SYNERGIZING MICROBIAL CULTURING, WHOLE GENOME SEQUENCING, ASYMMETRIC SYNTHESIS AND TANDEM MS FOR RECONSTRUCTION OF POLYKETIDE AND ALKALOID NATURAL PRODUCT BIOSYNTHESIS IN MARINE ACTINOMYCETE Nocardiosis sp. CMB-M0232

By
Norah Faihan Alqahtani

Submitted in partial fulfillment of the requirements for
the Degree of Doctor of Philosophy

Thesis Advisor: Rajesh Viswanathan, Ph.D.

Department of Chemistry
CASE WESTERN RESERVE UNIVERSITY
May, 2015
CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

(Norah Faihan Alqahtani)

candidate for the degree of (Doctor of Philosophy) *.

   Committee Chair

(Professor Robert Salomon)

   Committee Member

(Professor John Protasiewicz)

   Committee Member

(Professor Gregory Tochtrop)

   Committee Member

(Professor John Mieyal)

   Date of Defense

(Dec, 3rd 2014)

*We also certify that written approval has been obtained for any proprietary material contained therein.
Dedicated to

MyParents,

Naif (my husband),

Meshari and Fahad (my dear sons)

- you have shared this amazing and tough journey with me! Thanks!
Table of Contents

Table of Contents ........................................................................................................................................5

Abstract ......................................................................................................................................................10

Acknowledgment .......................................................................................................................................12

Acknowledgment (Arabic version) ...........................................................................................................15

List of Figures ..........................................................................................................................................17

List of Tables ............................................................................................................................................21

List of Schemes .......................................................................................................................................22

List of Symbols and Abbreviations .........................................................................................................23

CHAPTER 1: Synergistic Involving Genome Sequencing, Tandem Mass Spectrometry and Bio-Inspired Synthesis to Gain Biosynthetic Insights on the Genesis of Marine Actinomycetales Natural Products .................................................................27

1.1. Overview ..............................................................................................................................................28

1.2. Ocean-Derived Actinomycetes and their Importance in Metabolic Research .........................28

1.3. The Influence of Microbial Genomics on Drug Discovery ..........................................................36

1.4. Genome Mining for Secondary Metabolite Biosynthesis in Actinomycetes ....................40

1.5. An Integrated Approach to Complement Genome Sequencing for Nocardiopsis sp. CMB M0232 Reveals a Rich Biosynthetic Diversity .........................................................................................43
1.6. The Rise of Mass Spectrometry for Secondary Metabolic Pathway Studies

CHAPTER 2: Insights into Biogenesis of Nocardioazine Alkaloids

2.1. Why are Nocardioazines Important?

2.2. Biosynthetic Hypothesis towards Nocardioazine A and B

2.3. Results

2.3.1. Nocardiopsis sp. CMB-M0232 draft genome sequence and bioinformatics-based prediction of the noz gene cluster

2.3.2. Asymmetric synthesis of a suite of Noz pathway intermediates

2.3.3. Evaluation of the biosynthetic relevance of synthesized intermediates through NMR, LC-MS and HR-tandem MS reveals precursor-product relationships for nocardioazine B biosynthesis

2.4. Discussion

2.5. Conclusion

Acknowledgements and Author Contributions

CHAPTER 3: Cyclodipeptide Synthase-Catalyzed Chemoenzymatic Synthesis of Bioactive Natural Products Belonging to the Diketopiperazine Family

3.1. General Discussion on Microbial Biocatalysts for Organic Synthesis

3.2. NozA is a Cyclodipeptide Synthase Involved in DKP Formation in the Nocardioazine Pathway

Acknowledgements and Author Contributions
CHAPTER 4: Characterization of the *nsn* Biosynthetic Gene Cluster for the
Nocardiopsin Family from the Marine-Derived *Nocardiopsis sp*. CMB M0232 ..... 100

4.1. Introduction to Nocardiopsin Macrolides ......................................................... 101

4.2. Cloning and Sequencing of the Nocardiopsin Gene Cluster (*nsn*) ...................... 104

4.3. Bioinformatics Analyses Support Assembly of the Nocardiopsin Core by Polyketide Synthases NsnA-NsnD and Nonribosomal Peptide Synthetase NsnE ...... 106

4.4. A Cytochrome P450 and Epoxide Hydrolase are Proposed to Act Upon a Megasynthase-Tethered Nocardiopsin Precursor to Stage it for Macrolide Closure.. 108

4.5. Biosynthesis of the Tetrahydrofuran (THF) Group of Nocardiopsins A-D .......... 111

4.6. Pipecolic Acid Biosynthesis is Catalyzed by Lysine Cyclodeaminase Nsn .......... 113

4.7. The Novel Structural Features of Nocardiopsins are Underpinned by a Unique Biosynthetic Pathway................................................................. 117

Acknowledgements ........................................................................................................ 118

CHAPTER 5: Experimental Section ........................................................................ 119

5.1. Culturing of *Nocardiopsis sp*. CMB M0322 and Extraction of Secondary Metabolites ................................................................................................................. 120

5.2. *Nocardiopsis sp*. CMB-M0232 Genomic DNA Isolation, Draft Genome Sequencing and Pathway Annotation................................................................. 121

5.3. Cloning and Sequencing of the Nocardiopsin Gene Cluster (*nsn*) ............. 122

5.4. Bioinformatics Analyses of the Nocardiopsin Gene Cluster ............................ 123
5.5. Preparation of the Construct for Expression of Lysine Cyclodeaminase Gene NsnL in *E. coli*. ..................................................................................................................... 124

5.6. Overexpression and Purification of Lysine Cyclodeaminase NsnL .......... 125

5.7. Evaluation of the Function of NsnL Through in Vitro Assays and GC-FID ...... 127

5.8. General Experimental and Instrumentation for Asymmetric Synthesis and HPLC-MS-MS ...................................................................................................................................... 127

5.9. Structural Characterization of Synthetic and Biosynthetic Intermediates. ........ 128

5.10. Synthesis of cyclo-L-Trp-L-Trp DKP and cyclo-D-Trp-D-Trp DKP .......... 138

5.11. Asymmetric Synthesis of cyclo-C3-Me-L-Trp-L-Trp DKP (13) .............. 144

5.12. Asymmetric Synthesis of cyclo-C3-Me-L-Trp-N1'-Me-L-Trp DKP (16) ...... 153

5.13. Synthesis of cyclo-L-Trp-C3'-'prenyl-L-Trp DKP (12) and cyclo-L-Trp-N1'-Me-C3'-'prenyl-L-Trp DKP (15) ................................................................................................................................. 159

5.14. Asymmetric Synthesis of cyclo-L-Trp-N1'-Me-C3'-'prenyl-L-Trp DKP (15) .. 163

5.15. Conditions for HRMS, HPLC, LC-MS and MS2 Characterization and Identification of signature peaks. ........................................................................................................ 167

5.16. NMR for 13 from *Nocardiopsis* Extract ............................................. 186

5.17. Synthesis of Malonyl SNAc Thioester ..................................................... 189

5.18. Synthesis of Methyl Malonyl SNAc Thioester ...................................... 191

5.19. Synthesis of D/L-Trp CME Ester (47) ...................................................... 194

**APPENDIX** ........................................................................................................ 197

Appendix 1:Spectral Data .................................................................................... 198
Appendix 2: An Overview of the History and Current Relevance of Natural Products

A2.1.1. Why Natural Products are important? .............................................................. 312
A2.1.2. Are Marine Microbiota Sources for Novel Drugs? ........................................ 318
A2.1.3. Bioactive Metabolites from Marine Derived Actinomycetes ....................... 324
Bibliography ............................................................................................................. 329
ABSTRACT

Synergizing Microbial Culturing, Whole Genome Sequencing, Asymmetric Synthesis and Tandem MS for Reconstruction of Polyketide and Alkaloid Natural Product Biosynthesis in Marine Actinomycete Nocardiopsis sp. CMB-M0232

BY NORAH ALQAHTANI

Biosynthetic pathway engineering is rapidly growing by rationally harnessing the enzymatic potential of microbial systems. While marine cyanobacterial genus and a few gram positive bacteria (such as Saccharopolyspora erythraea) have offered an extensive array of promising biocatalysts with unique modular functions as polyketide synthases (PKS), non-ribosomal peptide synthetases (NPRS) or their hybrids, the biosynthetic potential of soil cyanobacteria and marine Nocardiopsis genus largely remain unexplored. Herein we study a marine actinomycete, Nocardiopsis sp. CMB-M0232 isolate as a model organism through an integrated approach involving genome sequencing, metabolic engineering, tandem-MS and asymmetric synthesis to reconstruct multiple biosynthetic pathways leading to PKS, NRPS, alkaloids and their hybrids as potential candidates for clinically relevant drug development. Chapter 1 and Appendix 2 introduce the reader to the significance of marine actinomycetes in the context of drug discovery and development. Chapters 2, 3 and 4 describe multiple recent success stories on the mechanistic investigations of novel biocatalytic systems discovered through this
integrated approach applied to *Nocardiopsis*, for the first time. Specifically, Chapter 2 describes the metabolic pathway to nocardioazines that are candidates for anti-tumor drug development through their participation as inhibitors of p-glycoprotein-mediated efflux. Their biosynthetic pathway is dissected. Chapter 3 builds on the *noz* gene cluster and NozA as a biocatalyst. Chapter 4 describes the *nsn*-coded modular megasynthase-based biosynthetic machinery for the assembly of the nocardiopsins. The nocardiopsins are polyketide-non-ribosomal peptide hybrid natural products that are structurally homologous to the rapamycins but are indeed distinct at select positions. Therefore, the nocardiopsin biosynthetic pathway is not only interesting and significant for the potential of creating new mTOR pathway analogs for their biological elucidation. But fundamentally is interesting for biosynthesis of heterocyclic rings such as tetrahydrofurans, pipecolates and proline-containing macrolide architectures. Chapter 5 describes the experimental details, and the corresponding NMR spectra are provided in Appendix 1. Overall, several novel lines of investigations are underway and the data gathered and presented herein constitutes the body of work for three distinct publications.
ACKNOWLEDGEMENTS

A dissertation is not the outcome of the efforts of a single mind. Many people have contributed to its development. At this time, I take the opportunity to acknowledge those who have made significant impact on my doctoral journey. First and foremost, I thank God for the numerous blessings He has bestowed throughout this dissertation journey. I would like to express my special appreciation to my advisor Prof. Rajesh Viswanathan, for being a tremendous mentor. You have been supportive since the days I began studying in the Chemistry department as a master’s student. I would like to thank you for encouraging my research and for allowing me to grow as a scientist. Ever since, you have supported me not only by providing assistance over almost six years, but also academically and emotionally through the rough road to finish this thesis. Your advice on both research as well as on my career have been priceless. I could not have asked for better role models, for being inspirational, supportive, and patient.

I would like to thank my committee members: Professors. Robert Salomon, John Protasiewicz, Gregory Tochtrop and John Mieyal, for creating a collegial environment for doing research and providing invaluable guidance and assistance. Very special thanks go out to my master’s advisor, Prof. Gregory Tochtrop, without whose motivation and encouragement, I would not have considered a graduate career at CWRU.

I greatly appreciate the timely and inspiring assistance from our collaborator Prof. Amy Lane (University of North Florida). Dr. Jonathan Karty (Indiana University) and Elle James. I thank the members of the Viswanathan group: Dr. Ormond Brathwaite, Dr.
Karthikeyan Thandavamurthy, Deepti Sharma, Brittney Bunn, Sarah Huff, Jade Moten and Deena Ibrahim, thank you for listening, offering me advice, exchanges of knowledge, skills, and venting of frustration during my graduate program, which helped enrich the experience and supporting me through this entire process. I thank the faculty and staff in the Chemistry Department who were always equipped with kind words of encouragement and advice as I progressed my way through the graduate process. Especially, graduate affairs coordinator, June Ilhan and Suzi Mason.

I am extremely grateful to my parents Mom and Dad, for their love, prayers, caring and sacrifices for educating and preparing me for my future and the continuous support they have given me throughout my time in graduate school; I could not have done it without them. I cannot begin to express my unfailing gratitude and love to my sisters and brothers, Hessah, Fahad, Haya, Sara, Muflih and Mohammed, who have supported me throughout this process and constantly encouraged me when the tasks seemed arduous and impossible.

As for my husband Naif, I find it difficult to express my appreciation because it is so boundless. He is my most enthusiastic cheerleader; he is my best friend; and he is an amazing husband and father. Without his love, support, and continued patience my contribution to this project would not be possible. He has been my biggest supporter through thick and thin, and I have blossomed because of that. I am grateful to my husband not just because he has given up so much to make my career a priority in our lives, but because he has seen me through the ups and downs of my entire research process.
I owe more than words could ever express to my lovely sons, Meshari and Fahad, for their sacrifice, patience, support and faith in me. They often had to endure my absence but they seldom complained about the evenings away from my family while attending classes, studying for exams, while doing experiments in lab (all at a time they needed my help doing their math homework or science projects). I hope they will enjoy the effort being completed. Naif, Meshari and Fahad, you have shared this entire amazing and tough journey with me, so it only seems right that I dedicate this dissertation to you. Finally, this research would not have been possible without the financial assistance and funding my scholarship from: Saudi Arabian Culture Mission and Ministry of Higher Education in Saudi Arabia and for that I thank Dr. Rashed Alsedran, and Dr. Elberier Mohammed.
ACKNOWLEDGEMENTS (Arabic version)

بسم الله الرحمن الرحيم

الحمد لله ذي المن والفضل والإحسان، حمدًا يليق بلجلائه وعظمته، وصلّى الله على خاتم الرسل، ومن لا بغي بعده، صلاة تقضي لنا بها الحاجات، وترفعنا بها أعلى الدرجات، وتبلغنا بها أقصى الغيابات من جميع الخيرات، في الحياة وبعد الممات، وله الشكر أولاً وأخيراً، على حسن توفيقه، وكرمه عونه، وعلى ما من وفتح به عليّ من إنجاز لهذه الرسالة، بعد أن يسر العسير، وذلّ الصعب، وفرزج الهم، وعلى تفضّلله عليّ بالذين كريمين شفاً لتي طريق العلم.

وكانت خير سند لي طيلة حياتي الدراسية من تشجيع ودعاية وصبر وعطاء، والذين العزيز فيجان مفلى القحطاني قدوني الأولى من الصغر ومن غرس في عمالي حب العلم والطموح، وإن لي مستحيل مع العلم والإصرار لطالما كنت الاب المختلف دائماً وأبداً الذي أفر واباهي به عامي جميع، فالاختلاف تميز الاب الذي يكون صديقاً قريباً في وقت الآلام، وأباب الحكيم وقت الرأي، المشورة دائماً تتفقالي واحوتي كبير هلمك، وحسن توجيهك بدون ضجر ولا تعب. أمي الغالية على مجلد القحطاني، امرأة التفاني والمشاعر والإشارة، انتي الامهات من بعدك، تختنق كلماتي وتعثر على اعتاب ماقدمتيه بصاحبة القلب الحنون والدعوات الصادقة، فانتي قد نذلت العمر واحمل السنين لتكوني الحصن المنيع والركن المكين لي ولأخوتي من غير الظروف والمحن، أمي يا متشبع الماضي، والإنسي الذي اIndependent به ليذل العقول والصعاب. انتي نعمه عظيمه من الله عليّ بما فعلتي إدرك يوماً ردها شكرها، والذي الحبيبين، وابن أوقدا في عمة الجهل شماعة مضيئة، وفتحما من مغالق الأبواب، باباً للخير مشرعاً، وأنظماً في أعماق الغفوات إدراكاً وحساً مرهفاً. امل ان يكون حصولي على هذا الدرجة العلميه هو نتاج غركما، وعلامة الأمر، وانغ وانتما تقفانه فهو احق لكما من نفسي.

على طبق من ذهب اهدي هذا النتاج إلى رفيق دربي ومسيري صديقي الغالي وزوجي العزيز نايف مسعود القحطاني، انت من شركاني التجربة لحظة بلحظة، وقاسمي الفرح والحزن في الغربة. كنت ومازلت دعمي المعنوي والنفسى فانت من يلمس لي العذر عند انشاعي الأيام، والليالي في دراستي بدون شكوى أو تذمر، انت من أحادتني و
بحبك واحترامك وتقديره، طيلة فترة الدراسة، ذلك من حبي أصدقه، ومن وفاني أغدقه، ومن شكري أجمله فلن اجد
لتضحيتك مثلك ولن استطيع ردك لإبالتكم ولا الدراست العلمية.

شكرني وحبك إلى فلذتي كبدتي مشاري وفهم لتحملهم معي المشق والجهد بانشغالي الدائم. امتد انجازي ومستقبلي
وشمحة حياتي اسلام الله ان بحفظكم ويرعاكم.

إلى جداتي الجزليات و شعفة شكره لكل الدعوات المخلصة والصادقة التي كانت ولا زالت تقطع سكون الليل.
جعلها الله في ميزان حسناتكم وأمد في عمركم بالعمل الصالح.

إلى اختي وصديقتي المميزة حصة صاحبة القلم الرشيق لكم تهاني بين أمثالك حروف الهجاء حرفًا يزلوه حرف ،
ولكم نفانت بين يدي مذكراتك الأقلام قلما ينزلوه قلم، لتحبي الامل في النفس المتعبة وتجلي هم الغربة وشوق الأهل.
والديار بكلمات تعذنا دائما بالنجاح لتحقيق الأمال.

وإلى اخي فيد بعيد الديار القريب بقليبه ونصبح فيه يشاركني ويشارك اخوتي الفرح بالانجاز والدعم حاال العثرات.

أتقدم بالشكر إلى إخواني واخوتي هيا ساره مفطح ومحمد. رققاط الورد والمسيرة في ظلم العلم لوقتكم بجانبي.
وعماتهم الصادقة.أيمن شركني السير في دروب المعرفة كونوا كما عهدتم خير خلف سلفا مما نناحكم الا
استمرار للنجاحي. وإلى اخواتي من ورتلكم هناك وآخائنا وعياء وإخواتي الصغيرات شكرنا لكم من الإعماق على
الدعاء الصادقات.

أخيرا وليس آخر إلى فيد وزوجه اسمه رحمهما الله رحمة واسعة كانت الاقدر اسرع قبل ان تنحل عيناي برويتكما
مجداً. رحمهما الله ومن عليكم بجنة عدن وذكرى طيبة لانتطاع.

الصلاة والسلام علي اشرف من وطئ الہرى محمد صلى الله عليه وسلم.
List of Figures

Figure 1: Structures of related compounds isolated from actinomycete ......................... 33
Figure 2: Chemical structure of Trabectedin ET-743 ...................................................... 34
Figure 3: Chemical structure of Kahalalide ..................................................................... 35
Figure 4: Chemical structures of the non-ribosomal peptides ......................................... 39
Figure 5: Chemical structures of Salinosporamide A and K, and Salinilactam A .......... 42
Figure 6: Illustration of an integrated cross-talk approach for exploiting Nocardiopsis
metabolic diversity ........................................................................................................... 43
Figure 7: Biosynthetic repertoire of marine-derived actinomycete Nocardiopsis sp. CMB
M0232. A. Genome draft. B. Metabolites that have been characterized, C. List of
bioinformatics-derived pathways identified. .................................................................... 44
Figure 8: Biosynthetic diversity of Nocardiopsis sp. CMB-M0232 in a chemical
representation ................................................................................................................... 45
Figure 9: Organization and bioinformatics-predicted functions of the two clusters of
Nocardiopsis sp. CMB-M0232 biosynthetic genes (noz) .................................................... 55
Figure 10: A. Biosynthetic events in the nocardioazine pathway and the genes encoding
for the respective steps. B. Precursors and products predicted for mid-stages of the
pathway involving C3-methylation, C3′-prenylation and N1′-methylation. ..................... 63
Figure 11: A. NOESY correlations for (16). B. 1H-1H LR-COSY correlations for (16) at 600 MHz. C. Chem-3D rendered ball and stick model of (16).

