C-50. By applying genomics guided bioinformatics, an approach that was adopted in configurational assignment of lobosamide¹⁵⁶ and salinilactam³³¹, remaining unassigned stereochemistry of the hydroxyl groups, conjugated olefin and aliphatic methyl groups were predicted based on stereoselectivity of the KRs of Mjm PKS.

Figure 26: Summary of diagnostic proton-proton coupling constants (red) and NOE correlations (blue arrows).

3.3.5. Analysis of meijiemycin BGC

Bioinformatic analysis of *mjm* BGC (**FIGURE 24A**) revealed insightful information on the biosynthetic mechanism of meijiemycin (**TABLE S 6**). The core structure of meijiemycin was generated from a Type 1 modular PKS assembly line which involved polyketide synthases encoded by genes *mjmA* to *mjml*. This eventual intermediate was released from the terminal ACP of Mjml via TE-catalyzed hydrolysis, yielding a linear molecule with a terminal carboxyl moiety. Structurally, it differs from typical polyene macrolides, which are generated from an intramolecular hydrolysis of the terminal β -ketoacyl thioester to yield a macrolide structure. Such polyene natural products with a terminal carboxyl group are rare in nature, with mycangimycin³¹⁷ being the only one found in literature. Although ECO-02031 (**FIGURE 21**) is also a linear polyene molecule generated by PKS enzymes, its terminal carboxyl group was condensed with an aminohydroxycyclopentenone to yield an eventual amide bond.³¹⁶ Similar to the biosynthesis of other polyene macrolides,^{53, 127, 332-334} OH-39 performed a favorable intramolecular nucleophilic attack on the carbonyl group at C-35, resulting in formation of a tetrahydropyran ring within meijiemycin (**FIGURE 28**). The intermediate was subjected to two more enzymatic tailoring modifications to yield the final meijiemycin molecule. First, an aliphatic methyl at C-38 was oxidized by MjmOI to yield a carboxyl group, a modification that was also found in the biosynthesis of polyene macrolides like amphotericin²⁰⁹, nystatin³³⁴, candididin (a carboxylated derivative of perimycin)³³⁵. Notably, MjmOI is homologous to AmphN

(identity/similarity: 67%/80%), which has been demonstrated to be responsible of carboxylation of C-16 methyl group of amphotericin.²⁰⁹ Finally, a perosamine moiety was transferred to the hydroxyl group at C-41 by a perosamine transferase (MjmSIII). This perosamine moiety was biosynthesized GDP-D-mannose via a two-step catalysis involving MjmSI and MjmSII (**FIGURE 27A**). Presence of a perosamine moiety is similar to perimycin yet rarely found in most polyene natural products. In contrast, mycosamine, a positional isomer of perosamine, is found in natamycin³³², amphotericin⁵³ and nystatin³³⁴. Biosynthesis of perosamine and mycosamine share a similar pathway which involves the initial conversion of GDP-D-mannose into GDP-4-keto-6-deoxy-D-mannose by a GDP-mannose 4,6-dehydratase. To yield perosamine, the keto group of GDP-4-keto-6-deoxy-D-mannose is subsequently converted by a perosamine synthase into an amino group.³³⁶ On the contrary, mycosamine biosynthesis involves an additional isomerization step to convert GDP-4-keto-6-deoxy-D-mannose to GDP-3-keto-6-deoxy-D-mannose prior to transamination by a mycosamine synthase. Till date, no isomerase has been detected for such function, thus, suggesting that this isomerization step could likely be a spontaneous reaction.³³⁷ Despite the structural similarity of perosamine and mycosamine, these two aminodeoxysugar synthases and transferases exhibit strict specificity to their cognate substrates,³³³ which explains the absence of polyene macrolides with alternate aminodeoxysugar isomers in native microbial hosts.

Figure 27: Proposed biosynthetic mechanism of meijiemyacin (Part 1). (A) Biosynthesis of GDP-D-perosamine from GDP-D-mannose. (B) Partial biosynthetic mechanism of meijiemyacin polyketide scaffold which covers until module 14.

Figure 28: Proposed biosynthetic mechanism of meijiemycin (Part 2). Polyketide intermediate is released via TE-catalyzed hydrolysis, followed by carboxylation and glycosylation to yield the final meijiemycin compound.

Detailed analysis of Mjm PKS was also performed to comprehend the biosynthesis of the core molecule along with predictions for the unresolved stereochemistry of several chiral centers in meijiemycin. Like most Type1 modular PKSs, biosynthesis of the polyketide intermediate occurred in a canonical manner through a linear, processive assembly of twenty-six rounds of chain elongation of an acetyl starter unit (**FIGURE 27 & FIGURE 28**). The substrate specificity of acyltransferase (AT) domains of Mjm PKS were predicted through identification of key residues.³³⁸ Due to the presence of a YASH motif, ATs of module 8, 17, 18, 19, 21 and 24 were predicted to be specific towards methylmalonyl-CoA, while the remaining HAFH motif containing ATs were specific towards malonyl-CoA (**FIGURE S 25**). That analysis was translated into chain extension of twenty-six acetyl groups with six methyl branches. The degree of reduction pertaining each round of chain elongation could be determined by further analysis of individual reductive domains (KR, DH, ER) within each module. Functional analysis of KR domains revealed that MjmKR10 was non-functional due to the absence of a key tyrosine residue in its catalytic region. For other functional KRs, stereocontrol of KR in dictating the configuration of α -substituent and β -hydroxyl group was analysed through identification of key residues within the domain.³³⁹ KRs were assigned accordingly as A-type or B-type, depending on the presence of a LDD loop (characteristic for B-type KR) and a tryptophan residue (characteristic for A-type KR) in their catalytic regions (**FIGURE S 26**).³³⁹ Notably, both MjmKR23 and MjmKR24 possess a GDS motif in place of a LDD loop, and are missing a A-type KR characteristic tryptophan in their catalytic region. While NOE correlations between H-55 and H-7 confirmed the presence of a *cis*-olefin generated by module 24, coupling constants of H-8 and H-9 and NOE correlation between H-9 and H-10 also confirmed that module 23 generated a *cis*-olefin. These NMR spectral evidence rendered support that MjmKR23 and MjmKR24 might be A1-type KRs. In addition, MjmKR8 and MjmKR19 are A2-type KRs, which means they can epimerize the α -methyl group from R to a S configuration. Hence, the methyl groups at C-38 and C-16 possess a S configuration prior to other chemical modification steps. DH domains of Mjm PKS, which catalysed dehydration to generate an olefin moiety, were also analyzed for their functionality (**FIGURE S 27**). MjmDH13 was found to be missing a catalytic histidine residue in the conserved HXXXGXXXXP motif, which resulted in its inactivity. Besides the hallmark HXXXGXXXXP motif, significant variation of the sequences of three other motifs within the DH domain sequence have also been found to disrupt its

functionality.³⁴⁰ In particular, MjmdH18 possesses a LLFSL motif in place of the typical LPFXW motif, which could likely be the reason for its non-functionality. Lastly, ER domains were detected in five PKS modules (module 13, 15, 16, 17, 18). However, due to inactivity of MjmdH13 and MjmdH18, only the olefins generated by module 15, 16 and 17 were subsequently reduced into aliphatic moieties.

Overall, the biosynthetic mechanism of meijiemyacin complements its NMR spectral analysis and HRMS data, which facilitates structural elucidation of a novel linear polyene molecule.

3.3.6. Preliminary fungal inhibition assay shows narrow-spectrum activity against *Candida albicans*

Preliminary fungal inhibition assay was performed on three fungal strains, with nystatin as an experimental control for comparison. Meijiemyacin was found to inhibit *Candida albicans* DF2162R, a clinical isolate from urinary tract infection (**TABLE 4**). However, no antagonistic effect was shown against two tested *Aspergillus* sp. strains. Additional testing against more fungal strains is required to ascertain the possibility of meijiemyacin as a narrow-spectrum therapeutic option.

Table 4: MIC values from preliminary fungal inhibition assay

Strains	MIC values (µg/mL)	
	Nystatin	Meijiemyacin
<i>Candida albicans</i> DF2162R	4-8	8-16
<i>Aspergillus westerdijkiae</i> NRRL 3174	32-64	>128
<i>Aspergillus fumigatus</i> 293	8-16	>128

3.4. Conclusion

In summary, discovery of meijiemyacin, a novel linear polyene natural product, signifies the continual prospect of natural products as a source of unique chemical scaffolds with therapeutic properties. The novelty of meijiemyacin is reinforced by the absence of any homologous gene clusters in all biological databases. Preliminary studies suggested that meijiemyacin exhibits antagonistic properties against *Candida albicans*. While additional assays are required to validate its suitability as a narrow-spectrum anti-*Candida* lead molecule, the ability to genetically manipulate *mjm* BGC prompts the potential of bioengineering other analogues of meijiemyacin.

In addition, genome mining of *Streptomyces* sp. SD50 also revealed vast biosynthetic potential of secondary metabolism that remains concealed thus far. As demonstrated in this study, instead of focusing on a single approach for uncovering new compounds, an integrated approach, which combines bioinformatics, culture medium optimization and genetic engineering, holds significant promise in extracting valuable molecules from Nature's 'dark chemical matter'. While genome mining has been regularly endorsed for dereplication and identification of novel BGCs, it possesses the capacity to aid ensuing structural elucidation of novel natural products. *In silico* functional analysis of individual proteins of a cluster can provide insights on structural features of a molecule, which facilitate interpretation of complex spectral data. Such an approach was implemented to hypothesize the stereochemistry of several chiral centers in meijiemyacin and another polyene molecule³³¹, and will likely gain increasing usage in natural product structural elucidation.³⁴¹

Chapter 4: Genomics-guided discovery of novel natural products from a *Micromonospora* strain.

4.1. Introduction

Before the turn of the Century, discovery of vast number of natural products from *Streptomyces* gradually established the reputation of actinomycetes as a prolific source of natural products with diverse therapeutic applications for Mankind. While the revelation of numerous cryptic clusters harbored by *Streptomyces* in the Post Genomics Era means that *Streptomyces* are still an unexhausted source of natural products, exploiting actinomycetes from under-represented genera can supplement the natural repertoire with novel chemical scaffolds that are taxonomically unique.²¹⁷⁻²¹⁸ These non-streptomycete actinomycetes are typically termed as rare actinomycetes because of their scarcity in isolation efforts as compared to *Streptomyces*. *Micromonospora* is one particular genus of non-streptomycete actinomycetes that grow as characteristic tiny colonies with distinctive single spores on their hyphae branches.³⁴² Morphologically similar to *Salinispora* but not marine obligate, this genus of actinomycetes possesses immense biosynthetic potential for novel natural products as bioinformatic analysis revealed one-fifth of its BGCs with no homology to known BGCs.³⁴³

Previous isolation efforts recovered more than twenty *Micromonospora* strains from three different sampling locations. Preliminary bioactivity screening of these isolated actinomycetes was performed against clinical isolates, which identified *Micromonospora* sp. MD118 as a bioactive strain with remarkable antagonistic effect against Gram-Positive and Gram-negative clinical isolates (**FIGURE S 28**). Previous studies on natural product biosynthesis and discovery in two aforementioned *Streptomyces* strains had proven the effectiveness of Genomics in assisting discovery and biosynthetic studies of novel natural products. Hence, a similar genomics guided approach was adopted for pursuing novel natural products from *Micromonospora* sp. MD118. As metabolite profiling of *Micromonospora* sp. MD118 fermented in a variety of culture media suggested that most of its BGCs were silent or weakly expressed under typical laboratory conditions, development of an integrative plasmid for constitutive overexpression of putative transcriptional activators was explored as an

option for increasing product titre. Also, improvement was made on existing pCRISPR-Cas9 plasmid to expedite selection of plasmid-free mutants.

Together, a genomics guided approach with implementation of synthetic biology tools enabled the discovery of two novel compounds – 1) Lumpurmycin, a novel benzoquinone ansamycin 2) sungeidine, a distinctive enediyne-derived molecule, from *Micromonospora* sp. MD118.

4.2. Methods and materials

4.2.1. *Micromonospora* sp. MD118 isolation and strain maintenance

Micromonospora sp. MD118 was isolated from Sungei Buloh Wetland Reserve, Singapore. Sediment samples collected from the mangrove swamps were treated in a similar manner as described for *Streptomyces* sp. SD85 in chapter 2. Following treatment and serial dilution, samples were spread onto modified Gauze's medium 2 agar (15 g/L Bacto agar, 5 g/L peptone, 10 g/L glucose, 3 g/L tryptone, top up to 1 L with brackish water from Sungei Buloh Wetland Reserve) that was supplemented with 50 µg/mL cycloheximide, 10 µg/mL nalixidic acid, 10 µg/mL novobiocin and 50 µg/mL nystatin. Isolation plates were incubated for up to three weeks at 28 °C. Individual colonies were obtained after a few rounds of streaking and the strain was subsequently maintained in MS agar at 30 °C.

4.2.2. Bioactivity screening of *Micromonospora* sp. MD118 via agar overlay assay

Antibacterial assay by overlay assay was carried out as mentioned previously.³⁴⁴ Actinomycetes isolated from mangrove were tested against 6 clinical isolates. These clinical isolates are Methicillin-resistant *Staphylococcus aureus* (MRSA) strain DR42412 from sputum, MRSA strain DB68004 from blood, *Pseudomonas aeruginosa* strain DU144476/07, *Candida albicans* strain DF0002672R, *Escherichia coli* strain Y9 and *Klebsiella pneumoniae* strain Y10. Actinomycete strains were cultured in CRM media (103 g/L of sucrose, 10.12 g/L of MgCl₂·6H₂O, 15 g/L of tryptic soy broth, 5 g/L of yeast extract and 10 g/L of glucose in distilled water) to exponential phase prior to the experiment. 2 µL of actinomycete liquid culture was inoculated on minimal media agar plate (0.5 g/L of L-asparagine, 0.655 g/L of K₂HPO₄·3H₂O, 0.2 g/L of MgSO₄·7H₂O, 0.01 g of FeSO₄·7H₂O, 20 g/L of instant ocean salts, 10 g/L of agar and 10 g/L of mannitol in distilled water, pH 7) and incubated at 30 °C for a week.

The test organisms were inoculated in 5 mL of Luria-Bertani (LB) media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl in distilled water) and incubated at 37 °C overnight with shaking. 1% of overnight culture was inoculated to 5 ml of fresh LB and incubated for a few hours at 37 °C until the OD₆₀₀ of the culture reaches 0.5. For the

test organism agar overlay, the freshly grown culture was diluted to a final OD₆₀₀ of 0.0025 in 12 mL of 0.7% LB agar and poured over the actinomycete-containing plate. The plates were incubated for 24 hr at 30 °C after cooling. Zones of inhibition around the actinomycete colonies were observed and measurements were recorded in millimeters (mm).

4.2.3. Construction of pSK152 and pHZ-cod9

pHZ-cod9 was constructed using pCRISPR-Cas9 as backbone vector. *codA*(sm) gene³⁴⁵ sequence was obtained as a synthetic DNA fragment (Bio basic, Canada). The DNA was PCR-amplified as a fragment assembled to a *Bst*BI digested pCRISPR-Cas9 via Gibson Assembly to yield pHZ-cod9. pHZ-cod9 retains the original cut sites used by pCRISPR-Cas9 for insertion of sgRNA cassette and HDR template, hence cloning procedure for performing CRISPR/Cas9 mediated genetic manipulation remains unchanged. Working concentration of 200 µg/mL of 5-fluorocytosine (5FC) was determined to be effective for counter selection.

pSK152 was constructed using pSET152, a conjugative and integrative plasmid that was derived from pUC18,³⁴⁶ as a backbone vector. *kasOp**-R_J-TAI(φC31)RBS was obtained as a synthetic DNA fragment (Bio Basic, Canada). The sequence was obtained in a previous study that explored the use of *riboJ*³⁴⁷ sequence as an insulator for modulating synthetic regulatory elements to quantitative expression of genes in a precise manner.³⁴⁸ The synthetic DNA fragment was amplified via PCR and assembled to a *Not*I digested pSET152 via Gibson Assembly to yield pSK152.

4.2.4. Genomic DNA isolation from *Micromonospora* sp. MD118

Genomic DNA of *Micromonospora* sp. MD118 was isolated in a similar manner as described for *Streptomyces* sp. SD50 in Chapter 3.

4.2.5. Genome sequencing of *Micromonospora* sp. MD118

Genome sequencing of *Micromonospora* sp. MD118 was performed with Single Molecule Real Time (SMRT) sequencing technology (Pacific Biosciences, California, USA) offered by PacBio RS II platform technology for obtaining long sequencing reads.

A single SMRT cell was used and the sequencing run was performed 3 times. Overall, the sequencing reads managed to generate a 120 X coverage of the genome. A read quality value of 85 was achieved with 81,582 zero-mode-waveguides (ZMWs). A mean polymerase read length of 13,894 bp was achieved with reads of inserts (ROI) readlength value of 10,425 bp. A single contig of size 6,761,541 bp was eventually obtained after assembly.

4.2.6. Construction of gene overexpression mutants for cluster activation

4.2.6.1. Construction of *Micromonospora* sp. MD118A *mmd2050* gene overexpression mutants to validate functionality of pSK152

mmd2050 gene was PCR-amplified from genomic DNA of *Micromonospora* sp. MD118 and assembled via Gibson Assembly with *NotI* digested pSK152, *NdeI/PacI* doubly digested pIJ12551, and *NotI* digested pSET152 to yield pSK152-*mmd2050*, pIJ15221-*mmd2050*, pSET152-*mmd2050*, respectively. Insert sequences were verified by DNA sequencing.

All three newly constructed plasmids were transformed into *E. coli* ET12567/pUZ8002 for conjugation. Conjugation was performed using spores of *Micromonospora* sp. MD118A as recipient and the procedure was similar to that described in Chapter 2. Correct exconjugants were determined and stability of integrated plasmid into the genome of *Micromonospora* sp. MD118A was determined by testing for apramycin susceptibility after 20 rounds of subculture in the absence of apramycin. As complementation of *mmd2050*, a phytoene synthase, to *Micromonospora* sp. MD118A could restore the biosynthesis of orange carotenoid pigment in the strain, colorimetric measurement was used to confirm expression of *mmd2050*.

4.2.6.2. Construction of *Micromonospora* sp. MD118A7, a *lpmR1R2X* triple gene overexpression mutant, and *Micromonospora* sp. MD118AE, a *sgdR2R7* double gene overexpression mutant

To construct *Micromonospora* sp. MD118A7, a *lpmR1R2X* triple gene overexpression mutant, plasmid pSK152-*lpmR1R2X* was created. Using genomic DNA of *Micromonospora* sp. MD118 as a template, *lpmR1R2* were PCR-amplified as a single fragment while *lpmX* was PCR-amplified as another separate fragment. The two

fragments were assembled into a *NotI* digested pSK152 to yield pSK152-*lpmR1R2X*. The plasmid design included the addition of a ribosomal binding site sequence upstream of *lpmX* to ensure its translation. Plasmid containing correct inserts was validated through DNA sequencing. pSK152-*lpmR1R2X* was subsequently conjugated into *Micromonospora* sp. MD118A to create *Micromonospora* sp. MD118A7.

To construct *Micromonospora* sp. MD118AE, a *sgdR2R7* double gene overexpression mutant, plasmid pSK152-*sgdR2R7* was created. Using genomic DNA of *Micromonospora* sp. MD118 as template, *sgdR2* was PCR-amplified as one fragment while *sgdR7* was PCR-amplified as another separate fragment. The two fragments were assembled into a *NotI* digested pSK152 to yield pSK152-*sgdR2R7*. The plasmid design included addition of a ribosomal binding site sequence upstream of *sgdR7* to ensure its translation. Plasmid containing correct inserts were validated through DNA sequencing and subsequently conjugated into *Micromonospora* sp. MD118A to create *Micromonospora* sp. MD118AE.

4.2.7. Metabolite profiling of *Micromonospora* sp. MD118

For agar plate fermentation, frozen spores of *Micromonospora* sp. MD118 were spread onto 7 different media (**TABLE S 7**) agar plates and incubated at 30 °C for 14 days. Agar containing mycelia was collected and blended with acetone for extraction of compounds from the agar bits and mycelia. The organic layers were dried under vacuum and re-dissolved in methanol. Shake flask liquid fermentation was performed using the same 7 media without agar. A log phase culture of *Micromonospora* sp. MD118 was inoculated into 50 mL of fresh medium at 4% inoculum and the liquid cultures were grown at 30 °C at 180 RPM for 5-8 days until late stationary phase of growth. Mycelium pellets were separated from culture broth by centrifugation. Mycelia and culture broth were extracted separately with methanol and ethyl acetate, respectively. The organic layers were dried under vacuum and re-dissolved in methanol.

HPLC analysis of the crude extract was performed with Grace® VisionHT™ C18-HL (4.6 mm x 250 mm, 5 µm) using an Agilent 1200 HPLC system equipped with DAD detector. A gradient elution with a binary solvent system at 1 mL/min was employed

using water as mobile phase A and acetonitrile as mobile phase B. Both mobile phases were supplemented with 0.1% formic acid. LC program was set at 0 min 10% B -> 5 min 20% B -> 35 min 70% B -> 50 min 90% B -> 60 min 100% B. Analytes were monitored at λ = 220 nm, 260 nm, 284 nm, 314 nm, 520 nm.

LC-MS was performed using Agilent 1260 LC system coupled with an Agilent 6550 Accurate-Mass QTOF spectrometer equipped with a Zorbax® Eclipse plus C18 (2.1 mm x 50 mm, 5 μ m). Water with 0.1% formic acid was set as mobile phase A while acetonitrile with 0.1% formic acid was set as mobile phase B. A flow rate of 0.3 mL/min was used and the gradient elution program was similar to that employed for HPLC-DAD analysis. Mass spectra were measured using positive and negative ion modes of ESI and conditions of MS were set accordingly (capillary voltage 3 kV; nebulizer 45 psig; source temperature 300 °C; gas flow 11 L/min; MS scan rate 1/s, MS/MS scan rate 5/s, collision energy of 10, 20 and 40 eV) with a mass range of 100 – 2000 Da. Autotuning of MS was performed prior to every sample analysis.

4.2.8. Construction of *Micromonospora* sp. MD118 mutants using CRISPR/Cas9 system

All cloning steps were carried out using *E. coli* TOP10 (Invitrogen, US). *Micromonospora* sp. MD118 spores were harvested from cultures grown on M2 medium agar (**FIGURE S 29**). Experimental procedures for construction of plasmid for CRISPR/Cas9 mediated gene mutation, conjugal transfer of plasmid from donor *E. coli* to *Micromonospora* sp. MD118, identification of true exconjugants, induction of *cas9* gene expression and plasmid curing were similar to those described in Chapter 2. For mutants which were created from pHZ-cod9 derivative plasmids, an additional counter selection step was implemented to facilitate identification of plasmid-free mutants. After incubation at 37 °C for 2-3 days to cure off plasmids, individual exconjugants were streaked onto agar plates containing 200 μ g/mL 5FC. Colonies that appeared after 2-3 days of incubation at 37 °C were spotted onto two types of agar plates – 1) with apramycin and 2) without apramycin to identify true plasmid-free mutants. Method to screen for successful mutants was similar to that described for Chapter 3.

4.2.8.1. Construction of *Micromonospora* sp. MD118A, a $\Delta mmd2050$ mutant to validate CRISPR/Cas9 functionality in the strain

To abolish production of sioxanthin related carotenoids, phytoene synthase gene (*mmd2050*) was deleted by introducing a frameshift deletion of 460 bp to generate *Micromonospora* sp. MD118 $\Delta mmd2050$ knockout mutant. Diagnostic PCR was performed to establish deletion of 460 bp in *mmd2050*. This mutant strain was termed *Micromonospora* sp. MD118A and was used for subsequent genetic engineering experiments and compound production.

4.2.8.2. Construction of $\Delta lpmB$ mutant to establish relationship between *lpm* BGC and lumpurmycin

lpmB gene was deleted by introducing a frameshift deletion of 730 bp to generate *Micromonospora* sp. MD118A $\Delta lpmB$ knockout mutant. Diagnostic PCR was performed to establish deletion of 730 bp *lpmB* gene (**FIGURE S 42**).

4.2.8.3. Construction of $\Delta sgdX1$, $\Delta sgdX2$, $\Delta sgdX3$ and $\Delta sgdU15$ single mutants for *in vivo* study of sungeidine biosynthesis.

To establish the relationship between sungeidine and *sgd* BGC, *sgdU15* gene, a conserved gene of the minimal enediyne PKS cassette was targeted for disruption. *sgdU15* gene was disrupted through an in-frame deletion of 570 bp, generating *Micromonospora* sp. MD118A $\Delta sgdU15$ knockout mutant (**FIGURE S 46**).

Upon validation that *sgd* BGC is responsible for biosynthesis of sungeidine, its biosynthetic mechanism was interrogated by constructing three different single gene knockout mutants to understand the role of these selected proteins and more importantly, to generate structural intermediates through interruption of complete biosynthesis of sungeidine. *Micromonospora* sp. MD118A $\Delta sgdX1$, $\Delta sgdX2$ and $\Delta sgdX3$ single knockout mutants were generated by disrupting their corresponding genes through in-frame deletion of 222 bp, 462 bp and 514 bp, respectively (**FIGURE S 46 TO FIGURE S 48**).

4.2.9. HPLC analysis of lumpurmycin and sungeidine analogs

Cultures were incubated at 30 °C for 7 days at 170 RPM and extracted as described above for metabolite profiling. HPLC analysis of crude extracts was performed with Agilent Zorbax Eclipse XDB-C18 (4.6 mm x 250 mm, 5 µm) using an Agilent 1200 HPLC system equipped with DAD detector. A gradient elution with a binary solvent system at 1 mL/min was employed using water as mobile phase A and acetonitrile as mobile phase B. Both mobile phases were supplemented with 0.1% formic acid. LC program was set at 0 min 10% B -> 5 min 20% B -> 35 min 70% B -> 50 min 90% B -> 60 min 100% B. Analytes were monitored at λ = 220 nm, 264 nm, 280 nm, 314 nm, 420 nm, 520 nm.

4.2.10. Large scale fermentation of *Micromonospora* sp. MD118A for compound isolation

Micromonospora sp. MD118A, a $\Delta mmd2050$ mutant of *Micromonospora* sp. MD118, was selected for large scale fermentation of lumpurmycin as it did not produce the orange carotenoid pigment. *Micromonospora* MD118AE, a *sgdR2R7* double gene overexpression mutant constructed from *Micromonospora* MD118A, was used for large scale fermentation of sungeidine and its intermediate (A1) as it was capable of higher product titre. Both strains were cultured separately in M5 (1% corn starch, 0.5% cotton seed flour, 0.005% CuSO₄·5H₂O, 0.1% CaCO₃, 0.00005% NaI) liquid medium (12 X 600 mL in 2 L baffled Erlenmeyer flasks and 12 X 900 mL in 5 L Erlenmeyer flasks) for 7 days at 30 °C, 220 RPM. Mycelia was harvested by centrifugation and extracted by stirring with 15 L of methanol for 0.5 hr at ambient temperature. The methanolic extract was filtered using a Buchner funnel. Spent medium was extracted with equal volume of ethyl acetate. All organic extracts were combined and dried using a rotary evaporator to afford a crude extract (~ 3 g). The crude extract was redissolved in 500 mL of methanol and extracted thrice with equal volumes of hexane to remove fatty material. After another round of evaporation, the dried crude was washed with ultrapure water to get rid of salts and redundant water miscible materials. Remaining crude was dissolved in 20% methanol and fractionated using reversed phase solid phase extraction (SPE). Empty cartridge was packed with 20 g of Agilent Bond Elut C18 resin and flushed with 100 mL of methanol to condition the resin. The resin was subsequently equilibrated with 10% acetonitrile. Crude extract was loaded into the

resin and eluted with 20% acetonitrile, 40% acetonitrile, 60% acetonitrile, 80% acetonitrile, 100% methanol in a stepwise manner. Sungeidine and its intermediate were eluted in 60% acetonitrile, while lumpurmycin was eluted in 100% methanol.

4.2.10.1. Lumpurmycin

The 100% methanol fraction which contained lumpurmycin was dried and soaked in minimal amount of DMF with sonication. The crude was subsequently diluted with aqueous methanol. Lumpurmycin was purified from the crude fraction using Shimadzu Prominence Preparative HPLC system equipped with an ACE C18 semi-prep column (10 mm X 250 mm, 5 μ m). A gradient elution program was set up with the following conditions: 0 min 60% aqueous methanol -> 40 min 100% methanol. A flow rate of 4 mL/min was employed for the reversed phase column chromatography procedure. Fractions containing pure lumpurmycin, which eluted at $t = 27$ min, were pooled and dried using a rotary evaporator. Final yield of NMR-pure lumpurmycin was 8 mg.

4.2.10.2. Sungeidine

The 60% acetonitrile fraction, which contained sungeidine and its intermediate, was dried and redissolved in methanol with sonication. The crude was subsequently diluted with aqueous acetonitrile. Sungeidine was purified from the crude fraction using Shimadzu Prominence Preparative HPLC system equipped with an ACE C18 semi-prep column (10 mm X 250 mm, 5 μ m). A gradient elution program was set up with the following conditions: 0 min 60% aqueous methanol -> 5 min 60% aqueous methanol -> 20 min 70% aqueous methanol -> 40 min 85% methanol. A flow rate of 4 mL/min was employed for the reversed phase column chromatography procedure. Fractions containing pure sungeidine, which eluted at $t = 27$ min, were pooled and dried using a rotary evaporator. Final yield of NMR-pure sungeidine was 1.6 mg. Fractions containing A1 were pooled and dried for subsequent purification.

4.2.10.3. Sungeidine intermediate (A1)

Partially pure A1 from previous chromatographic procedures was dissolved in minimal amount of DMF with sonication and diluted in aqueous methanol. The solute was subjected to another round of chromatography using the same set up as described for

sungeidine but with a different mobile phase elution system. A gradient elution program was set up with the following conditions: 0 min 40% aqueous acetonitrile -> 23 min 40% aqueous acetonitrile -> 30 min 45% aqueous acetonitrile -> 40 min 45% aqueous acetonitrile. Fractions containing pure A1, which eluted at $t = 22$ min, were pooled and dried using a rotary evaporator. Final yield of NMR-pure A1 was 2 mg.

4.2.11. Characterization of compounds

4.2.11.1. Lumpurmycin

Lumpurmycin was isolated as a reddish brown solid; UV (ACN) λ_{max} (log ϵ) 232 (3.9), 260 (3.60), 270 (3.68), 280 (3.59), 480 (2.27) nm; ^1H NMR (400 MHz in $\text{C}_5\text{D}_5\text{N}$) δ 6.96 (1H, d, $J = 11.3$ Hz, H-61), 6.85 (1H, br s, H-5), 6.61 (1H, dd, $J = 15.8, 10.1$ Hz, H-37), 6.54 (1H, dd, $J = 15.8, 11.2$ Hz, H-47), 6.51 (1H, dd, $J = 11.3, 11.3$ Hz, H-62), 6.37 (1H, dd, $J = 15.2, 10.6$ Hz, H-13), 6.31 (1H, m, H-36), 6.30 (1H, m, H-35), 6.30 (1H, m, H-34), 6.29 (1H, dd, overlapped, H-46), 6.06 (1H, dd, overlapped, H-38), 6.06 (1H, br s, H-3), 6.05 (1H, m, H-15), 6.04 (1H, m, H-33), 6.04 (1H, dd, overlapped, H-12), 6.01 (1H, dd, $J = 15.8, 6.3$ Hz, H-48), 5.95 (1H, dd, $J = 11.3, 6.0$ Hz, H-63), 5.91 (1H, dd, $J = 15.2, 7.4$ Hz, H-45), 5.78 (1H, dd, $J = 15.2, 6.7$ Hz, H-14), 5.56 (1H, dd, $J = 15.2, 7.8$ Hz, H-11), 5.10 (1H, q-like, $J = 6.3$ Hz, H-39), 4.84 (1H, m, H-49), 4.63 (1H, m, H-59), 4.58 (1H, m, H-43), 4.51 (1H, m, H-51), 4.45 (1H, m, H-57), 4.40 (1H, m, H-7), 4.36 (1H, m, H-19), 4.28 (1H, m, H-17), 4.18 (1H, m, H-67), 4.16 (1H, m, H-31), 4.15 (1H, m, H-69), 4.15 (1H, m, H-29), 4.15 (1H, m, H-27), 4.00 (1H, m, H-55), 3.90 (1H, m, H-65), 3.33 (1H, m, H-64), 3.20 (3H, s, OCH_3 -7), 3.09 (1H, dd, $J = 15.2, 7.8$ Hz, H_a -40), 3.02 (2H, s, H_2 -72), 2.96 (1H, dd, $J = 15.8, 8.7$ Hz, H_a -42), 2.92 (1H, dd, $J = 15.2, 6.4$ Hz, H_b -40), 2.88 (1H, dd, $J = 15.8, 4.2$ Hz, H_b -42), 2.70 (1H, br d, $J = 11.6$ Hz, H_a -70), 2.58 (1H, m, H_a -16), 2.57 (1H, m, H_a -44), 2.55 (1H, m, H_b -16), 2.50 (1H, m, H_b -44), 2.18 (1H, dd, $J = 12.3, 11.6$ Hz, H_b -70), 2.17 (1H, m, H_a -50), 2.11 (1H, m, H-66), 2.09 (1H, m, H_a -32), 2.09 (1H, m, H-58), 2.09 (1H, m, H-10), 2.04 (1H, m, H_a -56), 2.01 (1H, m, H_b -50), 2.01 (1H, m, H_a -30), 1.94 (1H, m, H_a -53), 1.92 (1H, m, H_b -30), 1.92 (1H, m, H_a -18), 1.91 (2H, m, H_2 -52), 1.90 (1H, m, H_a -28), 1.90 (1H, m, H_b -32), 1.90 (1H, m, H_b -56), 1.86 (1H, m, H-68), 1.86 (1H, m, H_b -18), 1.85 (3H, br s, CH_3 -60), 1.81 (1H, m, H-54), 1.78 (1H, m, H_b -28), 1.76 (1H, m, H_a -20), 1.69 (1H, m, H_a -25), 1.68 (1H, m, H_a -26), 1.68 (1H, m, H_b -53), 1.64 (1H, m, H_b -20), 1.63 (1H, m, H_b -26),

1.63 (1H, m, H_a-8), 1.62 (1H, m, H_a-21), 1.57 (1H, m, H_b-8), 1.52 (1H, m, H_a-9), 1.44 (1H, m, H-24), 1.44 (1H, m, H_b-21), 1.40 (1H, m, H_a-22), 1.40 (1H, m, H_b-25), 1.37 (1H, m, H_a-23), 1.35 (3H, d, *J* = 6.0 Hz, CH₃-66), 1.32 (1H, m, H_b-9), 1.31 (3H, d, *J* = 6.6 Hz, CH₃-58), 1.31 (1H, m, H_b-22), 1.18 (3H, d, *J* = 6.0 Hz, CH₃-68), 1.15 (1H, m, H_b-23), 1.09 (3H, d, *J* = 6.0 Hz, CH₃-64), 1.08 (3H, d, *J* = 6.0 Hz, CH₃-54), 0.92 (3H, d, *J* = 6.6 Hz, CH₃-10), 0.88 (3H, d, *J* = 6.2 Hz, CH₃-24); ¹³C NMR (400 MHz, in C₅D₅N) δ 209.0 (C-41), 186.2 (C-4), 184.9 (C-1), 170.3 (C-73), 149.7 (C-6), 144.5 (C-2), 141.7 (C-11), 140.2 (C-60), 137.2 (C-38), 136.8 (C-48), 134.4 (C-5), 134.2 (C-63), 133.3 (C-36), 133.0 (C-13), 133.0 (C-34), 132.9 (C-46), 132.2 (C-33), 131.1 (C-35), 130.6 (C-45), 130.4 (C-14), 130.0 (C-37), 129.9 (C-47), 128.5 (C-12), 124.2 (C-62), 121.1 (C-61), 100.9 (C-3), 97.0 (C-71), 79.7 (C-59), 78.0 (C-65), 77.0 (C-7), 75.9 (C-55), 75.4 (C-67), 74.7 (C-57), 72.5 (C-15), 71.8 (C-31), 71.5 (C-29), 71.4 (C-27), 71.4 (C-17), 71.2 (C-49), 70.6 (C-51), 69.5 (C-69), 68.3 (C-39), 67.8 (C-19), 67.7 (C-43), 57.2 (OCH₃-7), 52.0 (C-40), 51.5 (C-42), 48.1 (C-72), 45.9 (C-50), 45.5 (C-30), 45.1 (C-28), 44.5 (C-70), 44.0 (C-18), 43.7 (C-32), 42.3 (C-16), 41.7 (C-44), 41.2 (C-68), 41.1 (C-58), 40.2 (C-54), 38.8 (C-8), 38.1 (C-56), 37.6 (C-66), 37.3 (C-23), 36.9 (C-10), 36.7 (C-52), 36.2 (C-26), 35.0 (C-64), 33.6 (C-20), 33.2 (C-25), 33.2 (C-24), 32.6 (C-9), 28.6 (C-53), 27.5 (C-22), 26.4 (C-21), 20.3 (CH₃-10), 20.0 (CH₃-24), 18.9 (CH₃-64), 15.7 (CH₃-60), 15.7 (CH₃-54), 12.6 (CH₃-68), 9.0 (CH₃-66), 7.9 (CH₃-58). HRMS *m/z* 1480.9094 [M-H]⁻ (calculated for C₈₂H₁₃₀NO₂₂⁻ 1480.9085).