Figure 12: A. Extracted ion chromatogram (EIC) traces of synthetic standards. ESI TOF-MS² fragmentation data: B. cyclo-L-Trp-L-Trp DKP (5) (M+H)+; C. cyclo-L-Trp-C3′-prenyl-L-Trp DKP (12) (M+H)+; D. cyclo-C3-Me-L-Trp-L-Trp DKP (13) (M-H)-; E. cyclo-N1′-Me-L-Trp-L-Trp DKP (14) (M+H)+; F. cyclo-L-Trp-N1′-Me-C3′-prenyl-L-Trp DKP (15) (M+H)+; G. cyclo-C3-Me-L-Trp-N1′-Me-L-Trp DKP (16) and H. des-N1′-Me nocardioazone B (17).

Figure 13: A. Reverse phase HPLC traces for synthesized and extracted metabolites from Nocardiopsis sp. CMB-M0232. B. Culture of Nocardiopsis sp. CMB-M0232 at 7, 14 and 21 days. C. ESI-(−)-TOF-MS¹ and MS² spectrum of 13 from Nocardiopsis sp. CMB-M0232; D. ESI-(+)-TOF-MS¹ and MS² spectrum of 13 from Nocardiopsis sp. CMB-M0232.

Figure 14: Recently characterized noz cluster for the biosynthesis of nocardioazines.

Figure 15: A. Amino acid sequence alignment between NozA and Amir_4627 and bioinformatics model of NozA B. NozA-catalyzed formation of cyclo-L-Trp-L-Trp (5). Phyre 2.0 was used for rendering of the NozA structure.

Figure 16: NozA phylogenetic relationship.

Figure 17: Cloning of NozA for expression.

Figure 18: Enzyme assay for NozA.

Figure 19: Top: Structures of Streptomyces-derived pipecolate-containing macrolides. Bottom: Nocardiopsin family of actinobacterial pipecolate or proline-containing macrolides.
Figure 20: Organization of the \(\sim 77\) kB \(\text{nsn}\) nocardioopsin biosynthetic gene cluster. 104

Figure 21: Schematic of the nocardioopsin biosynthetic pathway... 108

Figure 22: A. Preliminary data displaying protein purification for NsnL, a L-lysine cyclodeaminase. B. Reconstitution assay for NsnL in nocardioopsin biosynthesis. 114

Figure 23: GC-FID analysis of products resulting from incubation of selected amino acid substrates with lysine cyclodeaminase NsnL. (a) Incubation of NsnL with L-lysine resulted in accumulation of L-pipecolic acid. (b) Incubation of NsnL with L-ornithine resulted in accumulation of L-proline. 115

Figure 24: Schematic of pipecolate (red) incorporation into the nocardioopsins polyketide chain. A. The presumed overall reaction and B. mechanism for NsnL, a lysine cyclodeaminase. 116

Figure 25: LC-ESI-TOF MS spectra for biosynthetic intermediates in their purely synthesized state. A: LC-ESI (+) ions from mixture of synthesized standards 5, 13, 15, and 16; B: LC-ESI (-) ions from mixture of synthesized standards 5, 13, 15, and 16. C. Extracted ion chromatograms for positive ion tandem mass spectra of 5, 8, 10, and 11 from biological extract. 171

Figure 26: Two diastereomers of \(\text{cyclo-L-Trp-C3\text{'}-a-prenyl-L-Trp DKP (12)}\) were synthesized and their LC traces were observed as distinct peaks at \(R_t\) 9.37 (minor) and 12.11 (major) min respectively. 171

Figure 27: \(\text{MS}^1\) and \(\text{MS}^2\) for two diastereomers of \(\text{cyclo-L-Trp-C3\text{'}-a-prenyl-L-Trp DKP (12)}\). 172
Figure 28: *cyclo*-L-Trp-L-Trp DKP (5). A. LC-ESI (-)-TOF- MS$^1$ and MS$^2$ spectrum from *Nocardiopsis* sp. CMB-M0232. B. ESI (-) MS$^2$ for 5 extracted from *Nocardiopsis* sp. CMB-M0232. C. ESI(-) MS$^2$ for synthetic 5. .......................................................... 178

Figure 29: LC- ESI (+)-TOF- MS$^1$ and MS$^2$ traces from *Nocardiopsis* sp. CMB-M0232 confirming presence of tabulated metabolites as given in Table 1 for (5) and (13) ....... 179

Figure 30: LC-ESI (+)-TOF- MS$^1$ and MS$^2$ traces from *Nocardiopsis* sp. CMB-M0232 confirming presence of tabulated metabolites as given in Table 1 for (16); (des-N1’-Me-nocardioazine B (17)) and (nocardioazine B (4)). .......................................................... 180

Figure 31. ESI (+)-TOF MS$^1$ and MS$^2$ spectra for (5) extracted from *Nocardiopsis* sp. CMB-M0232. ...................................................................................... 181

Figure 32: ESI (+)-TOF MS$^1$ and MS$^2$ spectra for (13) extracted from *Nocardiopsis* sp. CMB-M0232. *Cyclo-C3-Me-L-Trp-L-Trp* DKP (13). .................................................... 182

Figure 33. LC-ESI+ TOF MS$^1$ and MS$^2$ spectra for (nocardioazine B, 4) extracted from *Nocardiopsis* sp. CMB-M0232. ...................................................................................... 183

Figure 34. ESI (+)-TOF MS$^1$ and MS$^2$ spectra for nocardioazine B (4) from *Nocardiopsis* sp. CMB-M0232. ............................................................................................................ 184

Figure 35. TLC images of intermediates extracted from *Nocardiopsis* sp. CMB M0322 corresponding to alkaloidal fractions. .......................................................... 185

Figure 36. $^1$H NMR spectrum for 13 extracted from *Nocardiopsis* sp. CMB-M0232 after 21 days of culture followed by fractionation and purification. ............................................. 186

Figure 37. $^{13}$C NMR spectrum for 13 extracted from *Nocardiopsis* sp. CMB-M0232. 187

Figure 38: HPLC analyses of *cyclo*-L-Trp-L-Trp-DKP (5). .................................................... 188

Figure 39: HPLC analyses of *cyclo*-D-Trp-D-Trp-DKP (ent-5). .................................................... 188
List of Tables

Table 1: Predicted functions of putative nocardioazine biosynthetic enzymes based on bioinformatics analyses................................................................. 57

Table 2: LC-MS and MS1 and MS2 data for synthetic and extracted biosynthetic intermediates and products. ................................................................. 79

Table 3: Proposed functions of proteins from the Nocardiopsis sp. CMB-M0232 nocardopsin gene cluster (nsn) based on comparison of amino acid sequences with homologous enzymes............................................................................................................. 105

Table 4: List of chemical structures and corresponding numbers assigned................................................................. 131

Table 5: MS2 fragments for [M+H]+ peak at m/z = 373.1659 corresponding to cyclo-L-Trp-L-Trp DKP (5). .................................................................................... 173

Table 6: MS2 fragments for [M+H]+ peak at m/z = 387.1816 corresponding to cyclo-C3-Me-L-Trp-L-Trp DKP (13)......................................................................................................................... 174

Table 7: MS2 fragments for [M+H]+ peak at m/z = 401.1972 corresponding to cyclo-C3-Me-L-Trp-N1’-Me-L-Trp DKP (16)......................................................................................................................... 175

Table 8: MS2 fragments for [M+H]+ peak at m/z = 441.2285 corresponding to cyclo-L-Trp-C3’-nprenyl-L-Trp DKP (12) ......................................................................................................................... 176

Table 9: MS2 fragments for [M+H]+ peak at m/z = 455.2442 corresponding to Cyclo-L-Trp-N1’-Me-C3’-nprenyl-L-Trp DKP (15)......................................................................................................................... 177

Table 10: Novel metabolites produced by marine actinomycetes from 2003-2010. .................. 310
List of Schemes

Scheme 1: A. Structures of nocardioopsins A (1) and B (2), nocardioazines A (3) and B (4), cyclo-L-Trp-L-Trp (5) and cyclo-L-Trp-D-Trp (6). Nocardioazine skeletal numbering and ring labels are shown. B. Individual events of the proposed biosynthetic pathway to (3). Steps leading to (4) are the focus of this work................................................................. 52

Scheme 2: Synthesis of cyclo-L-Trp-L-Trp DKP (5) and cyclo-D-Trp-D-Trp DKP (ent-5).................................................................................................................................................. 60

Scheme 3: Synthesis of cyclo-C3-Me-L-Trp-L-Trp DKP (13) and cyclo-C3-Me-L-Trp-N1′-Me-L-Trp DKP (16)............................................................................................................... 67

Scheme 4: Synthesis of cyclo-L-Trp-C3′-"prenyl-L-Trp DKP (7) and cyclo-L-Trp-N1′-Me-C3′-"prenyl-L-Trp DKP (10).................................................................................................................................................. 71

Scheme 5: Synthesis of cyclo-N1′-Me-L-Trp-L-Trp DKP (14) and Des-N1′-Me-nocardioazine B (17)..................................................................................................................... 72

Scheme 6: Summary of the biosynthetic pathway for mid-stages of nocardioazine biosynthesis. Dotted lines show originally hypothesized possibilities and bold lines show the path that is evident from HPLC, LC-MS, MSn analyses for all relevant intermediates and by NMR (for 13 only)........................................................................................................................................ 81

Scheme 7: Synthesis of CME-esters ................................................................................................................................. 99
# List of Symbols and Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIBN</td>
<td>2,2’-azobisisobutyronitrile</td>
</tr>
<tr>
<td>anhyd.</td>
<td>anhydrous</td>
</tr>
<tr>
<td>AT</td>
<td>acyltransferase</td>
</tr>
<tr>
<td>ACP</td>
<td>acyl carrier protein</td>
</tr>
<tr>
<td>BINOL</td>
<td>1, 1’-bi-2, 2’-naphthol</td>
</tr>
<tr>
<td>BLASTP</td>
<td>protein basic local alignment search tool</td>
</tr>
<tr>
<td>(BOC)₂O</td>
<td>di-tert-butyl dicarbonate</td>
</tr>
<tr>
<td>BOP-Cl</td>
<td>bis(2-oxo-3-oxazolidinyl)phosphonic chloride</td>
</tr>
<tr>
<td>Calcd</td>
<td>calculated</td>
</tr>
<tr>
<td>Cbz-Cl</td>
<td>benzyloxycarbonyl chloride</td>
</tr>
<tr>
<td>CDPS</td>
<td>cyclodipeptide synthase</td>
</tr>
<tr>
<td>CI</td>
<td>chemical ionization</td>
</tr>
<tr>
<td>cm (NMR)</td>
<td>complex multiplet</td>
</tr>
<tr>
<td>cm⁻¹</td>
<td>reciprocal centimeters</td>
</tr>
<tr>
<td>CNBA</td>
<td>4-chloro-3-nitrobenzoic acid</td>
</tr>
<tr>
<td>¹³CNMR</td>
<td>carbon-13 nuclear magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>Cys</td>
<td>cysteine</td>
</tr>
<tr>
<td>d</td>
<td>doublet</td>
</tr>
<tr>
<td>dd</td>
<td>double doublet</td>
</tr>
<tr>
<td>ddd</td>
<td>double double doublet</td>
</tr>
<tr>
<td>DH</td>
<td>dehydratase</td>
</tr>
<tr>
<td>DKP</td>
<td>diketopiperazine</td>
</tr>
<tr>
<td>DMAc</td>
<td>N,N-dimethylacetamide</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-dimethylaminopyridine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>DMDO</td>
<td>dimethyldioxirane</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DMSO-d⁶</td>
<td>deuterated dimethyl sulfoxide</td>
</tr>
<tr>
<td>D₂O</td>
<td>deuterium oxide</td>
</tr>
<tr>
<td>dt</td>
<td>double triplet</td>
</tr>
<tr>
<td>D-Trp</td>
<td>D-tryptophan</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EH</td>
<td>enoyl reductase</td>
</tr>
<tr>
<td>FKBP</td>
<td>FK binding protein</td>
</tr>
<tr>
<td>FRB</td>
<td>FKBP12-rapamycin binding</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>HPLC</td>
<td>high-pressure liquid chromatography</td>
</tr>
<tr>
<td>HR-MS</td>
<td>high-resolution mass spectrometry</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>KR</td>
<td>ketoreductase</td>
</tr>
<tr>
<td>L</td>
<td>liter</td>
</tr>
<tr>
<td>LC-MS</td>
<td>liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>LiCl</td>
<td>lithium chloride</td>
</tr>
<tr>
<td>L-Trp</td>
<td>L-tryptophan</td>
</tr>
<tr>
<td>Lys</td>
<td>Lysine</td>
</tr>
<tr>
<td>m</td>
<td>meter</td>
</tr>
<tr>
<td>(M)⁺</td>
<td>molecular ion</td>
</tr>
<tr>
<td>MHz</td>
<td>mega Hertz</td>
</tr>
<tr>
<td>m/z</td>
<td>mass-to-charge ratio</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>mL</td>
<td>milliliter</td>
</tr>
</tbody>
</table>
mM  millimolar
mm  millimeter
mmol  millimole
mol  mole
Mol Wt.  molecular weight
Mp  melting point
mRNA  messenger RNA
MSD  mass spectrometric detector
mTOR  mammalian target of rapamycin
μL  microliter
μM  micromolar
μm  micron
μm²  square micron
NaPi  sodium phosphate buffer
N-Boc  N-tert-butoxycarbonyl
NCS  N-chlorosuccinimide
NMR  nuclear magnetic resonance spectroscopy
NP  natural product
NRPS  non-ribosomal peptide synthetase
PS  pyran synthase
PDE  phosphodiesterase
PKS  polyketide synthase
Rf  retention factor
RNA  ribonucleic acid
rpm  revolutions per minute
RT  room temperature
satd.  saturated
s  singlet
sec$^{-1}$ reciprocal second
Ser serine
t triplet
t-$ButOH$ t-Butyl alcohol
THF tetrahydrofuran
TLC thin layer chromatography

**Chemical Formula list:**

- **CDCl$_3$** deuteratedchloroform
- **CH$_2$Cl$_2$** dichloromethane
- **H$_2$O$_2$** Hydrogen peroxide
- **H$_2$SO$_4$** Sulfuric acid
- **HCl** Hydrochloric acid
- **KCl** potassium chloride
- **KOH** potassium hydroxid
- **NaCl** sodium chloride
- **NaH** sodium hydride
- **NaHCO$_3$** sodium bicarbonate
- **NaI** sodium iodide
- **Na$_2$SO$_4$** sodium sulfate
- **SOCl$_2$** thionyl chloride
CHAPTER 1

Synergistic Involving Genome Sequencing, Tandem Mass Spectrometry and Bio-Inspired Synthesis to Gain Biosynthetic Insights on the Genesis of Marine Actinomycetales Natural Products
1.1. Overview

Human history continues to be rewritten due to the seminal role played by biodiversity-derived products, called natural products (NPs). This chapter outlines a synergistic approach towards maximizing the biosynthetic potential of marine actinomycete Nocardiopsis. Overall this thesis is inspired by the rich history that the field of NPs have. Their place in history and current relevance are discussed in Appendix 2, for brevity. The remaining portion of this chapter sets the stage for biosynthesis of secondary metabolites in marine-actinomycete as the focus. In chapters 2, 3 and 4, data is presented that conveys the general message that gram positive actinobacteria has much more in store as drug discovery leads through novel biosynthetic mechanisms, than what we currently have discovered. The supplementary experimental details and spectral evidence for each experiment are provided in chapter 5 and appendix 1. The thesis is presented in a modular fashion, where each chapter could stand on its own and therefore the molecular references to structures, schematics and associated references are independent of each chapter.

1.2. Ocean-Derived Actinomycetes and their Importance in Metabolic Research

Gram-positive actinobacteria are responsible for almost half of the currently known bioactive secondary metabolites that are pharmacologically active compounds of unrivaled structural diversity,\(^1,^2,^3\) thus making them the most valuable prokaryotes from

---

an economical and biotechnological standpoint. Owing to their excellent track record in the discoveries of not only microbial natural products like immunosuppressive agents, but also, among a long list; tumor agents, insecticides, antioxidants, enzymes, and antibiotics like vancomycin, tetracycline, and erythromycin. There has been considerable effort put into the discovery and successful isolation of novel terrestrial actinomycetes for drug screening purposes over the last 50 years. With most life on earth finding its roots in the sea, and most of the earth’s surface being covered with water, it comes as no surprise that marine ecosystems have an extremely high and unrivaled biological diversity. The genetic and metabolic diversity of marine actinomycetes remains largely unknown mostly due to the unexplored distribution of actinomycetes in the oceans. However, it is speculated that these characteristics must be a reflection of the harsh conditions to which marine actinomycetes have had to adapt during their evolutionary history. Marine environments range from anaerobic conditions at below zero degrees centigrade on the deep sea floor, which have high pressure, minimum of light, variable salinity, to highly acidic conditions at temperatures of over 100 degrees centigrade near hydrothermal vents at the mid-ocean ridges. While terrestrial actinomycetes have acted as successful sources of drug discovery, until recently, the reverse has been true for marine actinomycetes. The skepticism behind marine actinomycetes stemmed from the fact that...

terrestrial bacteria are responsible for the production of resistant spores which get transported from land into the sea, where they stay dormant, but viable. As such the assumption, until recently, was that whatever actinomycetes were identified and isolated from marine natural products, were actually the wash-off from terrestrial actinomycetes. Such skepticism was fueled by findings like the discovery of similarities between the structures of *Mycale hentscheli* collected in Dunedin Harbor, South Island, New Zealand, and pederine, a toxin originally isolated from the terrestrial *Paederus* beetle in South America. Another example is ecteinascidin 743 that was isolated from the tunicate *Ecteinascidia turbinate*, and bears a striking resemblance to cyanoclaine and saframycin, which are derivatives of several terrestrial actinomycetes.

Over the last decade, evidence has come into light for the presence of indigenous marine actinomycetes. This has generate enough curiosity to spur the search for novel marine actinomycetes in the marine environment. Extensive research has unveiled the fact that the marine environment is almost an untapped source of microbes, and that contrary to popular belief, the oceans and seabed are not microbial deserts, but are inhabited by a profusion of microbes covering the 3 domains; bacteria, archaea, and eukarya. The earliest evidence for the presence of marine actinomycetes came from the description of *Rhodococcus marinonascene*, which was the first marine actinomycete to be characterized. This characterization, and the discovery that some strains displayed unique marine adaptations, while others were metabolically active in marine sediments, failed to provoke the search for new actinomycetes in the marine environment. This feat

---

was later accomplished by culture independent studies that found that indigenous marine actinomycetes from the genera *Actinomadura, Aeromicrobium, Dietzia, Gordonia, Marinophilus, Micromonospora, Nonomuraea, Rhodococcus, Saccharomonospora, Saccharopolyspora, Salinispora, Streptomyces, Solwaraspora, Williamsia,* and *Verrucosispora* exist in the oceans.\(^5\,\,6\,\,9\) Of these genera, *Streptomyces* represents the most diverse and bountiful genus of marine actinobacteria because of its array of species which differ in morphology, physiology and biochemical activities.\(^9\) *Streptomyces* are responsible for about 7600 bioactive metabolites of the over 10000 that have been discovered that are produced by actinomycetes, and are also the chief producers of antibiotics,\(^9\) thus making them of massive commercial value.\(^10\) *Streptomyces* are the source of the widely known trioxacarcins A-C, as well as D-F,\(^11\) which are natural antitumor antibiotics. Further in depth studies into traditional and new taxa of actinomycetes is warranted by the fact that predictions estimate that only about 10% of all natural products capable of being produced by the *Streptomyces* spp. have been discovered\(^12\) and characterized, thus leaving a massively untapped potential source of novel therapeutics.

The most intriguing discovery of all genera of marine actinomycetes was that of the genus *Salinispora*, the first new obligate marine actinomycete genus, and the unveiling of the widespread populations of this genus in marine sediments.\(^5\,\,12\) Initial

chemical studies of the *Salinispora* strain by Jensen et al (2005) revealed an unusual bicyclic $\beta$-lactone $\gamma$-lactam containing metabolite that they named salinosporamide A. Salinosporamide A inhibits the chymotrypsin like proteolytic activity of the mammalian 20S proteasome thus making it an important inhibitor for cancer chemotherapy.\(^{12}\) Salinosporamide A is related to the inhibitor omuralide (a non-marine transformation product of the microbial metabolite lactacystin\(^{13}\) but is about 35 times more potent.\(^ {12}\) In a different study conducted by Prudhomme *et al.*,\(^ {14}\) they found that salinosporamide A extracted from *Salinispora tropica* shows strong inhibitory activity against the erythrocytic stages of the of the *Plasmodium* parasite cycle, and possibly acts through the inhibition of the parasite 20S proteasome, and could have massive implications in the fight against malaria in the future.


Salinosporamide A is just one example of a profusion of marine actinomycetes that are leading to novel drug discoveries. The way was essentially paved for the discovery of novel anticancer compounds from marine actinomycetes by the discovery, development and commercialization of 1-β-D-arabinofuranosylcytosine (ARA-C), a key component of the pharmaceutical solution to myeloid leukemia. Marine actinomycetes from the family Micromonospora have also been found to be potent anticancer agents that target proteasome activity. Thiocoraline is a depsipeptide belonging to the family Micromonospora that inhibits RNA synthesis and is also cytotoxic against lung and colon cancer cell lines as well as melanoma, with preferential cytotoxic activity against colon cancer cell lines with defective p53 systems. Trabectedin (ET-743) (Figure 2) commercialized as Yondelin is produced by the tunicate Ecteinascidia turbinate and

---

interferes with the activation of the multidrug resistant pathway, thus disrupting the acquired resistance of cancer cells to natural drugs such as doxorubicin, and taxanes.\textsuperscript{15} Trabectedin was approved in 2007 in Europe for use in the treatment of soft-tissue sarcoma.\textsuperscript{16}

![Trabectedin ET-743](image)

\textbf{Figure 2:} Chemical structure of Trabectedin ET-743

One of the first marine actinomycete to enter clinical trials as an anticancer agent was the cyclic antiproliferative depsipeptide Didemnin B from the tunicate \textit{Trididemnum solidum}, which showed anticancer properties in a variety of studies, and has been studied in phase II clinical trials for use in the treatment of breast, ovarian, cervical, myeloma, glioblastoma/astrocytoma, and lung cancers.\textsuperscript{17} Despite testing against many cancer types, the clinical results were terminated in 1990 due to the toxicity of the compound. However the trials aided in the engineering of the related molecule aplidine, which interferes DNA and protein synthesis to trigger cell cycle arrest, as well inhibiting ornithine

decarboxylase, a critical enzyme in angiogenesis and tumor growth. The performance of the compound in preclinical trials triggered passage into phase I clinical trials for the treatment of solid tumors and lymphomas, which subsequently led to the assessment of the drug in stage II clinical trials for solid tumors. Another anticancer drug derived from marine actinomycetes is Kahalalide F (Figure 3), a family of natural depsipeptides which is extracted from the Hawaiian herbivorous mollusk *Elysia rufescens*. Kahalalide F shows potent cytotoxic activity *in vitro* against solid tumors, neuroblastoma, chondrosarcoma, and osteosarcoma, and *in vivo* against human prostate cancer via the arresting of the cell cycle at the G1-phase in a p53 independent manner.

**Figure 3:** Chemical structure of Kahalalide

---

1.3. The Influence of Microbial Genomics on Drug Discovery

Despite the success of microbial natural products over the last century as antibacterial, antifungal, antiparasitic, anticancer and immunosuppressive agents, pharmaceutical companies have withdrawn research efforts in novel microbial natural products.\textsuperscript{19} Several factors played a role in this switch, including a dramatic increase in frequency of rediscovery of already identified natural products, technical problems associated with the isolation, purification, and structure characterization of natural products from microbial fermentations, and finally, the emergence of combinatorial chemistry, with a great potential of identifying new compounds to screen for pharmacological activity.\textsuperscript{20} Isotopic labeling studies, as well as the tracking and incorporation of specific small molecule building blocks into natural products were characteristic of initial studies of the discovery of natural products from marine and terrestrial organisms.\textsuperscript{21}

The first marine actinomycete natural products for which a complete gene cluster was identified, sequenced and verified were enterocins (1-3), and wailupemycins (4-10).\textsuperscript{21} These polyketides were isolated from the screening of the marine actinomycete \textit{Streptomyces maritimus} genomic DNA clone library for type II Polyketide synthases, which are multi-enzyme complexes consisting minimally of two ketosynthase units and an acyl carrier protein (ACP).\textsuperscript{21} Studies of the genetic foundation of marine natural

\textsuperscript{20} Challis, L. G. Mining Microbial Genomes for New Natural Products and Biosynthetic Pathways. \textit{Microbiol.} \textbf{2008}, 154, 1555-1569.
product biosynthesis based on the identification and sequencing of individual gene clusters such as the ones responsible for the isolation of enterocins and wailupemycins, while a valuable resource for a deeper understanding of the biosynthesis of some natural products, failed to unearth the diversity of biosynthetic pathways from marine actinomycetes. For this reason, the reliance on marine natural product biosynthesis via sequencing of individual gene clusters experienced a decline in the 1990’s, which happened to coincide with the introduction and spread of mass microbial genome sequencing. This technique can obtain microbial genome sequences with unrivaled rapidity, and has greatly facilitated the identification of gene clusters in microbial genomes that are expected to encode natural product biosynthetic pathways. The emergence of genome sequencing as a means of exploring the biosynthetic pathways of marine organisms was made possible by advances in DNA sequencing technology in the late 90’s. However, the actual switch in focus from isotopic labeling studies to microbial genome sequencing was prompted by two landmark papers published in 1990 and 1991, by Katz and co-workers at Abbott laboratories in Chicago, and Ledley and co-workers at Cambridge University respectively, who individually cloned three giant genes of the enzyme that synthesizes the 14 membered macrolactone scaffold of the erythromycin antibiotics. The sequences of these giant genes unearthed the finding that 200 kDa erythromycin synthase consists of 7 modules of 3-6 individually folded protein domains, which are distributed across three proteins to form a multidomain enzyme that looks like an assembly line. Interestingly a subsequent finding that non-ribosomal

---

peptides including penicillin/cephalosporins, vancomycin, cyclosporine, and daptomycin are structurally compatible in terms of the multidomain enzyme, ultimately spurred the realization that a single model could explain the synthesis of thousands of diverse natural product scaffolds. There are over a dozen marine derived compounds in clinical development belonging to marine natural products, or their derivatives of microbial origin. Soblidotin, one of these marine derived compounds, is a derivative of dolastin 10 and was isolated from *Symploca* sp.
Figure 4: Chemical structures of the non-ribosomal peptides: cyclosporine, daptomycin, penicillin, and vancomycin, which are built on the same multidomain enzyme.
The genomic sequencing data of various living organisms has experienced a dramatic growth over the last decade. According to Zhao’s report, in 2011, there were only 5381 genome sequencing projects documented on the Genome on Line Database (GOLD), with 10298 genome projects documented of which 204 archaeal, 6087 bacterial and 2003 eukaryal genome projects were ongoing. As of July 7th 2014 however, there are 54,392 genome sequencing projects documented on GOLD of which 6398 complete projects are registered. 889 archaeal genomic projects are ongoing, as well as 35,801 bacterial, and 8381 eukaryal genome projects (http://www.genomesonline.org/). These figures best illustrate the growth of genome sequencing of microorganisms. High-throughput screening relies on a great diversity of compounds to get more hits for drug discovery. With the traditional activity-dependent screening continuing to identify compounds that are already known, the growth of genetic studies and discovery of new genomic sequencing techniques will play an important role in efforts to maximize the biodiversity of marine natural products in the future.22

1.4. Genome Mining for Secondary Metabolite Biosynthesis in Actinomycetes

Genetic engineering, genome mining, and combinatorial chemistry have been vital in this molecular age, in the transformation of genetic and enzymatic knowledge into marine natural products and ultimately therapeutic agents. The revelation that the genetic features responsible for the biosynthesis of almost all secondary metabolites are found in gene clusters formed by bacterial or fungal genomes has greatly facilitated the molecular cloning and the classification of biosynthetic genes.22 Knowledge and understanding garnered of genes involved in metabolite biosynthesis over the last two decades has made it possible to determine the chemical structure, however complex, of a secondary
metabolite that results from enzymes produced by the expression of a particular set of genomic sequences.\textsuperscript{24} This has been especially shown to be true for metabolites produced by multidomain enzymes\textsuperscript{23} such as type I polyketide synthases,\textsuperscript{24} or nonribosomal peptide synthases such as penicillin/cephalosporins, vancomycin, cyclosporine, and daptomycin,\textsuperscript{23,24} for which the gene organization is very reflective of the order of biosynthetic steps.\textsuperscript{24} The gene-to-compound approach that has been adopted in recent years consists of gene-based screening meant to explore the biosynthetic potential of microbial strains, followed by cloning in order to unveil new metabolites.\textsuperscript{22} An example of such an approach can be seen in the work of Zhao and co-workers (2011) at Dalian University of Technology in China. They isolated a gene cluster generally thought to be involved in the biosynthesis of glycopeptide antibiotics. By screening the marine actinobacteria strain library in their lab. Furthermore, they also obtained the genomic sequence of a new marine-derived \textit{Streptomyces sp} S187. Another example of genome mining to get biosynthetic products lies in the discovery of salinilactam A from the marine actinomycete \textit{Streptomyces tropica} strain CNB-440. The inspection of structure fragments of salinilactam A series compounds, which were identified thanks to the UV chromophores of polyene units, helped resolve and properly assemble the DNA sequences of \textit{slm} polyketide synthase in order to predict the complete structure of salinilactam A.\textsuperscript{22} A comparative genomic analysis of \textit{Streptomyces pacifica} led to the discovery of salinosporamide A, which bears a structural resemblance to salinosporamide.\textsuperscript{22} This was done via the identification of a draft genome of \textit{S. Pacifica}

which is related to a gene cluster in *S. tropica*, which is key in the biosynthesis of salinosporamide A.  

Figure 5: Chemical structures of Salinosporamide A and K, and Salinilactam A.

Aside from whole-sequence genome mining, genome scanning as an alternative method of discovering natural product biosynthetic gene clusters is efficient in the sense that a complete genome sequence is not sequenced. Genes for natural product biosynthesis aggregate to form clusters in a microbial genome, from which a small number of random genome sequence tags can be shotgun sequenced and analyzed. Their multiple genome sequence tags will represent any given gene cluster when analyzed. Probes are then made of genome sequence tags derived from genes that are deemed likely to be involved in the biosynthesis of natural products. This new gene to compound approach finds advantage in the fact that it can rapidly focus a large library of strains down to just the small group of strains that are most likely to produce new marine compounds. However, its success depends on a variety of factors including PCR conditions, the quality of template DNA, the design of suitable degenerate primers, and the selection of target genes for genome-based screening and sequence degeneracy.  

---

to its dependence on all these factors, genome scanning is restricted to the discovery of chemical with diverse structures.22

1.5. An Integrated Approach to Complement Genome Sequencing for Nocardiopsis sp. CMB M0232 Reveals a Rich Biosynthetic Diversity

Considering the fact that genome-sequencing and biosynthetic pathway research is creating refreshed opportunity for explanation of metabolite diversity residing in gram positive actinobacteria, the Viswanathan laboratory in collaboration with the Lane laboratory created a Roche 454-derived draft genome sequence for Nocardiopsis sp. CMB M0232 isolate. The pathway components reveal a diversity that could hardly have been predicted through direct metabolic extraction. For example, in conjunction with cyanobacterial genomes assembled and studied in the Viswanathan group, as illustrated in Figure 6, the diversity of metabolic pathways is multiplexed by adopting an integrated approach.

Figure 6: Illustration of an integrated cross-talk approach for exploiting Nocardiopsis metabolic diversity.
To date there are no fully sequenced genomes available on any domains for the genus *Nocardiopsis*, therefore we propose that an integrated approach would get us far greater diversity.

**Figure 7:** Biosynthetic repertoire of marine-derived actinomycete *Nocardiopsis* sp. CMB M0232, revealed through genome sequencing and bioinformatics. A. Genome draft. B. Metabolites that have been characterized, C. List of bioinformatics-derived pathways identified.

1. Cluster size is listed with ">" in cases where only a partial gene cluster has been sequenced.

2. Metabolically verified products of nocardioazine pathway; enzymology data are pending.
In summary of the draft genome of *Nocardiopsis* sp. CMB M0232, out of a 6.4 Mb total nucleotides sequenced in the assembly, we have firmly identified each of the 5597 putative open reading frames (ORFs). Among this set, multiple types of metabolic pathways belonging to 11 distinct families, each encoded by its respective gene cluster are established. Figure 8 illustrates the diversity in chemical language that resides in the genomic code of *Nocardiopsis*.

**Figure 8:** Biosynthetic diversity of *Nocardiopsis* sp. CMB-M0232 in a chemical representation.
In Chapter 4, we present the current status and ongoing progress towards the reconstitution of the PKS-NRPS-hybrid class of metabolic pathway, encoded by the \textit{nsn} gene cluster, leading to the Nocardiopsins, as shown in Figure 7B. In chapters 2 and 3, progress into biosynthetic pathway studies on alkaloids and cyclic dipeptides is discussed.

1.6. The Rise of Mass Spectrometry for Secondary Metabolic Pathway Studies

The recently emerging field of peptidogenomics offers a potential path to address the increasingly complicated situation of rising “un annotated” genome sequences in public databases. Due to time consuming discovery options, the diversity, distribution and biological functions of cyclic dipeptides are only now being explored. A novel and rapid method for identifying cyclic dipeptide gene clusters and classifying associated products was reported in 2012 by Kersten and co-workers, which relied on connecting the chemotypes of expressed dipeptides to their biosynthetic pathways. For ribosomal peptide natural product synthesis, the traditional classification based on bioactivity, producer, and structure has shifted towards biosynthesis.\textsuperscript{26} Per Kersten \textit{et al}. (2012), in ribosomal peptide natural product biosynthesis, a precursor gene that is encoded by the peptide sequence, is translated by the ribosome to consist of leader and core peptide regions, the former serves as scaffold and acts as the recognition sites for enzymes responsible for introducing post-translational changes to the machinery, and the latter is

made up of the primary sequence of the invented peptide natural product undergoing adjustment.

Cyclic dipeptides, owing to their simplicity, are the most common cyclic peptides occurring in nature.\textsuperscript{27} They exhibit a great variety of biological activities including antiviral, antibiotic, antimicrobial and antitumor properties, and it has been suggested that some of them may play an important role as chemical mediators of bacterial quorum-sensing and signaling systems.\textsuperscript{28} Owing to their reduced conformational flexibility and increased \textit{in vivo} stability, small cyclic dipeptides show more therapeutic potential \textit{in vivo}, than their linear counterparts that show great therapeutic potential \textit{in vitro} biological screening.\textsuperscript{28} For these reasons, it is of significant value to learn the MS fragmentation mechanism of more cyclic dipeptides for drug design purposes, and for studies dealing with the relationship between their structure and activity.\textsuperscript{27,28} Learning the fragmentation mechanisms of cyclic dipeptides could furnish information that is valuable for understanding of the fragmentation of larger and more intricate natural products, and other products of pharmaceutical value, that contain cyclic dipeptides as part of their structure.\textsuperscript{27}

Mass spectrometry with its high sensitivity and selectivity for natural products has emerged as an important and complimentary technique for the analysis of peptide natural products.\textsuperscript{26} Electrospray ionization tandem mass spectrometry (ESI-MS/MS) using low-energy collision-induced dissociation (CID) is the specific method of choice for fast

analysis of natural products, thanks to its ability to analyze compounds with medium or high polarity. To be able to employ this technique, an understanding of the relationship between each possible fragmentation pathway and the gas chemistry involved in each fragmentation step is a pre-requisite.

Based on these technical advances, we can summarize the field of “natural product metabolic research” to be a science in its Renaissance rather than a tradition that is obsolete. In the remaining portions of this thesis, I present specifically three sub-projects that underscore the importance of a synergistic approach. The next chapter details our efforts in characterizing the biosynthetic pathway towards PKS-NRPS-hybrid macrolides that belong to the Rapamycin family of macrolide anti-tumor antibiotics.
CHAPTER 2

Insights into Biogenesis of Nocardioazole Alkaloids
2.1. Why are Nocardioazines Important?

Marine actinomycetes continue to be rich sources of structurally diverse natural products and a significant number of these possess promising pharmacological properties.\(^{29}\) Recently, Capon and co-workers reported isolation and structural characterization of nocardiosins\(^{30}\) (e. g. 1 and 2, p. 48) and diketopiperazine (DKP) containing nocardioazine alkaloids\(^{31}\) (3-6) from the marine-derived actinomycete *Nocardiopsis* sp. CMB-M0232 (Scheme 1). Intriguingly, under low salinity fermentation conditions, gene regulatory mechanisms predominantly favour biosynthesis of hybrid polyketide and nonribosomal peptide-derived nocardiosins (path 1). Under relatively high salinity, DKPs including nocardioazines A and B (3 and 4), are dominant (path 2). Given the lack of studies on the characterization of biosynthetic intermediates, enzymes and corresponding genes, the molecular logic of nocardioazine assembly remains poorly understood.

2.2. Biosynthetic Hypothesis towards Nocardioazine A and B

Nocardioazines A and B possess a dimerized tryptophan DKP core. The skeleton comprises seven fused rings (A-B-C-D-C′-B′-A′) in a 6-5-5-6-5-5-6 diannulated manner forming a pyrroloindoindoline-DKP-pyrroloindoindoline assembly. Among DKP natural products, nocardioazines A and B (3 and 4) stand out as the only C-prenylated DKPs


reported from a bacterial source and the first indole-C3-normal prenylated DKP from any source. The co-isolation of cyclo-L-Trp-L-Trp DKP (5) and cyclo-L-Trp-D-Trp DKP (6) alongside 3 and 4 points to 5 or 6 (or one of their epimers, cyclo-D-Trp-D-Trp DKP, ent-5) as likely precursors for the more complex congeners 3 and 4.31 The first reported synthesis of nocardioazine B by Wang et al. corrected the originally assigned stereochemistry of the natural product, further pointing to the possibility of ent-5 as a likely intermediate.32

Scheme 1: A. Structures of nocardioypsins A (1) and B (2), nocardioazines A (3) and B (4), cyclo-L-Trp-L-Trp (5) and cyclo-L-Trp-D-Trp (6). Nocardioazine skeletal numbering and ring labels are shown. B. Individual events of the proposed biosynthetic pathway to 3. Steps leading to 4 are the focus of this work.
The latent symmetry present in the DKP ring system of 3 and 4 enables numbering (N1-N11 and N1′-N11′) of the skeletal constituents comprising the 6-5-5-6-5-5-6 skeleton, as illustrated in (Scheme 1A, p.48). Nature has further decorated the western half of the DKP core through a methyltransferase-catalyzed regio- and stereoselective indole C3-methylation event (potentially concomitant with a N11-C2 cyclization) resulting in the pyrroloindoline B-C ring fusion. The B′-C′ rings, on the eastern side of the DKP core, are functionalized through a prenyltransferase-catalyzed regio- and stereoselective indole C3′-normal prenylation event (with a 3′-1″ head-to-head connectivity). Similarly, a concomitant N11′-C2′ bond-forming cyclization seems likely, based on examples including a C3-prenylating FGaPT2, generating des-N1′-Me-nocardioazine B (4).

Further, an indole N1′-methylation (at C12) is observed as a likely event catalyzed by an N-methyltransferase leading to 3 and 4. Overall, these three pivotal biosynthetic events create the asymmetry in the two annulated pyrroloindoline moieties of 3 and 4.