4.2.11.2. Sungeidine

Sungeidine was isolated as a bright purple solid; UV (MeOH) $\lambda_{\max}$ (log ϵ) 228 (3.89), 250 (3.75), 294 (3.93), 312 (3.90), 418 (3.28), 500 (3.29), 536 (3.56), 572 (3.52) nm; ¹H NMR (700 MHz, in DMSO-*d*₆) δ 8.36 (1H, d, *J* = 9.2 Hz, H-19), 7.94 (1H, d, *J* = 8.2 Hz, H-11), 7.20 (1H, dd, *J* = 7.5, 7.5 Hz, H-27), 7.19 (1H, d, *J* = 7.5 Hz, H-28), 7.17 (1H, dd, *J* = 9.2, 2.3 Hz, H-18), 7.07 (1H, d, *J* = 7.5 Hz, H-26), 7.02 (1H, d, *J* = 2.3 Hz, H-16), 6.72 (1H, d, *J* = 8.2 Hz, H-10), 5.96 (1H, br s, H-5), 5.02 (1H, d, *J* = 4.4 Hz), 4.36 (1H, dd, *J* = 12.0, 6.8 Hz, H-9a), 3.61 (1H, d, *J* = 2.4 Hz, H-7), 3.50 (1H, m, H-8), 3.44 (1H, br d, *J* = 12.0 Hz, H-9b), 3.32 (1H, t, *J* = 4.4 Hz, H-3), 2.02 (3H, br s, H₃-30); ¹³C NMR (400 MHz, in DMSO-*d*₆) δ 197.8 (C-6), 187.1 (C-31), 158.3 (C-17), 158.3 (C-4), 152.6 (C-23), 136.5 (C-25), 136.3 (C-21), 133.9 (C-24), 132.8 (C-13), 132.3 (C-19), 131.8 (C-26), 131.0 (C-29), 130.6 (C-15), 129.9 (C-11), 129.0 (C-28), 128.1 (C-27),

125.6 (C-5), 124.2 (C-20), 122.2 (C-14), 120.3 (C-18), 118.3 (C-22), 117.1 (C-12), 107.0 (C-10), 103.5 (C-16), 58.2 (C-2), 53.5 (C-9), 52.3 (C-7), 44.6 (C-3), 40.5 (C-8), 23.5 (C-30). HRMS m/z 474.1173 $[M+H]^+$ (calculated for $C_{30}H_{19}NO_3SH^+$, 474.1163).

4.2.11.3. Sungeidine intermediate (A1)

A1 was isolated as a pale greenish yellow solid; UV (MeOH) λ_{max} (log ϵ) 232 (3.41), 282 (3.69), 290 (3.77), 326 (3.23), 374 (2.99) nm; 1H NMR (400 MHz, in DMSO- d_6) δ 8.36 (1H, d, J = 7.4 Hz, H-14), 7.75 (1H, d, J = 7.4 Hz, H-13), 7.58 (1H, br s, H-3), 6.76 (1H, d, J = 9.3 Hz, H-5), 6.43 (1H, s, OH-8), 6.12 (1H, br d, J = 9.3 Hz, H-6), 5.60 (1H, s, OH-7), 4.81 (1H, d, J = 8.2 Hz, H-8), 4.41 (1H, br d, J = 8.2 Hz, H-7); ^{13}C NMR (400 MHz, in DMSO- d_6) δ 195.1 (C-15), 139.5 (C-14), 136.3 (C-6), 136.3 (C-6), 134.5 (C-9), 133.3 (C-11), 133.3 (C-12), 130.8 (C-2), 130.1 (C-10), 126.8 (C-5), 125.9 (C-3), 125.0 (C-13), 108.2 (C-1), 74.4 (C-8), 71.8 (C-7). HRMS m/z 396.9389 $[M+H]^+$ (calculated for $C_{15}H_9IO_3SH^+$, 396.9395).

4.2.12. Plasmids constructed and primers used in this study

Plasmids involved in this study are listed in

TABLE S 8. Primers and other sequences used in this study are listed in

TABLE S 9 and **TABLE S 10**.

4.3. Results and discussion

4.3.1. *Micromonospora* sp. MD118 possesses a distinctive enediyne BGC and other uncharacterized BGCs.

Following observation that *Micromonospora* sp. MD118 exhibited antagonistic effect against several Gram-Positive and Gram-Negative clinical isolates (**FIGURE S 28**), attempts were made to isolate and identify the unknown antimicrobial natural product. Despite the ability to replicate bioactivity assay results, the compound appeared elusive on chemical analysis even after rigorous chemical extraction procedures. This led to speculation that the unknown compound could be a potent antimicrobial produced in traces below the limit of detection afforded by typical analytical HPLC setup. Interest in this compound prompted an alternate approach using a combination of genomics and synthetic biology for further investigation of *Micromonospora* sp. MD118.

Genome sequence of *Micromonospora* sp. MD118 subsequently revealed that it shares a 16S rRNA gene sequence identity of 99.9% with *Micromonospora aurantiaca* ATCC 27029(T). Considering its close relationship with *Micromonospora aurantiaca* ATCC 27029(T), *Micromonospora* sp. MD118 is likely to possess a circular genome (**FIGURE 29A**) as determined for its taxonomic neighbour.³⁴⁹ Preliminary mining of its genome for putative BGCs was performed using antiSMASH¹⁶¹. Notably, out of 18 BGCs (**TABLE S 11**) identified after curation, only one known BGC, homologous to previously characterized terpenoid associated sioxanthin BGC of *Salinispora tropica* CNB-440, was found,⁷³ indicating that *Micromonospora* sp. MD118 possesses significant potential for synthesizing novel secondary metabolites. Unsurprisingly, *Micromonospora* sp. MD118 shares many similar BGCs as identified in antiSMASH database¹⁶³ search of *Micromonospora aurantiaca* ATCC 27029(T). However, both strains do possess BGCs that are unique to themselves. Apart from having an additional phosphonate BGC (Cluster 3), *Micromonospora* sp. MD118 possesses a distinctive 10-membered enediyne BGC (sungeidine BGC) that was not found in the genomes of any of closely associated strains (*Micromonospora aurantiaca* ATCC 27029(T), *Micromonospora* sp. L5,³⁵⁰ *Micromonospora* sp. M42). Also, the sequence of opening reading frames flanking sungeidine BGC shared a striking homology across all three taxonomic neighbors, enabling the determination of cluster boundaries of the

BGC. Detailed analysis of sungeidine BGC uncover key differences when compared with other 10-membered enediyne BGCs (**FIGURE 41**), suggesting that it might synthesize a structurally unique molecule. Cluster 7 was identified as another cluster for further pursuit. This cluster was found in several other *Micromonospora* strains, including the aforementioned *Micromonospora aurantiaca* ATCC 27029(T), *Micromonospora* sp. L5 and *Micromonospora* sp. M42. Initial analysis of cluster 7 (lumpurmycin BGC) predicted a benzoquinone ansamycin natural product generated from a 3-amino-5-hydroxybenzoic acid (AHBA) starter unit. This compound turned out to be lumpurmycin as determined by further chemical analyses. Given the medical importance of benzoquinone ansamycins as antitumor agents by serving as heat shock protein inhibitors,³⁵¹ and other ansamycin natural products like rifamycin in the treatment against bacterial infections,³⁵² lumpurmycin was hypothesized to possess interesting therapeutic properties.

4.3.2. Generic tools developed for *Streptomyces* are compatible in *Micromonospora* sp. MD118 for genetic manipulation

Till date, most of the cloning and genetic manipulation tools for actinomycetes were developed for *Streptomyces*. Even then, successful genetic manipulation of some *Streptomyces* required conditions³⁵³⁻³⁵⁴ that deviated from typical procedures adopted for the model organism, *Streptomyces coelicolor* A3(2). Some actinomycetes isolated from the wild may possess strict regulatory system that restricts the uptake of DNA from its environment, thus, making them genetically intractable.

To assess the genetic tractability of *Micromonospora* sp. MD118 for in vivo genetic manipulation, one gene was selected for disruption via the CRISPR/Cas9 system that had been well-established in the lab. Prior to this experiment, CRISPR/Cas9 had only been demonstrated to work on one non-streptomycete actinomycete, which is *Actinoplanes* sp. SE50/110.³⁵⁵ Noting that *Micromonospora* sp. MD118 possesses a BGC (Cluster 5, **FIGURE 29B**) encoding biosynthetic genes responsible for generating the carotenoid moiety of sioxanthin and the fact that it produced an ubiquitous orange pigment resembling carotenoid compounds (**FIGURE S 29**), the phytoene synthase gene (*mmd2050*) in Cluster 5 was selected as target for disruption. Results were encouraging. Apart from exhibiting dissimilar morphological and growth characteristics to *Streptomyces* strains, disruption of *mmd2050* (**FIGURE 29D**) via CRISPR/Cas9

actualized the prospect of *in vivo* genetic manipulation of *Micromonospora* sp. MD118. Consequential abolishment of carotenoid associated orange pigment (**FIGURE 29C**) was also advantageous for subsequent isolation of other molecules. Given the redundancy of the orange pigments, *Micromonospora* sp. MD118 $\Delta mmd2050$ strain, termed as *Micromonospora* sp. MD118A, was subsequently used as a primary strain for successive experiments.

Figure 29: Genome map of *Micromonospora* sp. MD118 and generation of *Micromonospora* sp. MD118A, a Δ mmd2050 mutant. (A) Circular chromosome map of *Micromonospora* sp. MD118 with normalized GC plot depicted by inner ring in magenta and green corresponding to negative and positive values, respectively. Center ring illustrates normalized GC content of DNA sequence with positive values in yellow and negative values in black. Outer ring shows secondary metabolite BGCs representing different classes of BGCs – PKS (red), NRPS (yellow), hybrid PKS/NRPS (orange), RiPP (green), Terpene (blue), Others (black). (B) Partial sioxanthin BGC encoding carotenoid biosynthesis proteins of *Micromonospora* sp. MD118 is the only known BGC found in the strain. *mmd2050* (in bold red font), which encodes a phytoene synthase was targeted for disruption via CRISPR/Cas9. (C)

Abolishment of orange pigmentation after disruption of *mmd2050*. (D) Diagnostic PCR validated the disruption of *mmd2050*.

The compatibility of integrative vector with the strain was also tested using pSET152³⁵⁶, a conjugative and integrative actinomycete plasmid vector. Integrative plasmid vectors are useful for stable maintenance of introduced DNA without the need for selective pressure that replicative plasmids require. pSET152 possesses an *attP* sequence and a ϕ C31 derived integrase gene that recognizes a corresponding *attB* site on actinomycetal chromosome for site-directed integration of the entire plasmid into the genome.¹⁷⁹ Scanning of the genome of *Micromonospora* sp. MD118 revealed the presence of an *attB* site (**TABLE S 10**), prompting the prospect of integrating targeted genes into the strain for heterologous expression or gene overexpression studies *in vivo*. Expectations were matched by experimental outcome which demonstrated that pSET152 was stably maintained in *Micromonospora* sp. MD118 even after twenty rounds of subculture, spanning over a period of two months.

4.3.3. Modified vectors for efficient genetic manipulation of *Micromonospora* sp. MD118

4.3.3.1. pHZ-cod9, a modified pCRISPR-Cas9 plasmid with counterselectable marker facilitates selection of plasmid free mutants

In *Micromonospora* sp. MD118, successful gene disruption via CRISPR/Cas9 came at the expense of lengthy plasmid curing time. As compared with *Streptomyces* strains, exconjugants of *Micromonospora* sp. MD118 appeared to retain plasmids in spite of exposure to high temperature (39 °C) incubation and several rounds of subculture. Sequencing of plasmids revealed no mutational defects, prompting the possibilities that either the plasmid was effectively integrated into the chromosome or the strain was innately less sensitive to the temperature requirement conferred by pSG5³⁵⁷ replicon of pCRISPR-Cas9. Nonetheless, to overcome both scenarios would require a laborious process of picking of many colonies and subjecting them to apramycin susceptibility assay. Alternatively, a counterselectable marker could be introduced into pCRISPR-Cas9 to facilitate identification of plasmid-free colonies generated after plasmid curing procedure. The addition of a *codA* gene encoding cytosine deaminase (CodA) which converts 5FC to 5FU (**FIGURE 30B**), a toxic metabolite that confers cell death, could serve as a negative selection against plasmid bearing cells. CodA(sm),

a D314A mutant of the original CodA, was used instead of the original CoA as it possesses better substrate specificity and enzymatic efficiency than the wild-type protein.³⁵⁸ Such similar application of CodA(sm) as a counter selection marker in actinomycetes is not unprecedented,³⁴⁵ and has also been implemented in the design of another CRISPR/Cas9 based vector developed for actinomycetes.¹⁸² The modified pCRISPR-Cas9, containing *codA(sm)* gene for counterselection, was named as pHZ-cod9 (**FIGURE 30A**).

Figure 30: plasmid map of pHZ-cod9. (A) Vector map of pHZ-cod9. (B) *codA(sm)* gene encodes cytosine deaminase, which converts 5FC into 5FU, a toxic metabolite.

Counterselection efficiency of the newly constructed pHZ-cod9 was tested by subjecting *Micromonospora* sp. MD118A containing pHZ-cod9 with serial concentrations of 5FC. Low concentration of 5FC (<50 µg/mL) resulted in higher percentage of false positives (cells that contain plasmids). Effective working concentration of 5FC was determined to be 200 µg/mL as lower concentration of 5FC (<100 µg/mL) corresponded in higher percentage of false positives (plasmid bearing cells). Incubation of the plates with exconjugants that was exposed to 37 °C for 2-3 days prior to exposure 200 µg/mL of 5FC yielded more colonies than a similar procedure performed without the additional 37 °C incubation step, suggesting that curing of pSG5 replicon bearing plasmid in *Micromonospora* sp. MD118 was retarded rather than completely non-existent.

4.3.3.2. pSK152, an integrative and conjugative plasmid with synthetic modular elements for gene overexpression

Gene clusters in actinomycetes are typically under complex regulation with unresolved underlying mechanism. Conventional design of regulatory elements to drive gene overexpression in actinomycetes are subjected to factors that might affect consistency

in experimental outcomes. Translational efficiency of a target gene is not merely dependent on its transcription, but also other structural elements like 5' untranslated region (5' UTR), ribosomal binding site (RBS), and terminator sequence. Previously, 5' UTR had been shown to mediate ribosome binding on mRNA transcripts and affect their translation,³⁵⁹ and studies performed on *Streptomyces coelicolor* also demonstrated the influence of 5' UTR on both native genes³⁶⁰ and gene overexpression³⁶¹ experiments. The importance of implementing precise design of structural elements for genetic control in synthetic biology of actinomycetes was further supported by a recent study on *Streptomyces albus* J1074, which demonstrated the influence of RBS and 5' UTR in gene overexpression.³⁶²

Consequently, synthetic regulatory elements were considered for design of pSK152, an integrative and conjugative plasmid for gene overexpression in actinomycetes. The design utilized a promoter and RBS sequence that was separated by a ribozyme, RiboJ, which eliminates interference of both elements through autocatalytic cleavage of sequence upstream of the ribozyme.³⁴⁷ Drawing inspiration from the consistent quantitation of gene overexpression rendered by these modular synthetic regulatory elements,³⁴⁸ four different combinations of promoter-RiboJ-RBS cassettes were designed for future operon activation experiments via promoter replacement strategy. Out of these four cassettes, a *kasOp** - *riboJ* - RBS (from Φ C31 phage Tail Protein gene) cassette was inserted into a promoterless integrative and conjugative plasmid, pSET152, to create pSK152 (**FIGURE 31A**).

Efficiency of pSK152 was compared with pSET152 and pIJ12551 via overexpression of *mmd2050* gene. *mmd2050* gene overexpression plasmids were integrated into *Micromonospora* sp. MD118A to complement the loss of *mmd2050* in the native chromosome (**FIGURE 31B**). Growth on MS agar, which had previously been confirmed to activate sioxanthin biosynthesis in the native strain, demonstrated the ability of pSK152-*mmd2050* to complement the loss of *mmd2050* in *Micromonospora* sp. MD118A. Comparatively, colonies with integrated pIJ12551-*mmd2050* also exhibited intense orange color, while promoterless pSET152 yielded a faint orange color (**FIGURE 31C**).

Figure 31: Validation of pSK152, an integrative and conjugative plasmid for gene overexpression. (A) Vector map of pSK152. (B) Overexpression of *mmd2050* using newly constructed pSK152 in *Micromonospora* sp. MD118A. Disrupted *mmd2050* in the strain was complemented with a functional copy of *mmd2050* rendered by integrative plasmid bearing native *mmd2050* gene. (C) *mmd2050* overexpression mutants exhibited orange pigmentation, indicative of restoration of carotenoid biosynthesis.

Due to shortage of time, detailed transcriptional and translational analysis of *mmd2050* overexpression mutants was omitted and colorimetric assay was the sole option for determining expression efficiency of *mmd2050*. Production of orange pigment was still subjected to other regulatory elements within the entire biosynthetic pathway. Given that other genes in the BGC were regulated by native regulatory mechanisms, which were likely similar for all three different *mmd2050* overexpression mutants, production of orange pigment would be correlated to genetic control of *mmd2050*. However, in the event whereby expression of *mmd2050* exceeded the precursor production rate, the unbalanced metabolic flux would be limited by other proteins within the carotenoid biosynthetic pathway.

Nonetheless, validation of pSK152 and its compatibility in *Micromonospora* sp. MD118A fulfilled the intention of employing it for succeeding experiments in the strain.

4.3.4. Discovery of lumpurmycin

4.3.4.1. Constitutive overexpression of three genes *lpmR1*, *R2* and *X* elicited the production of lumpurmycins

The application of pSK152 rendered the possibility of cluster activation through overexpression of putative pathway activators a feasibility. Bioinformatic analysis of *lpm* BGC (**FIGURE 32A**, **TABLE S 12**) revealed two putative regulators, LpmR1 and LpmR2, which are members of the LuxR family transcriptional regulators. This family of proteins are widespread in actinomycetes, with diverse functions that are dependent on their domain architecture.³⁶³ LuxR proteins of PKS clusters typically serve as positive regulators of PKS based BGCs, like oligomycin,³⁶⁴ abyssomicin,¹⁹⁶ pimarinin,³⁶⁵ etc. Notably, two LuxR family proteins, GdmRI and GdmRII, had been found to serve as positive regulators in the biosynthesis of geldanamycin, a benzoquinone ansamycin.¹⁹³ Previously, constitutive overexpression of a LuxR positive regulator gene successfully activated the biosynthesis of a giant macrolide, stambomycin,³⁶⁶ while a similar strategy also activated the production of ansatrienes.³⁶⁷ Hence, a similar approach was adopted to explore the activation of *lpm* BGC through constitutive overexpression of *lpmR1* and *lpmR2* along with *lpmX*, a svp-type PPTase. Svp is a Sfp-type PPTase that was characterized previously and found to be a broad acting PPTase that phosphopantetheinylates the ACPs and PCPs of PKSs and NRPSs, respectively.³⁶⁸ Alignment of LpmX with Svp revealed that LpmX is homologous to the queried protein, sharing a %identity/%similarity value of 60%/69%. The significance of PPTase in secondary metabolism was demonstrated previously by activation of several cryptic BGCs,³⁶⁹ including the recent discovery of puromycins,³⁷⁰ through constitutive overexpression of PPTases. As such, constitutive overexpression of *lpmX* was considered as an approach for activating secondary metabolism in *Micromonospora* sp. MD118A, particularly *lpm* BGC. Instead of generating single gene overexpression mutants which required additional screening efforts, a plasmid construct containing all three genes (pSK152-*lpmR1R2X*) was generated as a rapid, preliminary attempt to assess the effects on the production of secondary metabolites.

Interestingly, colonies of *Micromonospora* sp. MD118A7, a triple gene (*lpmR1R2X*) overexpression mutant generated via conjugation of *Micromonospora* sp. MD118A

with pSK152-*lpmR1R2X*, exhibited a faint but distinctive reddish brown color that was absent from the empty vector bearing exconjugants (**FIGURE 32B**). Comparative metabolite profiling of *Micromonospora* sp. MD118A7 with MD118A revealed tiny resolved peaks in the chromatogram of *Micromonospora* sp. MD118A7 with nearly identical UV spectra. One of the peaks was eventually determined to be lumpurmycin. Also, a peculiar phenomenon was observed as traces of many unresolved, fuzzy peaks with similar UV spectra between $t = 21$ to 29 min resulted in a raised baseline between that region of the chromatogram (**FIGURE 32C**).

Encouraged by the identification of putative compounds from *lpm* BGC but hindered by the low yield afforded from MS agar fermentation, yield improvement through culture screening was performed. The strain was cultivated in six different media (**TABLE S 7**), under both liquid and solid conditions, to identify the optimal conditions for better yield. Overall, cultivation in liquid M5 medium was determined to generate the highest yield of the characteristic peaks identified earlier. Surprisingly, despite showing increase in yield for lumpurmycin and three other analogs, metabolite profiling of *Micromonospora* sp. MD118A7 fermented in liquid M5 also exhibited many unresolved peaks between $t = 21$ to 29 min. Abolishment of lumpurmycin and these peaks in *Micromonospora* sp. MD118A Δ *lpmB* mutant (**FIGURE 32C**) indicated that these peaks were compounds derived from the biosynthetic pathway of lumpurmycin gene cluster. In comparison, *Micromonospora* sp. MD118A produced lumpurmycin in higher quantity than *Micromonospora* sp. MD118A7 (**FIGURE 32C**), albeit at the expense of decreased yield of other analogs. Although the roles of these analogs were undetermined, this astonishing finding seemingly suggested a modification in coordination of lumpurmycin biosynthetic pathway as introduced by the overexpression of *lpmR1*, *lpmR2*, and *lpmX*. With priority given to yield instead of number of analogs, *Micromonospora* sp. MD118A was fermented in large scale for isolation of lumpurmycin.

Figure 32: Discovery of lumpurmycin through *lpmR1R2X* overexpression and culture condition screening. (A) Organization of putative lumpurmycin BGC. Genes with proposed functions in biosynthesis of lumpurmycin are annotated. (B) Exconjugants bearing pSK152-*lpmR1R2X* displayed a faint reddish brown pigmentation that was absent in control. (C) HPLC chromatogram of MD118A7 versus MD118A, wavelength (λ) = 284 nm is shown for comparison. Lumpurmycin is depicted by blue arrow. Blue asterisks and line represents analogs with similar UV spectrum. (D) UV spectrum of lumpurmycin.

4.3.4.2. Structural elucidation of lumpurmycin, a novel benzoquinone ansamycin

Figure 33: Lumpurmycin and other notable ansamycin natural products.

Lumpurmycin has a chemical formula of $C_{82}H_{131}NO_{22}$, which matches the m/z value of 1480.9094 $[M-H]^-$ obtained from (-)-HRMS analysis (**FIGURE S 30**). It is the largest ansamycin natural product to be discovered thus far (**FIGURE 33**). The ansamycin molecule possesses a tetrahydropyran ring and a benzoquinone moiety as determined by NMR spectral analysis, with a triene system that accounted for the characteristic UV maxima (260 nm, 270 nm, 282 nm) in its UV spectrum (**FIGURE 32D**). The structure of lumpurmycin was resolved with a series of 1D (1H , ^{13}C ^{13}C DEPT-135) and 2D (ROESY, HSQC, HMBC, COSY, NOESY) NMR experiments (**TABLE S 13**, **FIGURE S 31** to **FIGURE S 41**), which identified seven aliphatic methines, seven aliphatic methyl groups, seventeen oxymethines, twenty-two methylenes, nineteen vinyl protons and one vinyl methyl group.

Spin systems of lumpurmycin were connected through 1H - 1H COSY and HMBC correlations (**FIGURE 34**), except for three regions (C-20 to C-23; C-34 to C-36; C-52

to C-53), which were connected through bioinformatic predictions of Lpm PKS system. Geometry of vinyl protons were determined through J -coupling constants and NOE correlations (**FIGURE 35**) confirmed the stereochemistry of substituents on the tetrahydropyran ring within lumpurmycin. The magnitude of signals generated by lumpurmycin inevitably yielded overlapping signals, which impeded the confirmation of remaining chiral centers. Even though the molecule is stable, low yield of lumpurmycin (0.4 mg/L) hindered chemical derivatization of the molecule to resolve stereochemistry of unresolved chiral centers. Although X-ray crystal structure analysis would provide complete structural elucidation of lumpurmycin, attempts to crystal the compound ended in failure. The unresolved chiral centers were eventually determined through predictions provided by analysis of PKSs in the *lpm* BGC.

Figure 34: Diagnostic ^1H - ^1H COSY and HMBC correlations observed for lumpurmycin. Carbons numbered in red font were connected through analysis of Lpm PKS system.

Figure 35: Summary of diagnostic proton-proton coupling constants and NOE correlations observed for lumpurmycin. Stereochemistry of chiral carbons that could not be determined from J -based configuration and NOESY experiment are indicated with blue asterisk.

4.3.4.3. Proposed biosynthetic mechanism for lumpurmycin

Biosynthesis of lumpurmycin begins with the use of AHBA as a starter unit, which is tethered onto the ACP of LpmA's loading module. The biosynthetic origin of AHBA, which has been reviewed previously,³⁷¹ begins with UDP-D-glucose that is oxidized by LpmS1 into UDP-3-keto-D-glucose (**FIGURE 36A**). Using glutamine as an amino group donor, LpmS2, an AHBA synthase with proposed additional role in AHBA biosynthesis, subsequently transaminates the molecule into UDP-kanosamine. Subsequent hydrolysis of UDP-kanosamine into kanosamine is catalyzed by LpmS3, a phosphatase. This is followed by phosphorylation of kanosamine by a ROK family kinase, LpmS4, to yield kanosamine-6-phosphate, which undergoes a two-step process of isomerization and transketolation to yield an imino-erythrose-4-phosphate (IE4P) substrate. Although no isomerase has been identified within all characterized ansamycin BGCs, transketolation can be traced to the presence of transketolase within the BGCs encoding saliniketals and rifamycins in *Salinispora arenicola*.³⁷² Formation of (3,4-dideoxy-4-amino-D-arabino-seduheptulosonic acid 7-phosphate) aminoDAHP proceeds with an aminoDAHP synthase, utilizing IE4P as a substrate. Interestingly, *lpm* BGC does not possess an aminoDAHP synthase gene, along with the transketolase subunits, that are crucial for generating the required intermediates in the amino-shikimate pathway for biosynthesis of AHBA. However, screening of the genome of *Micromonospora* sp. MD118 revealed these enzymes in separate location of the genome. α and β subunits of transketolase are located in position (1822086 bp -1822769 bp) and (1822784 bp -1823659 bp), sharing homology with transketolase subunits Sare_1272 (%identity/%similarity: 61/69) and Sare_1273 (%identity/%similarity: 54/63), respectively. While an aminoDAHP synthase was found at position (117037 bp -118443 bp) of the genome, and is homologous to RifH (%identity/%similarity: 53/64), an aminoDAHP synthase involved in biosynthesis of rifamycin in *Amiclatopsis rifamycinica*.³⁷³ LpmS5 cyclizes aminoDAHP into (5-deoxy-5-aminodehydroquinic acid) aminoDHQ, which is dehydrated into 5-deoxy-5-aminodehydroshikimic acid (aminoDHS) by LpmS6. The intermediate is finally converted into AHBA by LpmS2.

Figure 36: Proposed biosynthesis of lumpurmycin (part 1). (A) Biosynthesis of AHBA. X depicts enzymes missing in *lpm* BGC but are located in other parts of the genome. (B) Proposed biosynthesis of lumpurmycin that is dissected into three portions. Numbering in red indicates module number and their corresponding role in the biosynthesis. Crossed domains are non-functional.

The starter unit AHBA undergoes thirty-three rounds of chain elongation catalyzed by the PKS system of Lpm BGC (LpmA to LpmM), followed by cyclization to form a benzoquinone ansamycin structure (**FIGURE 36B**). All KS domains were predicted to be active due to the presence of key catalytic residues (**FIGURE S 43B**). Analysis of AT domains of Lpm PKS modules was performed to predict their functionality and substrate specificity.³³⁸ Notably, two modules (LpmPKS19 and LpmPKS21) were predicted to be incapable of chain elongation due to the absence of AT domain in LpmPKS19 and the absence of crucial serine residue in the active site motif of LpmAT21 (**FIGURE S 43**). Conversely, two additional rounds of chain elongation in the polyketide chain suggested the possibility that two modules (LpmPKS9 and LpmPKS20) are iterative, catalyzing two rounds of chain elongation each in the

biosynthesis of lumpurmycin. Seven ATs (LpmAT7, 24, 26, 27, 29, 30, 31) were found to possess methylmalonyl-CoA specific YASH motif, indicating their presence for methylmalonyl-CoA as substrate for chain elongation (**FIGURE S 43A**). Remaining ATs possess malonyl-CoA specific HAFH motif, except LpmAT9, which has a FAAH motif (**FIGURE S 43B**). Consequently, substrate prediction for LpmAT9 was impossible. Sequence divergence in this hallmark motif has been reported in the biosynthesis of stambomycin, whereby a GAAH motif was found in SamAT13, resulting in substrate promiscuity.³⁶⁶ Inability to elucidate other lumpurmycin analogs meant that additional evidence is still required to determine the substrate selectivity of LpmAT9. However, based on the structure of lumpurmycin elucidated, the iterative LpmPKS9 is likely to utilize a methylmalonyl-CoA substrate for its initial round of chain elongation, followed by a malonyl-CoA substrate for its second round of chain elongation.

Figure 37: Proposed biosynthesis of lumpurmycin (part 2). Proposed biosynthesis of lumpurmycin structure generated by Lpm PKS. Final structure of lumpurmycin is generated from a PKS intermediate. Skeletal carbons in bold represents units condensed per round of elongation. Deviation from prediction of PKS and tailoring reactions are depicted in blue font.

Detailed analysis of KR domains of Lpm PKS revealed that LpmKR17 lacks a catalytic tyrosine residue, which makes it non-functional (**FIGURE S 44B**). This finding was reflected by the acetyl group provided by module 16, which was left unprocessed by

LpmPKS17. Key residues and motifs within the sequence of KRs were also analyzed to determine the stereochemistry of α -substituents and β -hydroxyl groups dictated by KR domains of Lpm BGC.^{339, 374} The stereochemistry of β -hydroxyl groups subsequently determines the geometry of olefin generated by a succeeding DH domain, if present. A-type and B-type KR dictate stereochemistry of β -hydroxyl groups in opposing manner (**FIGURE S 44A**). The presence of LDD-loop motif in B-type KRs differentiates them from A-type KRs, which possess a tryptophan residue within its catalytic region. KRs with (L/T/I/R)(A/D/E)(D) motif and lack tryptophan residue within their catalytic region were also classified as B-type KRs as the relaxed nature of LDD-loop motif had been observed in many previous instances,^{327, 375} even in KRs of *trans*-AT PKS, where only the second aspartic acid residue is conserved.³⁷⁶ Both types of KRs can be further classified to determine their influence on the stereochemistry of α -substituents. A2-type KR possesses a histidine residue at a specific position within their catalytic region, which is absent in A1-type KR. Whereas B2-type KR possesses a proline residue at another specific position within their catalytic region, which is absent in B1-type KR. Sequence alignment and classification of LpmKRs can be found in **FIGURE S 44B**. Stereocontrol dictated by LpmKRs was used to determine the stereochemistry of chiral centers that remained unresolved from NMR spectral analysis.

Analysis of DH domains of Lpm PKS revealed that LpmDH1 lacks a conserved catalytic histidine residue in the HXXXGXXXXP motif,³⁴⁰ rendering it non-functional (**FIGURE S 45**). Structural elucidation of lumpurmycin confirmed the prediction that LpmPKS1 partially reduced the β -keto group provided by AHBA into a hydroxyl moiety. Even though LpmDH3, 4, 28 and 29 possessed a HXXXDXXXXP motif (**FIGURE 28**), their functionalities were confirmed by the presence of olefins generated by dehydration of β -hydroxyl groups from preceding modules. Two discrepancies were observed, which deviated from the structure of lumpurmycin. Firstly, even though LpmDH10 was predicted to be active, lumpurmycin structure showed a hydroxyl group marked by blue asterisk in **FIGURE 37**, suggesting that dehydration did not occur. Secondly, lumpurmycin structure seems to suggest that the diene moiety generated by LpmPKS28 and LpmPKS29 was migrated a single carbon away from its original position (**FIGURE 37**). Analysis of LpmDH29 predicted a functional, normal dehydratase domain while no obvious genes of Lpm BGC was identified to justify this

unexpected diene shift from a biosynthetic perspective. Nonetheless, this rare phenomenon was observed previously in several transAT PKS derived natural products (rhizoxin,³⁷⁷ bacillaene,³⁷⁸ corallopyronin A³⁷⁹) and particularly in the biosynthesis of another ansamycin natural product, where similar migration of a diene resulted in formation of resultant structure of ansamitocin.³⁸⁰ Further studies confirmed that the unusual migration of diene within ansamitocin occurred during the third module of proansamitocin polyketide scaffold biosynthesis.³⁸¹ While the diene migration in lumpurmycin could likely occur in a manner similar to ansamitocin, further evidence is required to establish the biosynthetic logic behind the diene migration within lumpurmycin.

The nascent polyketide chain is likely released from LpmM via hydrolysis that is catalyzed by a terminal TE domain of LpmM. Formation of an ansamycin structure through cyclization of the linear polyketide chain likely occurs via TE-mediated lactamization. At present, it is unclear whether this process involves the additional standalone TE, LpmO3, found in the gene cluster. Two more tailoring reactions are required to complete the biosynthesis of lumpurmycin (**FIGURE 37**). One reaction involves methylation of the intermediate at 7-OH by LpmO2, an O-methyltransferase. The other involves oxidation of the phenolic moiety into a benzoquinone structure. The mechanism governing this oxidative procedure is still unclear. Previous study on geldanamycin suggested that oxidation of hydroquinone into benzoquinone from hydroquinone occurs spontaneously upon hydroxylation of the benzene ring of AHBA.³⁸² LpmO1, a FAD-dependent oxidoreductase that is homologous to GdmM from geldanamycin BGC,³⁸³ is putatively a hydroxylase that is responsible for hydroxylation of the benzene ring to form a hydroquinone structure.

4.3.5. Discovery of Sungeidine

4.3.5.1. Disruption of a *unbL* homologous gene, *sgdU15*, abolished production of sungeidine

Figure 38: Sungeidine and other 10-membered enediyne natural products. Enediyne cores are depicted in blue.

Amid optimizing fermentation conditions for lumpurmycin production, metabolite profile of *Micromonospora* sp. MD118A cultured in M5 medium revealed traces of a compound (sungeidine) that absorbed visible light with at $\lambda(\text{max}) = 500, 536, 572 \text{ nm}$ (FIGURE 39E). Bearing in mind that molecules capable of absorbing visible light at such wavelengths typically possess extensive conjugated π electron systems, the genome was screened for putative BGCs capable of producing similar chemical

moieties. Discounting both Cluster 5 and *lpm* BGC, a 10-membered enediyne BGC (Cluster 10, **TABLE S 11**, **FIGURE 39B**) was speculated to encode the mysterious compound. This cluster is homologous to other BGCs that encode anthraquinone fused 10-membered enediyne compounds like dynemicin A,³⁸⁴ uncialamycin,³⁸⁵ tiancimycin A³⁸⁶ and yangpumycin A³⁸⁷ (**FIGURE 38**) that are capable of absorbing visible light between the range of $\lambda = 540$ to 600 nm. To confirm the speculation that sungeidine is synthesized by Cluster 10 (*sgd* BGC), *sgdU15* was disrupted via CRISPR/Cas9 mediated genetic engineering. *artU15* is homologous to *unbL*, which is one of the five conserved genes [*unbL*, *unbV*, *unbU*, enediyne *pks* (*pksE*), thioesterase gene] found in all enediyne PKS BGCs,³⁸⁸⁻³⁸⁹ hence its disruption would prove detrimental to production of sungeidine if the compound is biosynthetically associated with *sgd* BGC. While *Micromonospora* sp. MD118A Δ *sgdU15* predictably failed to produce sungeidine, comparison of metabolite profiles also revealed abolishment of other peaks A1 and A2 (**FIGURE 39D**). The unique UV spectra exhibited by these peaks (**FIGURE 39E**) suggest that they might be analogs that are structurally distinctive from sungeidine. Although these compounds might be significant in providing more clues for resolving the ensuing mystery surrounding the biosynthetic mechanism of enediyne core or anthraquinone fused enediyne natural products, their yield was too low for obtaining adequate material with the current set up afforded in the laboratory.

4.3.5.2. Constitutive overexpression of *sgdR2* and *sgdR7* upregulated *sgd* BGC

Upregulation of *sgd* BGC was, thus, explored as a possible option for yield improvement. While replacement of native promoters with strong, constitutive ones had proven to be a useful strategy for activation or upregulation of weakly active clusters,^{200, 390-391} genes within *sgd* BGC are arranged into multiple operons that require time-consuming replacement of native promoters in a step-wise procedure. Instead, overexpression of putative positive regulators of the cluster was considered. As this strategy had previously facilitated the identification of lumpurmycin, it was hoped that a similar approach could increase the yield titre of sungeidine and its intermediates. Three transcriptional regulators (*SgdR2*, *R3*, *R7*) were identified from cluster, with *SgdR2* being a SARP family transcriptional regulator, *SgdR3* being a HxIR family transcriptional regulator and *SgdR7* being an AraC family transcriptional regulator. *SgdR7* shares homology with *SgcR2*, an AraC family transcriptional

regulator that had been established to positively regulate biosynthesis of C-1027, a nine-membered enediyne natural product,³⁹² while SgdR2 was identified as a putative positive regulator due to the fact that most SARP family regulators act as transcriptional activators of BGCs. These two genes were amplified into pSK152 to generate pSK152-*sgdR2R7* for overexpression. SgdR3 was not included as previous knockout study of SgcE1, which is homologous to SgdR3, suggested that it did not have any effect on C-1027 production.³⁹³

Gratifyingly, co-overexpression of *sgdR2* and *sgdR7* in *Micromonospora* sp. MD118AE, which refers to *Micromonospora* sp. MD118A with pSK152-*sgdR2R7* integrated within its genome (**FIGURE 39C**), resulted in an approximately 1-3 fold increase in production of sungeidine and its associated analogs (**FIGURE 39D**). The increase in yield, albeit small, was significant in facilitating observation of peaks A3 and A4 (**FIGURE 39E**), which were nearly negligible in the chromatogram of *Micromonospora* sp. MD118A (**FIGURE 39D**).

Figure 39: Identification of sungeidine BGC and metabolite profiling of associated mutants. (A) Structures of sungeidine and its intermediate, A1. **(B)** Organization of genes in *sgd* BGC. Genes in *sgd* BGC were annotated according to comparison with dynemicin BGC. **(C)** Integration of pSK152-*sgdR2R7* into the genome of *Micromonospora* sp, MD118 to generate *Micromonospora* sp. MD118AE. **(D)** HPLC profiling of mutants for yield comparison and identification of *sgd* BGC associated peaks, wavelength (λ) = 314 nm is shown for comparison. Peak corresponding to sungeidine is depicted in red, while all other identified intermediates are depicted in blue. **(E)** UV spectra of peaks associated with *sgd* BGC.