Nocardioazine A (3) has an additional isoprenoid-tethered DKP scaffold comprised of an 11-membered macrocycle (ring E) bridged between N1 and C3′ (tether numbered as 4″-3″-2″-1″). P-glycoprotein-mediated efflux pump (P-gp) inhibition is exhibited specifically by 3, by virtue of the macrocycle E.31 The first enantioselective synthesis of 3 (among other related alkaloids) was recently reported by Wang and Reisman and constituted an ingenious strategy towards synthetically assembling the

The biosynthetic pathway to nocardioazine A probably incorporates 4 as a reasonable intermediate, and employs a few additional oxidative transformations in tethering two annulated pyrroloindoline rings with a 5-carbon isoprenoid moiety. According to the hypothesis proposed by Raju et al\textsuperscript{31}, the C2′′-C3′′ olefinic bond is mono-oxidized into an oxirane and the C4′′ position participates in an interamolecular cyclization event with the indolic nitrogen of the B-ring on 7 to close the macrocycle. Herein we report the identification of a putative noz gene cluster encoding nocardioazine biosynthesis from the draft genome sequence of \textit{Nocardiopsis} sp. CMB-M0232. Annotations of enzymes in the noz pathway enabled generation of bioinformatically-predicted intermediates as candidates in the biosynthetic pathway to nocardioazine B. Employing a multi-pronged approach involving gene cluster annotations, asymmetric synthesis and high resolution tandem mass spectrometry, we report the \textit{in vivo} relevance of specific biosynthetic intermediates in the nocardioazine biosynthetic pathway leading up to 4. We present evidence that points to early stage assembly of \textit{cyclo}-Trp-Trp DKP as an intermediate that undergoes a regio and stereoselective C3-methyltransfer step resulting in the formation of a subsequent intermediate that is C3-normal prenylated and N1′- methylated by respective enzymes encoded in the noz pathway with reasonable promiscuity in the order of their occurrence. These results illuminate the specific precursor-product relationships in the nocardioazine alkaloid biosynthetic pathway and are expected to guide future genetic and enzymatic studies to further probe the noz pathway.

Figure 9: Organization and bioinformatics-predicted functions of the two clusters of *Nocardiopsis* sp. CMB-M0232 biosynthetic genes (*noz*) predicted to play roles in nocardioazine biosynthesis.

### 2.3. Results

#### 2.3.1. *Nocardiopsis* sp. CMB-M0232 draft genome sequence and bioinformatics-based prediction of the *noz* gene cluster

Sequencing and assembly of the *Nocardiopsis* sp. CMB-M0232 genome yielded a ~6.4 Mbp draft with >5500 open reading frames (ORFs). The hypothesized *noz* biosynthetic genes are clustered across two separate regions of the *Nocardiopsis* sp. CMB-M0232 chromosome (Figure 9, p.50). Bioinformatics analyses of these putative ORFs revealed candidate enzymes for nocardioazine biosynthesis (Table 1, p.52). BLASTP analyses of individual predicted ORFs in the entire draft genome revealed the presence of both putative nonribosomal peptide synthases (NRPSs) and a cyclodipeptide synthase (CDPS) as candidates for assembly of the DKP core during the early stage of nocardioazine biosynthesis (Scheme 1). The relatively higher level of complexity of the *noz* gene cluster is apparent through the bi-directionality the ORFs. However, bioinformatics analyses of adenylation domains from putative NRPSs revealed none that are predicted to accept two tryptophan substrates.\(^{35}\) Further, additional genes clustered with these putative NRPS-encoding genes were strongly suggestive of the biosynthesis of

---

hybrid polyketide synthase -nonribosomal peptide synthase (PKS-NRPS) products and other classes of secondary metabolites, rather than prenylated diketopiperazine alkaloids. Distinctly, a single putative CDPS (NozA) identified in the draft genome represents the most plausible candidate for assembly of cyclo-L-Trp-L-Trp DKP (5) and/or its enantiomer, ent-5 (Figure 9).
Table 1: Predicted functions of putative nocardioazine biosynthetic enzymes based on bioinformatics analyses. Two chromosomally distinct gene clusters (contig 1-2) encode these enzymes. # aa = number of amino acid residues; ID = % identity; Sim = % Similarity.

<table>
<thead>
<tr>
<th>Contig #1</th>
<th># aa</th>
<th>BLASTP annotation</th>
<th>NCBI accession number of homolog</th>
<th>organism</th>
<th>ID / Sim (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf1</td>
<td>421</td>
<td>hypothetical PLP-dependent protein</td>
<td>WP_012786924</td>
<td>Catenulispora acidiphila DSM 44928</td>
<td>51/62</td>
</tr>
<tr>
<td>orf2</td>
<td>402</td>
<td>transporter</td>
<td>AEF16056</td>
<td>Streptomyces vinaceudrappus NRRL 2363</td>
<td>58/72</td>
</tr>
<tr>
<td>nozA</td>
<td>234</td>
<td>cyclo-dipeptide synthase</td>
<td>YP_003102306</td>
<td>Actinosynnema mirum DSM 43827</td>
<td>35/54</td>
</tr>
<tr>
<td>nozD</td>
<td>385</td>
<td>cytochrome P450</td>
<td>KCP45129</td>
<td>Mycobacterium tuberculosis BTB09-382</td>
<td>30/42</td>
</tr>
<tr>
<td>nozE</td>
<td>398</td>
<td>cytochrome P450</td>
<td>WP_026248114</td>
<td>Streptomyces sp. MspMP-M5</td>
<td>36/52</td>
</tr>
<tr>
<td>orf6</td>
<td>825</td>
<td>adenylosuccinate synthase</td>
<td>WP_016473091</td>
<td>Streptomyces sp. HPH0547</td>
<td>57/70</td>
</tr>
<tr>
<td>orf7</td>
<td>307</td>
<td>XRE family transcriptional regulator</td>
<td>WP_020869817</td>
<td>Streptomyces rapamycinicus</td>
<td>80/89</td>
</tr>
<tr>
<td>orf8</td>
<td>362</td>
<td>ABC transporter</td>
<td>WP_013153583</td>
<td>Nocardiosis dassonvillei</td>
<td>67/77</td>
</tr>
<tr>
<td>orf9</td>
<td>289</td>
<td>ABC transporter</td>
<td>WP_017566834</td>
<td>Nocardiosis synnemataformans</td>
<td>76/88</td>
</tr>
<tr>
<td>orf10</td>
<td>277</td>
<td>hypothetical protein</td>
<td>WP_018657142</td>
<td>Actinomadura flavalba</td>
<td>59/69</td>
</tr>
<tr>
<td>orf11</td>
<td>136</td>
<td>MerR family transcriptional regulator</td>
<td>WP_017544945</td>
<td>Nocardiosis prasina</td>
<td>76/64</td>
</tr>
<tr>
<td>orf12</td>
<td>216</td>
<td>hypothetical protein</td>
<td>ERT00338</td>
<td>Sporothrix schenckii</td>
<td>36/52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contig #2</th>
<th># aa</th>
<th>BLASTP annotation</th>
<th>NCBI accession number of homolog</th>
<th>organism</th>
<th>ID / Sim (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf13</td>
<td>371</td>
<td>hypothetical protein</td>
<td>WP_017620974</td>
<td>Nocardiosis gilva YIM 90087</td>
<td>73/82</td>
</tr>
<tr>
<td>orf14</td>
<td>318</td>
<td>3-ketoacyl-ACP reductase</td>
<td>WP_018724690</td>
<td>Salinispora pacifica CNS055</td>
<td>63/74</td>
</tr>
<tr>
<td>orf15</td>
<td>456</td>
<td>family 1 glycosyltransferase</td>
<td>WP_017620972</td>
<td>Nocardiosis gilva YIM 90087</td>
<td>85/92</td>
</tr>
<tr>
<td>nozB</td>
<td>212</td>
<td>indole C3' and N1' methyltransferase</td>
<td>WP_026123683</td>
<td>Nocardiosis chromatogenes YIM 90109</td>
<td>84/92</td>
</tr>
<tr>
<td>nozC</td>
<td>358</td>
<td>indole C3' prenyltransferase</td>
<td>WP_017620970</td>
<td>Nocardiosis gilva YIM 90087</td>
<td>80/84</td>
</tr>
<tr>
<td>orf18</td>
<td>781</td>
<td>hypothetical protein</td>
<td>WP_017625718</td>
<td>Nocardiosis chromatogenes YIM 90109</td>
<td>68/80</td>
</tr>
</tbody>
</table>
Analyses of the *Nocardiopsis* sp. CMB-M0232 genome revealed a single putative prenyltransferase, NozC (Table 1), as the sole candidate for the unprecedented C3′-prenylation of the DKP core. The *nozC* gene is located within a cluster of biosynthetic genes chromosomally distinct from *nozA* (Figure 9). NozC shares homology with enzymes previously annotated as prenyltransferases but for which biosynthetic function has yet to be experimentally confirmed. However, little homology was noted between NozC and biochemically characterized prenyltransferases including the dimethylallyltryptophan syntheses FGaPT2 and AnaPT. This observation potentially explains by the unique regioselectivity of NozC as the sole prenyltransferase yielding C3′-normal prenylation. The *nozC* prenyltransferase gene is located within the same operon as *nozB*, which encodes a putative methyltransferase that is a candidate for C- and N-methylation of the DKP scaffold. Although the regioselectivity of NozB remains cryptic, BLASTP analyses revealed NozB possesses residues conserved among SAM-dependent methyltransferases.

The late stages of biosynthesis are hypothesized to encompass enzymes carrying out oxidation and cyclization of nocardioazine B (4) to yield nocardioazine A (3) as shown in Scheme 1B, p.48. Oxidation is hypothesized to functionalize 4 with a hydroxy leaving group at C-4″, thereby giving 7 and staging the molecule for intramolecular


cyclization to subsequently yield 3. An oxidative step is also predicted to accessorize the prenyl group with the epoxide moiety observed for 3. Candidates for these oxidative steps were found within the same operon as nozA (Figure 9 and Table 1, p.52). Both NozD and NozE share significant homology with biochemically characterized cytochrome P450s with roles in natural product biosynthesis. Specifically NozD shares 30% identity and 40% homology with Cyp121, the cytochrome P450 responsible for forming a C-C bond between aromatic carbons of the tyrosine residues of cyclo-L-Tyr-L-Tyr. NozE is also homologous to Cyp121 (38% identity; 52% homology). NozD and NozE are 49% identical and 59% homologous to one another. A cyclase is postulated to yield 3 from 7. However, this enzyme remains cryptic, as no candidate enzymes with homology to characterized cyclases were clustered with other predicted players in nocardioazine assembly. Several putative ORFs lacking homology to characterized enzymes were noted within the noz operons. This lack of homology precluded prediction of biosynthetic roles for these hypothetical enzymes (Table 1). It is plausible that one of these as-yet uncharacterized enzymes may catalyze ring closure of 7 to yield nocardioazine A (3). Also it is plausible that these putative ORFs may catalyze steps to tailor nocardioazine pathway products in order to generate newer congeners that have eluded isolation efforts to date.


2.3.2. Asymmetric synthesis of a suite of Noz pathway intermediates

In order to evaluate the \textit{in vivo} relevance of putative intermediates to the nocardioazine biosynthetic pathway, and to provide synthetic standards for the \textit{in vitro} characterization of NozA-catalyzed \textit{cyclo}-Trp-Trp formation, we constructed 5 and \textit{ent}-5. As shown in Scheme 2, p.55, \textit{cyclo}-L-Trp-L-Trp-DKP (5) was constructed through a four-step sequence, starting with protection of the amino functionality of L-Trp with a benzyloxycarbonyl (Cbz) group. Treatment with Cbz-Cl along with sodium bicarbonate-sodium carbonate in acetonitrile-water (2:3; v: v) as solvents, over 3 h resulted in 8 providing the western half of the DKP.

\begin{center}
\begin{tikzpicture}
  % Diagram code here
\end{tikzpicture}
\end{center}

\textbf{Scheme 2} : Synthesis of \textit{cyclo}-L-Trp-L-Trp DKP (5) and \textit{cyclo}-D-Trp-D-Trp DKP (\textit{ent}-5).
Similarly, treatment of L-Trp with thionyl chloride in methanol, refluxing for 18 h resulted in formation of the L-Trp methyl ester (9) in nearly quantitative yield, providing the eastern half of the DKP. BOP-Cl-mediated coupling of 8 and 9 in the presence of triethylamine as a base in THF resulted in amide 10 in 93% yield. BOP-Cl-mediated activation of the carboxylic acid functionality of 8 proved the most efficient for isolation of a high yield of amide product 10. Deprotection of the Cbz group in 10 under hydrogenating conditions in the presence of Pd-C in MeOH (with a trace amount of water) yielded deprotected amine precursor 15 which also contained an ester functionality as an intramolecular reactive partner. The DKP ring system was then formed through the treatment of 11 in 14M ammonia in methanol at 60 °C for 8 h resulting in cyclo-L-Trp-L-Trp DKP (5) in 95% yield. Likewise, an identical sequence was applied starting from D-Trp (through protection resulting in ent-8 and ester ent-9, followed by coupling to give ent-10, finally with deprotection-cyclization step) resulting in the formation of cyclo-D-Trp-D-Trp DKP (ent-5) in excellent overall yield. The signals in the 1H NMR spectrum of the product revealed no indication of an alternative diastereomer (arising from a mono epimerization event at C9/C9’), and therefore we obtained 5 in >19:1 diastereomeric ratio. Through chiral HPLC analysis and polarimetry, we confirmed the enantiomeric enrichment to be consistent with observing no racemization at either α-amino acid centers (Figure 31 and 32; chapter 5). Having synthesized cyclo-Trp-Trp-DKPs as both antipodes, we confirmed its biosynthetic relevance in the noz pathway through microbial culturing, followed by liquid chromatography-coupled high resolution-mass spectrometry (LC-MSMS) of Nocardiopsis sp. CMB-M0232, to be consistent with results published from Capon and
co-workers. Given that 5 is possibly accessed by \textit{Nocardiopsis} sp. toward construction of 3 and 4, the synthetic route outlined here serves to provide a reasonable access to 5 for future enzymology studies.

Inspired by the bioinformatics-predicted logic of nocardioazine enzymology, we assembled a suite of biosynthetic intermediates and aimed for verification of their relevance through a combination of LC-MS and tandem mass spectrometry. Specifically, the mid-stages of the pathway offered testable questions as outlined in Scheme 1B, p. 48. For example, if cyclo-Trp-DKPs are precursors for nocardioazine biosynthesis, and NozB and NozC are engaging in C3-methylation, N1'-methylation and C3'-prenylation events, is there a particular order to their occurrence? Are these enzymes metabolically-relevant towards creating specific precursors and products? Finally, could fermentation and extraction of metabolites directly from \textit{Nocardiopsis} sp. CMB-0232 followed by LC and tandem MS reveal insights into which early or mid-stage precursors are indeed relevant \textit{in vivo}? We set out to answer these questions by addressing the chemical synthesis and unambiguous establishment of the structures of these putative intermediates.

Given this prediction, and the relatively complex issue of optimizing alternate genetic hosts for successful reconstitution of the Noz biosynthetic gene cluster, we collectively, identified 5, \textit{ent-5}, 12, 13, 14, 15, 16 and 17 as chemical targets for asymmetric synthesis in order to evaluate their \textit{in vivo} presence and biosynthetic relevance in \textit{Nocardiopsis} sp. CMB-M0232 (Figure 10, p.58). NozB and NozC are expected to catalyze prenylation and methylation steps to yield six unique potential intermediates (12, 13, 14, 15, 16 and 17) depending on the order of reactions (Figure 10B, p. 58). Specifically, if \textit{Nocardiopsis} sp. CMB-M0232 were to employ an indole
C3′-prenyltransferase (hypothetical NozC) to install a dimethyl allyl group on DKP 5 then the product of this biosynthetic reaction is expected to be \textit{cyclo-L-Trp-C3′-prenyl-L-Trp} DKP (12). The formation of pyrroloindoline cycle (of 12) during this prenyltransfer step is based on fungal precedents such as FgaPT2.\textsuperscript{36}

\textbf{Figure 10:} A. Biosynthetic events in the nocardioazine pathway and the genes encoding for the respective steps. B. Precursors and products predicted for mid-stages of the pathway involving C3-methylation, C3′-prenylation and N1′-methylation.

Alternatively, if the indole C3-methyl transferase (NozB) were operative on the basic early-stage intermediate DKP 5, the product expected from this transformation
cyclo-C3-Me-L-Trp-L-Trp DKP (13). The corresponding N1’-methylated product from action of a methyltransferase (NozB, but regioselectively on the N1’ position) would be cyclo-L-Trp-N1’-Me-L-Trp DKP (14). Products 15, 16 and 17 encompass further increase of complexity through a subsequent enzymatic event. Their relevance in a combinatorial way is discussed in Scheme 5 (vide infra, p. 65). Also we envision the synthetic endeavour to set the foundation for future in vitro and in vivo reconstitution of individual steps catalysed by NozA, NozB and NozC in the nocardioazine pathway.

Due to the relative complexity of the proposed intermediates, regio- and stereoselective C3’-prenylation, C3-methylation and N1’-methylation presented challenges. As shown in Scheme 3, p. 61, 3-methyl indole (18) and enamide (19) served as reasonable starting points to employ an enantio- and diastereoselective indole-enamide [3+2] cycloaddition reaction in the presence of (S)-BINOL and tin(IV) chloride as a key step to install the C3-methyl functionality. En route to employing this step as a strategy towards assembling 13 and 16, enamide 19, was prepared from L-serine through the conversions involving the corresponding O-Boc derivative (see chapter 5). The [3+2] cycloaddition between 18 and 19 proceeded with a 12:1 diastereomeric ratio favouring the exo isomer 20a over the minor endo isomer 20b. Each diastereomer exhibited a 2:3 ratio of conformational isomers (caused by the Cbz group on N11 position) as revealed by the presence of equivalent sets of 1H NMR signals. The overall yield of the [3+2] cycloaddition product 20 is 61%. Considering the relative stereochemical disposition of substituents at C2, C3 and C9 in major exo isomer 20a, we initially attempted an LDA-
mediated deprotonation-reprotonation sequence to invert the C9 centre. By virtue of the lack of an allyl protecting group at N1 (that was present in a prior report\textsuperscript{34}), a retro-Michael addition occurred, resulting in degradation of 20a under strongly basic conditions. Therefore, a revision of appropriate conditions to achieve the correct relative stereochemistry was necessary. After careful screening and optimization, treatment with excess lithium hydroxide in a mixture of methanol, water and THF (1:1:1; v: v: v) affected this transformation efficiently to yield 20c. Concomitant to the epimerization, we observed base-mediated hydrolysis of the carboxymethyl ester functionality. It proved to be a beneficial outcome as the next step 	extit{en route} to 13 involved an amide bond to an L-Trp-containing partner 9. Similar to the conversion involving 8+9 to give 10, we observed that smooth peptide bond formation occured under BOP-Cl -mediated activation of 20c followed by nucleophilic participation of the amino functionality of 9 resulting in the coupled product 22 in 90% yield in THF as the solvent. Hydrogenative deprotection of the Cbz group of 22 was affected smoothly to result in 23. During this deprotection of the Cbz group under Pd-C, we observed direct intramolecular cyclization resulting in formation of 13 in ~10% efficiency. However, this low conversion rate motivated the employment of relatively stronger base\textsuperscript{42} involving methanolic NH$_3$ to result in the formation of cyclo-C3-Me-L-Trp-L-Trp DKP (13) in high efficiency through the participation of the secondary amino functional group in an internal nucleophilic substitution reaction. The overall synthetic sequence is 5 linear steps starting from 3-methyl indole (18). The overall yield for formation 13 was 24.9%. Similarly, the assembly of 16 began with 20a undergoing a sequential epimerization-hydrolysis event

under aqueous lithium hydroxide yielding \( 20c \). \( N1' \)-methylated L-Trp (21) was synthesized (from L-Trp, see chapter 5) for its engagement in a coupling step with \( 20c \).

Again, BOP-Cl activated \( 20c \), and under triethylamine and THF gave \( 24 \) as the product in 81% yield. Similar to the non-methylated counterpart \( 22 \), we could effect a hydrogenative deprotection followed by base-mediated intramolecular cyclization event on \( 24 \) to result in \( 16 \) (via \( 25 \)) in fairly high efficiency (91% yield) in 5 linear steps from commercially available \( 18 \). The overall yield for formation of \( 16 \) was 31.0%.
Scheme 3. Synthesis of cyclo-C3-Me-L-Trp-L-Trp DKP (13) and cyclo-C3-Me-L-Trp-N1'-Me-L-Trp DKP (16).
The establishment of relative stereochemical relationship along with the connectivity of substituents was verified for 16 (as a representative example of C3-methyl DKPs 13 and 16) using NOESY and long range (LR) $^1$H-$^1$H-COSY experiments as shown in Figure 11,p.63. Cyclo-C3-Me-L-Trp-$N^1$-Me-L-Trp DKP (16), possesses the 6-5-5-6 (A-B-C-D) ring system that is present both in nocardioazines A and B. As shown in Figure 11A, p.63, an NOE signal of 2.62% between C2-H and C3-Me indicates a cis relative stereochemistry at the B-C ring fusion (C2 – C3). An exo relationship of the pyrroloindoline ring system (B-C) was confirmed through the hydrogens at the C3-Me substituent displaying an NOE signal of 2.62% with the $\alpha$- hydrogen at C8 (H$_a$) of 16. The diastereotopic $\beta$-proton at C8 (H$_b$) displayed no NOE signal to the C3 methyl substituent. The L-Trp-derived stereocenter at C9 with a proton showed a 3.06% NOE with the $\alpha$- hydrogen at C8. The hydrogens at C2 and C9 displayed an NOE enhancement of 0.95%. Overall, these assignments were correlated with the COSY spectral assignment indicating the relative stereochemical disposition of substituents in 16 as shown in Figure 11C,p.63. The proton NMR signals assigned to C9, C8 H$_a$ and C8 H$_b$ were distinctly different from those at C9$'$H and C8$'$-H$_a$ and C8$'$-H$_b$ respectively. Overall, both diketopiperazines 13 and 16 displayed relative stereochemistry consistent with that observed in nocardioazines A and B.
Figure 11: A. NOESY correlations for (16). B. 1H-1H LR-COSY correlations for (16) at 600 MHz. Region between chemical shifts of 1.2 and 5.2 p is shown. C. Chem-3D rendered ball and stick model of (16) depicting relative stereochemistry of substituents across the central DKP ring.

We then aimed at cyclo-L-Trp-C3′-n-prenyl-L-Trp DKP (12) and its N1′-methylated variant cyclo-L-Trp-N1′-Me-C3′-n-prenyl-L-Trp DKP (15) as synthetic targets. Through a biomimetic prenylation method we published recently\(^4^3\), the methyl ester of L-tryptophan (9) served as a precursor to engage in a domino process initiated by a C3′-prenylation event (with prenyl bromide as the electrophile) subsequently resulting in a C-N bond forming pyrroloindoline cyclization, under acetate-solution conditions at room temperature, to result in the formation of 27a and 27b as a 4:1 mixture of exo and endo diastereomers (scheme 4, p.65). The overall yield for this transformation was 67%.

considering full recovery of unreacted 9. The fact that 27a and b were produced through a single biomimetic step provided direct route to the C3’-normal prenylated scaffold of nocardioazines. The stereochemical relationship between C3’-”prenyl substituent, C2’-H and C9’-carboxymethyl substituent for the major diastereomer 27a was established through NOESY correlations (see chapter 5). Unlike the C3-methylation step, the major diastereomer 27a arising from this one-step prenylation event mimics the relative stereochemistry of C9’-C2’-C3’ substitution pattern in nocardioazine A and B, thereby obviating the need for any epimerization of the α-amino acid stereocenter. Upon treatment of 27a with N-phthalyl-protected L-Trp-acid 29 (prepared in advance using a one-step protection reaction with phthalic anhydride, see chapter 5), under BOP-Cl activation and basic conditions, we obtained the coupled product 30 (comprising the carbon skeleton of target 12) in 90% yield. Gratifyingly, the coupled product 30 underwent a tandem sequence initiated by a hydrazine hydrate-mediated deprotection of the phthalyl group followed by an intramolecular cyclization in methanol-dichloromethane that resulted in cyclo-C3’-”prenyl-L-Trp-L-Trp DKP (12), 70% yield. NOESY experiment showed a 2.98% enhancement between C8’-H and olefinic C2”-H; 1.78% enhancement between protons on 2’ and 8’-α CH; and finally a 3.05% enhancement between protons on 8’β CH and 9’ positions. These NOEs confirmed the stereochemistry as cis across the DKP ring system and an overall exo arrangement for the B’-C’ pyrroloindoline ring fusion.
Scheme 4. Synthesis of cyclo-L-Trp-C3′-exo-prenyl-L-Trp DKP (7) and cyclo-L-Trp-N1′-Me-C3′-exo-prenyl-L-Trp DKP (10). \( a \) - % isolated yield based on recovered starting material.
Likewise, engagement of N1′-methylated-L-Trp carboxymethyl ester (26) in a one-step prenylation (under aqueous solution) using prenyl bromide resulted in 72% overall yield of C3′-prenylated 28a (major) and 28b (minor) based on recovery of unreacted 26. Similar to the formation of 30, upon subjecting 28a to a coupling reaction with 29 using BOP-Cl and triethylamine in THF, we obtained 31 that upon subjection to a hydrazine hydrate-mediated deprotection-cyclization sequence resulted in a B′-C′ ring-forming process leading to cyclo-L-Trp-N1′-Me-C3′-n-prenyl-L-Trp DKP (15) in 74% yield. The overall yields for formation of 12 and 15 were 42.4 % and 44.6% over 3 linear steps respectively. As shown in Scheme 5, using methods that were closely adapted from the synthesis of 26, we were able to mono-methylate the N1′ position of 5 to synthesize 14. Similar to the prenylation on 9 and 26, 13 underwent C3′-prenylation to result in des-N-Me-Nocardioazine B (17). The identities of 14 and 17 were verified by HPLC and tandem MS as shown in Figures 12 and 13, (p.69 and 71).

Scheme 5. Synthesis of cyclo-N1′-Me-L-Trp-L-Trp DKP (14) and Des-N1′-Me-nocardioazine B (17).
2.3.3. Evaluation of the biosynthetic relevance of synthesized intermediates through NMR, LC-MS and HR-tandem MS reveals precursor-product relationships for nocardioazone B biosynthesis

Dramatic advances in the application of MS-MS tools to connect molecules to individual gene clusters present in microbial systems are revolutionizing research in the post-genomics era.\textsuperscript{44} Further, use of MS\textsuperscript{1} and MS\textsuperscript{2} guided strategies can lead to illumination of biosynthetic relationships between multiple pathways encoded by respective gene cluster families.\textsuperscript{45} We applied MS\textsuperscript{1} and MS\textsuperscript{2} as tools in order to connect genomic and synthetic data with \textit{Nocardiopsis} biosynthetic intermediates from the noz pathway. EIC traces (Figure 12A, p. 69) indicated that HPLC-MS profiles uniquely separated and distinguished most of the synthesized intermediates. Unique signatures are observable in MS\textsuperscript{2} spectra for each putative biosynthetic metabolite (Figure 12B-H, p. 69). Specifically, Figure 12B shows the presence of cyclo-L-Trp-L-Trp DKP (5). Its [M+H]\textsuperscript{+} ion (at 373.1662 Da in positive ion mode ESI-MS profile) and its [M-H]\textsuperscript{-} ion (at 371.1530 Da in negative ion mode ESI-MS) are observed (see chapter 5). Characteristic Trp fragments corresponding to product ions at (Da: 242.0930; 214.0980; 169.0759; 159.0924; 130.0654 and 103.0547) were observed, all within error (± 5 ppm) as fingerprints of 5 through its MS\textsuperscript{2} fragmentation data presented in Figure 12B (and tabulated with mass accuracy and % relative abundance in Table 5,p.72). The product


molecular ions, arise out of neutral losses of a Trp moiety (129 Da), along with sequential loss of CO (28 Da) and/or HCONH₂ (45 Da), in various combinations. Neutral loss of HCN (27 Da) from ion at m/z = 130.0654 accounted for the presence of the m/z = 103.0547 ion. Likewise, we mapped the signatures of cyclo-L-Trp-C3′-prenyl-L-Trp DKP (12); cyclo-C3-Me-L-Trp- L-Trp DKP (13); cyclo-N1′-Me-L-Trp-L-Trp DKP (14); cyclo-L-Trp-C3′-prenyl-N1′-Me-L-Trp DKP (15); cyclo-C3-Me-L-Trp-N1′-Me-L-Trp DKP (16) and des-N1′-Me Nocardioazine B (17), as illustrated in Figure 12(p.69) and additionally elaborated in the experimental section, chapter 5.
Figure 12: A. Extracted ion chromatogram (EIC) traces of synthetic standards. ESI TOF-MS² fragmentation data: B. cyclo-L-Trp-L-Trp DKP (5) (M+H)+; C. cyclo-L-Trp-C3′-prenyl-L-Trp DKP (12) (M+H)+; D. cyclo-C3-Me-L-Trp-L-Trp DKP (13) (M-H)-; E. cyclo-N1′-Me-L-Trp-L-Trp DKP (14) (M+H)+; F. cyclo-L-Trp-N1′-Me-C3′-prenyl-L-Trp DKP (15) (M+H)+; G. cyclo-C3-Me-L-Trp-N1′-Me-L-Trp DKP (16) and H. des-N1′-Me nocardioazine B (17).
We then looked for signatures of 12, 13, 14, 15, 16 and 17 directly from cultures of *Nocardiopsis* sp. CMB-M0232 to probe their relevance as biosynthetic intermediates *in vivo*. Reverse-phase HPLC uniquely identified synthetic 12, 13, 15, 16 and 14 (Figure 13A, p. 71). Comparison of retention times of these synthetic compounds with the alkaloid fractions of *Nocardiopsis* sp. CMB-M0232 revealed that 12, 15 and 14 were not detectable biosynthetic products or intermediates. Interestingly, we detected the presence of three relevant compounds in *Nocardiopsis* sp. extracts, namely, 13, 16 and 17. Despite modifying several solvent conditions and flow rates, we were unable to distinctly separate 17 that overlapped in retention time with 13 and 16. As the cultures of *Nocardiopsis* matured over 21 days, the solution takes on a deeper red color (Figure 13B). Further supporting its biosynthetic relevance, TLC patterns of extracts suggested the presence of 13 (see Figure 34). We then asked if the MS² fragmentation pattern of 13 extracted from the bacterial culture matched with that of the synthetic standard. Gratifyingly, the ESI-TOF- MS¹ and MS² data for synthesized 13 overlapped precisely with that of the extracted metabolite (Figure 13C and D). The presence of an ion at m/z = 256.110 indicated that a C3-methyl group containing DKP from the precursor ion after loss of a neutral Trp unit (129 Da) was observed both in the synthesized standard as well as from the extract. Further, Both ¹H and ¹³C NMR analysis of LC-derived extracts confirmed the presence of the indicated peak at 11.8 min. to be 13 (Figure 28 and 29, chapter 5).
Figure 13: A. Reverse phase HPLC traces for synthesized and extracted metabolites from *Nocardiopsis* sp. CMB-M0232. B. Culture of *Nocardiopsis* sp. CMB-M0232 at 7, 14 and 21 days. C. ESI-(−)-TOF- MS<sup>1</sup> and MS<sup>2</sup> spectrum of 13 from *Nocardiopsis* sp. CMB-M0232 (top) matched with spectrum of synthesized 13 (bottom); D. ESI-(+)-TOF- MS<sup>1</sup> and MS<sup>2</sup> spectrum of 13 from *Nocardiopsis* sp. CMB-M0232 (top) matched with spectrum of synthesized 13 (bottom).
Table 2: LC-MS and MS\textsuperscript{1} and MS\textsuperscript{2} data for synthetic and extracted biosynthetic intermediates and products. NF – not found; for full detailed listing and corresponding formulas of molecular ions, see chapter 5. Note: metabolites with Mw = 369, 383, 482, 466, 468 and 452 were identified as DKPs in Capon's study.\textsuperscript{30,32} Ions represented in bold are identified from extracts of *Nocardiopsis* sp. CMB-M0232. * - EIC traces only, see chapter 5.

<table>
<thead>
<tr>
<th>Name of Metabolite</th>
<th>Molecular Formula</th>
<th>ESI HR-MS [M]\textsuperscript{+} [M+H]\textsuperscript{+} (expected)</th>
<th>LC retention time Found (synth.) (min)</th>
<th>MS\textsuperscript{2} Fragmentation Pattern</th>
<th>Biosynthetic Role</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>cyclo-L-Trp-L-Trp DKP (5) and cyclo-D-Trp-D-Trp DKP (ent-5)</td>
<td>C\textsubscript{22}H\textsubscript{20}N\textsubscript{4}O\textsubscript{2}</td>
<td>(372.1586) (373.1659) (371.1513) 373.1665 (+) and 371.1530 (-)</td>
<td>7.06 (7.06)</td>
<td>242.0925; 144.0805 and 130.0654</td>
<td>early stage intermediate</td>
<td>[M+H]\textsuperscript{+} and [M-H]\textsuperscript{-} observed in extracts Matches with synthetic standard</td>
</tr>
<tr>
<td>cyclo-L-Trp-C3\textsuperscript{-}'prenyl-L-Trp DKP (12)</td>
<td>C\textsubscript{27}H\textsubscript{28}N\textsubscript{4}O\textsubscript{2}</td>
<td>(440.2212) (441.2285) (439.2139) NF</td>
<td>NF (23.2)</td>
<td>373.1671; 242.0931; 113.0337; 198.1288; 183.1044; 130.0658</td>
<td>Mid stage product of C3\textsuperscript{-}prenyltransfer on 5</td>
<td>Not detected</td>
</tr>
<tr>
<td>cyclo-C3-Me-L-Trp-L-Trp DKP (13)</td>
<td>C\textsubscript{27}H\textsubscript{28}N\textsubscript{4}O\textsubscript{2}</td>
<td>(386.1743) (387.1816) (385.1670) 387.1825 385.1711</td>
<td>11.8 (11.8)</td>
<td>385.1690; 256.110; 130.065</td>
<td>Mid stage product of C3\textsuperscript{-}methyltransfer on 5</td>
<td>[M+H]\textsuperscript{+} observed in extracts Matches with synthetic standard</td>
</tr>
<tr>
<td>cyclo-N1\textsuperscript{-}'Me-L-Trp-L-Trp DKP (14)</td>
<td>C\textsubscript{23}H\textsubscript{22}N\textsubscript{4}O\textsubscript{2}</td>
<td>(386.1743) (387.1816) (385.1670) 387.1825 385.1711</td>
<td>NF (9.5)</td>
<td>399.1806; 212.1441; 144.0813; 130.0657</td>
<td>Mid stage product of N1\textsuperscript{-}M methyltransfer on 5</td>
<td>Not detected</td>
</tr>
<tr>
<td>cyclo-L-Trp-N1\textsuperscript{-}'Me-C3\textsuperscript{-}'prenyl-L-Trp DKP (15)</td>
<td>C\textsubscript{28}H\textsubscript{30}N\textsubscript{4}O\textsubscript{2}</td>
<td>(454.2369) (455.2442) (453.2296) NF</td>
<td>NF (18.0)</td>
<td>399.1806; 212.1441; 144.0813; 130.0657</td>
<td>Mid stage product of indole N1\textsuperscript{-}methyltransfer on 12</td>
<td>Not detected</td>
</tr>
<tr>
<td>cyclo-C3-Me-L-Trp-M1\textsuperscript{-}'Me-L-Trp DKP (16)</td>
<td>C\textsubscript{29}H\textsubscript{32}N\textsubscript{4}O\textsubscript{2}</td>
<td>(400.1899) (401.1972) (399.1826) 401.1972</td>
<td>12.2 (12.2)</td>
<td>401.1981; 256.1089; 184.0761; 144.0813</td>
<td>Mid stage indole N1\textsuperscript{-}methyltransferase product from 13</td>
<td>[M+H]\textsuperscript{+} observed in extracts Matches with synthetic standard</td>
</tr>
<tr>
<td>Des-N1\textsuperscript{-}'Me-nocardioazine B (17)</td>
<td>C\textsubscript{29}H\textsubscript{30}N\textsubscript{4}O\textsubscript{3}</td>
<td>(454.2369) (455.2442) 455.2442</td>
<td>12.6 (12.6)</td>
<td>256.1089; 184.0761; 144.0813</td>
<td>Putative precursor to secondary metabolite product 4</td>
<td>[M+H]\textsuperscript{+} observed in extracts</td>
</tr>
<tr>
<td>Nocardioazine B (4)</td>
<td>C\textsubscript{29}H\textsubscript{30}N\textsubscript{4}O\textsubscript{2}</td>
<td>(468.2525) (469.2598) 469.2695</td>
<td>18.44*</td>
<td>186.0914 156.0812 144.0805 and 130.0654</td>
<td>Putative precursor to secondary metabolite product 3</td>
<td>[M+H]\textsuperscript{+} observed in extracts</td>
</tr>
<tr>
<td>Nocardioazine A (3)</td>
<td>C\textsubscript{29}H\textsubscript{30}N\textsubscript{4}O\textsubscript{3}</td>
<td>(482.2318) (483.2396) 483.2396</td>
<td>8.80*</td>
<td>483.2396</td>
<td>Secondary metabolite product</td>
<td>[M+H]\textsuperscript{+} observed in extracts</td>
</tr>
</tbody>
</table>
Because N1′-methylation of 13 will produce 16 we anticipated the presence of 16 from cultured *Nocardiopsis* sp. CMB-M0232. Indeed, LC-ESI- (+) TOF MSMS analysis suggested the presence of 16 at Rt = 12.2 min (Figure 29 in chapter 5). HR-MS verification of its presence was obtained through observation of an ion at m/z = 401.1972 and furthermore, MS² fragmentation revealed the presence of characteristic ions at m/z = 256.109, 184.076, and 144.081 that matched well with synthesized 16 (Table 2, p72). The presence of des-N1′-Me nocardioazine B (17) was detected through the identification of a broad LC peak at Rt ~ 12.6 min that corresponded to an HR-MS signal at m/z = 455.2442 (Δ0 ppm). Its corresponding MS² spectrum revealed signature peaks at m/z 256.1089 (seen in fragmentation of 8+14 Da), 184.0761, and 144.0813 typically observed for all synthetic standards possessing the C3-methyl substitution and two Trp units of the DKP ring system (Table 2, p.72). Identification of additional metabolites from *Nocardiopsis* sp. fermentations including the target alkaloids 3 and 4 are outlined in chapter 5. Significantly 12, 14 and 15 were not detected in cultures of *Nocardiopsis* sp. CMB-M0232 as indicated by the conspicuous absence of signature prenylated ions at m/z = 198.129 (C₁₄H₁₆N⁺) observed only in 12 and a corresponding ion seen only for 15 at m/z = 212.144 (C₁₅H₁₈N⁺). Pathway specific metabolites identified through LC and their MS-MS fragmentation (in this study) are highlighted in green in Table 2, p.72. Upon consideration of individual enzymatic steps encoded by the genome specifically recruiting NozB and NozC, a biosynthetic pattern is evident as illustrated in Scheme 6, p.74. Catalysis of 5 to *ent*-5 remains elusive; however, downstream metabolic profiling indicated the operation to lead to production of 13 followed by 16 and / or 17 in a promiscuous manner, with their order of events being interchangeable.
Scheme 6: Summary of the biosynthetic pathway for mid-stages of nocardioazine biosynthesis. Dotted lines show originally hypothesized possibilities and bold lines show the path that is evident from HPLC, LC-MS, MS² analyses for all relevant intermediates and by NMR (for 13 only).
Upon consideration of the three enzymatic steps in a permutation manner, the multi-pronged approach reveals the relevance of \textit{C3-methylation as a step preceding the C3’-prenylation and N1’-methylation}. This is illustrated in Scheme 6, where dotted lines indicate originally hypothesized possible pathways, while bold lines indicate the pathway supported by experimental evidence from bio-guided synthesis and \textit{Nocardiopsis sp.} metabolite profiling.