4.3.5.3. structural elucidation of sungeidine and A1

Following isolation of adequate material through large scale cultivation of *Micromonospora* sp. MD118AE, structures of sungeidine and A1 were elucidated through mass spectral and NMR spectral analyses.

(+)-HRMS analysis of sungeidine revealed a m/z value of 474.1173 $[M+H]^+$ (**FIGURE S 55**), which matches its chemical formula of $C_{30}H_{19}NO_5S$ predicted from NMR spectral analysis (**TABLE S 15, FIGURE S 49 TO FIGURE S 54**). 2D NMR spectra (TOCSY, HSQC, HMBC, NOESY) identified an anthracene moiety with a hydroxyl at C-17 and a fused thiolactone attached to C-12 and C-14. Through HMBC and NOE correlations, the original 10-membered enediyne core was determined to be cyclized into 1,2,3-trisubstituted benzene that was bonded to C-21 of the anthracene (**FIGURE 40A AND FIGURE 40B**). A combination of TOCSY and HBMC correlations also revealed bonding between the anthracene with a pyrrolidine ring through the nitrogen, resulting in a tertiary amine. The only single methyl group identified was a vinyl methyl attached to C-4 through NOE and HMBC correlations.

Figure 40: 2D NMR correlations of sungeidine and A1. (A) TOCSY and key HMBC correlations of sungeidine. (B) Diagnostic NOE correlations of sungeidine. (C) 1H - 1H COSY and key HMBC correlations of A1.

A1, an intermediate of sungeidine, was determined to have a chemical formula of $C_{15}H_9IO_3S$ through (+)-HRMS, which revealed a m/z value of 396.9389 $[M+H]^+$ (**FIGURE S 62**). Presence of iodide was confirmed through MS/MS analysis of A1, which revealed a dominant fragment with loss of H_2O (378.9284 Da) and another fragment of lower abundance with losses of iodine and H_2O (252.0243 Da) (**FIGURE S 63**). NMR spectral analysis (**TABLE S 16, FIGURE S 56 TO FIGURE S 61**) elucidated the structure of A1, which possesses a thiolactone fused anthracene moiety as determined by HMBC and COSY correlations (**FIGURE 40C**). Unlike sungeidine, A1 possesses an additional hydroxyl group at C-8 and an iodide at C-1.

4.3.5.4. *sgd* BGC is unique from other identified anthraquinone-fused enediyne BGCs

At present, four anthraquinone-fused enediyne natural products (**FIGURE 38**) had been identified, along with their corresponding BGCs.^{386-387, 394} Tiancimycin A (Tnm), yangpumicin A (Ypm) and uncialamycin (Ucm) shares an identical enediyne core that is different from dynemicin A (Dyn), suggesting that processing of the enediyne core moiety might involve different enzymes that are absent in respective BGCs. Dyn BGC is also distinctively different from the *tnm*, *ypm* and *ucm* BGCs, possessing more genes than are absent in the other three clusters. Structurally, all four compounds possess an anthraquinone moiety, albeit with different substituents on one of its fused rings. Biosynthetic studies on *dyn* and *tnm* BGCs had subsequently established the role of Cytochrome P450 hydroxylases in the observed differential hydroxylation pattern of the anthraquinone ring.^{219, 386} However, difference in the enediyne core (**FIGURE 43B**) and more importantly, mechanism governing the biosynthesis of 10-membered enediyne core remains unresolved.

Sungeidine BGC

Yangpumicin BGC

Tiancimycin BGC

Uncialamycin BGC

Dynemicin BGC

Figure 41: Comparison of *sgd* BGC with other homologous enediyne BGCs. Genes annotated in red font represent those that were only found in either *sgd* BGC or the other homologous BGCs. All other genes are faded into background. Genes of all four homologous BGCs are annotated according to literature, while genes in *sgd* BGC are annotated according to comparison with *dyn* BGC.

Interestingly, unlike the four anthraquinone fused enediyne compounds, sungeidine possesses an anthracene moiety instead. Despite failure to recover any congeners of sungeidine with an intact enediyne core, structure of its enediyne core could be postulated retrospectively from the substrate prior to Bergman Cyclization (**FIGURE 43A**). The deduced enediyne core of sungeidine would be structurally distinctive from those found in other 10-membered enediyne natural products (**FIGURE 43B**). Notably, *ypm*, *ucm*, *tnm* and *dyn* BGCs all possess a set of three genes (*N*, *O*, *P*) which are missing in *sgd* BGC. *dynA1*, *dynA2*, *dynE13* are respectively homologous to *N*, *O* and *P* genes from *tnm*, *ucm* and *ypm* BGCs. Gene *N* was predicted to encode an ester cyclase from the *SnoaL* family of polyketide cyclases, which catalyzes aldol condensation.³⁹⁵ This protein could likely be involved in the fusion of anthraquinone with the 10-membered enediyne core. Gene *O* encodes a hypothetical protein whose function is unknown, while gene *P* encodes an FAD-dependent oxidoreductase, which could likely serve as an epoxidase that generates the epoxide ring found in the enediyne core of *Dyn*, *Ucm*, *Ypm* and *Tnm* (**FIGURE 38**). However, due to the lack of experimental evidence, the functions of these three proteins in the biosynthesis of anthraquinone fused enediyne compounds remained ambiguous. Another interesting observation is the presence of gene *M* only in *tnm*, *ypm* and *ucm* BGCs. Gene *M* encodes a Rieske (2Fe-2S) iron-sulfur domain protein, which are non-heme iron dependent oxygenases capable of catalysing unique tailoring reactions by mediating a single electron through its [2Fe-2S] cluster.³⁹⁶⁻³⁹⁷ It is unclear whether the absence of this gene in *dyn* BGC could be responsible for the difference in enediyne core of *Dyn* as compared with *Tnm*, *Ypm* and *Ucm*.

The absence of those genes in *sgd* BGC was complicated with the identification of five genes (*sgdX1*, *X2*, *X3*, *X4*, *X5*) (**TABLE S 14**) that are missing in the other four homologous BGCs. Both *sgdX1* and *sgdX4* encode hypothetical proteins with unknown functions, while *sgdX5* encodes an IS630 family transposase, which is unlikely to be involved in sungeidine biosynthesis. Particularly, *sgdX2* was predicted to encode a sulfotransferase. Despite the presence of a thiolactone moiety in both sungeidine and A1, this sulfotransferase is unlikely to be associated with that moiety as a similar thiolactone fused anthracene intermediate was recently uncovered in a biosynthetic study on dynemicin,³⁹⁸ which does not possess a homologous sulfotransferase gene in its BGC. Even so, *SgdX2* could be involved in biosynthesis

of sungeidine through an uncharacterized mechanism as similar sulfotransferases had been found in Cyanobacteria to activate polyketides for further processing.³⁹⁹ Finally, *sgdX3* encodes a MBL fold metallo-hydrolase protein with a N-terminal metallo- β -lactamase fold and a C-terminal alkyl sulfatase domain, suggesting a potential hydrolytic activity that might involve thioester cleavage.

4.3.5.5. Analysis of mutants revealed that SgdX2 is crucial for formation of sungeidine.

To understand the roles of SgdX1, SgdX2 and SgdX3 in the biosynthesis of sungeidine, all three corresponding genes were individually disrupted by in-frame deletion. Metabolite profiling of Δ *sgdX1* and Δ *sgdX3* mutants revealed the presence of previously identified peaks observed from the metabolite profile of overexpression strain *Micromonospora* sp. MD118AE. Along with the absence of new metabolites, HPLC findings (**FIGURE 42**) suggested that SgdX1 and SgdX3 are not essential for formation of sungeidine and the associated analogs. However, determination of their functions in *sgd* BGC, if any, would require further mechanistic studies. Unlike SgdX1 and SgdX3, disruption of SgdX2 revealed remarkable differences in its metabolite profile. The appearance of two new peaks (A5 and A6) (**FIGURE 42**) with UV spectra that resemble A3 and A4, respectively, coupled with the noticeable disappearance of sungeidine and A2 indicated that SgdX2 is essential for the formation of sungeidine and A2. This also suggested that A2, which possesses a UV spectrum similar to sungeidine, might be a metabolite at the later step of the biosynthetic pathway that is structurally related to sungeidine. Presence of A1 in the metabolite profile of Δ *sgdX2* also implied that SgdX2 is implicated in either the formation of enediyne core or the ligation of the anthracene moiety with the enediyne core. At present, it is unclear whether A5 and A6 serve as precursors or products of A3 and A4, respectively.

Comparison of metabolite profiles of crude extracts from 5 and 7 days old cultures of *Micromonospora* sp. MD118AE also provided useful insights on the temporal dynamics governing the biosynthesis of sungeidine and its associated analogs. Higher ratios of A3 and A4 in 5 days old fermentation culture of *Micromonospora* sp. MD118AE suggested that these two peaks were likely to occur in earlier stages of the pathway as compared to A1, A2 and sungeidine. Overall, distinctive difference in peak intensity between mutants and *Micromonospora* sp. MD118AE could likely be a

consequence of both the upregulation of *sgd* BGC in the *sgdR2R7* overexpression strain and pleiotropic effects generated by variation in growth.

Figure 42: Comparison of HPLC chromatogram of mutants and identification of *sgd* BGC associated peaks. Wavelength (λ) = 314 nm is shown for comparison. Peak corresponding to sungeidine is depicted in red, while all other identified intermediates are depicted in blue.

Previously, the PKSE of *dyn* BGC was shown to be responsible for biosynthesis of both the enediyne and anthraquinone cores,²¹⁹ with an anthracene intermediate that was subsequently elucidated.³⁹⁸ However, identity of the final intermediate that is common to biosynthetic pathways of both moieties remains hidden. While mechanistic studies with different PKSE-TE pairs in *E. coli* had consistently yield a heptaene molecule,⁴⁰⁰ inability to restore the production of 9-membered enediyne natural products for native host $\Delta pksE$ mutants through complementation with *pskE* from 10-membered enediyne BGC⁴² led to speculation that free heptaene might be a shunt product, rather than an intermediate involved in the biosynthesis of enediyne. Subsequent experiments with heterologous expression of calicheamicin PKSE (CalE8) in *E. coli* identified a β -hydroxy acid metabolite which suggested that **1**, a β -hydroxy thioester tethered to CalE8 was the true intermediate in enediyne biosynthesis.⁴⁰¹ That hypothesis was bolstered by more evidence from in vitro studies, which demonstrated that CalE8 accumulated **1**.⁴⁰² The structure of A1 and its absence in the metabolite profile of $\Delta sgdU15$ mutant suggested that A1 did not originate from the free heptaene (**FIGURE 39E**), which was observed at $t = 55$ min of HPLC analyses of all crudes. With the function of SgdU15 yet to be understood, the only certainty is that both A1 and a putative enediyne core of sungeidine both originates from **1**.

Figure 43: Proposed biosynthetic mechanism of sungeidine. (A) SgdE8 synthesizes a β -hydroxy thioester **1** that remains tethered to the PKS. Subsequent reactions generate **A1** and a enediyne core intermediate. X depicts an unknown leaving group. Formation of sungeidine might be promoted by spontaneous Bergman Cyclization of the 10-membered enediyne core. (B) Structures of 10-membered enediyne core from different 10-membered enediyne natural products.

4.4. Conclusion

To conclude, isolation of two novel compounds, lumpurmycin and sungeidine, from *Micromonospora* sp. MD118 demonstrates the rich biosynthetic potential of non-streptomycete actinomycetes in the discovery of novel natural products. In the past, most efforts on natural product discovery in rare actinomycetes had focus primarily on pleiotropic effect generated by random conditions. Such laborious techniques

generate low hit rates with undue risk of re-discovery of known compounds. While the application of synthetic biology tools in natural product discovery efforts in actinomycetes had largely been confined to *Streptomyces*, this work displays the prospect of such application in other actinomycetes to facilitate targeted natural product discovery in the Post Genomics Era. Given the importance of natural product biosynthesis in understanding protein functions and chemical logic of these molecules, these synthetic biology tools also enable manipulation of actinomycetes to interrogate different aspects of a biosynthetic pathway.

Structural elucidation of lumpurmycin and sungeidine is merely the first step forward. Considering the potency of benzoquinone ansamycins as antimicrobial and antitumor agents, additional biological assays are required to assess the potential of lumpurmycin as a therapeutical candidate.

The biosynthesis of sungeidine remains unresolved. To complicate matters, a congener possessing an intact enediyne core has proved elusive thus far. Efforts persist to interrogate selected genes in the cluster that might hopefully address some of these questions, particularly the mechanism governing the biosynthesis of the enediyne core.

Chapter 5: Future work

Overall, the projects uncovered three novel compounds with their corresponding BGCs, which were validated through genetic knockouts generated via CRISPR/Cas9. The discovery of sceliphrolactam BGC draws interest onto the plausible occurrence of a partially reducing, iterative PKS module within a large modular PKS system, which is unprecedented. However, time constraint hindered the achievements of both aims this dissertation. Whilst novel compounds were discovered, a comprehensive characterization of their biological activities and biosynthetic pathways were incomplete.

This dissertation demonstrates the potential of uncovering more novel compounds from actinomycetes in the Post-Genomic Era by complementing modern synthetic biology tools and bioinformatics with traditional natural product discovery workflow. Thought to have been depleted previously, genome mining of strains uncovered by the lab revealed numerous novel BGCs that encode unknown compounds. This finding is in agreement with comparative analyses of actinomycetal genomes, which shed light on their untapped secondary metabolite potential.²⁵³ The utilization of modern synthetic biology tools offers an attractive route to activate these cryptic BGCs without laborious screening of culture conditions for such intention. Given the unquestionable contribution of natural products in drug discovery, revival of natural product discovery from actinomycete will undoubtedly boost the drug pipeline for development of next-generation therapeutics.

As depicted by the number of BGCs present in the three Actinomycete strains studied, the potential for discovery of more compounds surpasses the outcomes of the current work. The establishment of a robust genetic engineering system in actinomycetes, coupled with recent upgrades in facilities for chemical isolation, means that *in silico* genome mining data can finally be translated into wet-lab outcomes. The lack of a chemical dereplication workflow can be compensated by targeting specific BGCs using the available synthetic biology tools. However, this approach is still subjected to the intactness of selected BGCs and genetic tractability of the microbe. With the accumulation of >30 genome sequences from an actinomycete collection exceeding 100 unique strains, identifying suitable BGCs is not a bottleneck in efforts to discover novel natural products. Instead, based on experimental observations, the growth

characteristics of wild-type actinomycete isolates often differ greatly from each other and genetic manipulation of each strain is subjected to detailed optimization of an otherwise, standardized experimental workflow. This aggravates an already compelling need to increase the throughput-ness of the experimental workflow in order to match the potential presented by *in silico* findings.

While the unpredictability of BGCs means that their intactness cannot be explicitly determined, the problem caused by genetic intractability can be resolved by development of tools for heterologous expression of BGCs. Notwithstanding the absence of essential but obscure elements that might be present in a native actinomycete strain, the increased finesse of an optimized heterologous expression host provides a feasible option for studying BGCs of interest. Ease of manipulation, both genetically and physically, offered by an optimized host also facilitates refactoring of BGCs for activation or gene functionality studies. As the lab is developing two actinomycete isolates as heterologous expression hosts for such purposes, the next objective will be to develop molecular cloning tools for assembly or direct capture of BGCs into suitable vectors.

Development of tools for heterologous expression of BGCs

Construction of genomic DNA library of selected strains is one strategy taken for conveying multiple BGCs into heterologous host for further studies. Generally, the size of BGCs can range widely from 10 kb to upwards of 200 kb. However, majority of BGCs containing modular thioester-active enzymes (PKS and NRPS) typically exceeds 50 kb in size, making shuttle bacterial artificial chromosome (BAC) vectors a more viable option for cloning and stable maintenance of BGCs. Construction of genomic DNA library of actinomycetes using a shuttle *Streptomyces-E. coli* BAC vector has already been demonstrated in the biosynthetic studies of notable compounds such as daptomycin and geldanamycin. Furthermore, such vectors can also be utilized to assemble BGCs from digested genomic DNA fragments via Linear plus Linear Homologous Recombination (LLHR) mediated cloning in a compatible *E. coli* strain that possesses the necessary recombinases RecE and RecT.²⁵² A suitable *Streptomyces-E.coli* BAC vector pBAC (pBACS3) has already been developed but remains utilized (²⁵² A suitable *Streptomyces-E.coli* BAC vector pBAC (pBACS3) has already been developed but remains unused (**Figure 44**).

Figure 44: Construction of pBACS3, a *Streptomyces*-*E. coli* shuttle vector for genomic DNA library construction. pBACe3.6 was used as a backbone BAC vector for modification. Fragment B contains OriT for conjugal transfer of the assembled plasmid from *E. coli* to *Streptomyces*, *aac(3)IV* gene which confers resistance to apramycin and phiC31 elements for recognition of attB site and subsequent integration into the chromosome of *Streptomyces*. Digestion with *Bam*HI will linearize the vector for subsequent ligation with similarly digested genomic DNA.

TAR cloning is another feasible strategy for direct capture or assembly of selected BGCs. This method harnesses the highly efficient homology directed recombination repair mechanism in *Saccharomyces cerevisiae* to capture or assemble linear DNA fragments containing overlapping ends to each other. Despite the low GC content of *S. cerevisiae* genomic DNA, TAR cloning has been successfully applied to high GC genomic DNA,²⁴⁴ proving that the homologous recombination repair system in yeast was not biased against high GC DNA. In recent years, TAR cloning has been developed as a tool to clone BGCs for biosynthetic studies via heterologous expression,^{191, 243, 245} thus offering an alternative platform for cloning of selected BGCs without the need for genomic DNA library construction. Apart from direct capture of DNA fragment from linearized DNA fragments, BGCs can be assembled with PCR amplified DNA fragments via a similar method termed as DNA assembler.²⁴⁶ With the development of CRISPR/Cas9 as a versatile genetic editing tool, the prospect of manipulating multiple sites of a gene cluster simultaneously via multiplex-CRISPR/Cas9 in *S. cerevisiae* further augments the application of TAR as a powerful precursor tool in by streamlining the process of refactoring targeted BGCs.³⁹¹ A *S. cerevisiae* – *E. coli* – *Streptomyces* shuttle would be necessary to enable TAR cloning in *S. cerevisiae*, stable maintenance in *E. coli* and heterologous expression in *Streptomyces* host. Using a similar vector backbone as the one used to construct

pBACS3, essential Yeast and *Streptomyces* elements were inserted accordingly to yield a resultant tripartite shuttle vector called pSTAR3 (**FIGURE 45**).

Figure 45: Construction of pSTAR3, a tripartite shuttle vector for TAR cloning and heterologous expression in *Streptomyces*. (A) Backbone vector pBACe3.6 is assembled with the necessary elements in a three-step cloning procedure to create the resultant vector pSTAR3. (B) A plasmid map of the resultant pSTAR3 vector.

Equipped with these vectors, the next step will involve elaborate optimization of experimental workflow to establish a throughput cloning platform for heterologous expression of multiple BGCs in concurrently developed *Streptomyces* hosts.

Determination of iterative PKS module in the biosynthesis of sceliphrolactam

As elaborated in chapter 2, one of the findings of this work is the identification of a BGC responsible for biosynthesis of sceliphrolactam. Despite evidence showing the lack of involvement of a trans PKS module, the identity of an iterative PKS module

within the cluster remains elusive. Two PKS modules, encoded by *sceQ* and *sceR* separately, were highlighted previously as candidates that might act iteratively to maintain the processivity of the entire PKS assembly line (**FIGURE 18**). One potential procedure to pinpoint the iterative functionality of *SceQ* or *SceR* involves interrogating both modules separately by altering the substrate specificity of their acyltransferase (AT) domains. The native AT domains of both modules contain a conserved HAFH motif (**FIGURE 46A**), which is specific for selecting malonyl-CoA as substrates (**FIGURE 46B**) for chain extension of the intermediate molecule.

Figure 46: Examination of substrate specificity of AT domains in *sce* BGC through sequence alignment. (A) Methylmalonyl-CoA specific AT contains YASH motif (red) while malonyl-CoA specific AT contains HAFH motif (blue). (B) Structures of methylmalonyl-CoA and methyl-CoA.

Alteration of substrate specificity of AT can be achieved either by swapping the entire AT domain for a previously identified methylmalonyl-CoA specific AT domain or via site-directed mutagenesis of the key HAFH motif to a YASH motif. Both methods are not without limitations. Entire AT domain swap usually disturbs the delicate interactions between KS and ACP domains of the same module, resulting in a drastic decrease in product titre.⁴⁰³ While site-directed mutagenesis approach is less disruptive to the entire PKS module architecture, it often results in promiscuity of the AT domain instead of definitive alteration of substrate specificity.⁴⁰⁴ Any attempt to alter the substrate specificity of AT will also require attention to the first half of post-AT linker region, which has been shown to correlate with substrate specificity of AT domain.⁴⁰⁵ Alteration of substrate specificity of AT can be achieved either by swapping the entire AT domain for a previously identified methylmalonyl-CoA specific AT domain or via site-directed mutagenesis of the key HAFH motif to a YASH motif.

Apart from maintaining the functionality of the mutated enzyme, issues pertaining to the interactions between the altered polyketide chain intermediate and downstream PKS enzymes must also be taken into consideration. Being the gatekeepers for

selecting nascent polyketide intermediates,⁴⁰⁶ KS domains of downstream PKS modules might dismiss the altered polyketide chain intermediate, leading to concomitant abrogation of compound production. Notably, apart from a single exception,⁴⁰⁷ most successful experiments involving AT domain engineering resulted from altering specificity to favour a smaller substrate (i.e. methylmalonyl-CoA to malonyl-CoA).^{287, 338, 403, 408-409} This is unsurprising as the reverse approach generates a bulkier polyketide chain intermediate, which presumably introduces steric hindrance that might obstruct its access to the binding pockets of downstream enzymes. Consequently, the risk of failure from such engineering attempts is higher.

Figure 47: Possible scenarios resulting from altered substrate specificity of SceQ-AT and SceR-AT domains. (A) If SceQ is iterative, two rounds of chain extension by SceQ will result in a di-methylated sceliphrolactam analog. (B) If SceQ is non-iterative, only a single round of chain extension will take place, eventually generating a methylated sceliphrolactam analog. (C) If SceR is iterative, two rounds of chain extension by SceR will result in a di-methylated sceliphrolactam analog. (D) If SceR is non-iterative, only a single round of chain extension will take place, eventually generating a methylated sceliphrolactam analog.

Nonetheless, if successful, alteration of SceQ-AT and SceR-AT substrate specificity will result in the production of various sceliphrolactam analogues. Even with drastic decreases in titre yield, the identity of these sceliphrolactam analogues will render substantial evidence into the role of SceQ and SceR in the biosynthesis of sceliphrolactam (FIGURE 47).

Activation of cryptic BGCs in *Streptomyces* sp. SD50

Excluding meijiemyacin, several other BGCs in *Streptomyces* sp. SD50 were also prioritized previously for isolation and determination of their encoded secondary metabolites.

One such cluster is C40 BGC (**FIGURE 48**) whereby detailed analysis revealed that it encodes a polyether natural product. It contains a set of genes (*EI*, *EII*, *EIII*) responsible for biosynthesis of ethylmalonyl-CoA, a substrate utilized by the PKS enzymes within the cluster for biosynthesis of the encoded polyketide skeleton. Despite not containing genes that might render unexpected features to the final polyether scaffold, the discovery of a novel polyether compound holds potential therapeutic significance. Before a widespread ban of antibiotic usage,⁴¹⁰ polyether compounds like salinomycin and lasalocid have always been associated as ionophore antibiotics in livestock and poultry.⁴¹¹⁻⁴¹² However, in recent times, new therapeutic purposes have been found for these polyether ionophores, prompting a revival of interest in this class of compounds.³²² One noteworthy example is salinomycin, which is currently under clinical trial as a anti-cancer therapeutic agent.⁴¹³ Also, this class of compounds have demonstrated toxicity against Plasmodium, indicating their potential as antimalarial drug.⁴¹⁴ In the latest discovery, nanchangmycin, a polyether compound, was found to be active against insect-specific flaviviruses such as Zika, West Nile, dengue and chikungunya.⁴¹⁵

After previous attempts to activate C40 BGC through variation of media and fermentation conditions ended in failure, a genetic manipulation approach was adopted. Detailed analysis of the BGC revealed a single regulatory protein gene, *c40R*, which encodes a LuxR family transcriptional regulator protein. LuxR regulators are functionally diverse and widely distributed among the genomes of Actinobacteria,³⁶³ with a number of them acting as positive regulators for PKS associated BGCs.^{193, 197, 416} Furthermore, SlnR, a LuxR family transcriptional regulator, was found to positively regulate salinomycin production.⁴¹⁷ LuxR regulators are functionally diverse and widely distributed among the genomes of Actinobacteria,³⁶³ with a number of them acting as positive regulators for PKS associated BGCs.^{193, 197, 416} Furthermore, SlnR, a LuxR family transcriptional regulator, was found to positively regulate salinomycin production.⁴¹⁷

Drawing inference from the aforementioned sources, C40R was hypothesized to be a positive regulator of C40 BGC hence, constitutive expression of c40R using a previously constructed integrative plasmid was performed accordingly. Unexpectedly, based on HPLC analysis of crudes, overexpression of C40R did not elicit production of any new compound. Currently, transcriptional analysis of the genes via RT-PCR is still necessary to ascertain the regulatory effect influenced by C40R on C40 BGC.

If C40R does not regulate C40 BGC positively, a promoter replacement strategy can still be adopted to activate the cluster. As all genes are clustered into two operons of opposing directions (*EII* to *OI*, *A* to *EIII*), replacing the intergenic region between gene *c40OI* and *c40A* with a bidirectional, constitutive promoter cassette will likely activate the transcription of all necessary genes in the operon.

Figure 48: Organization of putative C40 BGC. The cluster encodes genes that are likely responsible for biosynthesis of a polyether natural product.

C45 BGC is another BGC that is worth pursuing in future. Analysis of the cluster revealed homology to other spiro-tetronate polyketide BGCs. Spiro tetronate polyketide clusters possess a FkbH protein gene (**FIGURE 49**) which is responsible for biosynthesis of glyceryl-S-ACP, a substrate that reacts with the PKS intermediate through a two-step concerted Claisen condensation and lactonization to generate a tetronate moiety.⁴¹⁸ A subsequent intramolecular Diels-Alder reaction between a diene and a dienophile, which is contributed by the tetronate moiety, results in the formation of a spiro-tetronate. This class of polyketides demonstrates potent activity in various biological assays and serve as potential drug leads for cancer and metabolic disorders.³²¹

Figure 49: Organization of putative C45 BGC. The cluster encodes genes that are likely responsible for biosynthesis of a spirotetronate polyketide natural product. The *Streptomyces* antibiotic regulatory protein (SARP) gene of the BGC is denoted as *c45R*.

Given that the entire BGC is divided into multiple operons, refactoring of the entire BGC will require multiple promoter replacements in a step-wise manner. The time-consuming nature of this approach makes it unrealistic in practice. Another potential strategy involves the use of a method known as DNA assembler to assemble the entire BGC with PCR-amplified fragments with ends that are homologous to each other. This method will enable the implementation of constitutive promoters into specific sites during the PCR amplification procedure. Even though the cluster is ~78 kb in size, this approach is still feasible as TAR-based DNA assembly has been successfully used to even construct the entire genome (~582 kb) of *Mycoplasma genitalium* from PCR amplified DNA fragments.^{242, 419}

Analysis of other spirotetronate polyketide BGCs revealed a cluster specific regulatory protein that is homologous in all spirotetronate polyketide BGCs. This regulatory protein is a *Streptomyces* antibiotic regulatory protein (SARP) family protein that has been identified in many clusters as a positive regulator. In particular, a homologous SARP family protein (Abml), has been found to boost the production of abyssomicin, a spirotetronate polyketide, by 7.6 fold when overexpressed constitutively.¹⁹⁶ Similarly, constitutive overexpression of *c45R* can be considered to activate cluster 45.

Characterization of *sgd* BGC and mechanistic studies on enediyne biosynthesis

As described in chapter 4, two new compounds associated with the enediyne BGC - sungeidine and an associated intermediate A1, have been elucidated thus far. Sungeidine contains a unique tri-substituted nitrogen core that is unseen in all 10-membered enediyne natural products identified from literature. It also lacks an intact 10-membered enediyne moiety that is supposedly a feature of all 10-membered enediyne BGCs studied till date. Despite chemical screening of fermentation crudes

extracted under meticulous conditions, a congener of sungeidine with an intact enediyne core remains elusive. With so many proteins whose functions have yet to be characterized, many questions pertaining to the biosynthesis of sungeidine are still unanswerable. In summary, *in vivo* knock out mutants of *sgdX1*, *sgdX2* and *sgdX3*, three genes which are not homologous to other known 10-membered enediyne BGCs revealed that biosynthesis of sungeidine involves a sulfotransferase, SgdX2. However, the exact function of SgdX2 remains uncharacterized.

While comprehensive characterization of all genes within *sgd* BGC is practically impossible, elucidation of other associated metabolites (A2 to A6) could provide clues on the biosynthetic pathway. Additionally, efforts are also ongoing to knock out four more genes (*sgd15*, *sgd16*, *sgdU16*, *sgdE14*) *in vivo* via CRISPR/Cas9 genetic tool, which, hopefully, could supplement the structural data of associated metabolites and shed light on sungeidine biosynthesis. Nonetheless, it is sobering to note that more than a decade has passed since the discovery of the first enediyne BGC⁴²⁰ and substantial evidence is still lacking to support any proposed biosynthetic mechanisms. Development of CRISPR/Cas9 for genetic engineering has certainly facilitated *in vivo* studies of enediyne natural products, which was a challenge before the inception of a robust genetic manipulation tool. Hopefully, the ease of genetic manipulation of *Micromonospora* sp. MD118 using CRISPR/Cas9 could facilitate generation of gene knock outs that eventually provide the missing pieces for the puzzle behind the biosynthesis of enediyne natural products.

References

- Schulz, H., Beta oxidation of fatty acids. *Biochim. Biophys. Acta* **1991**, 1081 (2), 109-120.
- Bodner, G. M., Metabolism Part II: The tricarboxylic acid (TCA), citric acid, or Krebs cycle. *J. Chem. Educ.* **1986**, 63 (8), 673.
- Wyatt, M. A.; Wang, W.; Roux, C. M.; Beasley, F. C.; Heinrichs, D. E.; Dunman, P. M.; Magarvey, N. A., Staphylococcus aureus nonribosomal peptide secondary metabolites regulate virulence. *Science* **2010**, 329 (5989), 294-6.
- Waters, C. M.; Bassler, B. L., Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **2005**, 21, 319-46.
- Williams, P., Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiology* **2007**, 153, 3923-3938.
- BENNETT, R. N.; WALLSGROVE, R. M., Secondary metabolites in plant defence mechanisms. *New Phytol.* **1994**, 127 (4), 617-633.
- Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J., The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* **2015**, 14 (2), 111-129.
- Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A., Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies, and Future Directions. *J. Med. Chem.* **1994**, 37 (10), 1385-1401.
- Cragg, G. M.; Grothaus, P. G.; Newman, D. J., New Horizons for Old Drugs and Drug Leads. *J. Nat. Prod.* **2014**, 77 (3), 703-723.
- Patridge, E.; Gareiss, P.; Kinch, M. S.; Hoyer, D., An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov. Today* **2015**.
- Gu, J.; Gui, Y.; Chen, L.; Yuan, G.; Lu, H.-Z.; Xu, X., Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology. *PLoS One* **2013**, 8 (4), e62839.
- Kato, N.; Takahashi, S.; Nogawa, T.; Saito, T.; Osada, H., Construction of a microbial natural product library for chemical biology studies. *Curr. Opin. Chem. Biol.* **2012**, 16 (1-2), 101-8.
- Harvey, A. L., Natural products in drug discovery. *Drug Discov. Today* **2008**, 13 (19-20), 894-901.
- Pham, V. H.; Kim, J., Cultivation of unculturable soil bacteria. *Trends Biotechnol.* **2012**, 30 (9), 475-84.
- Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schaberle, T. F.; Hughes, D. E.; Epstein, S.; Jones, M.; Lazarides, L.; Steadman, V. A.; Cohen, D. R.; Felix, C. R.; Fetterman, K. A.; Millett, W. P.; Nitti, A. G.; Zullo, A. M.; Chen, C.; Lewis, K., A new antibiotic kills pathogens without detectable resistance. *Nature* **2015**, 517 (7535), 455-9.
- Berdy, J., Bioactive microbial metabolites. *J. Antibiot. (Tokyo)* **2005**, 58 (1), 1-26.
- Berdy, J., Thoughts and facts about antibiotics: where we are now and where we are heading. *J. Antibiot. (Tokyo)* **2012**, 65 (8), 385-95.
- Baltz, R. H., Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall? *SIM News* **2005**, 55, 186-196.
- Ziemert, N.; Lechner, A.; Wietz, M.; Millán-Aguiñaga, N.; Chavarria, K. L.; Jensen, P. R., Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proceedings of the National Academy of Sciences* **2014**, 111 (12), E1130-E1139.
- Wang, D. S.; Xue, Q. H.; Zhu, W. J.; Zhao, J.; Duan, J. L.; Shen, G. H., Microwave irradiation is a useful tool for improving isolation of actinomycetes from soil. *Microbiology* **2013**, 82 (1), 102-110.
- Tan, G. Y.; Ward, A. C.; Goodfellow, M., Exploration of Amycolatopsis diversity in soil using genus-specific primers and novel selective media. *Syst. Appl. Microbiol.* **2006**, 29 (7), 557-69.
- Wang, D. S.; Xue, Q. H.; Ma, Y. Y.; Wei, X. L.; Chen, J.; He, F., Oligotrophy is Helpful for the Isolation of Bioactive Actinomycetes. *Indian J. Microbiol.* **2014**, 54 (2), 178-84.

23. Hodges, T. W.; Slattery, M.; Olson, J. B., Unique actinomycetes from marine caves and coral reef sediments provide novel PKS and NRPS biosynthetic gene clusters. *Mar. Biotechnol. (N. Y.)* **2012**, *14* (3), 270-80.
24. Lee, S. D., Amycolatopsis jejuensis sp. nov. and Amycolatopsis halotolerans sp. nov., novel actinomycetes isolated from a natural cave. *Int. J. Syst. Evol. Microbiol.* **2006**, *56* (Pt 3), 549-53.
25. Gavrish, E.; Bollmann, A.; Epstein, S.; Lewis, K., A trap for in situ cultivation of filamentous actinobacteria. *J. Microbiol. Methods* **2008**, *72* (3), 257-62.
26. Awakawa, T.; Crusemann, M.; Munguia, J.; Ziemert, N.; Nizet, V.; Fenical, W.; Moore, B. S., Salinipyrone and Pacificanone Are Biosynthetic By-products of the Rosamicin Polyketide Synthase. *ChemBioChem* **2015**, *16* (10), 1443-7.
27. Shen, B., Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285-295.
28. Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D., The Phosphopantetheinyl Transferases: Catalysis of a Posttranslational Modification Crucial for Life. *Nat. Prod. Rep.* **2014**, *31* (1), 61-108.
29. He, J.; Hertweck, C., Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. *Chem. Biol.* **2003**, *10* (12), 1225-32.
30. Hong, H.; Sun, Y.; Zhou, Y.; Stephens, E.; Samborsky, M.; Leadlay, P. F., Evidence for an iterative module in chain elongation on the azalomycin polyketide synthase. *Beilstein J. Org. Chem.* **2016**, *12*, 2164-2172.
31. Wilkinson, B.; Foster, G.; Rudd, B. A.; Taylor, N. L.; Blackaby, A. P.; Sidebottom, P. J.; Cooper, D. J.; Dawson, M. J.; Buss, A. D.; Gaisser, S.; Bohm, I. U.; Rowe, C. J.; Cortes, J.; Leadlay, P. F.; Staunton, J., Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. *Chem. Biol.* **2000**, *7* (2), 111-7.
32. Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J., The Structure of Docking Domains in Modular Polyketide Synthases. *Chem. Biol.* **2003**, *10* (8), 723-731.
33. Cox, R. J.; Simpson, T. J., Fungal type I polyketide synthases. *Methods Enzymol.* **2009**, *459*, 49-78.
34. Zhang, Q.; Pang, B.; Ding, W.; Liu, W., Aromatic Polyketides Produced by Bacterial Iterative Type I Polyketide Synthases. *ACS Catalysis* **2013**, *3* (7), 1439-1447.
35. Sun, H. H.; Ho, C. L.; Ding, F. Q.; Soehano, I.; Liu, X. W.; Liang, Z. X., Synthesis of (R)-Mellein by a Partially Reducing Iterative Polyketide Synthase. *J. Am. Chem. Soc.* **2012**, *134* (29), 11924-11927.
36. Soehano, I.; Yang, L.; Ding, F.; Sun, H.; Low, Z. J.; Liu, X.; Liang, Z.-X., Insights into the programmed ketoreduction of partially reducing polyketide synthases: stereo- and substrate-specificity of the ketoreductase domain. *Org. Biomol. Chem.* **2014**, *12* (42), 8542-8549.
37. Helfrich, E. J. N.; Piel, J., Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **2016**, *33* (2), 231-316.
38. Wong, F. T.; Jin, X.; Mathews, I. I.; Cane, D. E.; Khosla, C., Structure and Mechanism of the trans-Acting Acyltransferase from the Disorazole Synthase (). *Biochemistry* **2011**, *50* (30), 6539-6548.
39. Jensen, K.; Niederkruger, H.; Zimmermann, K.; Vagstad, A. L.; Moldenhauer, J.; Brendel, N.; Frank, S.; Poplau, P.; Kohlhaas, C.; Townsend, C. A.; Oldiges, M.; Hertweck, C.; Piel, J., Polyketide proofreading by an acyltransferase-like enzyme. *Chem. Biol.* **2012**, *19* (3), 329-39.
40. Chen, X.; Ji, R.; Jiang, X.; Yang, R.; Liu, F.; Xin, Y., Iterative type I polyketide synthases involved in enediynes natural product biosynthesis. *IUBMB life* **2014**, *66* (9), 587-95.
41. Horsman, G. P.; Chen, Y.; Thorson, J. S.; Shen, B., Polyketide synthase chemistry does not direct biosynthetic divergence between 9- and 10-membered enediynes. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (25), 11331-11335.
42. Zhang, J.; Van Lanen, S. G.; Ju, J.; Liu, W.; Dorrestein, P. C.; Li, W.; Kelleher, N. L.; Shen, B., A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. *Proceedings of the National Academy of Sciences* **2008**, *105* (5), 1460-1465.

43. Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A., Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **2007**, *24* (1), 162-190.
44. Castaldo, G.; Zucko, J.; Heidelberger, S.; Vujaklija, D.; Hranueli, D.; Cullum, J.; Wattana-Amorn, P.; Crump, M. P.; Crosby, J.; Long, P. F., Proposed Arrangement of Proteins Forming a Bacterial Type II Polyketide Synthase. *Chem. Biol.* **2008**, *15* (11), 1156-1165.
45. Dreier, J.; Khosla, C., Mechanistic Analysis of a Type II Polyketide Synthase. Role of Conserved Residues in the β -Ketoacyl Synthase–Chain Length Factor Heterodimer. *Biochemistry* **2000**, *39* (8), 2088-2095.
46. Szu, P.-H.; Govindarajan, S.; Meehan, M. J.; Das, A.; Nguyen, D. D.; Dorrestein, P. C.; Minshull, J.; Khosla, C., Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. *Chem. Biol.* **2011**, *18* (8), 1021-1031.
47. Morita, H.; Abe, I.; Noguchi, H., 1.06 - Plant Type III PKS. In *Comprehensive Natural Products II*, Elsevier: Oxford, 2010; pp 171-225.
48. Saxena, P.; Yadav, G.; Mohanty, D.; Gokhale, R. S., A New Family of Type III Polyketide Synthases in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **2003**, *278* (45), 44780-44790.
49. Song, L.; Barona-Gomez, F.; Corre, C.; Xiang, L.; Udvary, D. W.; Austin, M. B.; Noel, J. P.; Moore, B. S.; Challis, G. L., Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by *Streptomyces coelicolor* genome mining. *J. Am. Chem. Soc.* **2006**, *128* (46), 14754-5.
50. Huang, H.; Hou, L.; Li, H.; Qiu, Y.; Ju, J.; Li, W., Activation of a plasmid-situated type III PKS gene cluster by deletion of a *wbl* gene in deepsea-derived *Streptomyces somaliensis* SCSIO ZH66. *Microb Cell Fact* **2016**, *15* (1), 116.
51. Weissman, K. J., Chapter 1 Introduction to Polyketide Biosynthesis. In *Methods Enzymol.*, Academic Press: 2009; Vol. 459, pp 3-16.
52. Ikeda, H.; Omura, S., Avermectin Biosynthesis. *Chem. Rev.* **1997**, *97* (7), 2591-2610.
53. Caffrey, P.; Lynch, S.; Flood, E.; Finnan, S.; Oliynyk, M., Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. *Chem. Biol.* **2001**, *8* (7), 713-23.
54. Sieber, S. A.; Marahiel, M. A., Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **2005**, *105* (2), 715-38.
55. Doroghazi, J. R.; Albright, J. C.; Goering, A. W.; Ju, K.-S.; Haines, R. R.; Tchalukov, K. A.; Labeda, D. P.; Kelleher, N. L.; Metcalf, W. W., A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, *10* (11), 963-968.
56. Du, L.; Shen, B., Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer *Streptomyces verticillus* ATCC 15003. *Chem. Biol.* **1999**, *6* (8), 507-517.
57. Wyckoff, E. E.; Smith, S. L.; Payne, S. M., VibD and VibH Are Required for Late Steps in Vibriobactin Biosynthesis in *Vibrio cholerae*. *J. Bacteriol.* **2001**, *183* (5), 1830-1834.
58. Ehmann, D. E.; Shaw-Reid, C. A.; Losey, H. C.; Walsh, C. T., The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proceedings of the National Academy of Sciences* **2000**, *97* (6), 2509-2514.
59. Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K., Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceedings of the National Academy of Sciences* **2014**, *111* (25), 9259-9264.
60. Fernandes, P.; Martens, E., Antibiotics in late clinical development. *Biochem. Pharmacol.* **2017**, *133*, 152-163.
61. Alder, J. D., Daptomycin: a new drug class for the treatment of Gram-positive infections. *Drugs of today (Barcelona, Spain : 1998)* **2005**, *41* (2), 81-90.
62. Rubinstein, E.; Keynan, Y., Vancomycin Revisited – 60 Years Later. *Frontiers in Public Health* **2014**, *2* (217).

63. Christianson, D. W., Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **2017**, *117* (17), 11570-11648.
64. Wendt, K. U.; Schulz, G. E., Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure* **1998**, *6* (2), 127-33.
65. Aaron, J. A.; Christianson, D. W., Trinuclear Metal Clusters in Catalysis by Terpenoid Synthases. *Pure and applied chemistry. Chimie pure et appliquee* **2010**, *82* (8), 1585-1597.
66. Pemberton, T. A.; Christianson, D. W., General Base-General Acid Catalysis by Terpenoid Cyclases(). *The Journal of antibiotics* **2016**, *69* (7), 486-493.
67. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H., Terpene synthases are widely distributed in bacteria. *Proceedings of the National Academy of Sciences* **2015**, *112* (3), 857-862.
68. Citron, C. A.; Gleitzmann, J.; Laurenzano, G.; Pukall, R.; Dickschat, J. S., Terpenoids are widespread in actinomycetes: a correlation of secondary metabolism and genome data. *ChemBioChem* **2012**, *13* (2), 202-14.
69. Dai, T.; Motohira, Y.; Itoh, N.; Kuzuyama, T.; Takahashi, S.; Seto, H., Cloning of the gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase from terpenoid antibiotic-producing *Streptomyces* strains. *Molecular and General Genetics MGG* **2000**, *262* (6), 957-964.
70. Kawasaki, T.; Hayashi, Y.; Kuzuyama, T.; Furihata, K.; Itoh, N.; Seto, H.; Dai, T., Biosynthesis of a Natural Polyketide-Isoprenoid Hybrid Compound, Furaquinocin A: Identification and Heterologous Expression of the Gene Cluster. *J. Bacteriol.* **2006**, *188* (4), 1236-1244.
71. Seto, H.; Orihara, N.; Furihata, K., Studies on the biosynthesis of terpenoids produced by actinomycetes. Part 4. Formation of BE-40644 by the mevalonate and nonmevalonate pathways. *Tetrahedron Lett.* **1998**, *39* (51), 9497-9500.
72. Smanski, M. J.; Yu, Z.; Casper, J.; Lin, S.; Peterson, R. M.; Chen, Y.; Wendt-Pienkowski, E.; Rajski, S. R.; Shen, B., Dedicated ent-kaurene and ent-atiseurene synthases for platensimycin and platencin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (33), 13498-503.
73. Richter, T. K. S.; Hughes, C. C.; Moore, B. S., Sioaxanthin, a novel glycosylated carotenoid reveals an unusual subclustered biosynthetic pathway. *Environ. Microbiol.* **2015**, *17* (6), 2158-2171.
74. Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K. D.; Fischbach, M. A.; Garavelli, J. S.; Goransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Muller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J.; Rebuffat, S.; Ross, R. P.; Sahl, H. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Sussmuth, R. D.; Tagg, J. R.; Tang, G. L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A., Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **2013**, *30* (1), 108-60.
75. Sivonen, K.; Leikoski, N.; Fewer, D. P.; Jokela, J., Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria. *Appl. Microbiol. Biotechnol.* **2010**, *86* (5), 1213-1225.
76. Oman, T. J.; van der Donk, W. A., Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* **2010**, *6* (1), 9-18.
77. Ovchinnikov, K. V.; Chi, H.; Mehmeti, I.; Holo, H.; Nes, I. F.; Diep, D. B., Novel Group of Leaderless Multi-peptide Bacteriocins from Gram-Positive Bacteria. *Appl. Environ. Microbiol.* **2016**, *82* (17), 5216-5224.
78. Rustem, K.; P., K. O., When the Leader Gets Loose: In Vivo Biosynthesis of a Leaderless Prenisin Is Stimulated by a trans - Acting Leader Peptide. *ChemBioChem* **2012**, *13* (16), 2433-2438.
79. Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A., Lasso Peptides: An Intriguing Class of Bacterial Natural Products. *Acc. Chem. Res.* **2015**, *48* (7), 1909-19.

80. Um, S.; Kim, Y.-J.; Kwon, H.; Wen, H.; Kim, S.-H.; Kwon, H. C.; Park, S.; Shin, J.; Oh, D.-C., Sungsanpin, a Lasso Peptide from a Deep-Sea Streptomyces. *J. Nat. Prod.* **2013**, *76* (5), 873-879.
81. Wang, H.; van der Donk, W. A., Biosynthesis of the Class III Lantipeptide Catenulipectin. *ACS Chem. Biol.* **2012**, *7* (9), 1529-1535.
82. Cheigh, C. I.; Pyun, Y. R., Nisin biosynthesis and its properties. *Biotechnol. Lett* **2005**, *27* (21), 1641-8.
83. Claesen, J.; Bibb, M. J., Biosynthesis and Regulation of Grisemycin, a New Member of the Linaridin Family of Ribosomally Synthesized Peptides Produced by *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **2011**, *193* (10), 2510-2516.
84. Hou, Y.; Tianero, M. D. B.; Kwan, J. C.; Wyche, T. P.; Michel, C. R.; Ellis, G. A.; Vazquez-Rivera, E.; Braun, D. R.; Rose, W. E.; Schmidt, E. W.; Bugni, T. S., Structure and Biosynthesis of the Antibiotic Bottromycin. *Org. Lett.* **2012**, *14* (19), 5050-5053.
85. Onaka, H.; Nakaho, M.; Hayashi, K.; Igarashi, Y.; Furumai, T., Cloning and characterization of the goadsporin biosynthetic gene cluster from *Streptomyces* sp. TP-A0584. *Microbiology* **2005**, *151* (Pt 12), 3923-33.
86. Hudson, G. A.; Zhang, Z.; Tietz, J. I.; Mitchell, D. A.; van der Donk, W. A., In Vitro Biosynthesis of the Core Scaffold of the Thiopeptide Thiomuracin. *J. Am. Chem. Soc.* **2015**, *137* (51), 16012-16015.
87. Kelly, W. L.; Pan, L.; Li, C., Thiostrepton biosynthesis: prototype for a new family of bacteriocins. *J. Am. Chem. Soc.* **2009**, *131* (12), 4327-34.
88. Maksimov, M. O.; Pan, S. J.; James Link, A., Lasso peptides: structure, function, biosynthesis, and engineering. *Nat. Prod. Rep.* **2012**, *29* (9), 996-1006.
89. Arndt, H. D.; Schoof, S.; Lu, J. Y., Thiopeptide antibiotic biosynthesis. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (37), 6770-3.
90. Heather, J. M.; Chain, B., The sequence of sequencers: The history of sequencing DNA. *Genomics* **2016**, *107* (1), 1-8.
91. Nawy, T., Single-cell sequencing. *Nat Meth* **2014**, *11* (1), 18-18.
92. Bachmann, B. O.; Van Lanen, S. G.; Baltz, R. H., Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 175-184.
93. Hindra; Huang, T.; Yang, D.; Rudolf, J. D.; Xie, P.; Xie, G.; Teng, Q.; Lohman, J. R.; Zhu, X.; Huang, Y.; Zhao, L.-X.; Jiang, Y.; Duan, Y.; Shen, B., Strain Prioritization for Natural Product Discovery by a High-Throughput Real-Time PCR Method. *J. Nat. Prod.* **2014**, *77* (10), 2296-2303.
94. Baltz, R. H., Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol.* **2008**, *8* (5), 557-63.
95. Ju, K. S.; Gao, J.; Doroghazi, J. R.; Wang, K. A.; Thibodeaux, C. J.; Li, S.; Metzger, E.; Fudala, J.; Su, J.; Zhang, J. K.; Lee, J.; Cioni, J. P.; Evans, B. S.; Hirota, R.; Labeda, D. P.; van der Donk, W. A.; Metcalf, W. W., Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc. Natl. Acad. Sci. U. S. A.* **2015**.
96. Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G., Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (22), 5601-5606.
97. Nielsen, K. F.; Månsson, M.; Rank, C.; Frisvad, J. C.; Larsen, T. O., Dereplication of Microbial Natural Products by LC-DAD-TOFMS. *J. Nat. Prod.* **2011**, *74* (11), 2338-2348.
98. Tawfike, A. F.; Viegelmann, C.; Edrada-Ebel, R., Metabolomics and dereplication strategies in natural products. *Methods Mol. Biol.* **2013**, *1055*, 227-44.
99. Wong, Weng R.; Oliver, Allen G.; Linington, Roger G., Development of Antibiotic Activity Profile Screening for the Classification and Discovery of Natural Product Antibiotics. *Chem. Biol.* **2012**, *19* (11), 1483-1495.
100. Singh, S. B.; Jayasuriya, H.; Ondeyka, J. G.; Herath, K. B.; Zhang, C.; Zink, D. L.; Tsou, N. N.; Ball, R. G.; Basilio, A.; Genilloud, O.; Diez, M. T.; Vicente, F.; Pelaez, F.; Young, K.; Wang, J., Isolation,

- Structure, and Absolute Stereochemistry of Platensimycin, A Broad Spectrum Antibiotic Discovered Using an Antisense Differential Sensitivity Strategy. *J. Am. Chem. Soc.* **2006**, *128* (36), 11916-11920.
101. Wang, J.; Kodali, S.; Lee, S. H.; Galgoci, A.; Painter, R.; Dorso, K.; Racine, F.; Motyl, M.; Hernandez, L.; Tinney, E.; Colletti, S. L.; Herath, K.; Cummings, R.; Salazar, O.; González, I.; Basilio, A.; Vicente, F.; Genilloud, O.; Pelaez, F.; Jayasuriya, H.; Young, K.; Cully, D. F.; Singh, S. B., Discovery of platencin, a dual FabF and FabH inhibitor with *in vivo* antibiotic properties. *Proceedings of the National Academy of Sciences* **2007**, *104* (18), 7612-7616.
 102. Pethe, K.; Sequeira, P. C.; Agarwalla, S.; Rhee, K.; Kuhen, K.; Phong, W. Y.; Patel, V.; Beer, D.; Walker, J. R.; Duraiswamy, J.; Jiricek, J.; Keller, T. H.; Chatterjee, A.; Tan, M. P.; Ujjini, M.; Rao, S. P. S.; Camacho, L.; Bifani, P.; Mak, P. A.; Ma, I.; Barnes, S. W.; Chen, Z.; Plouffe, D.; Thayalan, P.; Ng, S. H.; Au, M.; Lee, B. H.; Tan, B. H.; Ravindran, S.; Nanjundappa, M.; Lin, X.; Goh, A.; Lakshminarayana, S. B.; Shoen, C.; Cynamon, M.; Kreiswirth, B.; Dartois, V.; Peters, E. C.; Glynn, R.; Brenner, S.; Dick, T., A chemical genetic screen in *Mycobacterium tuberculosis* identifies carbon-source-dependent growth inhibitors devoid of *in vivo* efficacy. *Nature Communications* **2010**, *1*, 57.
 103. Stanley, S. A.; Grant, S. S.; Kawate, T.; Iwase, N.; Shimizu, M.; Wivagg, C.; Silvis, M.; Kazyanskaya, E.; Aquadro, J.; Golas, A.; Fitzgerald, M.; Dai, H.; Zhang, L.; Hung, D. T., Identification of Novel Inhibitors of *M. tuberculosis* Growth Using Whole Cell Based High-Throughput Screening. *ACS Chem. Biol.* **2012**, *7* (8), 1377-1384.
 104. de Jong, L. A.; Uges, D. R.; Franke, J. P.; Bischoff, R., Receptor-ligand binding assays: technologies and applications. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2005**, *829* (1-2), 1-25.
 105. Maffioli, S. I.; Zhang, Y.; Degen, D.; Carzaniga, T.; Del Gatto, G.; Serina, S.; Monciardini, P.; Mazzetti, C.; Guglierame, P.; Candiani, G.; Chiriac, A. I.; Facchetti, G.; Kaltofen, P.; Sahl, H.-G.; Dehò, G.; Donadio, S.; Ebright, R. H., Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase. *Cell* **2017**, *169* (7), 1240-1248.e23.
 106. Bergsdorf, C.; Ottl, J., Affinity-based screening techniques: their impact and benefit to increase the number of high quality leads. *Expert Opin Drug Discov* **2010**, *5* (11), 1095-107.
 107. Wu, W.-S.; Cheng, W.-C.; Cheng, T.-J. R.; Wong, C.-H., Affinity-Based Screen for Inhibitors of Bacterial Transglycosylase. *J. Am. Chem. Soc.* **2018**, *140* (8), 2752-2755.
 108. Miyazaki, Y.; Shibuya, M.; Sugawara, H.; Kawaguchi, O.; Hirsoe, C., Salinomycin, a new polyether antibiotic. *J. Antibiot. (Tokyo)* **1974**, *27* (11), 814-21.
 109. Gupta, P. B.; Onder, T. T.; Jiang, G.; Tao, K.; Kuperwasser, C.; Weinberg, R. A.; Lander, E. S., Identification of Selective Inhibitors of Cancer Stem Cells by High-Throughput Screening. *Cell* **2009**, *138* (4), 645-659.
 110. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonis, H. M.; Gerwick, W. H.; Dorrestein, P. C., Molecular networking as a dereplication strategy. *J. Nat. Prod.* **2013**, *76* (9), 1686-99.
 111. Duncan, Katherine R.; Crüsemann, M.; Lechner, A.; Sarkar, A.; Li, J.; Ziemert, N.; Wang, M.; Bandeira, N.; Moore, Bradley S.; Dorrestein, Pieter C.; Jensen, Paul R., Molecular Networking and Pattern-Based Genome Mining Improves Discovery of Biosynthetic Gene Clusters and their Products from *Salinispora* Species. *Chem. Biol.* **2015**, *22* (4), 460-471.
 112. Dashti, Y.; Grkovic, T.; Abdelmohsen, U. R.; Hentschel, U.; Quinn, R. J., Actinomycete Metabolome Induction/Suppression with N-Acetylglucosamine. *J. Nat. Prod.* **2017**, *80* (4), 828-836.
 113. Rigali, S.; Titgemeyer, F.; Barends, S.; Mulder, S.; Thomae, A. W.; Hopwood, D. A.; van Wezel, G. P., Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Reports* **2008**, *9* (7), 670-675.
 114. Świątek, M. A.; Tenconi, E.; Rigali, S.; van Wezel, G. P., Functional Analysis of the N-Acetylglucosamine Metabolic Genes of *Streptomyces coelicolor* and Role in Control of Development and Antibiotic Production. *J. Bacteriol.* **2012**, *194* (5), 1136-1144.

115. Rigali, S.; Nothaft, H.; Noens, E. E.; Schlicht, M.; Colson, S.; Muller, M.; Joris, B.; Koerten, H. K.; Hopwood, D. A.; Titgemeyer, F.; van Wezel, G. P., The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Mol. Microbiol.* **2006**, *61* (5), 1237-51.
116. Sanchez, S.; Chavez, A.; Forero, A.; Garcia-Huante, Y.; Romero, A.; Sanchez, M.; Rocha, D.; Sanchez, B.; Avalos, M.; Guzman-Trampe, S.; Rodriguez-Sanoja, R.; Langley, E.; Ruiz, B., Carbon source regulation of antibiotic production. *J. Antibiot. (Tokyo)* **2010**, *63* (8), 442-59.
117. Intra, B.; Euanorasetr, J.; Nihira, T.; Panbangred, W., Characterization of a gamma-butyrolactone synthetase gene homologue (stcA) involved in bafilomycin production and aerial mycelium formation in *Streptomyces* sp. SBI034. *Appl. Microbiol. Biotechnol.* **2016**, *100* (6), 2749-60.
118. Tan, G.-Y.; Peng, Y.; Lu, C.; Bai, L.; Zhong, J.-J., Engineering validamycin production by tandem deletion of γ -butyrolactone receptor genes in *Streptomyces hygroscopicus* 5008. *Metab. Eng.* **2015**, *28*, 74-81.
119. Kato, J.-y.; Funa, N.; Watanabe, H.; Ohnishi, Y.; Horinouchi, S., Biosynthesis of γ -butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. *Proceedings of the National Academy of Sciences* **2007**, *104* (7), 2378-2383.
120. Du, Y.-L.; Shen, X.-L.; Yu, P.; Bai, L.-Q.; Li, Y.-Q., Gamma-Butyrolactone Regulatory System of *Streptomyces chattanoogensis* Links Nutrient Utilization, Metabolism, and Development. *Appl. Environ. Microbiol.* **2011**, *77* (23), 8415-8426.
121. Imai, Y.; Sato, S.; Tanaka, Y.; Ochi, K.; Hosaka, T., Lincomycin at Subinhibitory Concentrations Potentiates Secondary Metabolite Production by *Streptomyces* spp. *Appl. Environ. Microbiol.* **2015**, *81* (11), 3869-3879.
122. Yoon, V.; Nodwell, J. R., Activating secondary metabolism with stress and chemicals. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 415-424.
123. Xu, F.; Nazari, B.; Moon, K.; Bushin, L. B.; Seyedsayamdost, M. R., Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074 Using High-Throughput Elicitor Screens. *J. Am. Chem. Soc.* **2017**, *139* (27), 9203-9212.
124. Cornforth, D. M.; Foster, K. R., Antibiotics and the art of bacterial war. *Proceedings of the National Academy of Sciences* **2015**, *112* (35), 10827-10828.
125. Chen, G.; Wang, G. Y.; Li, X.; Waters, B.; Davies, J., Enhanced production of microbial metabolites in the presence of dimethyl sulfoxide. *J. Antibiot. (Tokyo)* **2000**, *53* (10), 1145-53.
126. K., P. R., Small - molecule elicitation of microbial secondary metabolites. *Microbial Biotechnology* **2011**, *4* (4), 471-478.
127. Van Arnem, E. B.; Ruzzini, A. C.; Sit, C. S.; Horn, H.; Pinto-Tomás, A. A.; Currie, C. R.; Clardy, J., Selvamycin, an atypical antifungal polyene from two alternative genomic contexts. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (46), 12940-12945.
128. Koju, D.; Maharjan, S.; Dhakal, D.; Yoo, J. C.; Sohng, J. K., Effect of different biosynthetic precursors on the production of nargenicin A1 from metabolically engineered *Nocardia* sp. CS682. *J. Microbiol. Biotechnol.* **2012**, *22* (8), 1127-32.
129. Hiard, S.; Marée, R.; Colson, S.; Hoskisson, P. A.; Titgemeyer, F.; van Wezel, G. P.; Joris, B.; Wehenkel, L.; Sébastien, R., PREDetector: A new tool to identify regulatory elements in bacterial genomes. *Biochem. Biophys. Res. Commun.* **2007**, *357* (4), 861-864.
130. Traxler, M. F.; Watrous, J. D.; Alexandrov, T.; Dorrestein, P. C.; Kolter, R., Interspecies Interactions Stimulate Diversification of the *Streptomyces coelicolor* Secreted Metabolome. *mBio* **2013**, *4* (4).
131. Juana, P.; José, M. D.; F., B. A.; J., S. L.; Laura, S.; I., S. R., *Myxococcus xanthus* induces actinorhodin overproduction and aerial mycelium formation by *Streptomyces coelicolor*. *Microbial Biotechnology* **2011**, *4* (2), 175-183.
132. Slattery, M.; Rajbhandari, I.; Wesson, K., Competition-mediated antibiotic induction in the marine bacterium *Streptomyces tenjimariensis*. *Microb. Ecol.* **2001**, *41* (2), 90-96.

133. Asamizu, S.; Ozaki, T.; Teramoto, K.; Satoh, K.; Onaka, H., Killing of Mycolic Acid-Containing Bacteria Aborted Induction of Antibiotic Production by Streptomyces in Combined-Culture. *PLoS One* **2015**, *10* (11), e0142372.
134. Onaka, H.; Mori, Y.; Igarashi, Y.; Furumai, T., Mycolic Acid-Containing Bacteria Induce Natural-Product Biosynthesis in Streptomyces Species. *Appl. Environ. Microbiol.* **2011**, *77* (2), 400-406.
135. Onaka, H.; Ozaki, T.; Mori, Y.; Izawa, M.; Hayashi, S.; Asamizu, S., Mycolic acid-containing bacteria activate heterologous secondary metabolite expression in Streptomyces lividans. *The Journal Of Antibiotics* **2015**, *68*, 594.
136. Hoshino, S.; Okada, M.; Wakimoto, T.; Zhang, H.; Hayashi, F.; Onaka, H.; Abe, I., Niizalactams A–C, Multicyclic Macrolactams Isolated from Combined Culture of Streptomyces with Mycolic Acid-Containing Bacterium. *J. Nat. Prod.* **2015**, *78* (12), 3011-3017.
137. Dalmas, F. R.; Astarita, L.; DeFilippis, L.; Magel, E.; Fiedler, H.-P.; Bauer, R.; Hampp, R., Growth inhibition of an Araucaria angustifolia (Coniferopsida) fungal seed pathogen, Neofusicoccum parvum, by soil streptomycetes. *BMC Microbiol.* **2013**, *13* (1), 168.
138. Straight, P. D.; Willey, J. M.; Kolter, R., Interactions between Streptomyces coelicolor and Bacillus subtilis: Role of Surfactants in Raising Aerial Structures. *J. Bacteriol.* **2006**, *188* (13), 4918-4925.
139. Adnani, N.; Chevette, M. G.; Adibhatla, S. N.; Zhang, F.; Yu, Q.; Braun, D. R.; Nelson, J.; Simpkins, S. W.; McDonald, B. R.; Myers, C. L.; Piotrowski, J. S.; Thompson, C. J.; Currie, C. R.; Li, L.; Rajski, S. R.; Bugni, T. S., Coculture of Marine Invertebrate-Associated Bacteria and Interdisciplinary Technologies Enable Biosynthesis and Discovery of a New Antibiotic, Keyicin. *ACS Chem. Biol.* **2017**, *12* (12), 3093-3102.
140. Nai, C.; Meyer, V., From Axenic to Mixed Cultures: Technological Advances Accelerating a Paradigm Shift in Microbiology. *Trends Microbiol.* **2018**, *26* (6), 538-554.
141. Dashti, Y.; Grkovic, T.; Abdelmohsen, U.; Hentschel, U.; Quinn, R., Production of Induced Secondary Metabolites by a Co-Culture of Sponge-Associated Actinomycetes, Actinokineospora sp. EG49 and Nocardiosis sp. RV163. *Mar. Drugs* **2014**, *12* (5), 3046.
142. Sung, A. A.; Gromek, S. M.; Balunas, M. J., Upregulation and Identification of Antibiotic Activity of a Marine-Derived Streptomyces sp. via Co-Cultures with Human Pathogens. *Mar. Drugs* **2017**, *15* (8), 250.
143. Marmann, A.; Aly, A. H.; Lin, W.; Wang, B.; Proksch, P., Co-Cultivation—A Powerful Emerging Tool for Enhancing the Chemical Diversity of Microorganisms. *Mar. Drugs* **2014**, *12* (2), 1043-1065.
144. Bode, H. B.; Bethe, B.; Hofs, R.; Zeeck, A., Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* **2002**, *3* (7), 619-27.
145. Ripa, F. A.; Nikkon, F.; Zaman, S.; Khondkar, P., Optimal Conditions for Antimicrobial Metabolites Production from a New Streptomyces sp. RUPA-08PR Isolated from Bangladeshi Soil. *Mycobiology* **2009**, *37* (3), 211-214.
146. James, P. D.; Edwards, C.; Dawson, M., The effects of temperature, pH and growth rate on secondary metabolism in Streptomyces thermoviolaceus grown in a chemostat. *J. Gen. Microbiol.* **1991**, *137* (7), 1715-20.
147. Martín, J. F., Phosphate Control of the Biosynthesis of Antibiotics and Other Secondary Metabolites Is Mediated by the PhoR-PhoP System: an Unfinished Story. *J. Bacteriol.* **2004**, *186* (16), 5197-5201.
148. Voelker, F.; Altaba, S., Nitrogen source governs the patterns of growth and pristnamycin production in 'Streptomyces pristinaespiralis'. *Microbiology* **2001**, *147* (Pt 9), 2447-59.
149. Ng, Y. K.; Hodson, M. P.; Hewavitharana, A. K.; Bose, U.; Shaw, P. N.; Fuerst, J. A., Effects of salinity on antibiotic production in sponge-derived Salinispora actinobacteria. *J. Appl. Microbiol.* **2014**, *117* (1), 109-25.
150. Locatelli, F. M.; Goo, K.-S.; Ulanova, D., Effects of trace metal ions on secondary metabolism and the morphological development of streptomycetes. *Metallomics* **2016**, *8* (5), 469-480.

151. Arul Jose, P.; Sivakala, K. K.; Jebakumar, S. R. D., Formulation and Statistical Optimization of Culture Medium for Improved Production of Antimicrobial Compound by *Streptomyces* sp. JAJ06. *Int. J. Microbiol.* **2013**, *2013*, 526260.
152. Usha Kiranmayi, M.; Sudhakar, P.; Sreenivasulu, K.; Vijayalakshmi, M., Optimization of Culturing Conditions for Improved Production of Bioactive Metabolites by *Pseudonocardia* sp. VUK-10. *Mycobiology* **2011**, *39* (3), 174-181.
153. Padma, P. N.; Rao, A. B.; Yadav, J. S.; Reddy, G., Optimization of fermentation conditions for production of glycopeptide antibiotic vancomycin by *Amycolatopsis orientalis*. *Appl. Biochem. Biotechnol.* **2002**, *102-103* (1-6), 395-405.
154. Gao, X.; He, Q.; Jiang, Y.; Huang, L., Optimization of Nutrient and Fermentation Parameters for Antifungal Activity by *Streptomyces lavendulae* Xjy and Its Biocontrol Efficacies against *Fulvia fulva* and *Botryosphaeria dothidea*. *J. Phytopathol.* **2015**, *164* (3), 155-165.
155. Banskota, A. H.; McAlpine, J. B.; Sorensen, D.; Ibrahim, A.; Aouidate, M.; Pirae, M.; Alarco, A. M.; Farnet, C. M.; Zazopoulos, E., Genomic analyses lead to novel secondary metabolites. Part 3. ECO-0501, a novel antibacterial of a new class. *J. Antibiot. (Tokyo)* **2006**, *59* (9), 533-42.
156. Schulze, C. J.; Donia, M. S.; Siqueira-Neto, J. L.; Ray, D.; Raskatov, J. A.; Green, R. E.; McKerrow, J. H.; Fischbach, M. A.; Linington, R. G., Genome-Directed Lead Discovery: Biosynthesis, Structure Elucidation, and Biological Evaluation of Two Families of Polyene Macrolactams against *Trypanosoma brucei*. *ACS Chem. Biol.* **2015**, *10* (10), 2373-81.
157. Monciardini, P.; Iorio, M.; Maffioli, S.; Sosio, M.; Donadio, S., Discovering new bioactive molecules from microbial sources. *Microb Biotechnol* **2014**, *7* (3), 209-20.
158. Medema, M. H.; Fischbach, M. A., Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015**, *11* (9), 639-48.
159. Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T., antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **2013**, *41* (Web Server issue), W204-12.
160. Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; Cruz-Morales, P.; Duddela, S.; Dusterhus, S.; Edwards, D. J.; Fewer, D. P.; Garg, N.; Geiger, C.; Gomez-Escribano, J. P.; Greule, A.; Hadjithomas, M.; Haines, A. S.; Helfrich, E. J. N.; Hillwig, M. L.; Ishida, K.; Jones, A. C.; Jones, C. S.; Jungmann, K.; Kegler, C.; Kim, H. U.; Kotter, P.; Krug, D.; Masschelein, J.; Melnik, A. V.; Mantovani, S. M.; Monroe, E. A.; Moore, M.; Moss, N.; Nutzmam, H.-W.; Pan, G.; Pati, A.; Petras, D.; Reen, F. J.; Rosconi, F.; Rui, Z.; Tian, Z.; Tobias, N. J.; Tsunematsu, Y.; Wiemann, P.; Wyckoff, E.; Yan, X.; Yim, G.; Yu, F.; Xie, Y.; Aigle, B.; Apel, A. K.; Balibar, C. J.; Balskus, E. P.; Barona-Gomez, F.; Bechthold, A.; Bode, H. B.; Borriss, R.; Brady, S. F.; Brakhage, A. A.; Caffrey, P.; Cheng, Y.-Q.; Clardy, J.; Cox, R. J.; De Mot, R.; Donadio, S.; Donia, M. S.; van der Donk, W. A.; Dorrestein, P. C.; Doyle, S.; Driessen, A. J. M.; Ehling-Schulz, M.; Entian, K.-D.; Fischbach, M. A.; Gerwick, L.; Gerwick, W. H.; Gross, H.; Gust, B.; Hertweck, C.; Hofte, M.; Jensen, S. E.; Ju, J.; Katz, L.; Kaysser, L.; Klassen, J. L.; Keller, N. P.; Kormanec, J.; Kuipers, O. P.; Kuzuyama, T.; Kyrpides, N. C.; Kwon, H.-J.; Lautru, S.; Lavigne, R.; Lee, C. Y.; Lincuan, B.; Liu, X.; Liu, W.; Luzhetskyy, A.; Mahmud, T.; Mast, Y.; Mendez, C.; Metsa-Ketela, M.; Micklefield, J.; Mitchell, D. A.; Moore, B. S.; Moreira, L. M.; Muller, R.; Neilan, B. A.; Nett, M.; Nielsen, J.; O'Gara, F.; Oikawa, H.; Osbourn, A.; Osburne, M. S.; Ostash, B.; Payne, S. M.; Pernodet, J.-L.; Petrcek, M.; Piel, J.; Ploux, O.; Raaijmakers, J. M.; Salas, J. A.; Schmitt, E. K.; Scott, B.; Seipke, R. F.; Shen, B.; Sherman, D. H.; Sivonen, K.; Smanski, M. J.; Sosio, M.; Stegmann, E.; Sussmuth, R. D.; Tahlan, K.; Thomas, C. M.; Tang, Y.; Truman, A. W.; Viaud, M.; Walton, J. D.; Walsh, C. T.; Weber, T.; van Wezel, G. P.; Wilkinson, B.; Willey, J. M.; Wohlleben, W.; Wright, G. D.; Ziemert, N.; Zhang, C.; Zotchev, S. B.; Breitling, R.; Takano, E.; Glockner, F. O., Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015**, *11* (9), 625-631.
161. Blin, K.; Wolf, T.; Chevrete, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de Los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.;

- Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017**, *45* (W1), W36-w41.
162. Hadjithomas, M.; Chen, I. M.; Chu, K.; Ratner, A.; Palaniappan, K.; Szeto, E.; Huang, J.; Reddy, T. B.; Cimermancic, P.; Fischbach, M. A.; Ivanova, N. N.; Markowitz, V. M.; Kyrpides, N. C.; Pati, A., IMG-ABC: A Knowledge Base To Fuel Discovery of Biosynthetic Gene Clusters and Novel Secondary Metabolites. *MBio* **2015**, *6* (4), e00932.
163. Blin, K.; Medema, M. H.; Kottmann, R.; Lee, S. Y.; Weber, T., The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* **2017**, *45* (D1), D555-D559.
164. Lucas, X.; Senger, C.; Erxleben, A.; Gruning, B. A.; Doring, K.; Mosch, J.; Flemming, S.; Gunther, S., StreptomeDB: a resource for natural compounds isolated from Streptomyces species. *Nucleic Acids Res.* **2013**, *41* (Database issue), D1130-6.
165. Mostafa, M. V.; Dietmar, B.; Erich, L.; Grazyna, J. B.; M., T. C., Conjugative transfer functions of broad - host - range plasmid RK2 are coregulated with vegetative replication. *Mol. Microbiol.* **1992**, *6* (7), 907-920.
166. MacNeil, D. J., Characterization of a unique methyl-specific restriction system in Streptomyces avermitilis. *J. Bacteriol.* **1988**, *170* (12), 5607-5612.
167. MacNeil, D. J.; Gewain, K. M.; Ruby, C. L.; Dezeny, G.; Gibbons, P. H.; MacNeil, T., Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **1992**, *111* (1), 61-68.
168. Wohlleben, W.; Pühler, A., The Streptomyces ghanaensis low copy plasmid pSG2 and its use for vector construction. *Arch. Microbiol.* **1987**, *148* (4), 298-304.
169. Dyson, P. J.; Evans, M., pUCS75, a stable high-copy-number Streptomyces—Escherichia coli shuttle vector which facilitates subcloning from pUC plasmid and M 13 phage vectors. *Gene* **1996**, *171* (1), 71-73.
170. Kieser, T.; Hopwood, D. A.; Wright, H. M.; Thompson, C. J., pIJ101, a multi-copy broad host-range Streptomyces plasmid: Functional analysis and development of DNA cloning vectors. *Molecular and General Genetics MGG* **1982**, *185* (2), 223-238.
171. Fayed, B.; Younger, E.; Taylor, G.; Smith, M. C. M., A novel Streptomyces spp. integration vector derived from the S. venezuelae phage, SV1. *BMC Biotechnol.* **2014**, *14*, 51-51.
172. Phelan, R. M.; Sachs, D.; Petkiewicz, S. J.; Barajas, J. F.; Blake-Hedges, J. M.; Thompson, M. G.; Reider Apel, A.; Rasor, B. J.; Katz, L.; Keasling, J. D., Development of Next Generation Synthetic Biology Tools for Use in Streptomyces venezuelae. *ACS synthetic biology* **2017**, *6* (1), 159-166.
173. Kuhstoss, S.; Richardson, M. A.; Rao, R. N., Plasmid cloning vectors that integrate site-specifically in Streptomyces spp. *Gene* **1991**, *97* (1), 143-6.
174. Fogg, P. C. M.; Colloms, S.; Rosser, S.; Stark, M.; Smith, M. C. M., New Applications for Phage Integrases. *J. Mol. Biol.* **2014**, *426* (15), 2703-2716.
175. Gregory, M. A.; Till, R.; Smith, M. C., Integration site for Streptomyces phage phiBT1 and development of site-specific integrating vectors. *J. Bacteriol.* **2003**, *185* (17), 5320-3.
176. Keravala, A.; Calos, M. P., Site-specific chromosomal integration mediated by phiC31 integrase. *Methods Mol. Biol.* **2008**, *435*, 165-73.
177. Van Mellaert, L.; Mei, L.; Lammertyn, E.; Schacht, S.; Anne, J., Site-specific integration of bacteriophage VWB genome into Streptomyces venezuelae and construction of a VWB-based integrative vector. *Microbiology* **1998**, *144* (Pt 12), 3351-8.
178. Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F., PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (4), 1541-1546.
179. Herrmann, S.; Siegl, T.; Luzhetskaya, M.; Petzke, L.; Jilg, C.; Welle, E.; Erb, A.; Leadlay, P. F.; Bechthold, A.; Luzhetskyy, A., Site-Specific Recombination Strategies for Engineering Actinomycete Genomes. *Appl. Environ. Microbiol.* **2012**, *78* (6), 1804-1812.