\section*{2.4. Discussion}

Herein, our approaches to assemble the DKP core of 5 and follow it with regio- and stereoselective indole C3-methylation (18+19 to give 20) and indole C3’-prenylation (9/26 to give 27/28 respectively) are inspired by biosynthetic pathways that nature has encoded within the chromosome of \textit{Nocardiopsis}. For example, the assembly of the DKP core mediated by BOP-Cl-mediated coupling mimics the cyclodipeptide synthase (CDPS) assembly employed by multiple bacteria, including \textit{Nocardiopsis}, to dimerize various \(\alpha\)-amino acids. The C3-methylation of indoles derived from tryptophan moieties is observed in multiple natural product assemblies, although at this point, an enzyme carrying out selective transformation at the indole C3 position is not known.

Additionally, by virtue of employing a catalytic asymmetric version, the [3+2] cycloaddition methodology we employed to install the C3-methyl group en route to 13 and 16 is potentially useful due to the possibility of deriving either antipodal series. The direct prenylation of indoles under aqueous conditions is challenging. We derived our inspirations from the report of Bocchi et al., though that study was not specifically
targeting the nocardioazine skeleton. Additionally, the indole C3′-prenylation conditions we report herein resulted in a significant improvement in synthetic efficiency. A high degree of biomimicry is evident between our prenylation strategy and those of the indole prenyltransferases described in the literature leading to multiple alkaloid pathways encoded by the *Aspergillus* sp. and the *Bacillus subtilis* ComX pheromone synthesizing ComQ geranyl transferase enzyme.

Overall, the development of short and selective synthetic pathways to intermediates (Figure 10, p.58) enabled an LC-MSMS-based approach to delineate the biosynthetic circuitry leading to 4. As illustrated in Scheme 6 (p.74), it is evident that the methyltransferase step, possibly through the recruitment of a hypothesized C3-methyltransferase NozB, in *Nocardiopsis* sp. CMB-M0232 is successively processing methylations of 5 and 13. Genome sequencing and annotation points to NozA being a dedicated CDPS enzyme for biosynthesis of 5. Thus far, bioinformatics analyses by previous researchers have predicted more than 50 gene clusters to encode CDPS machinery for assembly of DKP natural products in various species spanning both prokaryotes and eukaryotes. However, far fewer of these CDPSs have been

biochemically characterized. Only a single experimentally characterized CDPS, Amir_4627 from *Actinosynnema mirum* has been established to yield cyclo-L-Trp-L-Trp DKP as the dominant product. Homologs of NozA are evident in a range of *Nocarditis* strains whose CDPS-containing gene clusters are available in publically deposited genomes. For example, the *Nocarditis alba* CDPS cluster encodes an enzyme (AlbC) possessing 40% sequence identity to NozA. Similarly, *Nocarditis lucitensis* also encodes a CDPS homologous to NozA (40% identity), although the DKP product of this CDPS remains unknown. To date, biochemically characterized CDPSs have been reported to catalyze the formation of DKPs exclusively from L-amino acids.

This is due to the mechanism of CDPSs, which employs aminoacyl-charged tRNAs from primary metabolism as substrates in catalyzing formation of the DKP scaffold. Hence, if nocardioazines A-B (3-4), featuring D-amino acid stereochemistry, are indeed CDPS-derived, then an unidentified isomerase is also expected as a required component of their biosynthetic pathway to isomerize cyclo-L-Trp-L-Trp DKP (5) into its antipode ent-5.

This possibility is particularly intriguing, given the observation of cyclo-L-Trp-L-Trp (5) from fermentations of *Nocarditis* sp CMB-M0232.

The possibility of the recruitment of a promiscuous indole C3′-normal prenyltransferase that could prenylate either 13 or 16 leading to 17 or 4 is plausible, though the substrate product relationship is uncertain as yet. Indole C3′-prenyltransfer is a well-studied biosynthetic step, and enzymes from a variety of fungal secondary


metabolic pathways are known. However, identification and functional studies of indole prenyltransferases from bacterial sources are only starting to be revealed, and no C3’-normal prenyltransferases have been characterized from any source. The fungus, *Neosartorya fischeri* recruits an enzyme (AnaPT, PDB ID: 4LD7_P) whose biosynthetic function closely resembles that expected for NozC. In the bacterium *Bacillus subtilis*, ComQ catalyzes C3-normal geranylation of an L-Trp derived peptide as part of biosynthesis of ComX, a metabolite implicated in quorum sensing. However, neither of these enzymes shares significant homology with NozC, the sole candidate prenyltransferase predicted through bioinformatics analyses of the *Nocardiopsis* sp. draft genome. This may potentially be explained by the unique regioselectivity of nocardioazine biosynthesis. To further evaluate the enzymatic underpinning of prenylation in nocardioazine biosynthesis, future work should focus on biochemical characterization of this putative prenyltransferase.

Various mechanisms for methyl transfer exist in nature, and at the moment the specific enzymology operative to perform N- and C-methylations remains unknown. Interestingly, the *A. mirum* gene cluster that produces cyclo-L-Trp-L-Trp via CDPS Amir_4627 also encodes a methyltransferase (Amir_4628), which in vitro experiments revealed transfers a methyl group to the two 2,5-DKP nitrogen in a subsequent fashion. Nocardioazine methyltransferase (e.g. NozB) mysteriously accommodates multiple orientations of the substrate, as the positions of methylation are C3 of one Trp unit and the N1’ of the other Trp unit. As shown in Scheme 1, we project (and corroborate the

proposal made in Capon’s study\textsuperscript{31}) the involvement of cytochrome P450s to mediate late stage oxidations and macrocyclization towards biosynthesis of Nocardioazine A (3). NozD and NozE, homologous to biochemically characterized cytochrome P450s, represent candidates for these steps. Overall, this predicted circuitry of enzymatic steps establishes the chemical structure validation necessary for future experimental characterization of regioselective and stereoselective biocatalysts involved in the noz pathway. Our currently presented synthesis of relevant biosynthetic intermediates is expected to expedite the \textit{in vitro} validation of individual enzymatic steps. While relatively simpler L-Trp-L-Trp DKP (as products of cyclodipeptide synthase biosynthesis) and other dimeric amino acid DKPs have been analysed through tandem mass spectrometry,\textsuperscript{27,54} complex DKPs like 12, 13, 14, 15, 16 and 17 thus far have not been investigated through mass spectrometry adding further importance to the present study.

2.5. Conclusion

Nocardioazines A and B (3 and 4), as the first indole-C3-normal prenylated DKPs from any biological source, present a poorly understood pathway. In this study, we laid the chemical foundations for understanding nocardioazine biosynthesis by synthesizing an exhaustive set of putative, bioinformatics-predicted intermediates. Structural verification through 1D and 2D NMR, and analyses through HPLC-MS and HR-MS methods established the framework for evaluation of the biological relevance of specific intermediates in the proposed noz (nocardioazine) pathway \textit{in vivo}. Upon comparing

HPLC and tandem mass spectrometry data between synthesized standards and alkaloid fractions extracted from *Nocardiopsis* sp. CMB-M0232, it is conclusively evident that indole C3-methylation leading to 13 is a biosynthetic event that precedes indole C3′-normal prenylation and a second methyl transfer to the N1′ position. In addition, through bioinformatics analyses of the draft genome of *Nocardiopsis* sp. CMB-M0232, a cyclodipeptide synthase (NozA) is predicted as a candidate for assembly of the cyclo-Trp-Trp precursor of the nocardioazine alkaloids. Addition efforts are necessary to unveil the complete genetic and enzymatic-underpinning of nocardioazine A and B biosynthesis.

**Acknowledgements and Author Contributions:**

RV and ALL designed the project. NFA and SKP performed asymmetric synthesis and characterized standards by higher order spectroscopy. NFA and SKP designed asymmetric synthesis of 5, 13, 14 and 17. SKP designed the asymmetric synthesis route to 12, 15 and 16. RV, NFA, ALL and EJ cultured Nocardiopsis sp. CMB-M0232. NFA extracted metabolites and JK characterized extracts with LC, MS1 and MS2 spectrometry. LC, MS1 and MS2 data analysis for metabolic and synthetic intermediates were executed by RV and JK. RV and ALL assembled genomic information of Nocardiopsis sp. CMB-M0232 and performed bioinformatic analyses. NFA and SKP wrote drafts of the manuscript that were edited by ALL and RV. RV, NFA and ALL acknowledge funding from SACM, UNF Transformational Learning Opportunity Award and a Research Corporation Cottrell College Science Award.
CHAPTER 3

Cyclodipeptide Synthase-Catalyzed Chemoenzymatic Synthesis of Bioactive Natural Products Belonging to the Diketopiperazine Family
3.1. General Discussion on Microbial Biocatalysts for Organic Synthesis

Microbial systems continue to inspire discovery of novel biocatalysts for the synthesis of organic molecules with unique structural and biological properties.\(^5^5\) Our initial efforts to probe nocardioazine biosynthesis focused on generation of *Nocardiopsis* sp. CMB-M0232 gene replacement mutants with the intention of employing these mutants as tools for determining biosynthetic intermediates. Like many actinomycetes, however, *Nocardiopsis* sp. proved resistant to genetic manipulation. Hence, we turned to alternative cloning, synthesis and mass spectrometry-centric strategies presented herein to experimentally establish nocardioazine biosynthetic intermediates predicted through bioinformatics analyses, and their corresponding enzyme functions.

2,5-Diketopiperazines of \(\alpha\)-amino acids are valuable structural cores that have not only inspired natural products research, but also have stimulated studies on structure and conformation of molecules.\(^5^6\) Asymmetric variants of this motif present synthetic challenges resulting in new reactions, catalysts, peptidomimetics, and chiral auxiliaries. Recent literature reports indicate continued interest in their total chemical synthesis and structure verification.\(^3^2,^3^4\) Their medicinal chemistry has revealed a range of disease therapeutics including \(\beta\)-turn mimetics, PDE inhibitors, oxytocin antagonists, cancer inhibitors and neoprotective agents.\(^5^6\) Extremely high antimicrotubule activity was observed for the DKP-derived didehydropiperazine-2, 5-dione motif of phenylahistins,


underscoring the medicinal importance of this core.\textsuperscript{57} While the more traditionally studied cyclodipeptides, $cyclo$-L-Pro-L-Val, $cyclo$-L-Pro-L-Phe and $cyclo$-L-Pro-L-Tyr, produced by microorganisms recently have been shown to have Lasi quorum-sensing (QS) properties\textsuperscript{58}, and to inhibit agr-mediated toxin biosynthesis\textsuperscript{59} in staphylococci\textsuperscript{60}, the cyclodipeptides of L or D-tryptophans (and many examples presented in this and related studies\textsuperscript{31, 32, 34}) are recent additions and therefore relatively underexplored from a biosynthetic and pharmacological standpoint. Bio-inspired synthesis of such a privileged core therefore has potential to allow new synthetic pathways towards creation of structural analogs through chemo enzymatic pathways and mutasynthesis.


Inspired by these preliminary findings, in this chapter we outline the development of a CDPS-based biocatalyst for the synthesis of novel diketopiperazines. Figure 14 illustrates the overall noz pathway and the presence of the L-Trp-L-Trp dimerizing enzyme NozA.

3.2. NozA is a cyclodipeptide synthase involved in DKP formation in the nocardioazaine pathway

The amino acid sequence of the putative CDPS, NozA, identified by bioinformatics analyses as the most plausible candidate for assembly of cyclo-L-Trp-L-Trp DKP (5) and/or its enantiomer, ent-5, was compared with sequences of known CDPSs. Amino acid sequence alignment revealed 35% identity between NozA and
Amir_4627 (NCBI Accession #YP_003102306; Figure 15), the sole characterized CDPS that incorporates two tryptophan residues into a DKP scaffold.61 NozA includes residues conserved among biochemically-characterized, catalytically functional CDPSs62 including Amir_462761 from Actinosynnema mirum. Beyond the conserved active site residues (highlighted in yellow in Fig. 15A), correlations are also apparent between NozA and Amir_4627 for residues implicated in recognition and binding of the enzyme to aminoacyl-charged tRNA substrates (highlighted brown). Similar predicted secondary and tertiary structural features are evident between the two enzymes (Figure 15A).

Confirming these bioinformatics-based predictions, we recently demonstrated that NozA catalysts formation of cyclo-L-Trp-L-Trp DKP (5) (unpublished results62).


62 James E.; Alqahtani, N.; Viswanathan R.; Lane A. *Manuscript on prepare*
Figure 15: A. Amino acid sequence alignment between NozA and Amir_4627 and bioinformatics model of NozA generated using Geneious. Clustal W was used for basic sequence alignment. B. NozA-catalyzed formation of cyclo-L-Trp-L-Trp (5). Phyre 2.0 was used for rendering of the NozA structure.
NozA, phylogenetically belongs to a family of predominantly uncharacterized CDPS enzymes. Their family tree is presented in Figure 16. Many of the candidate related members belong to other Nocardiopsis genera and therefore an underlying genetic make up in this actinobacteria reveals that they utilize L-Trp dimerizing activity for myriad biological purposes that remain unknown.

Figure 16: NozA phylogenetic relationship.
The biocatalytic potential residing in *Nocardiopsis* sp. CMB-M0232 in the form of NozA, prompted us to evaluate its biochemical function and establish preliminary enzymology. The biochemical evaluation of NozA as an enzyme involved an assay system developed through a combination of organic synthesis and gene cloning techniques. The synthesis illustrated in Scheme 2 of Chapter 2, p.55, provided a practical route to have testable quantities of L-Trp-L-Trp DKP as a synthetic standard for the enzymatic assay evaluation. The cloning strategy for *Nocardiopsis* NozA for expression in *E. coli* is described below in Figure 17.

**Figure 17:** Cloning of NozA for expression.
The gene was cloned and expressed in the *E. coli* M15[pREP4] cell line and the enzyme assay was reconstituted *in vivo* by introduction of positive and negative control experiments. Figure 18 illustrates the logical workflow for testing the enzyme function of NozA. First we ascertained that *E. coli* does not host a native DKP forming biosynthetic pathway, thereby making it a valid candidate for heterologous expression and enzyme assay. This fact was further corroborated in the recently published study by Giessen et al.\textsuperscript{61}
Figure 18: Enzyme assay for NozA.
The assay that was envisioned involves treatment of *E. coli* M15 cells with pQE-30-NozA vector as a positive control and those cells lacking the exogenous gene as a negative control (Figure 18). Upon subjecting the bacterial cells hosting the exogenous nozA gene Figure 18 shows HPLC traces (280 nm) for L-Trp-L-Trp (green), supernatants from *E. coli* M15 cells carrying CDPS-gene in the pQE30 vector (red), and supernatants from *E. coli* M15 carrying an empty vector (blue). Overlap between synthetic DKP and those in red trace at 22.5 mins are observed for supernatants from *E. coli* M15 cells carrying CDPS-gene in pQE30 vector, but not for the extract from the corresponding empty pQE30 vector. These results were further verified by high resolution LC/MS analyses to deduce potential molecular formulae and predict dipeptide structures. Overall we have confirmed the reconstituted product of the assay unambiguously to be L-Trp-L-Trp DKP.

Complete synthetic reconstitution of NozA as a biocatalyst for DKP synthesis is currently ongoing work in our laboratory. In order to reconstitute the enzyme activity completely in vitro, I synthesized the CME esters of L-Trp as illustrated below (scheme 7, p.90). Collectively 47 and ent 47 are candidates for future in vitro biocatalyst reconstitution experiments.
Scheme 7. Synthesis of CME-esters

Acknowledgements and Author Contributions:

RV and ALL designed the project. RV, NFA, ALL and EJ cultured *Nocardiopsis* sp. CMB-M0232. NFA extracted metabolites and JK characterized extracts with LC, MS\(^1\) and MS\(^2\) spectrometry. LC, MS\(^1\) and MS\(^2\) data analysis for metabolic and synthetic intermediates were executed by RV and JK. RV and ALL assembled genomic information of *Nocardiopsis* sp. CMB-M0232 and performed bioinformatic analyses of the gene cluster of *Nocardiopsis* sp. CMB-M0232. ALL and EJ did cloning and sequencing of the CDPS. NFA performed synthesis of CME-esters. RV, NFA and ALL acknowledge funding from SACM, UNF Transformational Learning Opportunity Award and a Research Corporation Cottrell College Science Award.
CHAPTER 4

Characterization of the nsn Biosynthetic Gene Cluster for the Nocardiopsin Family from the Marine-Derived Nocardiopsis sp. CMB M0232
4.1. Introduction to Nocardiopsin Macrolides

Rapamycin (1F), FK-506 (2F), and meridamycin (3F) are examples of a small family of natural products featuring a macrolide core and pipecolate group (Figure 19, p.93). Members of this family are noteworthy for their ability to bind immunophilin FKBPs, which interact with diverse proteins (e.g. mTOR) implicated in many intracellular signaling pathways. The formation of ternary complexes between these ligands, their respective cellular FKBPs, and FRBs (Rapamycin recognition domain of their receptors) create myriad downstream biological effects. Members of this family are used clinically as immunosuppressants to prevent transplanted organ rejection and for treatment of renal cell carcinoma. These families of metabolites further hold promise as future therapeutics for neurodegenerative diseases and other pathologies.

Despite prominent biomedical enthusiasm about the rapamycin/FK-506 natural product family, genome mining and traditional isolation-based approaches have afforded only a handful of unique skeletons (Figure 19).


Nocardiopsins A-D (4F-7F), isolated from marine-derived *Nocardiopsis* sp. CMB-M0232, are the only new representatives of this family reported during the past decade. The nocardiopsins feature several notable structural differences that distinguish them from previously reported natural products. Notably, the nocardiopsins lack α-ketoamide and hemiketal groups that play a critical role in the FKBP-binding capability of

*Figure 19:* Top: Structures of Streptomyces-derived pipecolate-containing macrolides. Bottom: Nocardiopsin family of actinobacterial pipecolate or proline-containing macrolides.
other family members (e.g. 1F-3F). The nocardiopsins (4F-7F) instead feature an alkene conjugated to the amide group, yet were found to bind FKBP12 with $K_D$ values in the low micromolar range. This mode of action suggests the possible involvement of the conjugated alkene moiety, however only preliminary evidence is available owing to a lack of sufficient quantities of 4F-7F. Further structurally distinguishing them from other rapamycin and FK-506 family members, all nocardiopsins feature a tetrahydrofuran (THF) group. The structures of 4F-7F also suggest that closure of the nocardiopsin macrolide occurs via a C6 hydroxy group of non-polyketide synthase (PKS) origin. This is in contrast to the vast majority of macrolide ring closures, which occur through the intramolecular reaction of a PKS derived hydroxy group with a PKS- or nonribosomal peptide synthase (NRPS)-tethered thioester. These structural features both distinguish the nocardiopsins from previously reported natural products and suggest that the nocardiopsin biosynthetic pathway includes key differences from the characterized pipicolate-polyketide pathways. Overall, biosynthetic pathway investigations are anticipated to provide novel insights into mechanisms of THF ring formations, offer a combinatorial pathway for analogs of nocardiopsins and also create opportunities for biomimetic syntheses. The knowledge gained on THF ring biosynthesis is also expected to provide new insights into how nature constructs similar ring systems towards polyether ladder toxins that are biosynthesized by dinoflagellates.


In this chapter we describe the cloning, sequencing, and bioinformatics analysis of the nocardiopsin biosynthetic gene cluster ($nsn$). We provide evidence for the function of this gene cluster in nocardiopsin biosynthesis through the \textit{in vitro} evaluation of the lysine cyclodeaminase homolog NsnL.

4.2. Cloning and sequencing of the nocardiopsin gene cluster ($nsn$)

Open reading frames (ORFs) predicted from the \textasciitilde 6.4 Mbp, \textasciitilde 1,200 contig draft genome sequence of \textit{Nocardiopsis} sp. CMB-M0232 were evaluated for homology to previously characterized PKSs and NRPSs,$^{72}$ revealing a partial gene cluster that is hypothesized to be responsible for nocardiopsin biosynthesis.