180. Fernández-Martínez, L. T.; Bibb, M. J., Use of the Meganuclease I-SceI of *Saccharomyces cerevisiae* to select for gene deletions in actinomycetes. *Sci. Rep.* **2014**, *4*, 7100.
181. Cobb, R. E.; Wang, Y.; Zhao, H., High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS synthetic biology* **2015**, *4* (6), 723-8.
182. Zeng, H.; Wen, S.; Xu, W.; He, Z.; Zhai, G.; Liu, Y.; Deng, Z.; Sun, Y., Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using CRISPR/Cas9-CodA(sm) combined system. *Appl. Microbiol. Biotechnol.* **2015**, 1-11.
183. Tong, Y.; Charusanti, P.; Zhang, L.; Weber, T.; Lee, S. Y., CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS synthetic biology* **2015**, *4* (9), 1020-9.
184. Liu, Y.; Tao, W.; Wen, S.; Li, Z.; Yang, A.; Deng, Z.; Sun, Y., In Vitro CRISPR/Cas9 System for Efficient Targeted DNA Editing. *mBio* **2015**, *6* (6).
185. Huang, H.; Zheng, G.; Jiang, W.; Hu, H.; Lu, Y., One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta biochimica et biophysica Sinica* **2015**, *47* (4), 231-43.
186. Horinouchi, S., Mining and polishing of the treasure trove in the bacterial genus streptomyces. *Biosci. Biotechnol. Biochem.* **2007**, *71* (2), 283-99.
187. Xu, J.; Zhang, J.; Zhuo, J.; Li, Y.; Tian, Y.; Tan, H., Activation and mechanism of a cryptic oviedomycin gene cluster via the disruption of a global regulatory gene, *adpA*, in *Streptomyces ansochromogenes*. *J. Biol. Chem.* **2017**, *292* (48), 19708-19720.
188. van Wezel, G. P.; McDowall, K. J., The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat. Prod. Rep.* **2011**, *28* (7), 1311-33.
189. McCormick, J. R.; Flardh, K., Signals and regulators that govern *Streptomyces* development. *FEMS Microbiol. Rev.* **2012**, *36* (1), 206-31.
190. Baltz, R. H., Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (2-3), 343-70.
191. Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S., Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (5), 1957-62.
192. Gomez, C.; Olano, C.; Mendez, C.; Salas, J. A., Three pathway-specific regulators control streptolydigin biosynthesis in *Streptomyces lydicus*. *Microbiology* **2012**, *158* (Pt 10), 2504-14.
193. He, W.; Lei, J.; Liu, Y.; Wang, Y., The LuxR family members GdmRI and GdmRII are positive regulators of geldanamycin biosynthesis in *Streptomyces hygroscopicus* 17997. *Arch. Microbiol.* **2008**, *189* (5), 501-10.
194. Jiang, M.; Yin, M.; Wu, S.; Han, X.; Ji, K.; Wen, M.; Lu, T., GdmRIII, a TetR Family Transcriptional Regulator, Controls Geldanamycin and Elaiophylin Biosynthesis in *Streptomyces autolyticus* CGMCC0516. *Sci. Rep.* **2017**, *7* (1), 4803.
195. Hou, B.; Lin, Y.; Wu, H.; Guo, M.; Petkovic, H.; Tao, L.; Zhu, X.; Ye, J.; Zhang, H., The Novel Transcriptional Regulator LmbU Promotes Lincomycin Biosynthesis through Regulating Expression of Its Target Genes in *Streptomyces lincolnensis*. *J. Bacteriol.* **2018**, *200* (2).
196. Tu, J.; Li, S.; Chen, J.; Song, Y.; Fu, S.; Ju, J.; Li, Q., Characterization and heterologous expression of the neoabyssomicin/abyssomicin biosynthetic gene cluster from *Streptomyces koyangensis* SCSIO 5802. *Microbial Cell Factories* **2018**, *17* (1), 28.
197. Olano, C.; García, I.; González, A.; Rodríguez, M.; Rozas, D.; Rubio, J.; Sánchez-Hidalgo, M.; Braña, A. F.; Méndez, C.; Salas, J. A., Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microbial Biotechnology* **2014**, *7* (3), 242-256.
198. Khalid, A.; Takagi, H.; Panthee, S.; Muroi, M.; Chappell, J.; Osada, H.; Takahashi, S., Development of a Terpenoid-Production Platform in *Streptomyces reveromyceticus* SN-593. *ACS synthetic biology* **2017**, *6* (12), 2339-2349.
199. Carlos, O.; Ignacio, G.; Aranzazu, G.; Miriam, R.; Daniel, R.; Julio, R.; Marina, S. H.; F., B. A.; Carmen, M.; A., S. J., Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microbial Biotechnology* **2014**, *7* (3), 242-256.

200. Zhang, M. M.; Wong, F. T.; Wang, Y.; Luo, S.; Lim, Y. H.; Heng, E.; Yeo, W. L.; Cobb, R. E.; Enghiad, B.; Ang, E. L.; Zhao, H., CRISPR–Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. *Nat. Chem. Biol.* **2017**, *13*, 607.
201. Liu, Y.; Wei, W.-P.; Ye, B.-C., High GC Content Cas9-Mediated Genome-Editing and Biosynthetic Gene Cluster Activation in *Saccharopolyspora erythraea*. *ACS synthetic biology* **2018**.
202. Pascolutti, M.; Quinn, R. J., Natural products as lead structures: chemical transformations to create lead-like libraries. *Drug Discovery Today* **2014**, *19* (3), 215-221.
203. Li, J.; Li, L.; Tian, Y.; Niu, G.; Tan, H., Hybrid antibiotics with the nikkomycin nucleoside and polyoxin peptidyl moieties. *Metab. Eng.* **2011**, *13* (3), 336-44.
204. Hoepfich, P. D., Clinical use of amphotericin B and derivatives: lore, mystique, and fact. *Clin. Infect. Dis.* **1992**, *14 Suppl 1*, S114-9.
205. Dixon, D. M.; McNeil, M. M.; Cohen, M. L.; Gellin, B. G.; La Montagne, J. R., Fungal infections: a growing threat. *Public Health Rep.* **1996**, *111* (3), 226-35.
206. Wiederhold, N. P., Antifungal resistance: current trends and future strategies to combat. *Infection and Drug Resistance* **2017**, *10*, 249-259.
207. De Poire, E.; Stephens, N.; Rawlings, B.; Caffrey, P., Engineered Biosynthesis of Disaccharide-Modified Polyene Macrolides. *Appl. Environ. Microbiol.* **2013**, *79* (19), 6156-6159.
208. Byrne, B.; Carmody, M.; Gibson, E.; Rawlings, B.; Caffrey, P., Biosynthesis of Deoxyamphotericins and Deoxyamphoteronolides by Engineered Strains of *Streptomyces nodosus*. *Chem. Biol.* **2003**, *10* (12), 1215-1224.
209. Carmody, M.; Murphy, B.; Byrne, B.; Power, P.; Rai, D.; Rawlings, B.; Caffrey, P., Biosynthesis of Amphotericin Derivatives Lacking Exocyclic Carboxyl Groups. *J. Biol. Chem.* **2005**, *280* (41), 34420-34426.
210. Brautaset, T.; Sletta, H.; Degnes, K. F.; Sekurova, O. N.; Bakke, I.; Volokhan, O.; Andreassen, T.; Ellingsen, T. E.; Zotchev, S. B., New Nystatin-Related Antifungal Polyene Macrolides with Altered Polyol Region Generated via Biosynthetic Engineering of *Streptomyces noursei*. *Appl. Environ. Microbiol.* **2011**, *77* (18), 6636-6643.
211. Davis, S. A.; Vincent, B. M.; Endo, M. M.; Whitesell, L.; Marchillo, K.; Andes, D. R.; Lindquist, S.; Burke, M. D., Nontoxic antimicrobials that evade drug resistance. *Nat. Chem. Biol.* **2015**, *11*, 481.
212. Sharma, S. V.; Tong, X.; Pubill-Ulldemolins, C.; Cartmell, C.; Bogosyan, E. J. A.; Rackham, E. J.; Marelli, E.; Hamed, R. B.; Goss, R. J. M., Living GenoChemetics by hyphenating synthetic biology and synthetic chemistry in vivo. *Nature Communications* **2017**, *8* (1), 229.
213. Guo, F.; Xiang, S.; Li, L.; Wang, B.; Rajasarkka, J.; Grondahl-Yli-Hannuksela, K.; Ai, G.; Metsä-Ketela, M.; Yang, K., Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metab. Eng.* **2015**, *28*, 134-142.
214. Sun, Y.-Q.; Busche, T.; Rückert, C.; Paulus, C.; Rebets, Y.; Novakova, R.; Kalinowski, J.; Luzhetskyy, A.; Kormanec, J.; Sekurova, O. N.; Zotchev, S. B., Development of a Biosensor Concept to Detect the Production of Cluster-Specific Secondary Metabolites. *ACS synthetic biology* **2017**, *6* (6), 1026-1033.
215. Ahmed, Y.; Rebets, Y.; Tokovenko, B.; Brötz, E.; Luzhetskyy, A., Identification of butenolide regulatory system controlling secondary metabolism in *Streptomyces albus* J1074. *Sci. Rep.* **2017**, *7* (1), 9784.
216. Zotchev, S. B., Genomics-Based Insights into the Evolution of Secondary Metabolite Biosynthesis in Actinomycete Bacteria. In *Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life*, Pontarotti, P., Ed. Springer International Publishing: Cham, 2014; pp 35-45.
217. ARUL JOSE, P.; Jebakumar, S. R. D., Non-streptomycete actinomycetes nourish the current microbial antibiotic drug discovery. *Front. Microbiol.* **2013**, *4* (240).
218. Tiwari, K.; Gupta, R. K., Rare actinomycetes: a potential storehouse for novel antibiotics. *Crit. Rev. Biotechnol.* **2012**, *32* (2), 108-32.

219. Cohen, D. R.; Townsend, C. A., A dual role for a polyketide synthase in dynemicin enediyne and anthraquinone biosynthesis. *Nat. Chem.* **2018**, *10* (2), 231-236.
220. Zaburannyi, N.; Rabyk, M.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A., Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics* **2014**, *15*, 97-97.
221. Winter, J. M.; Moffitt, M. C.; Zazopoulos, E.; McAlpine, J. B.; Dorrestein, P. C.; Moore, B. S., Molecular basis for chloronium-mediated meroterpene cyclization: cloning, sequencing, and heterologous expression of the napyradiomycin biosynthetic gene cluster. *J. Biol. Chem.* **2007**, *282* (22), 16362-8.
222. Kim, S. Y.; Zhao, P.; Igarashi, M.; Sawa, R.; Tomita, T.; Nishiyama, M.; Kuzuyama, T., Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two p450 hydroxylases. *Chem. Biol.* **2009**, *16* (7), 736-43.
223. Feng, Z.; Wang, L.; Rajski, S. R.; Xu, Z.; Coeffet-LeGal, M. F.; Shen, B., Engineered production of iso-migrastatin in heterologous *Streptomyces* hosts. *Bioorg. Med. Chem.* **2009**, *17* (6), 2147-53.
224. Lombo, F.; Velasco, A.; Castro, A.; de la Calle, F.; Brana, A. F.; Sanchez-Puelles, J. M.; Mendez, C.; Salas, J. A., Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two streptomyces species. *ChemBioChem* **2006**, *7* (2), 366-76.
225. Kallifidas, D.; Jiang, G.; Ding, Y.; Luesch, H., Rational engineering of *Streptomyces albus* J1074 for the overexpression of secondary metabolite gene clusters. *Microbial Cell Factories* **2018**, *17*, 25.
226. Campbell, W. C., History of avermectin and ivermectin, with notes on the history of other macrocyclic lactone antiparasitic agents. *Curr. Pharm. Biotechnol.* **2012**, *13* (6), 853-65.
227. Komatsu, M.; Uchiyama, T.; Ōmura, S.; Cane, D. E.; Ikeda, H., Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proceedings of the National Academy of Sciences* **2010**, *107* (6), 2646-2651.
228. Komatsu, M.; Komatsu, K.; Koiwai, H.; Yamada, Y.; Kozono, I.; Izumikawa, M.; Hashimoto, J.; Takagi, M.; Omura, S.; Shin-ya, K.; Cane, D. E.; Ikeda, H., Engineered *Streptomyces avermitilis* Host for Heterologous Expression of Biosynthetic Gene Cluster for Secondary Metabolites. *ACS synthetic biology* **2013**, *2* (7), 384-396.
229. Yamada, Y.; Arima, S.; Nagamitsu, T.; Johmoto, K.; Uekusa, H.; Eguchi, T.; Shin-ya, K.; Cane, D. E.; Ikeda, H., Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered *Streptomyces* host. *The Journal Of Antibiotics* **2015**, *68*, 385.
230. Gomez-Escribano, J. P.; Bibb, M. J., Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology* **2011**, *4* (2), 207-215.
231. Denis, F.; Brzezinski, R., A versatile shuttle cosmid vector for use in *Escherichia coli* and actinomycetes. *Gene* **1992**, *111* (1), 115-118.
232. Alduina, R.; Gallo, G., Artificial Chromosomes to Explore and to Exploit Biosynthetic Capabilities of Actinomycetes. *J. Biomed. Biotechnol.* **2012**, *2012*, 10.
233. Deng, Q.; Zhou, L.; Luo, M.; Deng, Z.; Zhao, C., Heterologous expression of Avermectins biosynthetic gene cluster by construction of a Bacterial Artificial Chromosome library of the producers. *Synthetic and Systems Biotechnology* **2017**, *2* (1), 59-64.
234. Alduina, R.; De Grazia, S.; Dolce, L.; Salerno, P.; Sosio, M.; Donadio, S.; Puglia, A. M., Artificial chromosome libraries of *Streptomyces coelicolor* A3(2) and *Planobispora rosea*. *FEMS Microbiol. Lett.* **2003**, *218* (1), 181-6.
235. Larionov, V.; Kouprina, N.; Graves, J.; Chen, X. N.; Korenberg, J. R.; Resnick, M. A., Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (1), 491-496.
236. Noskov, V. N.; Kouprina, N.; Leem, S. H.; Ouspenski, I.; Barrett, J. C.; Larionov, V., A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics* **2003**, *4* (1), 16.
237. Lee, N. C.; Larionov, V.; Kouprina, N., Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Res.* **2015**, *43* (8), e55.

238. Leem, S. H.; Noskov, V. N.; Park, J. E.; Kim, S. I.; Larionov, V.; Kouprina, N., Optimum conditions for selective isolation of genes from complex genomes by transformation-associated recombination cloning. *Nucleic Acids Res.* **2003**, *31* (6), e29.
239. Noskov, V. N.; Koriabine, M.; Solomon, G.; Randolph, M.; Barrett, J. C.; Leem, S. H.; Stubbs, L.; Kouprina, N.; Larionov, V., Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. *Nucleic Acids Res.* **2001**, *29* (6), e32-e32.
240. Raymond, C. K.; Sims, E. H.; Kas, A.; Spencer, D. H.; Kutayin, T. V.; Ivey, R. G.; Zhou, Y.; Kaul, R.; Clendenning, J. B.; Olson, M. V., Genetic variation at the O-antigen biosynthetic locus in *Pseudomonas aeruginosa*. *J. Bacteriol.* **2002**, *184* (13), 3614-22.
241. Benders, G. A.; Noskov, V. N.; Denisova, E. A.; Lartigue, C.; Gibson, D. G.; Assad-Garcia, N.; Chuang, R.-Y.; Carrera, W.; Moodie, M.; Algire, M. A.; Phan, Q.; Alperovich, N.; Vashee, S.; Merryman, C.; Venter, J. C.; Smith, H. O.; Glass, J. I.; Hutchison, C. A., Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.* **2010**, *38* (8), 2558-2569.
242. Gibson, D. G.; Benders, G. A.; Axelrod, K. C.; Zaveri, J.; Algire, M. A.; Moodie, M.; Montague, M. G.; Venter, J. C.; Smith, H. O.; Hutchison, C. A., 3rd, One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (51), 20404-9.
243. Kim, J. H.; Feng, Z.; Bauer, J. D.; Kallifidas, D.; Calle, P. Y.; Brady, S. F., Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. *Biopolymers* **2010**, *93* (9), 833-44.
244. Noskov, V. N.; Karas, B. J.; Young, L.; Chuang, R. Y.; Gibson, D. G.; Lin, Y. C.; Stam, J.; Yonemoto, I. T.; Suzuki, Y.; Andrews-Pfannkoch, C.; Glass, J. I.; Smith, H. O.; Hutchison, C. A., 3rd; Venter, J. C.; Weyman, P. D., Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS synthetic biology* **2012**, *1* (7), 267-73.
245. Bonet, B.; Teufel, R.; Crusemann, M.; Ziemert, N.; Moore, B. S., Direct capture and heterologous expression of *Salinispora* natural product genes for the biosynthesis of enterocin. *J. Nat. Prod.* **2015**, *78* (3), 539-42.
246. Shao, Z.; Zhao, H.; Zhao, H., DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **2009**, *37* (2), e16.
247. Bilyk, O.; Sekurova, O. N.; Zotchev, S. B.; Luzhetskyy, A., Cloning and Heterologous Expression of the Grecoacycline Biosynthetic Gene Cluster. *PLoS One* **2016**, *11* (7), e0158682.
248. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* **2009**, *6* (5), 343-345.
249. Gibson, D. G., Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **2011**, *498*, 349-61.
250. Li, L.; Zhao, Y.; Ruan, L.; Yang, S.; Ge, M.; Jiang, W.; Lu, Y., A stepwise increase in pristnamycin II biosynthesis by *Streptomyces pristinaespiralis* through combinatorial metabolic engineering. *Metab. Eng.* **2015**, *29*, 12-25.
251. Jiang, W.; Zhao, X.; Gabrieli, T.; Lou, C.; Ebenstein, Y.; Zhu, T. F., Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nature Communications* **2015**, *6*, 8101.
252. Yin, J.; Hoffmann, M.; Bian, X.; Tu, Q.; Yan, F.; Xia, L.; Ding, X.; Francis Stewart, A.; Müller, R.; Fu, J.; Zhang, Y., Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Sci. Rep.* **2015**, *5*, 15081.
253. Doroghazi, J. R.; Metcalf, W. W., Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **2013**, *14*, 611.
254. Otsuka, M.; Fujita, M.; Matsushima, Y.; Eguchi, T.; Shindo, K.; Kakinuma, K., Biosynthetic Pathway of Macrolactam Polyketide Glycoside Antitumor Antibiotic Vicenistatins. *Tetrahedron* **2000**, *56* (42), 8281-8286.

255. Hoshino, S.; Okada, M.; Wakimoto, T.; Zhang, H.; Hayashi, F.; Onaka, H.; Abe, I., Niizalactams A-C, Multicyclic Macrolactams Isolated from Combined Culture of Streptomyces with Mycolic Acid-Containing Bacterium. *J Nat Prod* **2015**, *78* (12), 3011-7.
256. Beemelmans, C.; Ramadhar, T. R.; Kim, K. H.; Klassen, J. L.; Cao, S.; Wyche, T. P.; Hou, Y.; Poulsen, M.; Bugni, T. S.; Currie, C. R.; Clardy, J., Macrotermycins A-D, Glycosylated Macrolactams from a Termite-Associated Amycolatopsis sp. M39. *Org Lett* **2017**, *19* (5), 1000-1003.
257. Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O., Mapping Microbial Response Metabolomes for Induced Natural Product Discovery. *ACS Chem. Biol.* **2015**, *10* (9), 1998-2006.
258. Takaishi, M.; Kudo, F.; Eguchi, T., A Unique Pathway for the 3-Aminobutyrate Starter Unit from L-Glutamate through β -Glutamate during Biosynthesis of the 24-Membered Macrolactam Antibiotic, Incadnine. *Org. Lett.* **2012**, *14* (17), 4591-4593.
259. Skellam, E. J.; Stewart, A. K.; Strangman, W. K.; Wright, J. L., Identification of micromonolactam, a new polyene macrocyclic lactam from two marine Micromonospora strains using chemical and molecular methods: clarification of the biosynthetic pathway from a glutamate starter unit. *J. Antibiot.* **2013**, *66* (7), 431-41.
260. Schulze, C. J.; Donia, M. S.; Siqueira-Neto, J. L.; Ray, D.; Raskatov, J. A.; Green, R. E.; McKerrow, J. H.; Fischbach, M. A.; Linington, R. G., Genome-Directed Lead Discovery: Biosynthesis, Structure Elucidation, and Biological Evaluation of Two Families of Polyene Macrolactams against Trypanosoma brucei. *ACS Chem. Biol.* **2015**, *10* (10), 2373-81.
261. Udworthy, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S., Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (25), 10376-10381.
262. Oh, D.-C.; Poulsen, M.; Currie, C. R.; Clardy, J., Sceliphrolactam, a Polyene Macrocyclic Lactam from a Wasp-Associated Streptomyces sp. *Org. Lett.* **2011**, *13* (4), 752-755.
263. Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L., MODULAR ORGANIZATION OF GENES REQUIRED FOR COMPLEX POLYKETIDE BIOSYNTHESIS. *Science* **1991**, *252* (5006), 675-679.
264. Walsh, C. T., Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility. *Science* **2004**, *303* (5665), 1805-1810.
265. Tripathi, G.; Rawal, S. K., Simple and efficient protocol for isolation of high molecular weight DNA from Streptomyces aureofaciens. *Biotech. Techniques* **1998**, *12* (8), 629-631.
266. Gust, B.; O'Rourke, S.; Bird, N.; Kieser, T. & Chater, K, Recombineering in Streptomyces coelicolor. *FEMS Online Protocols* **2006**.
267. Liu, D.; Coloe, S.; Baird, R.; Pedersen, J., Rapid Mini-Preparation of Fungal DNA for PCR. *J. Clin. Microbiol.* **2000**, *38* (1), 471-471.
268. Medema, M. H. e. a., Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015**, *11* (9), 625-631.
269. Bentley, S. D.; Chater, K. F.; Cerdano-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D., Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002**, *417*.
270. Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S., Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nat. Biotech.* **2003**, *21* (5), 526-531.
271. Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T., Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **1998**, *37* (6), 1585-1595.
272. Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D., The Phosphopantetheinyl Transferases: Catalysis of a Posttranslational Modification Crucial for Life. *Nat. Prdt. Rep.* **2014**, *31* (1), 61-108.
273. Murugan, E.; Liang, Z.-X., Evidence for a novel phosphopantetheinyl transferase domain in the polyketide synthase for enediyne biosynthesis. *FEBS Lett.* **2008**, *582*, 1097-1103.

274. Weber, T.; Laiple, K. J.; Pross, E. K.; Textor, A.; Grond, S.; Welzel, K.; Pelzer, S.; Vente, A.; Wohlleben, W., Molecular analysis of the kirromycin biosynthetic gene cluster revealed beta-alanine as precursor of the pyridone moiety. *Chem. Biol.* **2008**, *15* (2), 175-188.
275. Vicente, C. M.; Santos-Aberturas, J.; Payero, T. D.; Barreales, E. G.; de Pedro, A.; Aparicio, J. F., PAS-LuxR transcriptional control of filipin biosynthesis in *S. avermitilis*. *Appl. Microbiol. Biotechnol.* **2014**, *98* (22), 9311-24.
276. Gao, H.; Gruschow, S.; Barke, J.; Seipke, R. F.; Hill, L. M.; Orivel, J.; Yu, D. W.; Hutchings, M.; Goss, R. J. M., Filipins: the first antifungal "weed killers" identified from bacteria isolated from the trap-ant. *RSC Advances* **2014**, *4* (100), 57267-57270.
277. Gullo, V. P.; Zimmerman, S. B.; Dewey, R. S.; Hensens, O.; Cassidy, P. J.; Oiwa, R.; Omura, S., Factumycin, a new antibiotic (A40A): fermentation, isolation and antibacterial spectrum. *J. Antibiot.* **1982**, *35* (12), 1705-7.
278. Tong, Y.; Charusanti, P.; Zhang, L.; Weber, T.; Lee, S. Y., CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **2015**, *4* (9), 1020-9.
279. Huang, H.; Zheng, G.; Jiang, W.; Hu, H.; Lu, Y., One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta. Biochim. Biophys. Sin.* **2015**, *47* (4), 231-43.
280. Ogasawara, Y.; Katayama, K.; Minami, A.; Otsuka, M.; Eguchi, T.; Kakinuma, K., Cloning, sequencing, and functional analysis of the biosynthetic gene cluster of macrolactam antibiotic vicienistatin in *Streptomyces halstedii*. *Chem. Biol.* **2004**, *11* (1), 79-86.
281. Ogasawara, Y.; Kakinuma, K.; Eguchi, T., Involvement of glutamate mutase in the biosynthesis of the unique starter unit of the macrolactam polyketide antibiotic vicienistatin. *J. Antibiot.* **2005**, *58* (7), 468-72.
282. Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T., Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicienistatin biosynthesis. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (7), 1802-1807.
283. Shinohara, Y.; Kudo, F.; Eguchi, T., A Natural Protecting Group Strategy To Carry an Amino Acid Starter Unit in the Biosynthesis of Macrolactam Polyketide Antibiotics. *J. Am. Chem. Soc.* **2011**, *133* (45), 18134-18137.
284. Shinohara, Y.; Miyanaga, A.; Kudo, F.; Eguchi, T., The crystal structure of the amidohydrolase VinJ shows a unique hydrophobic tunnel for its interaction with polyketide substrates. *FEBS Lett.* **2014**, *588* (6), 995-1000.
285. Hong, H.; Samborsky, M.; Lindner, F.; Leadlay, P. F., An Amidohydrolase Provides the Missing Link in the Biosynthesis of Amino Marginolactone Antibiotics. *Angew. Chem. Int. Ed.* **2016**, *55* (3), 1118-1123.
286. Liew, C. W.; Nilsson, M.; Chen, M. W.; Sun, H.; Cornvik, T.; Liang, Z.-X.; Lescar, J., Crystal structure of the acyltransferase domain of the iterative polyketide synthase in enediyne biosynthesis. *J. Biol. Chem.* **2012**.
287. Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R., Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry* **2001**, *40* (51), 15464-15470.
288. Haydock, S. F.; Aparicio, J. F.; Molnar, I.; Schwecke, T.; Khaw, L. E.; Konig, A.; Marsden, A. F. A.; Galloway, I. S.; Staunton, J.; Leadlay, P. F., Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **1995**, *374* (2), 246-248.
289. Keatinge-Clay, A. T., A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **2007**, *14* (8), 898-908.
290. Siskos, A. P.; Baerga-Ortiz, A.; Bali, S.; Stein, V.; Mamdani, H.; Spiteller, D.; Popovic, B.; Spencer, J. B.; Staunton, J.; Weissman, K. J.; Leadlay, P. F., Molecular basis of Celmer's rules: stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases. *Chem. Biol.* **2005**, *12* (10), 1145-53.

291. Soehano, I.; Yang, L.; Ding, F.; Sun, H.; Low, Z. J.; Liu, X.; Liang, Z.-X., Insights into the programmed ketoreduction of partially reducing polyketide synthases: stereo- and substrate-specificity of the ketoreductase domain. *Org. Biomol. Chem.* **2014**, *12* (42), 8542-8549.
292. Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H., Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains. *ACS Chem. Biol.* **2009**, *4* (1), 41-52.
293. Olano, C.; Wilkinson, B.; Moss, S. J.; Brana, A. F.; Mendez, C.; Leadlay, P. F.; Salas, J. A., Evidence from engineered gene fusions for the repeated use of a module in a modular polyketide synthase. *Chem. Comm.* **2003**, (22), 2780-2.
294. Tatsuno, S.; Arakawa, K.; Kinashi, H., Analysis of modular-iterative mixed biosynthesis of lankacidin by heterologous expression and gene fusion. *J. Antibiot.* **2007**, *60* (11), 700-8.
295. Olano, C.; Wilkinson, B.; Sanchez, C.; Moss, S. J.; Sheridan, R.; Math, V.; Weston, A. J.; Brana, A. F.; Martin, C. J.; Oliynyk, M.; Mendez, C.; Leadlay, P. F.; Salas, J. A., Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tu4055: cluster analysis and assignment of functions. *Chem. Biol.* **2004**, *11* (1), 87-97.
296. Gaitatzis, N.; Silakowski, B.; Kunze, B.; Nordsiek, G.; Blocker, H.; Hofle, G.; Muller, R., The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem.* **2002**, *277* (15), 13082-90.
297. Chen, H.; Du, L., Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl. Microbiol. Biotechnol.* **2016**, *100* (2), 541-57.
298. Xu, W.; Zhai, G.; Liu, Y.; Li, Y.; Shi, Y.; Hong, K.; Hong, H.; Leadlay, P. F.; Deng, Z.; Sun, Y., An Iterative Module in the Azalomycin F Polyketide Synthase Contains a Switchable Enoylreductase Domain. *Angew Chem. Int. Ed. Engl.* **2017**, *56* (20), 5503-5506.
299. Shao, L.; Qu, X.-D.; Jia, X.-Y.; Zhao, Q.-F.; Tian, Z.-H.; Wang, M.; Tang, G.-L.; Liu, W., Cloning and characterization of a bacterial iterative type I polyketide synthase gene encoding the 6-methylsalicylic acid synthase. *Biochem. Biophys. Res. Commun.* **2006**, *345* (1), 133-139.
300. Van Lanen, S. G.; Oh, T. J.; Liu, W.; Wendt-Pienkowski, E.; Shen, B., Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. *J. Am. Chem. Soc.* **2007**, *129* (43), 13082-13094.
301. Lu, W.; Roongsawang, N.; Mahmud, T., Biosynthetic studies and genetic engineering of pactamycin analogs with improved selectivity toward malarial parasites. *Chem. Biol.* **2011**, *18* (4), 425-31.
302. Katz, L.; Baltz, R. H., Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (2-3), 155-76.
303. Cragg, G. M.; Newman, D. J., NATURAL PRODUCTS: A CONTINUING SOURCE OF NOVEL DRUG LEADS. *Biochim. Biophys. Acta* **2013**, *1830* (6), 3670-3695.
304. Gould, K., Antibiotics: from prehistory to the present day. *J. Antimicrob. Chemother.* **2016**, *71* (3), 572-575.
305. Doroghazi, J. R.; Buckley, D. H., Widespread homologous recombination within and between *Streptomyces* species. *The ISME journal* **2010**, *4* (9), 1136-43.
306. Low, Z. J.; Pang, L. M.; Ding, Y.; Cheang, Q. W.; Le Mai Hoang, K.; Thi Tran, H.; Li, J.; Liu, X.-W.; Kanagasundaram, Y.; Yang, L.; Liang, Z.-X., Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a *Streptomyces* strain isolated from mangrove sediment. *Sci. Rep.* **2018**, *8* (1), 1594.
307. Warnock, D. W., Fungal diseases: an evolving public health challenge. *Med. Mycol.* **2006**, *44* (8), 697-705.
308. Pospiech, A.; Neumann, B., A versatile quick-prep of genomic DNA from gram-positive bacteria. *Trends Genet.* **1995**, *11* (6), 217-8.
309. Koren, S.; Walenz, B. P.; Berlin, K.; Miller, J. R.; Bergman, N. H.; Phillippy, A. M.; Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **2017**, *27* (5), 722-736.

310. Walker, B. J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C. A.; Zeng, Q.; Wortman, J.; Young, S. K.; Earl, A. M., Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **2014**, *9* (11), e112963.
311. Wiegand, I.; Hilpert, K.; Hancock, R. E., Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **2008**, *3* (2), 163-75.
312. Lambert, R. J.; Pearson, J., Susceptibility testing: accurate and reproducible minimum inhibitory concentration (MIC) and non-inhibitory concentration (NIC) values. *J. Appl. Microbiol.* **2000**, *88* (5), 784-90.
313. Wang, X.-J.; Yan, Y.-J.; Zhang, B.; An, J.; Wang, J.-J.; Tian, J.; Jiang, L.; Chen, Y.-H.; Huang, S.-X.; Yin, M.; Zhang, J.; Gao, A.-L.; Liu, C.-X.; Zhu, Z.-X.; Xiang, W.-S., Genome Sequence of the Milbemycin-Producing Bacterium *Streptomyces bingchenggensis*. *J. Bacteriol.* **2010**, *192* (17), 4526-4527.
314. Lee, I.; Ouk Kim, Y.; Park, S. C.; Chun, J., OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **2016**, *66* (2), 1100-1103.
315. Wang, X.-J.; Zhang, B.; Yan, Y.-J.; An, J.; Zhang, J.; Liu, C.-X.; Xiang, W.-S., Characterization and analysis of an industrial strain of *Streptomyces bingchenggensis* by genome sequencing and gene microarray. *Genome* **2013**, *56* (11), 677-689.
316. McAlpine, J. B.; Bachmann, B. O.; Pirae, M.; Tremblay, S.; Alarco, A.-M.; Zazopoulos, E.; Farnet, C. M., Microbial Genomics as a Guide to Drug Discovery and Structural Elucidation: ECO-02301, a Novel Antifungal Agent, as an Example. *J. Nat. Prod.* **2005**, *68* (4), 493-496.
317. Oh, D.-C.; Scott, J. J.; Currie, C. R.; Clardy, J., Mycangimycin, a Polyene Peroxide from a Mutualist *Streptomyces* sp. *Org. Lett.* **2009**, *11* (3), 633-636.
318. Kim, D.-G.; Moon, K.; Kim, S.-H.; Park, S.-H.; Park, S.; Lee, S. K.; Oh, K.-B.; Shin, J.; Oh, D.-C., Bahamaolides A and B, Antifungal Polyene Polyol Macrolides from the Marine Actinomycete *Streptomyces* sp. *J. Nat. Prod.* **2012**, *75* (5), 959-967.
319. Narasimhachari, N.; Swami, M. B., Dermostatin: a revised hexaene structure. *J. Antibiot. (Tokyo)* **1970**, *23* (11), 566.
320. Villa, T. G.; Veiga-Crespo, P., *Antimicrobial Compounds. [electronic resource] : Current Strategies and New Alternatives*. Berlin, Heidelberg : Springer Berlin Heidelberg : Imprint: Springer, 2014.: 2014.
321. Lacoske, M. H.; Theodorakis, E. A., Spirotetronate Polyketides as Leads in Drug Discovery. *J. Nat. Prod.* **2015**, *78* (3), 562-575.
322. Wang, H.; Liu, N.; Xi, L.; Rong, X.; Ruan, J.; Huang, Y., Genetic Screening Strategy for Rapid Access to Polyether Ionophore Producers and Products in Actinomycetes. *Appl. Environ. Microbiol.* **2011**, *77* (10), 3433-3442.
323. Mori, T.; Takahashi, K.; Kashiwabara, M.; Uemura, D.; Katayama, C.; Iwadare, S.; Shizuri, Y.; Mitomo, R.; Nakano, F.; Matsuzaki, A., Structure of oxazolomycin, a novel β -lactone antibiotic. *Tetrahedron Lett.* **1985**, *26* (8), 1073-1076.
324. Liu, T.; Lin, X.; Zhou, X.; Deng, Z.; Cane, D. E., Mechanism of Thioesterase-Catalyzed Chain Release in the Biosynthesis of the Polyether Antibiotic Nanchangmycin. *Chem. Biol.* **2008**, *15* (5), 449-458.
325. Zhao, C.; Coughlin, J. M.; Ju, J.; Zhu, D.; Wendt-Pienkowski, E.; Zhou, X.; Wang, Z.; Shen, B.; Deng, Z., Oxazolomycin Biosynthesis in *Streptomyces albus* JA3453 Featuring an "Acyltransferase-less" Type I Polyketide Synthase That Incorporates Two Distinct Extender Units. *J. Biol. Chem.* **2010**, *285* (26), 20097-20108.
326. Sun, Y.; Zhou, X.; Dong, H.; Tu, G.; Wang, M.; Wang, B.; Deng, Z., A complete gene cluster from *Streptomyces nanchangensis* NS3226 encoding biosynthesis of the polyether ionophore nanchangmycin. *Chem. Biol.* **2003**, *10* (5), 431-41.
327. Oh, D.-C.; Poulsen, M.; Currie, C. R.; Clardy, J., Sceliphrolactam, a Polyene Macrocyclic Lactam from a Wasp-Associated *Streptomyces* sp. *Org. Lett.* **2011**, *13* (4), 752-755.

328. Ganis, P.; Avitabile, G.; Mechliniski, W.; Schaffner, C. P., Polyene macrolide antibiotic amphotericin B. Crystal structure of the N-iodoacetyl derivative. *J. Am. Chem. Soc.* **1971**, *93* (18), 4560-4564.
329. Rychnovsky Scott, D.; Richardson Timothy, I., Relative and Absolute Configuration of Filipin III. *Angewandte Chemie International Edition in English* **1995**, *34* (11), 1227-1230.
330. Kim, D. G.; Moon, K.; Kim, S. H.; Park, S. H.; Park, S.; Lee, S. K.; Oh, K. B.; Shin, J.; Oh, D. C., Bahamaolides A and B, antifungal polyene polyol macrolides from the marine actinomycete *Streptomyces* sp. *J. Nat. Prod.* **2012**, *75* (5), 959-67.
331. Udworthy, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S., Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (25), 10376-10381.
332. Aparicio, J. F.; Colina, A. J.; Ceballos, E.; Martin, J. F., The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimaricin. A new polyketide synthase organization encoded by two subclusters separated by functionalization genes. *J. Biol. Chem.* **1999**, *274* (15), 10133-9.
333. Hutchinson, E.; Murphy, B.; Dunne, T.; Breen, C.; Rawlings, B.; Caffrey, P., Redesign of polyene macrolide glycosylation: engineered biosynthesis of 19-(O)-perosaminyl-amphoteronolide B. *Chem. Biol.* **2010**, *17* (2), 174-82.
334. Brautaset, T.; Sekurova, O. N.; Sletta, H.; Ellingsen, T. E.; StrLm, A. R.; Valla, S.; Zotchev, S. B., Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. *Chem. Biol.* **2000**, *7* (6), 395-403.
335. Gil, J. A.; Campelo-Diez, A. B., Candicidin biosynthesis in *Streptomyces griseus*. *Appl. Microbiol. Biotechnol.* **2003**, *60* (6), 633-42.
336. Stroehrer, U. H.; Karageorgos, L. E.; Brown, M. H.; Morona, R.; Manning, P. A., A putative pathway for perosamine biosynthesis is the first function encoded within the rfb region of *Vibrio cholerae* O1. *Gene* **1995**, *166* (1), 33-42.
337. Aparicio, J. F.; Caffrey, P.; Gil, J. A.; Zotchev, S. B., Polyene antibiotic biosynthesis gene clusters. *Appl. Microbiol. Biotechnol.* **2003**, *61* (3), 179-88.
338. Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortes, J.; Rudd, B. A.; Staunton, J.; Leadlay, P. F., Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **2003**, *30* (8), 489-94.
339. Keatinge-Clay, A. T., A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **2007**, *14* (8), 898-908.
340. Keatinge-Clay, A., Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **2008**, *384* (4), 941-53.
341. Tietz, J. I.; Mitchell, D. A., Using Genomics for Natural Product Structure Elucidation. *Curr. Top. Med. Chem.* **2016**, *16* (15), 1645-94.
342. Bergey, D. H.; Whitman, W. B.; Goodfellow, M.; Kämpfer, P.; Busse, H.-J. r., Bergey's manual of systematic bacteriology. Vol. 5, Vol. 5. **2012**.
343. Carro, L.; Nouioui, I.; Sangal, V.; Meier-Kolthoff, J. P.; Trujillo, M. E.; Montero-Calasanz, M. d. C.; Sahin, N.; Smith, D. L.; Kim, K. E.; Peluso, P.; Deshpande, S.; Woyke, T.; Shapiro, N.; Kyrpides, N. C.; Klenk, H.-P.; Göker, M.; Goodfellow, M., Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential. *Sci. Rep.* **2018**, *8* (1), 525.
344. Brady, S. F., Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat. Protoc.* **2007**, *2* (5), 1297-305.
345. Dubeau, M.-P.; Ghinet, M. G.; Jacques, P.-É.; Clermont, N.; Beaulieu, C.; Brzezinski, R., Cytosine Deaminase as a Negative Selection Marker for Gene Disruption and Replacement in the Genus *Streptomyces* and Other Actinobacteria. *Appl. Environ. Microbiol.* **2009**, *75* (4), 1211-1214.
346. Luzhetskii, A. N.; Ostash, B. E.; Fedorenko, V. A., [Interspecies conjugation of *Escherichia coli*-*Streptomyces globisporus* 1912 using integrative plasmid pSET152 and its derivatives]. *Genetika* **2001**, *37* (10), 1340-7.

347. Lou, C.; Stanton, B.; Chen, Y.-J.; Munsky, B.; Voigt, C. A., Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* **2012**, *30* (11), 1137-1142.
348. Bai, C.; Zhang, Y.; Zhao, X.; Hu, Y.; Xiang, S.; Miao, J.; Lou, C.; Zhang, L., Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*. *Proceedings of the National Academy of Sciences* **2015**, *112* (39), 12181-12186.
349. Kirby, R., Chromosome diversity and similarity within the Actinomycetales. *FEMS Microbiol. Lett.* **2011**, *319* (1), 1-10.
350. Hirsch, A. M.; Alvarado, J.; Bruce, D.; Chertkov, O.; De Hoff, P. L.; Detter, J. C.; Fujishige, N. A.; Goodwin, L. A.; Han, J.; Han, S.; Ivanova, N.; Land, M. L.; Lum, M. R.; Milani-Nejad, N.; Nolan, M.; Pati, A.; Pitluck, S.; Tran, S. S.; Woyke, T.; Valdés, M., Complete Genome Sequence of Micromonospora Strain L5, a Potential Plant-Growth-Regulating Actinomycete, Originally Isolated from Casuarina equisetifolia Root Nodules. *Genome Announcements* **2013**, *1* (5).
351. Sanderson, S.; Valenti, M.; Gowan, S.; Patterson, L.; Ahmad, Z.; Workman, P.; Eccles, S. A., Benzoquinone ansamycin heat shock protein 90 inhibitors modulate multiple functions required for tumor angiogenesis. *Mol. Cancer Ther.* **2006**, *5* (3), 522-32.
352. Parenti, F.; Lancini, G., CHAPTER 27 - Rifamycins A2 - Finch, Roger G. In *Antibiotic and Chemotherapy (Ninth Edition)*, Greenwood, D.; Norrby, S. R.; Whitley, R. J., Eds. Content Repository Only! London, 2010; pp 326-333.
353. Netzker, T.; Schroeckh, V.; Gregory, M. A.; Flak, M.; Krespach, M. K. C.; Leadlay, P. F.; Brakhage, A. A., An Efficient Method To Generate Gene Deletion Mutants of the Rapamycin-Producing Bacterium *Streptomyces iranensis* HM 35. *Appl. Environ. Microbiol.* **2016**, *82* (12), 3481-3492.
354. Galm, U.; Wang, L.; Wendt-Pienkowski, E.; Yang, R.; Liu, W.; Tao, M.; Coughlin, J. M.; Shen, B., In vivo manipulation of the bleomycin biosynthetic gene cluster in *Streptomyces verticillus* ATCC15003 revealing new insights into its biosynthetic pathway. *J. Biol. Chem.* **2008**, *283* (42), 28236-45.
355. **!!! INVALID CITATION !!! {}.**
356. Flett, F.; Mersinias, V.; Smith, C. P., High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* **1997**, *155* (2), 223-9.
357. Muth, G.; Farr, M.; Hartmann, V.; Wohlleben, W., *Streptomyces ghanaensis* plasmid pSG5: nucleotide sequence analysis of the self-transmissible minimal replicon and characterization of the replication mode. *Plasmid* **1995**, *33* (2), 113-26.
358. Mahan, S. D.; Ireton, G. C.; Stoddard, B. L.; Black, M. E., Alanine-Scanning Mutagenesis Reveals a Cytosine Deaminase Mutant with Altered Substrate Preference. *Biochemistry* **2004**, *43* (28), 8957-8964.
359. Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R.; Weissman, J. S., Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **2009**, *324* (5924), 218-23.
360. Jeong, Y.; Kim, J.-N.; Kim, M. W.; Bucca, G.; Cho, S.; Yoon, Y. J.; Kim, B.-G.; Roe, J.-H.; Kim, S. C.; Smith, C. P.; Cho, B.-K., The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). *Nature Communications* **2016**, *7*, 11605.
361. Yi, J. S.; Kim, M. W.; Kim, M.; Jeong, Y.; Kim, E. J.; Cho, B. K.; Kim, B. G., A Novel Approach for Gene Expression Optimization through Native Promoter and 5' UTR Combinations Based on RNA-seq, Ribo-seq, and TSS-seq of *Streptomyces coelicolor*. *ACS synthetic biology* **2017**, *6* (3), 555-565.
362. Horbal, L.; Siegl, T.; Luzhetskyy, A., A set of synthetic versatile genetic control elements for the efficient expression of genes in Actinobacteria. *Sci. Rep.* **2018**, *8* (1), 491.
363. Santos, C. L.; Correia-Neves, M.; Moradas-Ferreira, P.; Mendes, M. V., A Walk into the LuxR Regulators of Actinobacteria: Phylogenomic Distribution and Functional Diversity. *PLoS One* **2012**, *7* (10), e46758.

364. Yu, Q.; Bai, L.; Zhou, X.; Deng, Z., Inactivation of the positive LuxR-type oligomycin biosynthesis regulators OlmRI and OlmRII increases avermectin production in *Streptomyces avermitilis*. *Chin. Sci. Bull.* **2012**, *57* (8), 869-876.
365. Santos-Aberturas, J.; Vicente, C. M.; Payero, T. D.; Martín-Sánchez, L.; Cañibano, C.; Martín, J. F.; Aparicio, J. F., Hierarchical Control on Polyene Macrolide Biosynthesis: PimR Modulates Pimaricin Production via the PAS-LuxR Transcriptional Activator PimM. *PLoS One* **2012**, *7* (6), e38536.
366. Laureti, L.; Song, L.; Huang, S.; Corre, C.; Leblond, P.; Challis, G. L.; Aigle, B., Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (15), 6258-63.
367. Xie, C.; Deng, J. J.; Wang, H. X., Identification of AstG1, A LAL family regulator that positively controls ansatrienins production in *Streptomyces* sp. XZQH13. *Curr. Microbiol.* **2015**, *70* (6), 859-64.
368. Sanchez, C.; Du, L.; Edwards, D. J.; Toney, M. D.; Shen, B., Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. *Chem. Biol.* **2001**, *8* (7), 725-38.
369. Zhang, B.; Tian, W.; Wang, S.; Yan, X.; Jia, X.; Pierens, G. K.; Chen, W.; Ma, H.; Deng, Z.; Qu, X., Activation of Natural Products Biosynthetic Pathways via a Protein Modification Level Regulation. *ACS Chem. Biol.* **2017**, *12* (7), 1732-1736.
370. Yan, X.; Zhang, B.; Tian, W.; Dai, Q.; Zheng, X.; Hu, K.; Liu, X.; Deng, Z.; Qu, X., Puromycin A, B and C, cryptic nucleosides identified from *Streptomyces alboniger* NRRL B-1832 by PPTase-based activation. *Synthetic and Systems Biotechnology* **2018**, *3* (1), 76-80.
371. Floss, H. G.; Yu, T.-W.; Arakawa, K., The biosynthesis of 3-amino-5-hydroxybenzoic acid (AHBA), the precursor of mC7N units in ansamycin and mitomycin antibiotics: a review. *The Journal Of Antibiotics* **2010**, *64*, 35.
372. Wilson, M. C.; Gulder, T. A. M.; Mahmud, T.; Moore, B. S., Shared biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the sare1259 encoded cytochrome P450. *J. Am. Chem. Soc.* **2010**, *132* (36), 12757-12765.
373. August, P. R.; Tang, L.; Yoon, Y. J.; Ning, S.; Müller, R.; Yu, T.-W.; Taylor, M.; Hoffmann, D.; Kim, C.-G.; Zhang, X.; Hutchinson, C. R.; Floss, H. G., Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *ansamycin* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem. Biol.* **1998**, *5* (2), 69-79.
374. Weissman, K. J., Polyketide stereocontrol: a study in chemical biology. *Beilstein J. Org. Chem.* **2017**, *13*, 348-371.
375. Soehano, I.; Yang, L.; Ding, F.; Sun, H.; Low, Z. J.; Liu, X.; Liang, Z. X., Insights into the programmed ketoreduction of partially reducing polyketide synthases: stereo- and substrate-specificity of the ketoreductase domain. *Org. Biomol. Chem.* **2014**, *12* (42), 8542-9.
376. K., P. S.; Jianting, Z.; J., A. A.; Madeline, E. D.; T., K.-C. A., Structural and functional studies of a trans-acyltransferase polyketide assembly line enzyme that catalyzes stereoselective α - and β -ketoreduction. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82* (9), 2067-2077.
377. Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C., Functionally distinct modules operate two consecutive $\alpha,\beta \rightarrow \beta,\gamma$ double-bond shifts in the rhizoxin polyketide assembly line. *Angew. Chem. Int. Ed. Engl.* **2010**, *49* (8), 1460-4.
378. Jana, M.; G., G. D. C.; R., A. C.; K., B. S.; Kathrin, S.; D., S. R.; Marianne, E.; Harald, G.; Gerhard, B.; Jörn, P., The Final Steps of Bacillaene Biosynthesis in *Bacillus amyloliquefaciens* FZB42: Direct Evidence for β,γ Dehydration by a trans-Acyltransferase Polyketide Synthase. *Angew. Chem. Int. Ed.* **2010**, *49* (8), 1465-1467.
379. Lohr, F.; Jenniches, I.; Frizler, M.; Meehan, M. J.; Sylvester, M.; Schmitz, A.; Gutschow, M.; Dorrestein, P. C.; König, G. M.; Schaberle, T. F., $\alpha,\beta \rightarrow \beta,\gamma$ double bond migration in coralopyronin A biosynthesis. *Chemical Science* **2013**, *4* (11), 4175-4180.

380. Yu, T. W.; Bai, L.; Clade, D.; Hoffmann, D.; Toelzer, S.; Trinh, K. Q.; Xu, J.; Moss, S. J.; Leistner, E.; Floss, H. G., The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (12), 7968-73.
381. Taft, F.; Brünjes, M.; Knobloch, T.; Floss, H. G.; Kirschning, A., Timing of the $\Delta^{10,12}$ - $\Delta^{11,13}$ Double Bond Migration During Ansamitocin Biosynthesis in *Actinosynnema pretiosum*. *J. Am. Chem. Soc.* **2009**, *131* (11), 3812-3813.
382. Dongho, L.; Kyeong, L.; Fu, C. X.; Tien, D. N.; K., B. S.; Mijeoung, L.; Chul, S. J.; Woncheol, K.; Kap, J. J.; Suk, L. J.; Choong-Hwan, L.; Jeong-Hyung, L.; Young-Soo, H.; Joon, L. J., Biosynthesis of the Heat-Shock Protein 90 Inhibitor Geldanamycin: New Insight into the Formation of the Benzoquinone Moiety. *ChemBioChem* **2006**, *7* (2), 246-248.
383. Rascher, A.; Hu, Z.; Buchanan, G. O.; Reid, R.; Hutchinson, C. R., Insights into the Biosynthesis of the Benzoquinone Ansamycins Geldanamycin and Herbimycin, Obtained by Gene Sequencing and Disruption. *Appl. Environ. Microbiol.* **2005**, *71* (8), 4862-4871.
384. Konishi, M.; Ohkuma, H.; Tsuno, T.; Oki, T.; VanDuyne, G. D.; Clardy, J., Crystal and molecular structure of dynemicin A: a novel 1,5-diyne-3-ene antitumor antibiotic. *J. Am. Chem. Soc.* **1990**, *112* (9), 3715-3716.
385. Davies, J.; Wang, H.; Taylor, T.; Warabi, K.; Huang, X.-H.; Andersen, R. J., Uncialamycin, A New Eneidyne Antibiotic. *Org. Lett.* **2005**, *7* (23), 5233-5236.
386. Yan, X.; Ge, H.; Huang, T.; Hindra; Yang, D.; Teng, Q.; Crnovčić, I.; Li, X.; Rudolf, J. D.; Lohman, J. R.; Gansemans, Y.; Zhu, X.; Huang, Y.; Zhao, L.-X.; Jiang, Y.; Van Nieuwerburgh, F.; Rader, C.; Duan, Y.; Shen, B., Strain Prioritization and Genome Mining for Eneidyne Natural Products. *mBio* **2016**, *7* (6).
387. Yan, X.; Chen, J.-J.; Adhikari, A.; Yang, D.; Crnovcic, I.; Wang, N.; Chang, C.-Y.; Rader, C.; Shen, B., Genome Mining of *Micromonospora yangpuensis* DSM 45577 as a Producer of an Anthraquinone-Fused Eneidyne. *Org. Lett.* **2017**, *19* (22), 6192-6195.
388. Zazopoulos, E.; Huang, K.; Staffa, A.; Liu, W.; Bachmann, B. O.; Nonaka, K.; Ahlert, J.; Thorson, J. S.; Shen, B.; Farnet, C. M., A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nat. Biotechnol.* **2003**, *21* (2), 187-90.
389. Liu, W.; Ahlert, J.; Gao, Q.; Wendt-Pienkowski, E.; Shen, B.; Thorson, J. S., Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model. *Proceedings of the National Academy of Sciences* **2003**, *100* (21), 11959-11963.
390. Luo, Y.; Huang, H.; Liang, J.; Wang, M.; Lu, L.; Shao, Z.; Cobb, R. E.; Zhao, H., Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nature Communications* **2013**, *4*, 2894.
391. Kang, H. S.; Charlop-Powers, Z.; Brady, S. F., Multiplexed CRISPR/Cas9- and TAR-Mediated Promoter Engineering of Natural Product Biosynthetic Gene Clusters in Yeast. *ACS synthetic biology* **2016**, *5* (9), 1002-10.
392. Chen, Y.; Yin, M.; Horsman, G. P.; Huang, S.; Shen, B., Manipulation of pathway regulation in *Streptomyces globisporus* for overproduction of the enediyne antitumor antibiotic C-1027. *The Journal of antibiotics* **2010**, *63* (8), 482-485.
393. Chen, Y.; Yin, M.; Horsman, G. P.; Shen, B., Improvement of the Eneidyne Antitumor antibiotic C-1027 Production by Manipulating its Biosynthetic Pathway Regulation in *Streptomyces globisporus*. *J. Nat. Prod.* **2011**, *74* (3), 420-424.
394. Gao, Q.; Thorson, J. S., The biosynthetic genes encoding for the production of the dynemicin enediyne core in *Micromonospora chersina* ATCC53710. *FEMS Microbiol. Lett.* **2008**, *282* (1), 105-114.
395. Sultana, A.; Kallio, P.; Jansson, A.; Wang, J. S.; Niemi, J.; Mantsala, P.; Schneider, G., Structure of the polyketide cyclase SnoL reveals a novel mechanism for enzymatic aldol condensation. *EMBO J.* **2004**, *23* (9), 1911-21.
396. Tang, M.-C.; Zou, Y.; Watanabe, K.; Walsh, C. T.; Tang, Y., Oxidative Cyclization in Natural Product Biosynthesis. *Chem. Rev.* **2017**, *117* (8), 5226-5333.

397. Schneider, D.; Schmidt, C. L., Multiple Rieske proteins in prokaryotes: Where and why? *Biochim. Biophys. Acta* **2005**, *1710* (1), 1-12.
398. Cohen, D. R.; Townsend, C. A., Characterization of an Anthracene Intermediate in Dynemicin Biosynthesis. *Angew. Chem. Int. Ed. Engl.* **2018**.
399. McCarthy, J. G.; Eisman, E. B.; Kulkarni, S.; Gerwick, L.; Gerwick, W. H.; Wipf, P.; Sherman, D. H.; Smith, J. L., Structural basis of functional group activation by sulfotransferases in complex metabolic pathways. *ACS Chem. Biol.* **2012**, *7* (12), 1994-2003.
400. Horsman, G. P.; Chen, Y. H.; Thorson, J. S.; Shen, B., Polyketide synthase chemistry does not direct biosynthetic divergence between 9-and 10-membered enediynes. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (25), 11331-11335.
401. Katherine, B.; A., T. C., Environmental Control of the Calicheamicin Polyketide Synthase Leads to Detection of a Programmed Octaketide and a Proposal for Eneidyne Biosynthesis. *Angew. Chem. Int. Ed.* **2012**, *51* (45), 11316-11319.
402. Belecki, K.; Townsend, C. A., Biochemical Determination of Enzyme-Bound Metabolites: Preferential Accumulation of a Programmed Octaketide on the Eneidyne Polyketide Synthase CalE8. *J. Am. Chem. Soc.* **2013**, *135* (38), 14339-14348.
403. Lau, J.; Fu, H.; Cane, D. E.; Khosla, C., Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry* **1999**, *38* (5), 1643-51.
404. Dunn, B. J.; Cane, D. E.; Khosla, C., Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* **2013**, *52* (11), 1839-41.
405. Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozono, I.; Kawahara, T.; Okada, M.; Awakawa, T.; Ito, T.; Asakawa, Y.; Ueki, M.; Takahashi, S.; Osada, H.; Wakimoto, T.; Ikeda, H.; Shin - ya, K.; Abe, I., Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chem. Int. Ed.* **2017**, *56* (7), 1740-1745.
406. Murphy, A. C.; Hong, H.; Vance, S.; Broadhurst, R. W.; Leadlay, P. F., Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chem. Commun.* **2016**, *52* (54), 8373-8376.
407. Stassi, D. L.; Kakavas, S. J.; Reynolds, K. A.; Gunawardana, G.; Swanson, S.; Zeidner, D.; Jackson, M.; Liu, H.; Buko, A.; Katz, L., Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (13), 7305-7309.
408. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G., Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (5), 1846-51.
409. Ruan, X.; Pereda, A.; Stassi, D. L.; Zeidner, D.; Summers, R. G.; Jackson, M.; Shivakumar, A.; Kakavas, S.; Staver, M. J.; Donadio, S.; Katz, L., Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J. Bacteriol.* **1997**, *179* (20), 6416-25.
410. Hao, H.; Cheng, G.; Iqbal, Z.; Ai, X.; Hussain, H. I.; Huang, L.; Dai, M.; Wang, Y.; Liu, Z.; Yuan, Z., Benefits and risks of antimicrobial use in food-producing animals. *Front. Microbiol.* **2014**, *5* (288).
411. Delfino, J.; Mathison, G. W.; Smith, M. W., Effect of Lasalocid on Feedlot Performance and Energy Partitioning in Cattle2. *J. Anim. Sci.* **1988**, *66* (1), 136-150.
412. Nagaraja, T. G.; Taylor, M. B., Susceptibility and resistance of ruminal bacteria to antimicrobial feed additives. *Appl. Environ. Microbiol.* **1987**, *53* (7), 1620-1625.
413. Zhou, S.; Wang, F.; Wong, E. T.; Fonkem, E.; Hsieh, T. C.; Wu, J. M.; Wu, E., Salinomycin: a novel anti-cancer agent with known anti-coccidial activities. *Curr. Med. Chem.* **2013**, *20* (33), 4095-101.
414. D'Alessandro, S.; Corbett, Y.; Ilboudo, D.; Misiano, P.; Dahiya, N.; Abay, S. M.; Habluetzel, A.; Grande, R.; Gismondo, M. R.; Decherig, K.; Koolen, K.; Sauerwein, R. W.; Taramelli, D.; Basilico, N.; Parapini, S., Salinomycin and other ionophores as new class of antimalarial drugs with transmission blocking activity. *Antimicrob. Agents Chemother.* **2015**.

415. Rausch, K.; Hackett, B.; Weinbren, N.; Reeder, S.; Sadovsky, Y.; Hunter, C.; Schultz, D. C.; Coyne, C.; Cherry, S., Screening bioactives reveals nanchangmycin as a broad spectrum antiviral active against Zika virus. *Cell reports* **2017**, *18* (3), 804-815.
416. Cui, H.; Ni, X.; Liu, S.; Wang, J.; Sun, Z.; Ren, J.; Su, J.; Chen, G.; Xia, H., Characterization of three positive regulators for tetramycin biosynthesis in *Streptomyces ahygroscopicus*. *FEMS Microbiol. Lett.* **2016**, *363* (12), fnw109-fnw109.
417. Jiang, C.; Wang, H.; Kang, Q.; Liu, J.; Bai, L., Cloning and characterization of the polyether salinomycin biosynthesis gene cluster of *Streptomyces albus* XM211. *Appl. Environ. Microbiol.* **2012**, *78* (4), 994-1003.
418. Sun, Y.; Hong, H.; Gillies, F.; Spencer, J. B.; Leadlay, P. F., Glycerol-S-acyl carrier protein as an intermediate in the biosynthesis of tetronate antibiotics. *ChemBioChem* **2008**, *9* (1), 150-6.
419. Gibson, D. G.; Benders, G. A.; Andrews-Pfannkoch, C.; Denisova, E. A.; Baden-Tillson, H.; Zaveri, J.; Stockwell, T. B.; Brownley, A.; Thomas, D. W.; Algire, M. A.; Merryman, C.; Young, L.; Noskov, V. N.; Glass, J. I.; Venter, J. C.; Hutchison, C. A., 3rd; Smith, H. O., Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **2008**, *319* (5867), 1215-20.
420. Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Csisny, A.; Whitwam, R. E.; Farnet, C. M.; Thorson, J. S., The Calicheamicin Gene Cluster and Its Iterative Type I Enediyne PKS. *Science* **2002**, *297* (5584), 1173-1176.
421. Thattai, M.; Burak, Y.; Shraiman, B. I., The origins of specificity in polyketide synthase protein interactions. *PLoS Comput. Biol.* **2007**, *3* (9), 1827-35.
422. Chen, S.; Xue, Y. Q.; Sherman, D. H.; Reynolds, K. A., Mechanisms of molecular recognition in the pikromycin polyketide synthase. *Chem. Biol.* **2000**, *7* (12), 907-918.
423. Haydock, S. F.; Aparicio, J. F.; Molnár, I.; Schwecke, T.; Khaw, L. E.; König, A.; Marsden, A. F. A.; Galloway, I. S.; Staunton, J.; Leadlay, P. F., Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **1995**, *374* (2), 246-248.

Appendix

Supplementary tables and figures for chapter 2

Table S 1: Primers and sequences used for study discussed in Chapter 2.

Primer name	Sequence (5' -> 3')	Purpose
Ctg1_20 sgRNA F	CATGCCATGGgtcgcgcatatcaccacggGTTTTAGAGCTAGAAATAGC	sgRNA amplification
Ctg1_20 sgRNA2 F	CATGCCATGGcccgtcgcgcatatcaccaGTTTTAGAGCTAGAAATAGC	
Ctg1_20 sgRNA3 F	CATGCCATGGcccgcctccgtggtgatatcGTTTTAGAGCTAGAAATAGC	
Ctg1_20 sgRNA4 F	CATGCCATGGcccgtggtgatatccggcgacGTTTTAGAGCTAGAAATAGC	
Ctg1_20 sgRNA5 F	CATGCCATGGcgcggacgtgttcgacccagGTTTTAGAGCTAGAAATAGC	
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	sgRNA sequencing
sgRNA check F	CACAGGAAACAGCTATGACC	
sgRNA check R2	TCAGCCCAGATCCCCGATCC	
sgRNA check R	ACCCCCATTCAAGAACAGC	
SceN left F	TCGTGGAAGGCACTAGAAGGTGCATGCCTACCTGCTGACC	Homologous recombination template construction for pCRISPR-Cas9-SceN
SceN left R	TTGCGCTGGAAGCGGTAGGTTGAGCTGCATGTCGAGGTA	
SceN right F	ACCTACCGCTTCCAGCGCAA	
SceN right R	GGTCGATCCCCGCATATAGGGCGAAGTCACCGTCGTGGT	
pcrispr-cas9 seq	TGAGGCTTGCAGGGAGTCAA	HDR template sequencing
pcrispr-cas9 seq R	CGTCGCTCTCTGGCAAAGCT	
SceN check F	GCTCGGCTCGGTGAAGTCGA	Validation of sceN deletion
SceN check R	AAGAGCACGCGGCCGGAGAT	
Boundary 1F	TTCCAGTTCTTTATGGCGGTC	Boundary between sceP and sceQ
Boundary 1R	GAAGAAGTCGTGGTCAACT	
Boundary 2F	CAGGAAATGGGCTTCGACTC	Boundary between sceQ and sceR
Boundary 2R	GAGACGATGCTGCCGATAT	
Boundary 3F	TTCGTGCTCTTCTCCTCCCTC	Boundary between sceR and sceS
Boundary 3R	GATGCCAGGGTGTACGAGAT	
SceQ-R left F40	TCGTGGAAGGCACTAGAAGGTGGAACCTGCACGAACTCAC	Homologous recombination template construction for pCRISPR-Cas9-SceQ-R
SceQ-R left R40	CGGAGCTTGTGCGCGGTGCCCTGGACTCCCAGTTCGTTGT	
SceQ-R right F20	GGCACCGCCGACAAGCTCCG	
SceQ-R right R40	GGTCGATCCCCGCATATAGGGATCGAGCCGAGCCACATGG	
SceQ-R check F2	AAATGGGCTTCGACTCGCTC	PCR screening of sceQ-R fusion mutant
SceQ-R check R2	GAAGCCGGACACCACGTCGA	
SceQ-R check F2	AAATGGGCTTCGACTCGCTC	Sequencing of PCR amplified fragment of sceQ-R fusion mutant
SceQ-R check R2	GAAGCCGGACACCACGTCGA	
SceQ-R spacer 1	gagtcagtgatggaagagc	protospacer sequence
SceQ-R spacer 2	agtccagtgatggaagacc	
gapdhp(EL) promoter	GCTGCTCCTTCGGTCGGACGTGCGTCTACGGGCACCTTACCGCA GCCGTCGGCTGTGCGACACGGACGGATCGGGCGAACTGGCCGA TGCTGGGAGAAGCGCGCTGCTGTACGGCGCGCACCGGGTGGCG AGCCCCTCGGCGAGCGGTGTGAACTTCTGTGAATGGCCTGTTC GGTTGCTTTTTTATACGGCTGCCAGATAAGGCTTGACGATCTG GGCGGTACCGCTATGATCGGGGCGTTCTGCAATTCTTAGTGC GAGTATCTGAAAGGGGATACGC	Promoter sequence for dual sgRNA synthetic construct
oop terminator	ACTCCATCTGGATTTGTTTCAGAACGCTCGGTTGCCGCCGGGCGT TTTTTA	terminator sequence for dual sgRNA synthetic construct

Figure S 1. Prediction of stereochemistry of β -hydroxyl group and olefin structures through KR fingerprint. (a) Geometry of olefin structure from dehydration by DH is dependent on chirality of β -hydroxyl group generated by upstream KR domain. (b) Comparison of KR fingerprint from *sce* PKS modules with A-Type and B-Type KR examples obtained from erythromycin PKS.

Figure S 2: Verification of sequence boundaries between *sceP* to *sceS*. (A) Three sets of primers were designed to amplify the sequence between *sceP* to *sceQ*, *sceQ* to *sceR* and *sceR* to *sceS*. (B) Band sizes of PCR products were in accordance to predicted DNA sizes based on *sce* gene cluster sequence. Sequences of the PCR products were confirmed via DNA sequencing.

Figure S 3. ^1H NMR spectrum of scilipholactam in $\text{DMSO}-d_6$, 500 MHz

Figure S 4. ^{13}C NMR spectrum of scilipholactam in $\text{DMSO}-d_6$, 500 MHz.

Figure S 9. 1D ^1H NOESY spectrum of sceliphrolactam ($\text{DMSO}-d_6$, 500 MHz) with selective excitation of H-15 to establish the geometry of olefin connecting H-15 and H-14.

Figure S 10. Mass spectrum of sceliphrolactam. The data was collected using Agilent G6540 QTOF (+ESI scan) Frag = 215.0V.

Figure S 11. Analysis of docking domains and effect of *sceR-Q* gene fusion on sceliphrolactam biosynthesis. Multiple sequence alignment of the docking domains of Sce PKS system. Each row shows an interacting head (H) – tail (T) pair, with the key residues involved in specific domain-domain interaction highlighted in red and green according to previous studies ^{292, 421}. The two residues connected by dash lines link are matching pairs that engage in charge-charge or hydrophobic interaction. The compatible H1-T1 type docking domain pair for *sceQ*-*SceR* are shown in the box. According to the classification of PKS docking domains ^{292, 421}, the paired docking domains for *SceO*/*SceQ*, *SceQ*/*SceR*, *SceR*/*SceS* belong to the H1-T1 type; whereas the docking domains for *SceN*/*SceO* and *SceS*/*SceT* belong to the H2-T2 type. The docking domains of *SceQ* and *SceR* share high sequence similarity with those of the *PikAII*/*PikAIII* pair from pikromycin biosynthesis ^{292, 422}, with the matching Asp/Arg pair and non-polar residues (Fig. 5A). The compatibility of the docking domains suggests that *SceQ* and *SceR* are able to interact with each other through the docking domains.

Table S 2: ¹H and ¹³C NMR data for sceliphrolactam (DMSO-*d*₆, 500 MHz)

no.	δ _H	δ _C	Type
1		172.4	C
2	4.99 s	93.7	CH
3		165.6	C
3-OH	13.68 s		
4	5.84 d (15.0)	121.6	CH
5	6.53 d (15.0)	138.2	CH
6		134.5	C
7	5.85 d (13.5)	135.8	CH
8	6.36 dd (13.5, 12.0)	(128.3) ^a	CH
9	5.42 dd (13.5, 7.5)	136.7	CH
10	4.04 dd (9.0, 7.5)	70.1	CH
10-OH	4.78 br.s		
11	3.79 br.d (9.0)	(74.5) ^a	CH
11-OH	4.78 br.s		
12	4.25 br.s	79.6	CH
12-OH	4.99 s		
13		(201.7) ^b	C
14	6.21 m	119.9	CH
15	6.63 m	146.5	CH
16	7.39 t (14.0)	125.5	CH
17	6.67 m	143.5	CH
18		135.3	C
19	6.24 m	135.9	CH
20	6.21 m	135.7	CH
21	6.19 m	127.4	CH
22	5.81 dd (15.0, 11.0)	131.4	CH
23	5.36 dd (15.0, 9.0)	138.1	CH
24	2.20 m	(39.3) ^a	CH

25a	3.04 m	(43.9) ^a	CH
25b	2.97 m		
25-NH	7.74 s		
26	1.80 s	12.0	CH ₃
27	1.60 s	12.0	CH ₃
28	1.00 d (6.0)	16.2	CH ₃

^aFrom HSQC correlation

^b From HBMC correlation

Supplementary tables and figures for chapter 3

Table S 3: Primers and sequences used in chapter 3

Primer name	Sequence (5' -> 3')	Purpose
Nanchang ctg20-33 sgRNA F	CATGCCATGGcgcgcccgcgtgaaccgcaGTTTTAGAGCTAGAAAT AGC	sgRNA amplification for <i>nanA1</i> KO
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	
oxazolomycin ctg6-179 sgRNA F	CATGCCATGGtcgctggcgtgcacaagagGTTTTAGAGCTAGAAAT AGC	sgRNA amplification for <i>ozmH</i> KO
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	
MjmE sgrna F sgRNA R	CATGCCATGGTGGACCGCCATCGTGAGCACGTTTTAGAGCT AGAAATAGC ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	sgRNA amplification for <i>mjmE</i> KO
MjmB sgrna F sgRNA R	CATGCCATGGcggggagcagatctgctcctGTTTTAGAGCTAGAAATA GC ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	sgRNA amplification for <i>mjmB</i> KO
sgRNA check F sgRNA check R2 sgRNA check R	CACAGGAAACAGCTATGACC TCAGCCCAGATCCCCGATCC ACCCCCATTCAAGAACAGC	sgRNA sequencing
Nanchang left F40 Nanchang left R40 Nanchang right F20	TCGTGGAAGGCACTAGAAGGTGAACCTGATCCTGGCCGAG GTGACATTGGATACGACGGGAGGTGACCCAGCAGTTCGTC CCCGTCGTATCCAATGTCAC	Homologous recombination template construction for pCRISPR- Cas9- <i>nanA1</i>
Nanchang right R40 oxazolo left F40 oxazolo left R40 oxazolo right F20	GGTCGATCCCCGCATATAGGGCTGATCCCGAAGAACTCCG TCGTGGAAGGCACTAGAAGGGAACATCGGGTCGAAGGAGG GCTGCGCTATCTGTACGTGCAACAGCTTCCTCGCCCACTA GCACGTACAGATAGCGCAGC	Homologous recombination template construction for pCRISPR- Cas9- <i>ozmH</i>
oxazolo right R40 mjmE left F40 mjmE left R40 mjmE right F20	GGTCGATCCCCGCATATAGGCGGCGTGGTCTATCTCGAGA TCGTGGAAGGCACTAGAAGGTCCACGGTCTGACCCAGCTC GGTGCGCTGTGCTGGAAGAGAGTTCTGGTCCGCCGTCGA A TCTTCCAGCGACAGCGCACC	Homologous recombination template construction for pCRISPR- Cas9- <i>mjmE</i>
mjmE right R40 mjmB left F40 mjmB left R40 mjmB right F20	GGTCGATCCCCGCATATAGGCAAGATGGTGATGGCGCTGC TCGTGGAAGGCACTAGAAGGTGAAGTCAACCTCGGTCCAC CGGAGCTTGTGCGCGGTGCCCTGGACTCCCAGTTCGTTGT GGCACCGCCGACAAGCTCCG	Homologous recombination template construction for pCRISPR- Cas9- <i>mjmB</i>
mjmB right R40 pcrispr-cas9 seq	GGTCGATCCCCGCATATAGGGATCGAGCCGAGCCACATGG TGAGGCTTGACAGGGAGTCAA	HDR template sequencing
pcrispr-cas9 seq R Nan check F	CGTCGCTCTCTGGCAAAGCT GTACTGCCGGATCGGTCCAC	Validation of <i>nanA1</i> deletion
Nan check R Oxa check F	CATGCCGGTCAACGAGGCGT GAGTCGATGCCGTAGTGCTC	Validation of <i>ozmH</i> deletion
Oxa check R mjmE check F	CGATGAACTCCTGCGGCT GCGATTCCAGGTCTTCCTGT	Validation of <i>mjmE</i> deletion
mjmE check R mjmB check F	ATGTGCACACCATCCTGGAG GTGTCTCGTTCGGGATCAG	Validation of <i>mjmB</i> deletion
mjmB check R	GGCCTTCTCATGCCTGTTCT	

Table S 4: Secondary metabolite gene clusters in *Streptomyces* sp. SD50 genome

BGC	Position		Type/Product
	From	To	
Cluster 1	122918	143841	Terpene
Cluster 2	388501	408890	Nucleoside
Cluster 3	430122	478446	T1pks
Cluster 4	586696	637027	T1pks
Cluster 5	638528	648902	Bacteriocin
Cluster 6	667666	689600	Terpene
Cluster 7	853374	957226	Otherks-T1pks-Nrps
Cluster 8	1030942	1052012	Terpene
Cluster 9	1062669	1173121	Ladderane-Nrps
Cluster 10	1185319	1258492	Otherks-Nrps
Cluster 11	1278724	1334443	T1pks-Nrps
Cluster 12	1357922	1380669	Lantipeptide
Cluster 13	1474504	1517260	Other
Cluster 14	1604525	1625685	Terpene
Cluster 15	1648072	1659007	Butyrolactone
Cluster 16	1940080	1962353	Terpene
Cluster 17	2081381	2176350	T1pks-Nrps
Cluster 18	2513766	2540521	Terpene
Cluster 19	2603342	2625864	Lasso peptide
Cluster 20	2693123	2735656	T2pks
Cluster 21	3073481	3084812	Bacteriocin
Cluster 22	3109297	3158215	Nrps
Cluster 23	3194559	3242442	T1pks
Cluster 24	3412148	3424175	Siderophore
Cluster 25	3694864	3715937	Terpene
Cluster 26	4077349	4120090	Other
Cluster 27	4723209	4754553	Linaridin-Siderophore
Cluster 28	6802456	6824967	Lasso peptide
Cluster 29	7062010	7104450	Arylpolyene-Ladderane
Cluster 30	7104662	7159690	Nrps

Cluster 31	7155041	7237573	T2 PKS - angucycline type compound
Cluster 32	7296121	7320947	Lantipeptide
Cluster 33	7330892	7359517	Lantipeptide
Cluster 34	7952919	7981590	Lantipeptide
Cluster 35	8155548	8196588	Otherks
Cluster 36	8614363	8624767	Ectoine
Cluster 37	8775756	8789857	Siderophore
Cluster 38	8883765	9018151	Nanchangmycin
Cluster 39	9238288	9272082	T1pks-Nrps
Cluster 40	9436066	9535378	T1pks cluster encoding polyether compound
Cluster 41	9665751	9677027	Terpene
cluster 42	9677922	9748929	transATpks-NRPS
cluster 43	9971965	10017381	T1pks
cluster 44	10052934	10159414	Oxazolomycin
cluster 45	10193304	10256113	Nrps
cluster 46	10352306	10487208	T1pks cluster encoding spirotetronate compound
cluster 47	10644012	10707030	Bacteriocin-T1 pks
cluster 48	10896943	10943203	Otherks
cluster 49	10965266	11302609	Meilingmycin
cluster 50	11323061	11490171	Meijiemycin

Figure S 12: Inactivation of *nanA1* gene. Diagnostic PCR was performed to select for mutants with deletion of 751 bases from *nanA1* gene.