![Figure 20: Organization of the \textasciitilde 77 kB ($nsn$) nocardiopsin biosynthetic gene cluster.](image)

Cosmid clones carrying \textit{Nocardiopsis} sp. genomic DNA were screened for these genes, and three overlapping cosmids were sequenced and assembled to yield a single 119 kB contiguous sequence, from which 77 kB was annotated as the nocardiopsin gene

cluster (*nsn*, Figure 20). The *nsn* cluster encodes enzymes homologous to characterized biosynthetic enzymes (Table 3), including polyketide synthases (*Nsna*-*Nsnd*), a nonribosomal peptide synthase (*Nsne*), cytochrome P450s (*Nsnf; Nsnh*), an epoxide hydrolase (*Nsng*), and a lysine cyclodeaminase (*Nsnl*). The cluster also encodes proteins with predicted roles in pathway regulation (*Nsnr1*-*Nsnr4*) and antibiotic transport (*Nsnt1*-*Nsnt2*). The fact that the transporter and regulatory genes surround the cluster, indicates that we identified an intact full cluster for the putative pathway to the nocardiopsins.

**Table 3**: Proposed functions of proteins from the *Nocardiopsis* sp. CMB-M0232 nocardiopsin gene cluster (*nsn*) based on comparison of amino acid sequences with homologous enzymes. (*nsn* cluster deposited in NCBI as accession #TBD.)

<table>
<thead>
<tr>
<th>gene</th>
<th>translated protein size (#aa)</th>
<th>annotation</th>
<th>protein homolog (NCBI accession number)</th>
<th>identity/similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf(-3)</td>
<td>408</td>
<td>adenylylsulfate kinase</td>
<td>MB27_02760 (KHD79007)</td>
<td>47/58</td>
</tr>
<tr>
<td>orf(-2)</td>
<td>644</td>
<td>phosphogluconate dehydratase</td>
<td>RHU231_200054 (CDZ87071)</td>
<td>77/83</td>
</tr>
<tr>
<td>orf(-1)</td>
<td>145</td>
<td>(2Fe-2S)-binding protein</td>
<td>SSIG_01081 (EWS90718)</td>
<td>57/67</td>
</tr>
<tr>
<td>nsnl</td>
<td>249</td>
<td>oxidoreductase</td>
<td>NBRGN_038_01510 (GAJ81480)</td>
<td>91/94</td>
</tr>
<tr>
<td>nsnH</td>
<td>414</td>
<td>cytochrome P450</td>
<td>AveE (BAJ84477)</td>
<td>31/45</td>
</tr>
<tr>
<td>nsnT1</td>
<td>213</td>
<td>ABC transporter protein</td>
<td>IL38_13665 (KGI81075)</td>
<td>56/71</td>
</tr>
<tr>
<td>nsnT2</td>
<td>150</td>
<td>transporter protein</td>
<td>RradSPS_0169 (AHY45452)</td>
<td>80/88</td>
</tr>
<tr>
<td>orf7</td>
<td>873</td>
<td>ATPase</td>
<td>IL38_11110 (KGI81478)</td>
<td>61/77</td>
</tr>
<tr>
<td>nsnF</td>
<td>359</td>
<td>cytochrome P450</td>
<td>GfsF (BAJ16472)</td>
<td>44/55</td>
</tr>
<tr>
<td>nsnL</td>
<td>345</td>
<td>lysine cyclodeaminase</td>
<td>PipA (CBW45757)</td>
<td>55/70</td>
</tr>
<tr>
<td>nsnE</td>
<td>1544</td>
<td>nonribosomal peptide synthase</td>
<td>MerP (ABC87508)</td>
<td>49/62</td>
</tr>
<tr>
<td>nsnA</td>
<td>6409</td>
<td>polyketide synthase</td>
<td>PldAI (BAH02268)</td>
<td>50/61</td>
</tr>
<tr>
<td>nsnB</td>
<td>5166</td>
<td>polyketide synthase</td>
<td>MerC (ABJ97439)</td>
<td>51/61</td>
</tr>
</tbody>
</table>
4.3. Bioinformatics analyses support assembly of the nocardiopsin core by polyketide synthases NsnA-NsnD and nonribosomal peptide synthetase NsnE

The *nsn* cluster encodes four ORFs (NsnA-NsnD) sharing homology with biochemically characterized modular type I PKSs (Figure 20, Table 3). Bioinformatics analyses revealed that the four PKSs encompass a predicted loading module as well as ten extender modules, corresponding to the number of extension cycles expected to afford the nocardiopsin polyketide core (Figure 21). Domains within each extender module and the substrate preference of each acyltransferase (AT) domain are consistent

---

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession</th>
<th>Function</th>
<th>Description</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>nsnC</em></td>
<td>4937</td>
<td>polyketide synthase</td>
<td>ApoS7 (AEP40934)</td>
<td>49/60</td>
</tr>
<tr>
<td><em>nsnD</em></td>
<td>1801</td>
<td>polyketide synthase</td>
<td>LipPKS3 (ABB05104)</td>
<td>51/61</td>
</tr>
<tr>
<td><em>nsnG</em></td>
<td>145</td>
<td>epoxide hydrolase</td>
<td>Rv2740 (2BNG_A)</td>
<td>49/57</td>
</tr>
<tr>
<td><em>nsnR3</em></td>
<td>230</td>
<td>LuxR transcription regulator</td>
<td>AURJ3M (ACD75765)</td>
<td>62/72</td>
</tr>
<tr>
<td><em>orf17</em></td>
<td>307</td>
<td>(2Fe-2S)-binding protein</td>
<td>DF19_12420 (KDN77345)</td>
<td>52/66</td>
</tr>
<tr>
<td><em>orf18</em></td>
<td>431</td>
<td>glycosyl transferase</td>
<td>MPTA5024_33555 (ETK31740)</td>
<td>54/64</td>
</tr>
<tr>
<td><em>nsnR4</em></td>
<td>133</td>
<td>XRE family transcription regulator</td>
<td>M877_14005 (EST28853)</td>
<td>57/64</td>
</tr>
<tr>
<td><em>orf20</em></td>
<td>204</td>
<td>protease</td>
<td>Q760_03435 (KGM01157)</td>
<td>70/79</td>
</tr>
<tr>
<td><em>orf21</em></td>
<td>227</td>
<td>protease</td>
<td>BU52_27470 (KES04021)</td>
<td>66/81</td>
</tr>
<tr>
<td><em>orf22</em></td>
<td>257</td>
<td>uracil DNA glycosylase</td>
<td>LK10_00325 (KHL05635)</td>
<td>74/83</td>
</tr>
</tbody>
</table>

---

with the proposed bioinformatics model (Figure 21). Each predicted ketoreductase (KR), dehydratase (DH), and enoyl reductase (EH) domain contains amino acid sequence signatures that correspond to common catalytically active PKS domains. Together, these analyses support biosynthesis of the nocardiopsin polyketide core via a prototypical type I modular PKS, with assembly proceeding co-linearly to yield the parent polyketide chain of 4F-7F. Following PKS module 10, pipecolic acid (as in 4F-5F) or proline (as in 6F-7F) is incorporated into the polyketide chain through reactions catalyzed by the NRPS homolog NsnE. Sequence homology of the adenylation (A) domain of NsnE with those of NRPSs known to act upon pipecolic acid substrates supported the role of NsnE in pipecolate incorporation. Like previously reported pipecolate-incorporating NRPSs, NsnE exhibits relaxed substrate specificity, as evidenced by structural differences observed for the amino acid moieties of 4F-5F vs. 6F-7F. NsnE includes two predicted condensation (C) domains (Figure 21). Analogous to the biosynthetic pathways for 1F-3F, one of these C domains is proposed to catalyze transfer of the polyketide chain to the pipecolate group while the other is expected to catalyze closure of the macrolide, with concomitant release of the macrolide from the NRPS.


**Figure 21:** Schematic of the nocardiopsin biosynthetic pathway. (top) Domain organization and substrate specificity for each PKS module, as proposed by bioinformatics analyses. 90 (bottom) To afford the C6 hydroxy group required for macrolide closure, the C5-C6 alkene formed via PKS module 2 is proposed to be converted to an epoxide by NsnF. This epoxide is then opened via epoxide hydrolase NsnG to yield a C5-C6 vicinal diol suitable as a substrate for NRPS-catalyzed ring closure. Epoxide formation and opening is illustrated as occurring while the linear precursor is tethered to the NRPS; however, the actual timing of these steps within the PKS-NRPS assembly line remains to be determined.

### 4.4. A cytochrome P450 and epoxide hydrolase are proposed to act upon a megasynthesistethered nocardiopsin precursor to stage it for macrolide closure

The structures of nocardiopsins A-D (4F-7F) imply that macrolide ring closure occurs through the reaction of a C6 hydroxy group with the NRPS-tethered pipecolyl or prolinyl group (Figure 19). For all previously characterized pipecolate-polyketide family pathways as well as the majority of macrolides, the requisite hydroxy group is formed by reduction of a β-ketothioester intermediate by a PKS KR domain.71 However, KR-
catalyzed hydroxylation at C6 of the linear nocardiopsin precursor is not expected. Instead, module 2 of the PKS assembly line is expected to produce a C5-C6 alkene group (Figure 21), suggesting a unique route to stage the molecule for macrolide closure.

The nsn cluster encodes two cytochrome P450 homologs (NsnF and NsnH) as well as an epoxide hydrolase homolog (NsnG) as the most plausible candidates for staging the megasynthase-tethered nocardiopsin precursor with the hydroxy group needed for ring closure. Comparison of NsnF with previously characterized polyketide-producing cytochrome P450s revealed that NsnF is most similar to several epoxidases (e.g. 55% similarity to GfsF from the FD-891 pathway). NsnH is instead more similar to P450s with roles in unactivated carbon hydroxylation (e.g. 45% similarity to RosC from the rosamicin pathway)\textsuperscript{76} and tetrahydrofuran biosynthesis (e.g. 45% similarity to AveE from the avermectin pathway)\textsuperscript{77}.

The homology of NsnF to biochemically characterized epoxidases, along with observation of an epoxide hydrolase homolog (NsnG) within the cluster, implicate these two enzymes as the most plausible candidates to stage the linear precursor for ring closure (Figure 21). NsnF is proposed to act upon the linear PKS-NRPS tethered nocardiopsin precursor, catalyzing epoxidation of the C5-C6 alkene. NsnG is then proposed to catalyze nucleophilic ring opening to transform the C5-C6 epoxide into a vicinal diol (Figure 21).

NsnG is most closely related to members of the limonene epoxide hydrolase family (e.g. 57% similarity to Rv2740). This family of epoxide hydrolases is highly

\textsuperscript{77} Chang-Hong, P.; Matsuzaki, K.; Ikeda, H.; Tanaka, H.; Omura, S. Production of 6, 8a-seco-6, 8a-deoxy Derivatives of Avermectins by a Mutant Strain of \textit{Streptomyces avermitilis}. \textit{J. Antibiot.} \textbf{1995}, 48, 92-94.
distinct from classic epoxide hydrolases involved in secondary metabolism. Classic epoxide hydrolases feature a highly conserved α/β hydrolase fold and catalyze a two-step mechanism that includes a covalent enzyme-substrate intermediate. Instead, X-ray crystallographic and computational studies of two limonene epoxide hydrolases, Rv2740 (PDB accession #2BNG) and LEH (PDB accession #1NWW), revealed this unique class of enzymes feature a distinct structure and unique conserved active site triad that catalyzes epoxide ring opening through a one-step general acid-catalyzed mechanism.  

Amino acid sequence alignment of NsnG with Rv2740 and LEH revealed that NsnG also possesses this active site triad, supporting that NsnG catalyzes epoxide opening through a mechanism analogous to that of previously characterized limonene epoxide hydrolases. However, NsnG must possess fundamental differences from previously characterized members of this enzyme family, since NsnG is the first that apparently acts upon a megasynthase-tethered substrate.

To probe the in vivo functions of nsnF and nsnG, we attempted to eliminate these genes from Nocardiopsis sp. CMBM0232 using the PCR-targeted gene replacement strategy pioneered by Gust and co-workers, as well as insertional mutagenesis approaches analogous to those employed by Zotchev and co-workers within the Nocardiopsis genus. Unfortunately, Nocardiopsis sp. CMB-M0232 proved entirely resistant to genetic manipulation via these methods.


The sequential two-enzyme pathway (NsnF and NsnG) proposed to poise the linear nocardiopsin precursor for NRPS-catalyzed cyclization is highly unusual. Both enzymes are expected to act upon the nocardiopsin precursor chain prior to its release from the PKS-NRPS megasynthase (NsnA-NsnE). Few tailoring enzymes acting upon megasynthase-tethered substrates have been reported.\(^{80}\) As the most analogous example, the stambomycin pathway was recently reported as the first to stage an acyclic polyketide for macrolide formation via cytochrome P450-catalyzed hydroxylation of an unactivated carbon.\(^{81}\) Based upon analysis of the *nsn* gene cluster, nocardiopsin biosynthesis instead employs the sequential action of *both* a cytochrome P450 epoxidase and an epoxide hydrolase to stage the polyketide for ring closure (Figure 21). While epoxide formation and hydrolysis are relatively common polyketide tailoring steps, to our knowledge, the current study offers the first example of a biosynthetic pathway that appears to apply these combined steps to poise a molecule for macrolide formation. This apparently coordinated action of a PKS-NRPS megasynthase with a P450 and epoxide hydrolase highlights the ingenuity of nature in designing enzymes to collaboratively assemble complex molecular structures and suggests that these enzymes may represent valuable tools for chemoenzymatic polyketide synthesis.

### 4.5. Biosynthesis of the tetrahydrofuran (THF) group of nocardiopsins A-D


The nocardiopsins (4F-7F) feature a THF group that is absent in all other natural products of the rapamycin and FK-506 family. Previous studies have demonstrated the biosynthesis of cyclic ethers through a variety of intriguing mechanisms. These include biosynthesis of the nonactin THF group by an enoyl CoA hydratase homolog,\textsuperscript{82} and the biosynthesis of polyether ladder natural products such as lasalocid through a cascade of epoxide ring opening cyclization reactions.\textsuperscript{83} Pyran group formation has also recently been demonstrated via a pyran synthase (PS) domain within a trans-AT PKS assembly line (e.g. in pederin biosynthesis).\textsuperscript{84} Analogously, within cis-AT PKS assembly lines, a DH domain from the ambruticin pathway was recently demonstrated to catalyze pyran formation in addition to dehydration. A unique cyclase domain within the PKS has also been suggested to play a role in pyran functionalization.

Based upon nocardiopsin structures (4-7) as well as bioinformatics analyses of the nocardiopsin PKS assembly line and other putative enzymes encoded by the nsn cluster (Figure 20; Figure 21; Table 3), each of these previously described routes to THF formation appear highly unlikely.

As a more plausible alternative, a handful of cytochrome P450s have also been implicated in cyclic ether biosynthesis. Examination of ORFs encoded by the \textit{nsn} cluster revealed NsnH, a cytochrome P450 homolog, as one potential candidate for catalyzing

THF formation. NsnH shares 43% homology with AurH. AurH is a bifunctional
cytochrome P450 from the aureothin pathway, and catalyzes the hydroxylation of an
unactivated carbon atom as well as subsequent THF formation via a ring closure reaction
involving this hydroxy and an alkene group.\textsuperscript{85} NsnH also shares homology with AveE
(45% similarity), which was previously postulated to catalyze formation of the
avermectin THF group by a mechanism similar to that of AurH.\textsuperscript{Error! Bookmark not defined.}
Analyses of nocardiopsin PKS domains (i.e. within modules 4-5, Figure 21) suggest that
the THF precursor features hydroxy groups at both C9 and C11. Hence, if NsnH is indeed
operative in THF formation, it is expected to function via a mechanism distinct from that
of AveE and AurH.

Another possibility is that THF functionalization of nocardiopsin requires NsnH
in addition to other enzyme(s) or, alternatively, that THF formation does not require
NsnH. Of note, NsnI is encoded immediately upstream from NsnH and shares homology
with several GenBank enzymes annotated as oxidoreductases. Unfortunately, none of
these NsnI homologs have been biochemically characterized, leaving the role of NsnI in
THF biosynthesis ambiguous. Future synthetic and enzymatic studies will aim to
determine the roles of NsnH and/or NsnI in formation of the THF moiety that
differentiates nocardiopsins from other pipecolate-polyketide natural products.

4.6. Pipecolic acid biosynthesis is catalyzed by lysine cyclodeaminase Nsn

Previous studies demonstrated that 4F-5F (featuring a pipecolate group) exhibited
significant binding with FKBP12 while 6F-7F (featuring a proline group) were inactive.\textsuperscript{30}

\textsuperscript{85} He, J.; Müller, M.; Hertweck, C. Formation of the Aureothin Tetrahydrofuran Ring by a Bifunctional
Hence, the pipecolate moiety appears essential to the biological activity of nocardiopsins. NsnL, a lysine cyclodeaminase homolog (62% similarity to RapL from rapamycin pathway), was hypothesized to catalyze the formation of L-pipecolic acid from L-lysine in the nocardiopsin pathway.

To establish the function of NsnL and further connect enzymes from the nsn cluster to nocardiopsin biogenesis, NsnL (39.4 kDa) was recombinantly expressed as a His-fusion protein in *E. coli* and purified (Figure 22A). The typical yield of NsnL was ~15 mg per liter of culture.

![Figure 22: A. Preliminary data displaying protein purification for NsnL, a L-lysine cyclodeaminase. B. Reconstitution assay for NsnL in nocardiopsin biosynthesis.](image)

In order for the function of NsnL to be confirmed, the reconstitution assay for the enzyme involved L-lysine as a substrate, with the enzyme NsnL. Furthermore, NsnL was incubated with selected individual substrates (i.e. L-lysine, D-lysine, and L-ornithine) to probe its function and substrate specificity in vitro. Reaction products were derivatized.

---

and evaluated by gas chromatography with flame ionization detection (GC-FID). These analyses revealed that NsnL catalyzed the conversion of L-lysine to L-pipecolic acid consistent with the preliminary assay (Figure 23a). D-lysine was not accepted as a substrate (data not shown). This established NsnL as a source of pipecolic acid for incorporation into nocardipins by NRPS NsnE (Figure 21), and support that the enzyme is stereospecific. NsnL also catalyzed the conversion of L-ornithine to L-proline, suggesting promiscuity with respect to the substrate carbon chain length (Figure 23b).

![Figure 23](image_url)

**Figure 23:** GC-FID analysis of products resulting from incubation of selected amino acid substrates with lysine cyclodeaminase NsnL. (a) Incubation of NsnL with L-lysine resulted in accumulation of L-pipecolic acid, which was not detected in L-lysine or NsnL only controls. (b) Incubation of NsnL with L-ornithine resulted in accumulation of L-proline, which was not detected in L-ornithine or NsnL only controls. (I.S. = norvaline internal standard; * denotes L-pipecolic acid signal; ** denotes L-proline signal, as determined by comparison with synthetic standards.)

NsnL is expected to catalyze these cyclization reactions through a mechanism analogous to that established for homologous ornithine cyclodeaminases yielding proline in primary metabolism and that proposed for RapL, which catalyzes the formation of L-pipecolic acid during rapamycin biosynthesis. In this mechanism, the α-amino group of lysine or ornithine is first oxidized to an imine, with nicotinamide adenine dinucleotide
(NAD+) as a cofactor. The terminal amine then attacks the imine carbon to generate a cyclic tetrahedral intermediate (figure 24). The loss of ammonia from this intermediate then gives a cyclic imino acid, which is reduced by NADH to yield pipecolic acid and reform NAD+. Supporting this mechanism, NsnL catalyzed the formation of pipecolic acid or proline only when NAD+ was added to in vitro reactions.