Disruption of *ozmH* gene

Figure S 13: Inactivation of *ozmH* gene. Diagnostic PCR was performed to select for mutants with deletion of 692 bp from *ozmH* gene.

Disruption of *mjmB* gene

Figure S 14: Inactivation of *mjmB* gene. Diagnostic PCR was performed to select for mutants with deletion of 926 bases from *mjmB* gene.

Figure S 15: Inactivation of *mjmE* gene. Diagnostic PCR was performed to select for mutants with deletion of 772 bases from *mjmE* gene.

Table S 5: ^1H and ^{13}C NMR spectroscopic data of meijiemycin.

Position	δ_{H} (multiplicity, J in Hz)		δ_{C}
	in DMSO- d_6	in pyridine- d_5	in DMSO- d_6
1			168.0
2	5.77 (d, 15.1)	6.36 (d, 15.6)	124.7
3	6.72 (ddd, 15.1, 7.0, 6.7)	7.51 (ddd, 15.6, 7.0, 6.7)	145.2
4	Ha: 2.39 (m) Hb: 2.24 (m)	Ha: 2.82 (ddd, 15.8, 7.5, 7.0) Hb: 2.56 (ddd, 15.8, 6.7, 4.5)	38.5
5	4.68 (dd, 6.9, 5.8)	5.23 (dd, 7.5, 4.5)	67.4
6			141.3
7	6.30 (br d, overlapped)	6.62 (br d, 11.0)	121.6
8	6.12 (dd, 11.2, 11.2)	6.43 (dd, 11.0, 11.0)	122.3
9	5.88 (dd, 11.6, 11.2)	6.08 (dd, 11.6, 11.0)	129.1
10	6.51 (dd, 15.1, 11.6)	6.91 (dd, 15.4, 11.6)	125.3
11	5.73 (dd, 15.1, 6.3)	6.12 (dd, 15.4, 7.0)	139.2
12	2.26 (m)	2.65 (m)	43.7

13	3.63 (m)	4.41 (m)	70.8
14	Ha: 1.38 (m)	Ha: 2.07 (m)	37.1
	Hb: 1.22 (m)	Hb: 1.87 (m)	
15	3.96 (m)	4.75 (m)	68.9
16	1.78 (m)	2.40 (m)	42.2
17	3.31 (br d, 7.6)	3.87 (br d, 7.9)	78.8
18	1.75 (m)	2.07 (m)	34.5
19	3.45 (br d, 9.2)	3.94 (br d, 9.2)	74.1
20	1.38 (m)	1.76 (m)	36.3
21	Ha: 1.67 (m)	Ha: 2.06 (m)	32.9
	Hb: 0.98 (m)	Hb: 1.25 (m)	
22	Ha: 1.34 (m)	1.22~1.35	26.5
	Hb: 1.16 (m)	1.22~1.35	
23	1.33 (2H, m)	1.36 (2H, m)	29.8
24	1.33 (2H, m)	1.36 (2H, m)	29.8
25	1.96 (2H, overlapped)	2.02 (2H, overlapped)	32.7
26	5.41 (m)	5.58 (dt, 15.0, 7.5)	132.4
27	5.41 (m)	5.71 (dt, 15.0, 7.5)	127.3
28	Ha: 2.09 (m)	Ha: 2.45 (m)	41.3
	Hb: 2.03 (m)	Hb: 2.41 (m)	
29	3.62 (m)	4.21 (m)	69.4
30	1.45 (2H, m)	1.96 (2H, m)	44.6
31	3.77 (m)	4.52 (m)	67.6
32	1.49 (2H, m)	1.90 (2H, m)	45.8
33	4.20 (m)	5.16 (m)	66.3
34	Ha: 1.67 (m)	Ha: 2.17 (dd, 11.0, 4.9)	46.6
	Hb: 1.58 (m)	Hb: 2.07 (dd, overlapped)	
35			97.6
36	Ha: 1.90 (m)	Ha: 2.73 (m)	44.5
	Hb: 1.19 (m)	Hb: 1.96 (m)	
37	3.96 (m)	5.23 (m)	65.9
38	1.95 (dd, overlapped)	2.98 (dd, 10.5, 10.5)	58.1
39	3.81 (m)	4.93 (m)	66.5
40	Ha: 1.86 (m)	Ha: 2.46 (m)	38.9
	Hb: 1.52 (m)	Hb: 2.39 (m)	
41	4.32 (m)	5.13 (m)	76.1
42	5.64 (m)	6.10 (dt, 15.0, 6.3)	134.6
43	6.30 (dd, overlapped)	6.74 (dd, 15.0, 10.7)	133.1
44	6.22~6.30	6.25 (dd, overlapped)	131.1~133.9
45	6.22~6.30	6.21~6.40	131.1~133.9
46	6.22~6.30	6.21~6.40	131.1~133.9
47	6.22~6.30	6.21~6.40	131.1~133.9
48	6.22~6.30	6.21~6.40	131.1~133.9
49	6.22~6.30	6.21~6.40	131.1~133.9
50	6.22~6.30	6.21~6.40	131.1~133.9

51	6.31 (dd, overlapped)	6.21~6.40	133.9
52	6.13 (dd, 15.1, 10.7)	6.14 (dd, overlapped)	131.5
53	5.75 (m)	5.69 (m)	130.7
54	1.75 (d, 7.0)	1.65 (d, 7.0)	18.7
55	1.76 (br s)	2.05 (br s)	18.8
56	0.98 (d, 6.7)	1.27 (d, 6.7)	17.2
57	0.69 (d, 6.3)	0.98 (d, 6.6)	12.1
58	0.87 (d, 6.7)	1.26 (d, 6.7)	11.4
59	0.70 (d, 6.3)	0.85 (d, 6.3)	15.8
60			174.7
1'	4.36 (br s)	4.95 (br s)	98.9
2'	3.63 (dd, overlapped)	4.39 (br s)	70.6
3'	3.18 (dd, overlapped)	3.81 (dd, overlapped)	73.3
4'	2.46 (dd, overlapped)	3.30 (dd, 10.2, 10.2)	54.7
5'	3.05 (m)	3.59 (m)	72.2
6'	1.13 (d, 6.0)	1.46 (d, 5.4)	18.6

Figure S 16: ^1H NMR spectrum (400 MHz, in $\text{DMSO}-d_6$) of meijiemycin.

Figure S 17: ^{13}C NMR spectrum (400 MHz, in DMSO-d_6) of meijiemycin.

Figure S 18: ^1H - ^1H COSY NMR spectrum (400 MHz, in DMSO-d_6) of meijiemycin.

Figure S 19: HSQC NMR spectrum (400 MHz, in DMSO- d_6) of meijiemycin.

Figure S 20: HMBC NMR spectrum (500 MHz, in DMSO-*d*₆) of meijiemycin.

Figure S 21: NOESY NMR spectrum (400 MHz, in DMSO- d_6) of meijiemycin.

Figure S 22: ^1H NMR spectrum (400 MHz, in $\text{C}_5\text{D}_5\text{N}$) of meijiemycin.

Figure S 23: ^1H - ^1H COSY NMR spectrum (400 MHz, in $\text{C}_5\text{D}_5\text{N}$) of meijiemycin.

Figure S 24: High resolution mass spectrum of meijiemycin.

Table S 6: Genes involved in meijiemycin biosynthesis and their putative functions

Gene	Size (amino acid)	Closest homolog [source]	Identity/Similarity (%)	proposed function	UniProt ID
<i>ctg1-17</i>	335	NADPH:quinone reductase [<i>Streptomyces griseochromogenes</i>]	85/93	unknown	A0A1B1AY81
<i>mjmR</i>	224	LuxR family transcriptional regulator [<i>Streptomyces bingchenggensis</i> BCW-1]	97/98	Transcriptional regulator	D7C2C5
<i>mjmSI</i>	345	GDP-mannose 4,6-dehydratase [<i>Streptomyces bingchenggensis</i> BCW-1]	98/99	GDP-mannose 4,6-dehydratase	D7C2C4
<i>mjmB</i>	9637	Beta-ketoacyl synthase [<i>Streptomyces</i> sp. NBS 14/10]	95/96	Polyketide synthase	A0A229GXM4
<i>mjmOII</i>	63	Ferredoxin-1 [<i>Streptomyces</i> sp. NBS 14/10]	98/98	Electron transfer	A0A229GYB0
<i>mjmOI</i>	391	Cytochrome P450 [<i>Streptomyces bingchenggensis</i> BCW-1]	98/99	Carboxylase	D7C2C0
<i>mjmSIII</i>	467	MGT family glycosyltransferase [<i>Streptomyces</i> sp. NBS 14/10]	99/100	Glycosyltransferase	A0A229GX50
<i>mjmU</i>	35	Flavin oxidoreductase (<i>Pandoraea</i> sp. ISTKB)	57/61	unknown	A0A1E3LP46
<i>mjmC</i>	7495	Polyketide synthase (<i>Streptomyces</i> sp. NBS 14/10)	95/96	Polyketide synthase	A0A229GX27
<i>mjmD</i>	1748	Uncharacterized protein (<i>Streptomyces</i> sp. NBS 14/10)	96/97	Polyketide synthase	A0A229GXC1
<i>mjmE</i>	9157	Modular polyketide synthase [<i>Streptomyces bingchenggensis</i> BCW-1]	96/97	Polyketide synthase	D7BRE1
<i>mjmF</i>	3633	Modular polyketide synthase [<i>Streptomyces bingchenggensis</i> BCW-1]	93/95	Polyketide synthase	D7BRE0
<i>mjmG</i>	1371	RifB protein [<i>Streptomyces bingchenggensis</i> BCW-1]	95/97	Polyketide synthase	D7BRD9
<i>mjmH</i>	431	Beta-ketoacyl synthase [<i>Streptomyces</i> sp. NBS 14/10]	77/84	Polyketide synthase	A0A229G640
<i>mjmI</i>	3589	Polyketide synthase [<i>Streptomyces</i> sp. NBS 14/10]	95/96	Polyketide synthase	A0A229H3V3
<i>mjmSII</i>	351	Perosamine synthetase [<i>Streptomyces bingchenggensis</i> BCW-1]	97/98	Perosamine synthetase	D7BRD5
<i>mjmA</i>	9914	Short-chain dehydrogenase [<i>Streptomyces</i> sp. NBS 14/10]	94/96	Polyketide synthase	A0A229H394
<i>ctg1-34</i>	136	Glyoxalase [<i>Streptomyces</i> sp. NBS 14/10]	95/97	unknown	A0A229H307

A

B

	Active Site		Substrate specificity motif	
mjmATL	DDAELLHHTAYTQPALFAHVALFRLLESWGTTADFVGHSGIGAVAAAHVAGALSADAC	...	PASVVVAGDEGAVEDVVAHFADSGRRTKRLVSHAFHSPRMDAMLDFRQVVTLSITAP	mjmATL
mjmAT1	-DADALDRTEYAQPALFALEVALFRLAESWGIVPDLLVGHSGIGELAAAHVAGVLSLDAC	...	PSSVVLSDGETVADIAAHFEARGRKATRLVSHAFHSPIMAPMLADFRAAVERLAQEP	mjmAT1
mjmAT2	AEAEALLDRTEFTQPALFAVEVALFRLLESWGITPRFVGHSGIGELAAAHVAGVLSLDAC	...	PEAAVITAGDESAVRDVADEIAARGRKATRLVSHAFHSPIMDPMDDFRTVAEGLSYHAP	mjmAT2
mjmAT3	AEAEALLDRTEFTQPALFAVEVALFRLLESWGITPDYLVGHSGIGELAAAHVAGVLSLDAC	...	PGSVVVS GAERQVLEIADRMAALGRKTSRLVSHAFHSPIMDPMDDFRTVAEGLSYQPP	mjmAT3
mjmAT4	-DEGLLNRTTEFAQPALFAVEVALFRLLESWGIPDYLGHSGIGELAAAHVAGVLSLDAC	...	PGSVVVS GAEPQVLEIADRMAALGRKTSRLVSHAFHSPIMDPMDDFRTVAEGLSYRPP	mjmAT4
mjmAT5	PEAALLDRITAYAPALFAVEVALFRLLESWGIPDFLAGHSGIGELAAAHVAGVLSLDAC	...	PGSLVVS GAETALEIAERFAAQRKTSRLVSHAFHSPIMAPMLNEFRTVAEGLSYQPP	mjmAT5
mjmAT6	PEAEALLDITGTQPALFAVEVALFRLLESWGIPDFLAGHSGIGELAAAHVAGVLSLDAC	...	PSAVVVGDEDAVAEIEAHFAGEGRKTKRLVSHAFHSPRMEPMLEEFRGVVAELSPQAP	mjmAT6
mjmAT7	PEAEALLDITGTQPALFAVEVALFRLLESWGIPDFLAGHSGIGELAAAHVAGVLSLDAC	...	PSALVLSGDEAVLAVAHFRLGRKTKRLVSHAFHSPRMDAMLDGFRVAQGLDYAAP	mjmAT7
mjmAT8	-GAPSLERVDVQPVSWAVMVS LAALWRS CGVEPDVGHSGIGELAAAHVAGVLSLDAC	...	TRSVVVS GDEALDELIGELTGEVIRIRIAVDYASHSAHVEELHEELLERLAPLSPRTA	mjmAT8
mjmAT9	AEAEALLDRITGTQPALFAVEVALFRLLESWGITADHLTGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITSGDEAVVLEIAGRFAGEGRKTRRLTVSHAFHSPIMEPMLAEFRQVAGLSYEAP	mjmAT9
mjmAT10	AEAEALLDRITGTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PTATVITAGDEAVARIAARLESRGRKTRRLTVSHAFHSPIMEPMLPEFATVAEHLAYQEP	mjmAT10
mjmAT11	AEAEALLDETAWTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITSGDEAVVLEIAARLESRGRKTRRLTVSHAFHSPIMEPMLAEFRQVAGLSYEAP	mjmAT11
mjmAT12	PEAEALLDETGTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITAGDETAVLEIAARYEEQGRKTRRLTVSHAFHSPIMDTMLAEFAEVARGISYEAP	mjmAT12
mjmAT13	ADAALLDETGTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITAGEEAQVQVAARLEAEGRKTRRLTVSHAFHSPIMDPMDDFRTVAEGLSYQEP	mjmAT13
mjmAT14	ADAALLDETGTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITAGDETAVSEIAARFEGEGRKTRRLTVSHAFHSPIMEPMLDGFRRVARDISYEAP	mjmAT14
mjmAT15	VEAGLLDETGTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PSSVVVAGEEAAVLEVAAMFEGELGRKTRRLTVSHAFHSPIMDPMDDFRTVAEGLSYQEP	mjmAT15
mjmAT16	VEAGLLDETGTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITAGDEAVVLEVAARFEAEGRKTRRLTVSHAFHSPIMDPMDDFRTVAEGLSYQEP	mjmAT16
mjmAT17	-GAPSLERVDVQPVSWAVMVS LAALWRS CGVEPDVGHSGIGELAAAHVAGVLSLDAC	...	PRSVVVS GDEALDELIGELTGEVIRIRIAVDYASHSAHVEELHEELLERLAPLSPRTA	mjmAT17
mjmAT18	-GAPSLERVDVQPVSWAVMVS LAALWRS CGVEPDVGHSGIGELAAAHVAGVLSLDAC	...	PRSVVVS GDEALDELIGELTGEVIRIRIAVDYASHSAHVEELHEELLERLAPLSPRTA	mjmAT18
mjmAT19	-GAPSLERVDVQPVSWAVMVS LAALWRS CGVEPDVGHSGIGELAAAHVAGVLSLDAC	...	PRSVVVS GDEALDELIGELTGEVIRIRIAVDYASHSAHVEELHEELLERLAPLSPRTA	mjmAT19
mjmAT20	AEAEALLDETAWTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITSGDEAVVLEIAGRFAGEGRKTRRLTVSHAFHSPIMEPMLAEFRQVAGLSYEAP	mjmAT20
mjmAT21	-GAPSLERVDVQPVSWAVMVS LAALWRS CGVEPDVGHSGIGELAAAHVAGVLSLDAC	...	PGSVVVS GDAQALDELIGELTGEVIRIRIAVDYASHSAHVEELHEELLERLAPLSPRTA	mjmAT21
mjmAT22	-DQSLDRITMYAQALFALEVALFRLLESWGTRPDYLLGHSGIGELAAAHVAGVLSLDAC	...	PDVSVVS GDEAVVLEIAARFEAEGRKTRRLTVSHAFHSPIMEPMLAEFRQVAGLSYEAP	mjmAT22
mjmAT23	-DDQILLDRITDFAQALFALEVALFRLLESWGTRPDYLLGHSGIGELAAAHVAGVLSLDAC	...	PRSVVVS GDEAVVLEIAARFEAEGRKTRRLTVSHAFHSPIMEPMLAEFRQVAGLSYEAP	mjmAT23
mjmAT24	-DQALLDRITDFAQALFALEVALFRLLESWGTRPDYLLGHSGIGELAAAHVAGVLSLDAC	...	PSSVVVS GDEAVVLEIAARFEAEGRKTRRLTVSHAFHSPIMEPMLAEFRQVAGLSYEAP	mjmAT24
mjmAT25	AEAEALLDETGTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITSGDEAVVLEIAARFEAEGRKTRRLTVSHAFHSPIMEPMLAEFRQVAGLSYEAP	mjmAT25
mjmAT26	PEAEALLDRITGTQPALFAVEVALFRLLESWGIVPDYLAGHSGIGELAAAHVAGVLSLDAC	...	PRSVVITAGDEAVVLEIAARFEAEGRKTRRLTVSHAFHSPIMEPMLAEFRQVAGLSYEAP	mjmAT26

Figure S 25: AT domain analysis of Mjm PKS. (A) Precursor units utilized by acyltransferase as substrate for chain elongation. (B) Analysis of AT domains show that all of them possess a GHSXG motif⁴²³ at the active site. Malonyl coA specificity is determined by HAFH motif in red font while YASH motif in blue font selects for methylmalonyl CoA as substrate.³³⁸

Figure S 26: KR domain analysis of Mjm PKS. (A) KR exerts stereocontrol over the geometry of hydroxyl and succeeding olefin moieties. A-type KR reduces β -keto group into a S-hydroxyl group that is subsequently dehydrated into a *cis*-olefin moiety, while B-type KR reduces β -keto group into a R-hydroxyl group that is subsequently dehydrated into a *trans*-olefin moiety. A2-type KR is differentiated from A1-type KR by presence of histidine residue in the catalytic region (depicted by dark blue column), while B2-type KR is differentiated from B1-type KR by presence of proline in the catalytic region (depicted by green column). (B) KR domains of Mjm PKS which possesses a tryptophan residue (in red) in their catalytic region were annotated as Type A-type KRs, while those that possesses a LDD loop were annotated as B-type KRs.³³⁹ MjmKR23 and MjmKR24, which contain a GDS motif in place of a LDD loop and lack a tryptophan residue (highlighted in yellow) in their catalytic regions, were determined to be A-type KRs through NMR spectral analysis of meijiemycin.

HXXXGXXXXP		GYXGYXPF	
motif		motif	
mjmDH1	VLS--TTAQPWLAQHVVAGQVLF	GTGTFVELAVRAGDEADQGVRELTITAPLALPEHG	***QWAYARTWPPADAVFLDLGGFYERRS-AGFDGYGAPGLRVLRALWQTEDAVLAEVALPEH
mjmDH2	RLS--LRTHPWLAQHVVAGVLLF	GTGTFVELAVRAGDEADQGVRELTITAPLVLPERGA	***DTAFEAQAQWPTTGATFIEDLGLDYLRA-ELGFGYGLPLGLRAAWRRGAEVFAEVALPEG
mjmDH3	RLSPRLRSHAWLAQHRVSRVLF	GTGTFVELAVRAGDEADQGVRELTITAPLVLPERGA	***GHWGFWVWPPEAGAGVPLAGCYERRA-ELGFGYGLPLGLRAAWRRGQDVFAEVALPDA
mjmDH4	RLS--TRSHPWLAQHRVSRVLF	GTGTFVELAVRAGDEADQGVRELTITAPLVLPERGS	***DAMPATQWPPGATGVALLDGLFYERRA-AGFDGYGAPGLRAAWRLGELFAEVALPER
mjmDH5	RLS--VGSHPWLAQHVVGRVLF	GTGTFLELATRAADQGVCDRVEELTIAAPLVLPQGA	***ATATPLVWPPEASAEHIVSDCYDRA-DJGFGYGLPLGLRAAWRRGDELFAEVALPER
mjmDH6	RLS--LRTHPWLAQHVRSVTILPA	GTGTFLELATRAADQGVCDRVEELTIAAPLALPEHG	***DIAFTDAQWPPADGAPLALDGLDYLRYRQAGLEYGFGYGLPLGLRAAWRRGADTFAEVALEPS
mjmDH13	RLS--LTTHPWLAQHVVGQVCF	GTGTFLELVVHAADQAGCRIDELSTIAAPVLPERGA	***VADFDTTAWPPHAGTDLGLYEHYER-RAJGFGYAPGLGLRAVWRGAEVFAEVALPED
mjmDH14	RLS--VATHPWLAQHAAGGVVLF	GTGTFLELAIRAGDQVGGCGVEELTIAAPVLPERGA	***---SQSGAWPPPEATIEVDLGGFYERRA-EGFGYGVFGYGLRAVWRREDEIFAEVVLDEQ
mjmDH15	RLS--LRSHPWLAQHAVGVVLF	GTGTFLELAIRAGDQGVCDRVEELTIAAPVLPERDA	***PAMGTWPPTEATFVLELDGLGYKFFRA-EGFGYGLPLGLRAVWRQREDEIFAEVALPEQ
mjmDH16	RLS--LRSHPWLAQHAVGVVLF	GTGTFLELAIRAGDQVGCRIEETLAEPLVLPERGA	***FDASVWPPEGAVALLDLGGFYDRA-EGFGYGLPLGLRAWGRGDEVFAEVALEEQ
mjmDH17	RLS--LRSHPWLAQHAVGVVLF	GTGTFLELAIRAGDQVCDLVDELTLAPLVLERDA	***FDGAVWPPEGAVALMDDEYLARIA-DQFGYGVFGYGLRAVWRQGVDEVFAEVALPEQ
mjmDH18	RLS--LRSHPWLAQHAVGVVLF	GTGTFLELAIRAGDQVCDVDELTLAPLVLERDA	***FDATVWPPTGGAVVMDGGEFYERRA-EGGVGFGYGLRAVWRTEDEVFAEVALPEQ
mjmDH22	RLS--TRATPWLAQHVAGSVLF	GTGTFVELALHAGHEVGCPLLEETIAAPVLPREGG	***EPLTAWPPAGADPIDLGGYDLIRA-ERSGYGVGAPGLRAAWRLGDEVFADIELQE
mjmDH23	RLS--REGQGVVVHETIGAVLLF	GTGTFVELVVRAGDEVGGCVGEELMGAPLVVPRRG	***ADLTQWPPPEGAEDVAGAYDGR-EGFGYGVGAPGLRAAWRRGDEVFVAEVLDES
mjmDH24	RLS--REGQGVVVHETIGAVLLF	GTGTFVELVVRAGDEVGGCVGEELMGAPLVVPRRG	***ADLTQWPPFARATELDGLDGYRFLD-EGHYGVGAPGLRAAWRRGDEVFVAEVLDES
mjmDH26	RLS--VGSHPWLAQHVGGTALF	GTGTFLELATRAADQVGCGRVEELTIAEPLAIGERST	***PDARGAAMPWPGAEFVEVDGIFYERRA-ALGFGYGVFGYGLRAVWRRGDEVFAAALPEE

Supplementary tables and figures for chapter 4

Figure S 28: Agar overlay bioactivity assay of *Micromonospora* sp. MD118. (A) The strain exhibited antagonistic effect against all tested clinical isolates, with potent effect against MRSA strain DB68004 (obtained from blood). (B) Radius of the zone of inhibition was measured and recorded accordingly.

Table S 7: List of medium recipes for fermentation of *Micromonospora* sp. MD118

Name of medium	Composition
Instant ocean MM	0.5 g of L-asparagine, 0.655 g of $K_2HPO_4 \cdot 3H_2O$, 0.2 g of $MgSO_4 \cdot 7H_2O$, 0.01 g of $FeSO_4 \cdot 7H_2O$, 20 g of instant ocean salts, 10 g of agar and 10 g of mannitol in 1 L of distilled water, pH 7 (15 g/L bacto agar, for agar)
M1	0.75% starch, 0.5% soybean meal, 0.12% K_2HPO_4 , 0.4% $CaCO_3$, 0.003% $FeSO_4$ and 0.0001% $CoCl_2$ (15 g/L bacto agar, for agar)
M2	1% glucose, 2% soluble starch 0.5% yeast extract, 0.5% tryptone, and 0.1% $CaCO_3$ in tap water, pH 7.6 (15 g/L bacto agar, for agar)
M3	0.5% glucose, 2% soluble starch, 1.5% soybean meal, 0.5% yeast extract, 0.2% tryptone, 0.4% $CaCO_3$, 0.2% NaCl, 0.05% KH_2PO_4 , pH 7.0 (15 g/L bacto agar, for agar)
M4	2% sucrose, 0.5% molasses, 0.2% peptone, 0.25% $CaCO_3$, 0.01% $FeSO_4 \cdot 7H_2O$, 0.02% $MgSO_4 \cdot 7H_2O$, and 0.01% KI (15 g/L bacto agar, for agar)
M5	1% corn starch, 0.5% cotton seed flour, 0.005% $CuSO_4 \cdot 5H_2O$, 0.1% $CaCO_3$, 0.00005% NaI (15 g/L bacto agar, for agar)
M6	0.5 g/L K_2HPO_4 , 3 g/L casein, 10 g/L corn starch, 1 g/L peptone, 10 g/L malt extract, 1 g/L yeast extract, distilled water, pH 7.5 (15 g/L bacto agar, for agar)

Table S 8: List of plasmids

Name	Description	Reference
pSET152	<i>ori</i> (pUC18), <i>lacZα</i> , <i>int</i> (φC31), <i>attP</i> (φC31), <i>oriT</i> (RP4), <i>aac(3)-IV</i>	Luzhetskii, Ostash et al. 2001
pIJ12551	A pSET152 derivative with <i>ermEp*</i>	Sherwood, Hesketh et al. 2013
pSK152	A pSET152 derivative with <i>kasOp*</i> and TAI(φC31) RBS separated by <i>riboJ</i> sequence	This work
pCRISPR-Cas9	A conjugative and replicative CRISPR/Cas9 plasmid for actinomycetes, <i>ApraR</i>	Tong, Charusanti, Zhang, Weber, & Lee, 2015
pHZ-cod9	A pCRISPR-Cas9 derivative with <i>codA(sm)</i> gene for counter selection	This work
pSK152- <i>mmd2050</i>	pSK152 with <i>mmd2050</i> under control of synthetic promoter-RBS cassette [<i>kasOp*</i> -Rj-TAI(φC31) RBS]	This work
pIJ12551- <i>mmd2050</i>	pIJ12551 with <i>mmd2050</i> under control of <i>ermEp*</i>	This work
pSET152- <i>mmd2050</i>	pSET152 with <i>mmd2050</i> in the absence of promoter	This work
pSK152- <i>lpmR1R2X</i>	pSK152 with <i>lpmR1</i> , <i>lpmR2</i> and <i>lpmX</i> under control of synthetic promoter-RBS cassette [<i>kasOp*</i> -Rj-TAI(φC31) RBS]	This work
pSK152- <i>sgdR2R7</i>	pSK152 with <i>sgdR2</i> and <i>sgdR7</i> under control of synthetic promoter-RBS cassette [<i>kasOp*</i> -Rj-TAI(φC31) RBS]	This work
pCRISPR-Cas9- <i>mmd2050</i>	pCRISPR-Cas9 plasmid for generating <i>Δmmd2050</i>	This work
pHZ-cod9- <i>sgdX3</i>	pHZ-cod9 plasmid derivative for generating <i>ΔsgdX3</i>	This work
pHZ-cod9- <i>sgdU15</i>	pHZ-cod9 plasmid derivative for generating <i>ΔsgdU15</i>	This work
pHZ-cod9- <i>sgdX1</i>	pHZ-cod9 plasmid derivative for generating <i>ΔsgdX1</i>	This work
pHZ-cod9- <i>sgdX2</i>	pHZ-cod9 plasmid derivative for generating <i>ΔsgdX2</i>	This work
pHZ-cod9- <i>lpmB</i>	pHZ-cod9 plasmid derivative for generating <i>ΔlpmB</i>	This work

Table S 9: List of primers used in this study

Primer name	Sequence (5' -> 3')	Purpose
mmd2050 sgna F	CATGCCATGGtggtcgatcgattcgtcgcGTTTTAGAGCTAGAAATAG	sgRNA amplification for sioxanthin BGC KO
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	

lpmB sgRNA F	CATGCCATGGAGAGCCCCATCGGCGGTGCCGTTTTAGAGCTA GAAATAGC	sgRNA amplification for lumpurmycin BGC <i>lpmB</i> KO
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	
sulfo sgRNA F	CATGCCATGGTCCGTCCGTCCCCGAAATGGGTTTTAGAGCTA GAAATAGC	sgRNA amplification for enediynes BGC <i>sgdX2</i> KO
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	
ctg1-14sgRNA F	CATGCCATGGCGCCTTGGTGTCCGGTTCCAGTTTTAGAGCTA GAAATAGC	sgRNA amplification for enediynes BGC <i>sgdX1</i> gene KO
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	
Metallo sgRNA F	CATGCCATGGagaagtcgccgcagcacaggGTTTTAGAGCTAGAAAT AGC	sgRNA amplification for enediynes BGC <i>sgdX3</i> gene
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	
ctg1-28 dynU15 sgRNA F	CATGCCATGGcgacgtggatcaggtggatgTTTTAGAGCTAGAAATA GC	sgRNA amplification for enediynes BGC <i>sgdU15</i> gene
sgRNA R	ACGCCTACGTAAAAAAGCACCGACTCGGTGCC	
sgRNA check F	CACAGGAAACAGCTATGACC	sgRNA sequencing
sgRNA check R2	TCAGCCCAGATCCCCGATCC	
sgRNA check R	ACCCCCATTCAAGAACAGC	
mmd2050 left F40	TCGTCTGAAGGCACTAGAAAGGAAGTTGCTCATCATGCCCCGC	Homologous recombination template construction for pCRISPR-Cas9- <i>mmd2050</i>
mmd2050 left R40	GAAGTTGGTGAGCTGGAAGGGTCATAGGCAGCGGTGAT	
mmd2050 right F20	CCTTCCAGCTCACCAACTTC	
mmd2050 right R40	GGTCGATCCCCGCATATAGGGTGGATGTCGTCGTACCGAT	
4156 left F40	TCGTCTGAAGGCACTAGAAAGGACTACCTCGGTCTCGGCAAC	Homologous recombination template construction for pHZ- cod9- <i>lpmB</i>
4156 left R40	TTCCAGCATCGGCTCCATCACACCAACGACCACGACACAT	
4156 right F20	TGATGGAGCCGATGCTGGAA	
4156 right R40	GGTCGATCCCCGCATATAGGTATTCGACGCCAGACGGGA	
sulfo left F40	TCGTCTGAAGGCACTAGAAAGGAAGGGAGAGGGCAAGTGACC	Homologous recombination template construction for pHZ- cod9- <i>sgdX2</i>
sulfo left R40	CACACCGAGGAACCTACACATGAGAATCTCCTTCGTGTGG	
sulfo right F20	TGTGTAAGTTCTCGGTGTG	
sulfo right R40	GGTCGATCCCCGCATATAGGGTAGAAGTCGCCGCAGCACA	
ctg1-14 left F40	TCGTCTGAAGGCACTAGAAAGGTCAAGTAGCATGACAGCCGGA G	Homologous recombination template construction for pHZ- cod9- <i>sgdX1</i>
ctg1-14 left R40	CGGTGATCCAGGTGACGTACGCGTACGGCGTGTGTTGAT	
ctg1-14 right F20	GTACGTACCTGGATCACCG	
ctg1-14 right R40	GGTCGATCCCCGCATATAGGGATCAGCGTGAGTTTGTGCGC	
Metallo left F40	TCGTCTGAAGGCACTAGAAAGGCGGAGGCGAAGGTCATCCTG	Homologous recombination template construction for pCRISPR-Cas9- <i>sgdX3</i>
Metallo left R40	CATCATCGCCACCACACCGTGAGGACGTGGTCGATGTGGC	
Metallo right F20	ACGGTGTGGTGCGATGATG	
Metallo right R40	GGTCGATCCCCGCATATAGGGGATGATCGCACTGTTCCGC	
ctg1-28 dynU15 left F40	TCGTCTGAAGGCACTAGAAAGGAGGTGCTGGTGTCCACGCTG	Homologous recombination template construction for pCRISPR-Cas9-ctg1- 28 <i>sgdU15</i>
ctg1-28 dynU15 left R40	AAGGCCACACCCTGACCGAGGTTGAAGCCGTCGAGGAAGG	
ctg1-28 dynU15 right F20	CTCGGTGAGGGTGTGGCCTT	
ctg1-28 dynU15 right R40	GGTCGATCCCCGCATATAGGCTGCATCCGCATCACCGGGT	
pcrispr-cas9 seq	TGAGGCTTGAGGGAGTCAA	HDR template sequencing
pcrispr-cas9 seq R	CGTCGCTCTCTGGCAAAGCT	
phytoene synthase 2050 check F	GATCTCGGTGAGCGATTCTGA	Validation of <i>mmd2050</i> gene deletion
phytoene synthase 2050 check R	CACCTCGTACTCGATCAGCT	
4156 check F	CGGTATCTCGGGGACCAATG	Validation of <i>lpmB</i> gene deletion
4156 check R	ACGCGAACTCTTCCAGCATC	
Sulfo check F	TACTTCAACGGCGTCTCACC	

Sulfo check R	GAAGGCATCCTGGTCGTTGT	Validation of <i>sgdX2</i> gene deletion
ctg1-14 check F	GGTGAGGGAGCACAGATGAG	Validation of <i>sgdX1</i> gene deletion
ctg1-14 check R	GTGGGAGTCAGGATCAGGTG	
Metallo check F	TACAAGGGTGAAGCTTCGCCG	Validation of <i>sgdX3</i> gene deletion
Metallo check R	GCGGACGATGAACTGCGGAT	
dynU15 check F	CTCGCCGACTTCACGGTC	Validation of <i>sgdU15</i> deletion
dynU15 check R	GGATCAGAGCGACGCGAG	
promoter cassette F	GTCGACTCTAGAGGATCCGCgtgttcacattcgaaccgtct	PCR amplification of kasO* _p – riboJ – TAI(φ31) RBS cassette for construction of pSK152
promoter cassette R	ACGAATTCGATATCGCGCGCGggccgcaaaaccctaggtcgtc	
pSK152 sio F	CGACGACCTAGGGGTTTTGCATGGACGCGGATCTGACCGC	Construction of pSK152- <i>mmd2050</i> for overexpression of gene <i>mmd2050</i>
pSK152 sio R	TACGAATTCGATATCGCGCGCTCATCCGGCCAGTGCCGGTT	
pSET152 sio F	GTCGACTCTAGAGGATCCGCATGGACGCGGATCTGACCGC	Construction of pSET152- <i>mmd2050</i> for overexpression of gene <i>mmd2050</i>
pSET152 sio R	ACGAATTCGATATCGCGCGCTCATCCGGCCAGTGCCGGTT	
pIJ12551 sio F	GTCTAGAACAGGAGGCCCCATATGATGGACGCGGATCTGACCGC	Construction of pIJ12551- <i>mmd2050</i> for overexpression of gene <i>mmd2050</i>
pIJ12551 sio R	ATGCAGGTCGACTCTAGTTATCATCCGGCCAGTGCCGGTT	
pIJ12551 sio F	GTCTAGAACAGGAGGCCCCATATGATGGACGCGGATCTGACCGC	Construction of pIJ12551- <i>mmd2050</i> for overexpression of gene <i>mmd2050</i>
pIJ12551 sio R	ATGCAGGTCGACTCTAGTTATCATCCGGCCAGTGCCGGTT	
codA(sm) F	TCCTATTCGGAAGTTCCTATTCTCGAGGTTGACATCTTTTG	PCR amplification of codA(sm) gene for construction of pHZ-cod9
codA(sm) R	GCGTCGCTTGGTCGGTCATTTATGCTCAGCGCTTGTAGTC	
Md118 cluE 4330 F	CGACGACCTAGGGGTTTTGCGTGTTGCGTGCGGCAATTGT	PCR amplification of <i>sgdR2</i> (fragment 1) for pSK152- <i>sgdR2R7</i>
Md118 cluE 4330 R	atggacactccttacttagaTCATAGGTTTGTTCGCCGACAGA	
Md118 cluE 4329 F	tctaagtaaggagtgccatATGCCTGCCGTCCACGAAAC	PCR amplification of gene <i>sgdR7</i> (fragment 2) for pSK152- <i>sgdR2R7</i>
Md118 cluE 4329 R	ACGAATTCGATATCGCGCGCTGGACCCGGTTCCACCGAC	
Md118 clu7 4172 F	CGACGACCTAGGGGTTTTGCATGGAGCTGACCGAGCGTTC	PCR amplification of <i>lpmR2</i> gene (fragment 1) for construction of pSK152- <i>lpmR1R2X</i>
Md118 clu7 4172 R	gcgtatccccttcagatacTGTTCCGTGCGGAAGCCCGA	
Md118 clu7 4171 F	gtatctgaaaggggatacgcGTGTTGGTTCAGCAAGAGGA	PCR amplification of <i>lpmR1</i> gene (fragment 2) for construction of pSK152- <i>lpmR1R2X</i>
Md118 clu7 4171 R	atggacactccttacttagaTCAGGAGGTGGCGCTCTGTG	
Md118 clu7 4175 F	tctaagtaaggagtgccatGTGATCGAGCAGATCCTGCC	PCR amplification of <i>lpmX</i> gene (fragment 3) for construction of pSK152- <i>lpmR1R2X</i>
Md118 clu7 4175 R	ACGAATTCGATATCGCGCGCTTAGGACGCGACCGCCGGGA	

Table S 10: List of sequences used in this study.

Sequence name	Sequence (5' -> 3')	Purpose
<i>Micromonospora</i> sp. MD118 φC31 <i>attB</i> site sequence	CGGTGCGGGTGCCAGGAGGTGCCCTTCGGCTCGCCCGGCGCGTAGTC GACC	Site specific integration by φC31 integrase

psk152 kasO* _p – riboJ – TAI(ϕ31) RBS cassette	GTCGACTCTAGAGGATCCGCTgttcacattcgaaccgtctctgctTTGACAacatgctgtgc ggtgttgTAAAGTcgtggccAagctgtcaccggatgtgctttccggtctgatgagtcggtgaggacgaa acagcctctacaaataattttgttaacgacgacctaggggttttcgcccGCGCGCGATATCGAA TTCGT	promoter- riboJ-RBS cassette for construction of pSK152, an integrative overexpress ion plasmid
<i>codA</i> (sm) gene with promoter sequence	ATGCCGAACAACGCCCTGCAGACCATCATCAACGCCCGCCTGCCGGGC GAGGAGGGCCTGTGGCAGATCCACCTCCAGGACGGCAAGATCTCCGC CATCGACGCCAGTCCGGCGTCATGCCGATCACCAGAGAACTCCCTGGA CGCCGAGCAGGGCCTGGTTCATCCCGCCGTTCTGTCGAGCCGCACATCC ACCTGGACACCAACCCAGACCGCCGGCCAGCCGAACTGGAACCACTCC GGCACCCCTGTTTCGAGGGCATCGAGCGCTGGGCCGAGCGCAAGGCCCT GCTGACCCACGACGACGTGAAGCAGCGCGCCTGGCAGACCCCTGAAGT GGCAGATCGCCAACGGCATCCAGCACGTCCGCACCCACGTGACGTGT CCGACGCCACCCCTGACCGCCCTGAAGGCCATGCTGGAGGTCAAGCAG GAGGTCGCCCCCTGGATCGACCTGCAGATCGTCGCCTTCCCGCAGGAG GGCATCCTGTCTTACCCGAACGGCGAGGCCCTGCTGGAGGAGGCCCT GCGCCTGGGCGCCGACGTGGTGGGCGCCATCCCGCACTTCGAGTTCA CCCGCGAGTACGGCGTCGAGTCCCTGCACAAGACCTTCGCCCTCGCCC AGAAGTACGACCGCCTGATCGACGTCCACTGCGACGAGATCGACGACG AGCAGTCCCGCTTCGTCGAGACCGTCGCGCCGCCCTGGCCCAACACGAG GGCATGGGCGCCCGCGTCACCGCCTCCACACCAACCGCCATGCACTC CTACAACGGCGCCTACACCTCCCGCCTGTTCCGCCTGCTGAAGATGTC GGGCATCAACTTCGTCGCCAACCCTGGTCAACATCCACCTGCAGGG CCGCTTCGACACCTACCCGAAGCGCCGCGGCATCACCCGCGTCAAGGA GATGCTGGAGTCCGGCATCAACGTCTGCTTCGGCCACGACGCCGTGTT CGACCCGTGGTACCCGCTGGGCACCGCCAACATGCTGCAGGTCTCTGCA CATGGGCCTGCACGTCTGCCAGCTGATGGGCTACGGCCAGATCAACGA CGGCCTGAACCTGATCACCCACCACTCCGCCCGCACCCCTGAACCTGCA GGACTACGGCATCGCCGCCGGCAACTCCGCCAACCTGATCATCCTGCC GGCCGAGAACGGCTTCGACGCCCTGCGCCGCCAGGTCCCGGTCCGCT ACTCCGTCCGCGGGCGGCAAGGTATCGCCTCCACCCAGCCGGGCCAG ACCACCGTCTACCTGGAGCAGCCGGAGGCCATCGACTACAAGCGCTGA	<i>codA</i> (sm) gene with promoter sequence for construction of pHZ9- <i>codA</i>

Table S 11: List of putative BGCs identified from *Micromonospora* sp. MD118.

BGC	Position From	To	Type/Product
cluster 1	24332	45282	Terpene
cluster 2	1547269	1588330	T3pks
cluster 3	1683110	1723967	Phosphonate
cluster 4	1730905	1790422	Nrps
cluster 5	2194495	2215421	Sioxanthin associated BGC
cluster 6	2295055	2316020	Terpene
cluster 7	4476334	4696438	Lumpurmycin BGC
cluster 8	4767256	4826083	Siderophore-NRPS
cluster 9	4826368	4859509	T1pks-NRPS
cluster 10	4859627	4895098	sungeidine BGC
cluster 11	5021415	5071651	T1pks-Nrps
cluster 12	5552330	5573532	Terpene
cluster 13	5767757	5810328	T2pks

cluster 14	5982942	6054675	Oligosaccharide-Terpene-Nrps
cluster 15	6108007	6148294	Nrps
cluster 16	6254140	6353496	T3pks-Nrps
cluster 17	6582319	6611897	Bacteriocin-Terpene
cluster 18	6749616	6761541	Terpene

Table S 12: List of genes in *lpm* BGC and their proposed functions.

Gene	Size (amino acid)	Closest homolog [source]	Identity/Similarity (%)	proposed function	UniProt ID
<i>lpmU1</i>	186	NADPH-dependent FMN reductase [<i>Micromonospora aurantiaca</i> strain ATCC 27029]	100/100	unknown	D9T6U4
<i>lpmS4</i>	313	ROK family kinase [<i>Micromonospora</i> sp. M42]	99/99	Kinase	W7VRF2
<i>lpmU2</i>	118	DoxX family protein [<i>Micromonospora</i> sp. TSRI0369]	96/97	unknown	A0A1Q5G PC0
<i>lpmS6</i>	144	3-dehydroquinone dehydratase [<i>Micromonospora marina</i>]	97/100	aDHQ dehydratase	A0A1C4Z VP3
<i>lpmS2</i>	388	3-amino-5-hydroxybenzoic acid synthase [<i>Micromonospora</i> sp. WMMB235]	99/99	AHBA synthase	A0A1S1W EV9
<i>lpmS1</i>	377	Oxidoreductase [<i>Micromonospora</i> sp. TSRI0369]	99/99	Oxidoreductase	A0A1Q5G PC3
<i>lpmS3</i>	232	Phosphoglycolate phosphatase [<i>Micromonospora</i> sp. WMMB235]	99/99	Phosphatase	A0A1S1W F18
<i>lpmU3</i>	1234	Thioester reductase [<i>Micromonospora</i> sp. TSRI0369]	99/99	unknown	A0A1Q5G PA2
<i>lpmU4</i>	1228	Amino acid adenylation domain protein [<i>Micromonospora aurantiaca</i> strain ATCC 27029]	97/98	unknown	D9T6V2
<i>lpmU5</i>	397	Uncharacterized protein [<i>Micromonospora</i> sp. TSRI0369]	99/100	unknown	A0A1Q5G PC7
<i>lpmS5</i>	346	3-dehydroquinone synthase [<i>Micromonospora</i> sp. M42]	98/99	ADHQ synthase	W7VNM6
<i>lpmU6</i>	275	Uncharacterized protein [<i>Micromonospora</i> sp. II]	97/97	unknown	A0A1E2U 890
<i>lpmA</i>	4971	AMP-dependent synthetase [<i>Micromonospora</i> sp. TSRI0369]	98/99	PKS	A0A1Q5G P98
<i>lpmO1</i>	521	3-(3-hydroxyphenyl)propionate hydroxylase [<i>Micromonospora</i> sp. TSRI0369]	99/99	FAD dependent oxidoreductase	A0A1Q5G PB0
<i>lpmO2</i>	296	Methyltransferase [<i>Micromonospora</i> sp. TSRI0369]	99/100	O-methyltransferase	A0A1Q5G PC6
<i>lpmO3</i>	254	Cadicidin biosynthesis thioesterase [<i>Micromonospora</i> sp. M42]	99/99	Thioesterase	W7W770
<i>lpmB</i>	1855	Beta-ketoacyl synthase [<i>Micromonospora</i> sp. TSRI0369]	96/97	PKS	A0A1Q5G PB3
<i>lpmC</i>	6442	Beta-ketoacyl synthase [<i>Micromonospora aurantiaca</i> ATCC 27029]	94/96	PKS	D9T6W1
<i>lpmU7</i>	283	Metallophosphoesterase [<i>Micromonospora aurantiaca</i> ATCC 27029]	98/98	Unknown	D9T6W2

<i>lpmD</i>	7523	Acyl transferase [<i>Micromonospora</i> sp. TSRI0369]	98/99	PKS	A0A1Q5H K81
<i>lpmE</i>	9947	Beta-ketoacyl synthase [<i>Micromonospora</i> sp. TSRI0369]	96/98	PKS	A0A1Q5H LZ1
<i>lpmF</i>	3368	Beta-ketoacyl synthase [<i>Micromonospora</i> sp. TSRI0369]	99/99	PKS	A0A1Q5H K77
<i>lpmG</i>	4878	Acyl transferase [<i>Micromonospora</i> sp. TSRI0369]	97/98	PKS	A0A1Q5H K75
<i>lpmH</i>	3745	Acyl transferase [<i>Micromonospora</i> sp. TSRI0369]	97/98	PKS	A0A1Q5H K84
<i>lpmI</i>	4711	Beta-ketoacyl synthase [<i>Micromonospora aurantiaca</i> ATCC 27029]	97/98	PKS	D9T6W8
<i>lpmJ</i>	3652	Polyketide synthase [<i>Micromonospora</i> sp. TSRI0369]	97/98	PKS	A0A1Q5H K95
<i>lpmK</i>	1665	Acyl transferase [<i>Micromonospora</i> sp. TSRI0369]	99/99	PKS	A0A1Q5H KA0
<i>lpmL</i>	3165	Polyketide synthase [<i>Micromonospora</i> sp. TSRI0369]	99/99	PKS	A0A1Q5H KC0
<i>lpmM</i>	1827	6-deoxyerythronolide-B synthase [<i>Micromonospora</i> sp. TSRI0369]	99/99	PKS	A0A1Q5H KA9
<i>lpmT1</i>	351	Bacteriocin ABC transporter ATP-binding protein [<i>Micromonospora</i> sp. II]	99/99	transporter protein	A0A1E2U 7Q3
<i>lpmT2</i>	280	Transport permease protein [<i>Micromonospora</i> sp. II]	99/100	transporter protein	A0A1E2U 848
<i>lpmR1</i>	941	LuxR family transcriptional regulator [<i>Micromonospora</i> sp. M42]	98/99	transcriptional regulator	W7VM91
<i>lpmR2</i>	952	LuxR family transcriptional regulator [<i>Micromonospora</i> sp. TSRI0369]	98/99	transcriptional regulator	A0A1Q5H K85
<i>lpmU8</i>	458	AMP-dependent synthetase [<i>Micromonospora</i> sp. II]	99/99	unknown	A0A1E2U 8U5
<i>lpmU9</i>	166	DNA-binding protein [<i>Micromonospora</i> sp. TSRI0369]	98/98	unknown	A0A1Q5H KA6
<i>lpmX</i>	221	4'-phosphopantetheinyl transferase EntD (Siderophore biosynthesis) [<i>Micromonospora aurantiaca</i>]	98/99	Svp-type PPTase	A0A1C6S J55
<i>lpmU10</i>	471	AMP-dependent synthetase and ligase [<i>Micromonospora aurantiaca</i> ATCC 27029]	98/98	unknown	D9T6Y0
<i>lpmS7</i>	222	Phosphoglycolate phosphatase [<i>Micromonospora</i> sp. TSRI0369]	99/100	Phosphatase	A0A1Q5H KD4
<i>lpmT3</i>	314	Multidrug ABC transporter ATP-binding protein [<i>Micromonospora</i> sp. II]	100/100	transporter protein	A0A1E2U 806
<i>lpmT4</i>	246	Transport permease protein [<i>Micromonospora</i> sp. II]	99/100	transporter protein	A0A1E2U 7R5
<i>lpmU11</i>	81	Uncharacterized protein [<i>Micromonospora</i> sp. TSRI0369]	99/100	unknown	A0A1Q5H KA4

Figure S 29: *Micromonospora* sp. MD118 grown on 6 different agar media.

Figure S 30: (-)-HRMS of lumpurmycin.

Table S 13: ^1H and ^{13}C NMR spectroscopic data of lumpurmycin (in $\text{C}_5\text{D}_5\text{N}$, 400 MHz)^a

Position	δ_{H} (mult., J in Hz)	δ_{C} , multi.	Position	δ_{H} (mult., J in Hz)	δ_{C} , multi.
1		184.9, C	40	3.09, dd (15.2, 7.8)	52.0, CH_2
2		144.5, C		2.92, dd (15.2, 6.4)	
3	6.06, br s	100.9, CH	41		209.0, C
4		186.2, C	42	2.96, dd (15.8, 8.7)	51.5, CH_2
5	6.85, br s	134.4, CH		2.88, dd (15.8, 4.5)	
6		149.7, C	43	4.58, m	67.7, CH
7	4.40, m	77.0, CH	44	2.57, m; 2.50, m	41.7, CH_2

7-OMe	3.20, s	57.2, CH ₃	45	5.91, dd (15.2, 7.4)	130.6, CH
8	1.63, m; 1.57, m	38.8, CH ₂	46	6.29, dd, overlapped	132.9, CH
9	1.52, m; 1.32, m	32.6, CH ₂	47	6.54, dd (15.8, 11.2)	129.9, CH
10	2.09, m	36.9, CH	48	6.01, dd (15.8, 6.3)	136.8, CH
10-Me	0.92, d (6.6)	20.3, CH ₃	49	4.84, m	71.2, CH
11	5.56, dd (15.2, 7.8)	141.7, CH	50	2.17, m; 2.01, m	45.9, CH ₂
12	6.04, dd, overlapped	128.5, CH	51	4.51, m	70.6, CH
13	6.37, dd (15.2, 10.6)	133.0, CH	52	1.91, m	36.7, CH ₂
14	5.78, dd (15.2, 6.7)	130.4, CH	53	1.94, m; 1.68, m	28.6, CH ₂
15	6.05, m	72.5, CH	54	1.81, m	40.2, CH
16	2.58, m; 2.55, m	42.3, CH ₂	54-Me	1.08, d (6.0)	15.7, CH ₃
17	4.28, m	71.4, CH	55	4.00, m	75.9, CH
18	1.92, m; 1.86, m	44.0, CH ₂	56	2.04, m; 1.90, m	38.1, CH ₂
19	4.36, m	67.8, CH	57	4.45, m	74.7, CH
20	1.76, m; 1.64, m	33.6 ^b , CH ₂	58	2.09, m	41.1, CH
21	1.62, m; 1.44, m	26.4, CH ₂	58-Me	1.31, d (6.6)	7.9, CH ₃
22	1.40, m; 1.31, m	27.5, CH ₂	59	4.63, m	79.7, CH
23	1.37, m; 1.15, m	37.3, CH ₂	60		140.2, C
24	1.44, m	33.2, CH	60-Me	1.85, br s	15.7, CH ₃
24-Me	0.88, d (6.2)	20.0, CH ₃	61	6.96, d (11.3)	121.1, CH
25	1.69, m; 1.40, m	33.2 ^b , CH ₂	62	6.51, dd (11.3, 11.3)	124.2, CH
26	1.68, m; 1.63, m	36.2, CH ₂	63	5.95, dd (11.3, 6.0)	134.2, CH
27	4.15, m	71.4 ^c , CH	64	3.33, m	35.0, CH
28	1.90, m; 1.78, m	45.1, CH ₂	64-Me	1.09, d (6.0)	18.9, CH ₃
29	4.15, m	71.5 ^c , CH	65	3.90, m	78.0, CH
30	2.01, m; 1.92, m	45.5, CH ₂	66	2.11, m	37.6, CH
31	4.16, m	71.8 ^c , CH	66-Me	1.35, d (6.0)	9.0, CH ₃
32	2.09, m; 1.90, m	43.7, CH ₂	67	4.18, m	75.4, CH
33	6.04, m	132.2, CH	68	1.86, m	41.2, CH
34	6.30, m	133.0, CH	68-Me	1.18, d (6.0)	12.6, CH ₃
35	6.30, m	131.1, CH	69	4.15, m	69.5, CH
36	6.31, m	133.3, CH	70	2.70, br d (11.6)	44.5, CH ₂
37	6.61, dd (15.8, 10.1)	130.0, CH		2.18, dd (12.3, 11.6)	
38	6.06, dd, overlapped	137.2, CH	71		97.0, C
39	5.10, q-like (6.3)	68.3, CH	72	3.02, s	48.1, CH ₂
			73		170.3, C

^a Assignments were made by a combination of 1D and 2D NMR experiments.

^{b-c} Signals with the same superscript might be interchangeable.

Figure S 31: ¹H NMR spectrum (400 MHz, in C₅D₅N) of lumpurmycin.

Figure S 32: ¹³C NMR spectrum (400 MHz, in C₅D₅N) of lumpurmycin.

Figure S 33: ^{13}C DEPT-135 NMR (400 MHz, in $\text{C}_5\text{D}_5\text{N}$) of lumpurmycin.

Figure S 34: HSQC NMR spectrum (400 MHz, in $\text{C}_5\text{D}_5\text{N}$) of lumpurmycin.

Figure S 35: COSY NMR spectrum (400 MHz, in C₅D₅N) of lumpurmycin.

Figure S 36: HMBC NMR spectrum (400 MHz, in C₅D₅N) of lumpurmycin.

Figure S 37: NOESY NMR spectrum (400 MHz, in C_5D_5N) of lumpurmycin.

Figure S 38: ROESY NMR spectrum (400 MHz, in C_5D_5N) of lumpurmycin.

Figure S 39: ^1H NMR spectrum (400 MHz, in $\text{DMSO}-d_6$) of lumpurmycin.

Figure S 40: ^{13}C NMR spectrum (400 MHz, in $\text{DMSO}-d_6$) of lumpurmycin.

Figure S 41: HMBC NMR spectrum (400 MHz, in DMSO- d_6) of lumpurmycin.

Figure S 42: Inactivation of *lpmB* gene. Diagnostic PCR was performed to select for mutants with deletion of 730 bases from *lpmB* gene.

Figure S 43: AT domain analysis of Lpm PKS. (A) Precursor units utilized by acyltransferase as substrate for chain elongation. (B) Analysis of AT domains show that all of them possess a GHSXG motif⁴²³ at the active site. Malonyl coA specificity is determined by HAFH motif in red font while YASH motif in blue font selects for methylmalonyl CoA as substrate.³³⁸

Figure S 44: KR domain analysis of Lpm PKS. (A) KR exerts stereocontrol over the geometry of hydroxyl and succeeding olefin moieties. A-type KR reduces β -keto group into a S-hydroxyl group that is subsequently dehydrated into a *cis*-olefin moiety, while B-type KR reduces β -keto group into a R-hydroxyl group that is subsequently dehydrated into a *trans*-olefin moiety. A2-type KR is differentiated from A1-type KR by presence of histidine residue in the catalytic region (depicted by dark blue column), while B2-type KR is differentiated from B1-type KR by presence of proline in the catalytic region (depicted by green column). (B) KR domains of Lpm which possesses a tryptophan residue (in red) in their catalytic region were annotated as Type A-type KRs, while those that possesses a LDD loop were annotated as B-type KRs.³³⁹

	HXXXGXXXXP motif	GYXGPXF motif	
LpmDH1	AGGPGVAGLADAGHPLLGGLLAVAGTQGVCTGRIATATSPWLAA	-----AGELARAANAADRLVVDVRLPDDARI	LpmDH1
LpmDH2	PGDVGAAGLDDAGHPLLGASVVLAEESTGHLFTGRLSADAQFWLAD	GYAYGPAFQGLRAVWTSQDEVYAEVELPAAQR	LpmDH2
LpmDH3	VETVPGLGVRPVDHALLGAARVADDDVRVFTGEISVLTHPWLAEH	GVEYGPFLFRGLRAVWRTGDEVCAEVAVPDG	LpmDH3
LpmDH4	PADVTAAGLTPTGHPHLLGAVFRSADGDEVTLTGRLSVETQFWLAE	GFAYGPAFSGLRAAWRVGADVYAEVALPDPA	LpmDH4
LpmDH8	VGDVSGAGLGVAGHPLLGAAVRLAGDDEVLTGRLSTSTHPWLAD	GLSYGPVFQGVRRRAWTADGGAYAEVALPDGAA	LpmDH8
LpmDH9	TGDIVSGAGLGAAGHPLLGAAVRLAGDDEVLTGRLSASTHPWLAD	GLAYGPAFRLRLRWTSQDEVYAEVLPDEAA	LpmDH9
LpmDH10	AGDVSGAGLGVAEHGLLGAAVDLAGDDEVLTGRLSPSTTHPWLA	GLSYGPAFQGLRRVWTSQEDDVFAEVLPPDAA	LpmDH10
LpmDH13	SADVSGAGLGVARHPLLGAAVSVAGEDMVLTGRLSVSTHPWLAD	GLAYGPVFRGLRRRAWTADGEVFAEVALPDEV	LpmDH13
LpmDH14	TADVSGAGLGVAGHPLLGAAVSVAGADMVLTGRLSLTTHVWLAD	GLAYGPVFRGLRRVWTSQDEVFAEVALPDEVA	LpmDH14
LpmDH15	TGDIVSGAGLGRTHPLLGATVDLAGDDEMVLTLGRLSLATHPWLA	GLTYGPAFQGLQRAWVGRDELFAEVALPEAAA	LpmDH15
LpmDH19	RSLLPAVGLDVVDHPLLSAVADLPDEGLLFTGRITLTTHPWLA	GYHYGPTFRALRAAWRSGRDILAEVALDADAS	LpmDH19
LpmDH20	WTDVSGAGLGAAGHPLLGAAVSVAGEDMVLTGRLSTSTHAWLA	GLAYGPVFRGLRRVWTSQDEVFAEVALPDEVA	LpmDH20
LpmDH24	VRDVSGAGLQDTGHPHLLSAAVFPVADRTVLTGRLSVRTHPWLA	GLGYGPFRLRLRWTSQDEVYAEVALPEEAS	LpmDH24
LpmDH28	TGDIVTAGLVRPAQHPLLGAAVGVAAHAEGLFTGRLARRTHPWIC	GYDYGPSTALRSAAWRGDDLFAELALPEAPA	LpmDH28
LpmDH29	AQDVSSAGLQDAGHPLLSATLTFPDSERMVLTGRLSVRTHPWLE	GFGYGPFRGLRAAWIRDDVFAEVALPAGHH	LpmDH29

	DXXX(H/Q) motif	LPFXW motif	
LpmDH1	PDAERYGVHPLVDAALRLPLAGFAS	PTARLLPTEWAGVRRHAGVAHALRVCLAPAGADAVTL	LpmDH1
LpmDH2	VDADRYGLHPALDAALRAAHFGGLAA	AGQLLLPFAWNGVSLHATGATLRVRLTRVGPALIALTVAD	LpmDH2
LpmDH3	TEVAGFGLHPALDAALHIVGLLSDA	ASGVLLPFAWTVTLTSAVGTALRVARRSGG-GVALTLTDPAG	LpmDH3
LpmDH4	ADAGGYGLHPALDAALHIAVGLGIV	PGEAMPPFAWSDVTLTSAHGADTLRVRLTPTGPDVMSLVADATG	LpmDH4
LpmDH8	GEAGAFGVHPALDAALHIVALLLEAE	TGGPKVPFAFTGVQVHASGARTLRVRLTRDGS-GVRLVACDGG	LpmDH8
LpmDH9	GDADGFAVHPALDAALHIVGLLLA-E	SGGPKVPFAFTGVQVHASGARTLRVRLTRDGS-RVRLVACDESG	LpmDH9
LpmDH10	AEAARFGVHPALDAALHPIGLLPGAE	TGGPKVPFAFTGVQVHASGARLLRVRLTRDGS-AVRLTARDEAG	LpmDH10
LpmDH13	GDPAFGVHPALDAALHIVGLLPADQ	GGGPKVPFAFTGVQVHAGARTLRVRLTASGN-SVRLVAADETG	LpmDH13
LpmDH14	TEAARFGVHPALDAALHPIGLLPETE	SGGPKVPFAFTGVQVYASGARVLRVRLSRTGS-AVRLVACDESG	LpmDH14
LpmDH15	TESAGFGVHPALDAALHPIGLLLGGE	SSGPKVPFAFTGVQVYASGARVLRVRLSRTGS-GLRLAAYDEVG	LpmDH15
LpmDH19	TDTTAFSVHPALDAALHIAIGASATDG	GIRPGLPFTWISGVITRPTARELRIRLRATDAGSVALTVDGTA	LpmDH19
LpmDH20	GDAGFGVHPALDAALHPIGLLPGSE	SGGARVPFAFTGVQVYASGARVLRVRLTRDGS-AVRLVACDESG	LpmDH20
LpmDH24	ADASRFGLHPALDAALHIVGLLTLG	-GAGMPFAFTSVLHAGATRLRLRPLCPDPTVAVTLADTAG	LpmDH24
LpmDH28	AEAGGYGLHPALDAALHIVGLLTPPAG	PGYGLPFSWIDVALHGAGAAARVRVSHGTGPDSSISLLTDTVTG	LpmDH28
LpmDH29	AEAGRFGLHPALDAALHIVGLLTPPAG	AGVPLPFAWTVTLTSAVGTALRVRLTASGN-SVRLVAADETG	LpmDH29

Figure S 45: DH domain analysis of Lpm PKS. All mjm DH domains possess the crucial HXXXGXXXXP motif involved in the catalytic site,³⁴⁰ except for LpmDH1, 3, 4, 28, 29. LpmDH1 lacks the catalytic histidine residue, hence is non-functional. Structural analysis of lumpurmcin confirmed that LpmDH 3, 4, 28 and 29 are functional.

Table S 14: List of genes in *sgd* BGC and their proposed functions.

Gene	Size (amino acid)	Closest homolog [source]	Identity/ Similarity (%)	proposed function	UniProt ID
<i>sgdA5</i>	350	Methyltransferase domain-containing protein [<i>Streptosporangium subroseum</i>]	54/64	SAM dependent methyltransferase	A0A239NGV0
<i>sgdU21</i>	189	Uncharacterized protein [<i>Micromonospora chailaphumensis</i>]	77/86	Hypothetical protein	A0A1C4W0Q8
<i>sgdU20</i>	190	Uncharacterized protein [<i>Micromonospora</i> sp. CB01531]	47/57	Hypothetical protein	A0A1Q5BEB8
<i>sgd17</i>	399	Uncharacterized protein (<i>Amycolatopsis coloradensis</i>)	48/58	Ketone reductase	A0A1R0KZ46
<i>sgdR3</i>	421	Ubiquinone biosynthesis protein (<i>Saccharopolyspora shandongensis</i>)	61/72	Hypothetical protein/Transcriptional regulator	A0A1H2QMM1
<i>sgd16</i>	463	PBS lyase (<i>Amycolatopsis coloradensis</i>)	84/91	PBS lyase	A0A1R0KZ63
<i>sgd15</i>	214	Uncharacterized protein (<i>Streptomyces</i> sp. CB03234)	85/91	Hypothetical protein	A0A125SA03
<i>sgdU16</i>	286	Uncharacterized protein (<i>Amycolatopsis coloradensis</i>)	50/61	Enediynes self sacrifice resistance protein	A0A1R0KZD6
<i>sgdU8</i>	237	Transcriptional regulator, HxIR family (<i>Micromonospora yangpuensis</i>)	57/70	Transcriptional regulator	A0A1C6UU66

<i>sgdX1</i>	155	Uncharacterized protein (<i>Streptomyces</i> sp. RV15)	59/68	Hypothetical protein	A0A117RYF8
<i>sgd14</i>	347	Uncharacterized protein (<i>Streptomyces</i> sp. RV15)	75/84	Hypothetical protein	A0A117RYI1
<i>sgdE10</i>	417	Uncharacterized protein (<i>Streptomyces</i> sp. RV15)	66/77	Cytochrome P450	A0A117RYX8
<i>sgdT5</i>	328	ABC transporter (<i>Streptomyces</i> sp. RV15)	64/75	ABC transporter protein	A0A101UT78
<i>sgdT8</i>	282	Transport permease protein (<i>Streptomyces</i> sp. RV15)	72/85	ABC-2 Transporter protein	A0A117RYB3
<i>sgdX2</i>	253	Uncharacterized protein (<i>Streptomyces</i> sp. RV15)	53/73	Sulfotransferase	A0A101UT94
<i>sgdX3</i>	438	MBL fold metallo-hydrolase (<i>Streptomyces</i> sp. RV15)	57/72	Metallo-beta-lactamase family protein	A0A101UTE4
<i>sgdE14</i>	122	Uncharacterized protein (<i>Streptomyces</i> sp. RV15)	64/82	Glyoxalase/dioxygenase	A0A101UT77
<i>sgdX4</i>	236	Uncharacterized protein (<i>Streptomyces</i> sp. RV15)	69/77	Hypothetical protein	A0A101UT83
<i>sgdO6</i>	295	Uncharacterized protein (<i>Streptomyces</i> sp. RV15)	52/65	O-methyltransferase	A0A101UT91
<i>sgd19</i>	437	Uncharacterized protein (<i>Streptomyces</i> sp. RV15)	54/68	Cytochrome P450	A0A101UTD0
<i>sgdR7</i>	249	Uncharacterized protein [<i>Couchioplanes caeruleus</i>]	45/56	Transcriptional regulator	A0A1K0FU55
<i>sgdR2</i>	627	Predicted ATPase (<i>Micromonospora citrea</i>)	45/58	Transcriptional regulator	A0A1C6VYX0
<i>sgdX5</i>	341	Transposase (<i>Micromonospora crema</i>)	57/72	IS630 family transposase	A0A1N5U972
<i>sgdU15</i>	353	Uncharacterized protein [<i>Couchioplanes caeruleus</i>]	50/57	UnbL (unknown conserved protein in enediynes biosynthesis)	A0A1K0FTV9
<i>sgdU14</i>	626	DynU14 (<i>Micromonospora chersina</i>)	63/72	UnbV (unknown conserved protein in enediynes biosynthesis)	Q84HJ0
<i>sgdT3</i>	313	Uncharacterized protein (<i>Micromonospora yangpuensis</i>)	66/77	UnbU (unknown conserved protein in enediynes biosynthesis)	A0A1C6UU53
<i>sgdE8</i>	1901	Enediynes polyketide synthase [<i>Couchioplanes caeruleus</i>]	54/64	Enediynes Polyketide Synthase	A0A1K0GXS2
<i>sgdE7</i>	152	DynE7 (<i>Micromonospora chersina</i>)	65/74	Thioesterase	Q84HI7
<i>sgdS6</i>	224	Uncharacterized protein (<i>Micromonospora citrea</i>)	49/60	NADPH-dependent F420 oxidoreductase	A0A1C6W033
<i>sgdU6</i>	308	Uncharacterized protein (<i>Micromonospora haikouensis</i>)	57/72	Hypothetical protein	A0A1C4Y6G0

Figure S 46: Inactivation of *sgdU15* and *sgdX3* genes. Diagnostic PCR was performed to select for mutants with deletion of 570 bases and 514 bases from *sgdU15* and *sgdX3* genes, respectively.

Figure S 47: Inactivation of *sgdX1* gene. Diagnostic PCR was performed to select for mutants with deletion of 222 bases from *sgdX1* gene.

Figure S 48: Inactivation of *sgdX2* gene. Diagnostic PCR was performed to select for mutants with deletion of 462 bases from *sgdX2* gene.

Table S 15: NMR data of sungeidine (in DMSO-*d*₆)^a

Position	δ_C , multi. ^b	δ_H (J in Hz) ^c	HMBC ^c
2	58.2, CH	5.02 (d, 4.4)	C-3, 8, 23, 25, 29, 9
3	44.6, CH	3.32 (t-like, 4.4)	C-2, 4, 5, 6, 7, 8, 24, 30
4	158.3, C		
5	125.6, CH	5.96 (br s)	C-3, 6, 7, 30
6	197.8, C		
7	52.3, CH	3.61 (d, 2.4)	C-3, 6, 8, 24, 28, 29
8	40.5, CH	3.50 (m)	C-2, 29
9	53.5, CH ₂	Ha: 4.36 (dd, 12.0, 6.8) Hb: 3.44 (br d, 12.0)	C-2, 7, 8 C-3, 7, 8, 23
10	107.0, CH	6.72 (d, 8.2)	C-12, 21, 22, 23, 31
11	129.9, CH	7.94 (d, 8.2)	C-30, 13, 23, 31
12	117.1, C		
13	132.8, C		
14	122.2, C		
15	130.6, C		
16	103.5, CH	7.02 (d, 2.3)	C-14, 17, 18, 20
17	158.3, C		
18	120.3, CH	7.17 (dd, 9.2, 2.3)	C-16, 17, 20
19	132.3, CH	8.36 (d, 9.2)	C-15, 16, 17, 21

20	124.2, C		
21	136.3, C		
22	118.3, C		
23	152.6, C		
24	133.9, C		
25	136.5, C		
26	131.8, CH	7.07 (d, 7.5)	C-24, 25, 28
27	128.1, CH	7.20 (dd, 7.5, 7.5)	C-25
28	129.0, CH	7.19 (d, 7.5)	C-7, 24
29	131.0, C		
30	23.5, CH ₃	2.02 (3H, br s)	C-2, 3, 4, 5, 6
31	187.1, C		

^a Assignment were made by a combination of 1D and 2D NMR experiments. ^b Measured at 400 MHz.

^c Measured at 700 MHz.

Figure S 49: ¹H NMR spectrum (700 MHz, in DMSO-*d*₆) of sungeidine.

Figure S 50: ^{13}C NMR spectrum (400 MHz, in $\text{DMSO}-d_6$) of sungeidine.

Figure S 51: HSQC NMR spectrum (700 MHz, in $\text{DMSO}-d_6$) of sungeidine.

Figure S 52: TOCSY NMR spectrum (700 MHz, in DMSO- d_6) of sungeidine.

Figure S 53: HMBC NMR spectrum (700 MHz, in DMSO- d_6) of sungeidine.

Figure S 54: NOESY NMR spectrum (700 MHz, in DMSO- d_6) of sungeidine.

Figure S 55: High resolution mass spectrum of sungeidine.

Table S 16: NMR data of A1 (in DMSO- d_6 , 400 MHz)^a

Position	δ_c , multi.	δ_H (J in Hz)	COSY	HMBC
1	108.2, C			
2	130.8, C			
3	125.9, CH	7.58 (br s)		C-1, 5, 9, 11
4	136.3, C			
5	126.8, CH	6.76 (d, 9.3)	H-6, 7	C-3, 7, 9

6	136.3, CH	6.12 (br d, 9.3)	H-5, 7	C-4, 8
7	71.8, CH	4.41 (br d, 8.2)	H-5, 6, 8	C-5, 8
7-OH		5.60 (s)	H-7	
8	74.4, CH	4.81 (d, 8.2)	H-7	C-7, 9, 10
8-OH		6.43 (s)	H-8	
9	134.5, C			
10	130.1, C			
11	133.3, C			
12	133.3, C			
13	125.0, CH	7.75 (d, 7.4)	H-14	C-1, 12, 15
14	139.5, CH	8.36 (d, 7.4)	H-13	C-1, 2, 12
15	195.1, C			

^a Assignment were made by a combination of 1D and 2D NMR experiments.

Figure S 56: ¹H NMR spectrum (400 MHz, in DMSO-*d*₆) of A1.

Figure S 57: ^{13}C NMR spectrum (400 MHz, in $\text{DMSO}-d_6$) of A1.

Figure S 58: HSQC NMR spectrum (400 MHz, in $\text{DMSO}-d_6$) of A1.

Figure S 59: COSY NMR spectrum (400 MHz, in DMSO- d_6) of A1.

Figure S 60: HMBC NMR spectrum (400 MHz, in DMSO- d_6) of A1.

Figure S 61: ROESY NMR spectrum (400 MHz, in DMSO- d_6) of A1.

Figure S 62: High resolution mass spectrum of A1.

Figure S 63: High resolution MS/MS of A1 with collision induced dissociation set at 40 eV. Blue dot depicts product ion.

Publications

1. Low, Z. J.; Pang, L. M.; Ding, Y.; Cheang, Q. W.; Le Mai Hoang, K.; Thi Tran, H.; Li, J.; Liu, X.-W.; Kanagasundaram, Y.; Yang, L.; Liang, Z.-X., Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a *Streptomyces* strain isolated from mangrove sediment. *Sci. Rep.* **2018**, 8 (1), 1594.