Figure 24: Schematic of pipecolate (red) incorporation into the nocardiosins polyketide chain. A. The presumed overall reaction and B. mechanism for NsnL, a lysine cyclodeaminase.
4.7. The novel structural features of nocardiopsins are underpinned by a unique biosynthetic pathway, which offers opportunities for future study

The discovery of nocardiopsins A-D (4F-7F) by Capon and co-workers marked the first novel members of the pipecolate-polyketide class of FKBP modulators in over a decade and significantly enhanced the structural diversity of this class of medicinally promising natural products.\textsuperscript{30} Establishment of the nocardiopsin biosynthetic gene cluster (\textit{nsn}), presented in the current work, eludes to a number of unique pathway features. The further study of these is anticipated afford increased understanding of Nature’s array of polyketide tailoring mechanisms as well as potential biocatalysts for the synthesis of biologically active polyketides. Of particular interest are the two enzymes, an epoxidase (\textit{NsnF}) and epoxide hydrolase (\textit{NsnG}) homolog, expected to act sequentially on megasynthase-tethered polyketide intermediates to equip these molecules with hydroxy groups necessary for macrolide closure. Given that macrolides are important determinants of biological activities for a variety of polyketides,\textsuperscript{71} the characterization of \textit{NsnF} and \textit{NsnG} may offer novel strategies for the engineering of designer polyketide macrolides. Further, additional enzymes encoded by the \textit{nsn} pathway (particularly \textit{NsnH} and/or \textit{NsnI}) represent candidates for biosynthesis of the THF moiety of 4F-7F and suggest a unique route to the formation of this biosynthetically and biomedically interesting group. To facilitate future mechanistic investigations as well as probable combinatorial biosynthetic approaches to create nocardiopsin analogues for clinical investigations, we have established the synthetic pathways for a suite of cell-permeable SNAC-esters of malonyl, methyl malonyl systems (38–45, see Chapter 5).
Acknowledgements:

We are grateful to Prof. Robert Capon for providing *Nocardiopsis* sp. CMB-M0232 and to Prof. Sergey B. Zotchev for providing pKE5. ALL acknowledges funding from a Research Corporation Cottrell College Science Award, University of North Florida (UNF) Transformational Learning Opportunity Awards, UNF Academic Affairs Scholarship Grants, and a UNF Faculty Fellowship Award.
CHAPTER 5

Experimental Section
5.1. Culturing of Nocardiopsis sp. CMB M0322 and Extraction of Secondary Metabolites

Nocardiopsis sp. strain CMB-M0232, originally-isolated by the Capon group from a sediment sample obtained from South Molle Island from a depth of 55 m\(^87\) was provided by Prof. Robert Capon (University of Queensland), and maintained following previously described procedures as pure strains of individual colonies. This organism was maintained and propagated for further culturing using standard microbiological techniques. Nocardiopsis sp. CMB-M0232 were grown to dense colonies on a single agar plate (comprising 25 mL of 1% starch, 0.4% yeast extract, 0.2% peptone, 1.8% agar, and 0.0005% rifampicin) under incubation at 27 °C for four weeks. For larger laboratory culture and for generating extracts, a frozen glycerol stock culture (1.2 mL) of Nocardiopsis sp. (CMB-M0232) was used to inoculate a 250 mL Schott flask containing 80 mL of M1 broth (1% starch, 0.4% yeast extract and 0.2% peptone dissolved in deionized water, supplemented with 3% (by weight) ocean salt (Instant Ocean\textsuperscript{®}, USA). The flask was shaken at 225 rpm in a rotary shaking incubator for ~6-14 d at 27 °C depending on the maturity of each inoculation, as measured by OD\(_{600}\). An aliquot of this seed culture (5.0 mL, average OD\(_{600} = 0.6\)) was used to inoculate each of six 2 L Schott flasks containing 500 mL of M1 broth, and fermentation was continued for a further 8-21 d (at 27 °C and with rotary shaking at 225 rpm). Following fermentation, the culture was extracted with an equal volume of EtOAc (i.e. 500 mL per flask) and the combined

organic phase concentrated in vacuo to yield a crude extract (250 mg). The crude extract was triturated sequentially with hexane (25 mL), CH$_2$Cl$_2$ (25 mL) and MeOH (25 mL), to afford individual fractions of 71.2 mg, 56.3 mg and 13.6 mg respectively. Fractions were concentrated to dryness in vacuo, and the CH$_2$Cl$_2$ fraction was subsequently subjected to HPLC fractionation (Zorbax CN 5 μm, 250 × 9.4 mm column, 4 mL/min gradient elution from 60% H$_2$O/MeOH to 100% MeOH over 55 min, with a hold at 100% MeCN for 5 min) to yield multiple metabolites as described below. Genomic DNA was isolated from Nocardiopsis sp. CMB M0322 and was subjected to sequencing.

5.2. Nocardiopsis sp. CMB-M0232 genomic DNA isolation, draft genome sequencing and pathway annotation.

Genomic DNA (gDNA) was isolated by standard procedures. Sequencing was conducted on this gDNA by Cofactor Genomics (St. Louis, MO) using Illumina with paired-end 80 base pair reads and 454 with single-end 400 bp reads. A second assembly was independently performed by Beckman Coulter Genomics and the overall number of contiguous sequence now stands at 754. Initial assembly yielded 1,223 contiguous sequence with an average length of 5.2 kB. ORFs were predicted using GeneMark. The initial assembly was deemed sufficient for elucidation of gene clusters presented in this thesis. To locate candidate enzymes for nocardiopsin biosynthesis, ORFs were searched for homology to representative modular PKSs and NRPSs using BLASTP in Geneious.


BLASTP searches were employed to determine ORFs with homology to those in the NCBI database for alkaloidal pathways in Chapters 3 and 4.

5.3. Cloning and sequencing of the nocardiopsin gene cluster (nsn).

Individual contiguous sequence from the draft genome sequence included only partial NRPS and PKS gene clusters. To complete the nsn pathway, a cosmid library of gDNA was constructed and screened for these gene clusters. To prepare the cosmid library, gDNA was digested with Sau3AI to yield ~30-50kB gDNA fragments. These gDNA fragments were ligated into SuperCos vector following manufacturer guidelines (Agilent Technologies). Ligations were packaged with MaxPlax lambda packaging extracts (EpiCentre), and introduced into E. coli XL1-MRF Blue following manufacturer directions. The resulting ~2,000 member gDNA cosmid library was stored in 96-well microtiter plates at -80 °C.

This cosmid clone library was systematically screened by PCR to isolate members carrying fragments of the nsn gene cluster. Library members within individual microtiter plates were pooled by stamping culture aliquots from microtiter plates onto LB agar, incubating overnight at 37 °C, and scraping colonies into 100 μL of sterile water. These pooled samples were used as templates for PCR experiments; primers were designed to amplify putative portions of the nsn cluster from the draft genome sequence. Each PCR contained 20.2 μL of molecular biology grade water, 0.5 μL of template, 0.5 μL of dNTPs (200 μM of each dNTP), 2.5 μL of 10X ThermoPol buffer, 1 μL DMSO, 0.1 μL each of NRPSPipe_F and NRPSPipe_R primers (0.4 μM each, and 0.1 μL of ThermoPol.
Taq DNA polymerase (New England Biolabs). PCR was conducted with an initial
denaturation cycle of 94 °C for 3:00, followed by 30 cycles of 94 °C for 45s, 60 °C for 30
s, and 72 °C for 60 s, a final extension cycle of 72 °C for 5 min, and a hold at 4 °C until
evaluation of products by agarose gel electrophoresis.

After library plates containing cosmid(s) of interest were identified by PCR,
library members from individual rows and columns of these plates were prepared, pooled,
and again evaluated by PCR to reveal specific microtiter plate wells containing cosmids
of interest. End sequencing was completed on pure cosmids carrying targeted genes, and
each end sequence was evaluated for homology to genes predicted to encode natural
product biosynthesis. These end sequences were used to design additional primers to
extend the gene cluster by further library screening, analogous to above and using
primers described in Supporting Information. Ultimately, three overlapping cosmids were
subjected to Sanger and 454 sequencing and assembly by Operon Genomics (Huntsville,
AL) to yield a single contig.

5.4. Bioinformatics analyses of the no cardiopsin gene cluster.90

ORFs within the nsn gene cluster were proposed using FramePlot 91 and
FGENESB (Softberry).92 Each ORF was evaluated for homology to previously reported

90 a. Bachmann, B. O.; Ravel, J. Chapter 8. Methods for in Silico Prediction of Microbial Polyketide and
Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data. Methods in Enzymology. 2009,
458, 181-217; b. Yadav, G.; Gokhale, R. S.; Mohanty, D. SEARCHPKS: A Program for Detection and
Overmars, L.; Siezen, R.J.; Francke, c. Classification of the Adenylation and Acyl-Transferase Activity of
18, e62136.
91 Ishikawa, J.; Hotta, K. FramePlot: a New Implementation of the Frame Analysis for Predicting Protein-
Coding Regions in Bacterial DNA with a High G + C Content FEMS. Microbiol. Lett. 1999, 174, 251-253
proteins using BLASTP. Individual modules and domains within PKSs and the NRPS were predicted using the tool developed by Bachmann and Ravel.\textsuperscript{90} AT and C domain substrate preference was assigned using methods developed by Khayatt et al.\textsuperscript{90} and the SEARCHPKS program by Yadav et al.\textsuperscript{90} Amino acid alignments were completed using ClustalW.\textsuperscript{93}

5.5. Preparation of the construct for expression of lysine cyclodeaminase gene NsnL in \textit{E. coli}.

\textsuperscript{92} Solovyev, V.; Salamov, A. \textit{Automatic annotation of microbial genomes and metagenomic sequences, Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies} (Ed.: R. W. Li), Nova Science Publishers, 2011, pp. 61-78.

The gene encoding NsnL was PCR amplified using forward and reverse primers (NsnL_pETF and NsnL_pETR, Supporting Information) incorporating NspV and NcoI restriction sites for cloning into the corresponding sites of the pET30a(+) expression vector to yield a fusion protein with an N-terminal hexahistidine tag. The PCR mixture was analogous to that described above, except that Roche High Fidelity Taq DNA polymerase and corresponding buffer was utilized. PCR was completed with an initial denaturation cycle of 94 °C for 2:00, followed by 15 cycles of 94 °C for 45s, 55 °C for 45s, and 72 °C for 60 s, 10 cycles of 94 °C for 45s, 60 °C for 45 s, and 72 °C for 60 s, and a final extension cycle of 72 °C for 5 min. PCR products were held at 4 °C until analysis by agarose gel electrophoresis. The ~1.1 kb PCR product was ligated into pGEM-T vector (Promega) and the resulting construct propagated in *E. coli* JM109. The construct was digested with NspV and NcoI to release the *nsnL* PCR product with sticky ends for ligation into pET30a(+) (Novagen). The *nsnL* gene was ligated into equivalently digested pET30a(+), the resulting construct propagated in JM109, and then introduced into *E. coli* BL21(DE3) for gene expression. DNA sequencing confirmed the *nsnL* gene insert sequence.

5.6. Overexpression and purification of lysine cyclodeaminase NsnL.

A 6 h culture of BL21(DE3) carrying pET30a(+)/*nsnL* was diluted by 1:1000 in LB containing kanamycin (30 μg/mL) and incubated for ~4 h at 200 rpm and 37 °C to achieve an optical density (OD600) of ~0.5-0.6. The culture was supplemented with 1 mM IPTG to induce expression of *nsnL*; cultures were incubated for an additional 13 h at 15 °C with shaking at 250 rpm. Cells were pelleted by centrifugation and stored at -80 °C.
until protein purification. To purify NsnL, 1 g wet mass of cells were lysed by mixing with 4 mL BPER (Thermo Scientific) supplemented with lysozyme (1.5 mg/mL) and DNase (1.5 U/mL, Qiagen) and incubating at 30 °C for 30 min. Resulting lysates were clarified by centrifugation at 12,000 rpm for 30 min at 4 °C. The supernatant was passed through a 0.4 μm filter and supplemented to achieve a final concentration of 400 mM NaCl, 10% glycerol, and 2 mM imidazole in 25 mM Tris (pH 8). To batch bind NsnL, the solution (4 mL) was incubated with 0.5 mL of Ni agarose slurry (Qiagen) by shaking on ice at 200 rpm for 2 h. This mixture was poured into a column, and 1.25 mL of the above buffer was used to wash the column. The column was then eluted sequentially with 250 μL aliquots of the above buffer supplemented with final concentrations of 5 mM, 20 mM, 40 mM, 60 mM, and 200 mM imidazole, respectively. Finally, the column was eluted with 1.25 mL of buffer containing 500 mM imidazole. In this final fraction, NsnL (39.4 kDa) was enriched to >90% purity, as determined by SDS-PAGE (Supporting Information). The buffer was exchanged by two-step dialysis at 4 °C, first with 500 mL of 50 mM Tris buffer (pH 8) containing 100 mM NaCl, 1 mM EDTA, and 10% glycerol, followed by dialysis with 500 mL of 50 mM Tris buffer (pH 8) containing 100 mM NaCl, 10% glycerol, and 1 mM TCEP. NsnL was concentrated by ~4-fold using a 9 kD cutoff centrifugal concentrator, the protein concentration was determined by the BCA method, and NsnL was stored at -80 °C. The typical yield of NsnL was ~15 mg per L of culture.
5.7. Evaluation of the function of NsnL through in vitro assays and GC-FID.

Treatment and control assays were prepared in a final reaction volume of 100 μL containing 100 mM potassium phosphate buffer (pH 8), 400 μM NAD+, and 3 mM DTT. In addition to these components, treatment samples contained 17.5 μM NsnL as well as 10 mM of L-lysine, D-lysine, or L-ornithine. Control samples contained individual amino acid substrates in the absence of NsnL, or NsnL in the absence of added substrate. Assays analogous to those described above were also setup without the addition of NAD+ to evaluate the effect of NAD+ addition on NsnL activity.

Assays were incubated overnight (~15 h) at 30 °C with shaking at ~200 rpm. Resulting reaction products were immediately extracted, derivatized, and subjected to GC-FID using the EZ:Faast GC-FID kit (Phenomenex) and following manufacturer recommendations. Synthetic L-pipecolic acid (Sigma) and L-proline (Sigma) were equivalently derivatized and used as standards.

5.8. General Experimental and Instrumentation for Asymmetric Synthesis and HPLC-MS-MS

Reagents, Solvents and Glassware. All small-scale dry reactions were carried out under a blanket of nitrogen, using standard syringe-septum, and cannulation techniques.94 AIBN was recrystallized from ether and stored at 0-5 °C in an amber bottle. Dry THF was obtained by distillation over sodium-benzophenone ketyl. Dry triethyl amine and diisobutyl ethyl amine was obtained after distillation over KOH. Dry

dichloromethane and dry DMF were prepared by distilling over calcium hydride.
Anhydrous ether and hexanes were obtained from an m-Braun solvent purification system (charged with A2 alumina as a desiccant). All other solvents were purified according to specific literature procedures.

**Chromatography.** Analytical thin-layer chromatography (TLC) was performed with silica Gel 60 Å (230-400 mesh) specifically to monitor the progress of each chemical reaction and used as a guide for purification of the ensuing mixtures. These were conducted on glass plates (7.5 x 2.5 and 7.5 x 5.0 cm) coated with silica gel G containing 13% calcium sulfate as binder or on pre-coated 0.2 mm thick 60 F254 silica plates and various combinations of ethyl acetate and hexane were used as eluent. Visualization of spots after TLC was accomplished by exposure to iodine vapour and/or UV light (254 nm). All compounds were purified using flash column chromatography (Silica gel grade: 200-400 mesh, 40-63 μm) at medium pressure (20 psi). Preparatory thin-layer chromatography (TLC) (to obtain purified compounds) for select products were performed on glass plates (7.5 x 2.5 and 7.5 x 5.0 cm) coated with 60 Å silica gel. Yields refer to compounds isolated to analytical purity after chromatography.

5.9. **Structural Characterization of Synthetic and Biosynthetic Intermediates.**

NMR spectroscopic analyses (1H NMR, 13C NMR) were conducted for all new compounds. 1H (400 MHz) and 13C (100 MHz) spectra were recorded on a 400 MHz

---


spectrometer, with the exception of a few compounds recorded on a 600 MHz spectrophotometer ($^1$H: 600 MHz and $^{13}$C: 150 MHz). Pertinent frequency is specifically reported for each compound. Chemical shift values (δ) for NMR spectra are reported in parts per million (ppm) relative to the residual (indicated) solvent peak (CDCl$_3$ or CD$_3$OD). Additional peaks other than the compound in question, if any, are calibrated based on reported values for trace impurities. Coupling constants are reported in Hz. Data for $^1$H NMR are reported as follows: chemical shift (δ, ppm), multiplicity (s = singlet, brs = broad singlet, d = doublet, t = triplet, q = quartet, ddd = double double doublet, m = multiplet, cm = complex multiplet), integration corresponding to the number of protons followed by coupling constants in Hz. For $^{13}$C NMR spectra, the nature of the carbons (C, CH, CH$_2$ or CH$_3$) was determined by recording the Distortionless Enhancement by Polarization Transfer (DEPT) experiment, and notations are provided in parentheses. $^{13}$C NMR data is reported in parts per million (δ) relative to the residual (indicated) solvent peak. All melting points for solids were determined on a Buchi B-540 instrument and are reported uncorrected. pH determination was performed with a standard pH meter. IR spectra were recorded on a FT-IR spectrophotometer. Chiroptical measurements ([α]$_D$) were obtained on a polarimeter in a 100 × 2 mm cell. Chiral HPLC analyses for enantio-enriched synthetic intermediates were performed using a Shimadzu LC-20-AT Series separations module equipped with Shimadzu SPD-M20A PDA (photo diode array) multiple wavelength detector (180nm-800nm). The overall system, CBM-20 was controlled using LC Solutions software. Raw data was plotted using Origin®

software program after exporting absorbance data as an ASCII-formatted file. Analytical separations of enantioenriched mixtures were carried out on Daicel® (normal phase) AS chiral column. High-resolution mass spectrometry (HRMS) data for synthetic compounds reported herein were obtained by direct infusion of methanolic solutions on a HDMS QTOF mass spectrometer. Accurate LC-MS-MS data of biological extracts were recorded with a Waters Acquity I-Class UPLC system and a Waters Synapt G2 HDMS mass spectrometer as described below.
Table 4. List of chemical structures and corresponding numbers assigned.

<table>
<thead>
<tr>
<th>Compound #</th>
<th>Structure</th>
<th>Published / reference name given in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>nocardiopsin A</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>nocardiopsin B</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>nocardioazine A</td>
</tr>
<tr>
<td>4</td>
<td><img src="image4.png" alt="Structure 4" /></td>
<td>nocardioazine B</td>
</tr>
<tr>
<td>5* and (ent-5)*</td>
<td><img src="image5.png" alt="Structure 5" /></td>
<td><em>cyclo</em>-L-Trp-L-Trp DKP and <em>cyclo</em>-D-Trp-D-Trp DKP</td>
</tr>
</tbody>
</table>

Note: Synthetic targets assembled in this study as putative intermediates in nocardioazine biosynthetic pathway are denoted by *. 
<table>
<thead>
<tr>
<th></th>
<th><img src="image" alt="Chemical Structure" /></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>cyclo-L-Trp-D-Trp DKP</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Late stage epoxide intermediate</td>
<td></td>
</tr>
<tr>
<td>8 and ent-8</td>
<td><img src="image" alt="Chemical Structure" /></td>
<td>N-Cbz-L-Trp-acid or N-Cbz-D-Trp-acid</td>
</tr>
<tr>
<td>9 and ent-9</td>
<td><img src="image" alt="Chemical Structure" /></td>
<td>L-Trp methyl ester or D-Trp methyl ester</td>
</tr>
<tr>
<td>10 and ent-10</td>
<td><img src="image" alt="Chemical Structure" /></td>
<td>L-Trp-N-Cbz-L-Trp-COOMe-dimer or D-Trp-N-Cbz-D-Trp-COOMe-dimer</td>
</tr>
<tr>
<td>11 and ent-11</td>
<td><img src="image" alt="Chemical Structure" /></td>
<td>L-Trp-L-Trp-COOMe-dimer or D-Trp-D-Trp-COOMe-dimer</td>
</tr>
<tr>
<td>12*</td>
<td><img src="image" alt="Chemical Structure" /></td>
<td>cyclo-L-Trp- C3′-&quot;prenyl-L-Trp DKP</td>
</tr>
<tr>
<td>13*</td>
<td><img src="image" alt="Chemical Structure" /></td>
<td>cyclo-C3-Me-L-Trp -L-Trp DKP</td>
</tr>
<tr>
<td></td>
<td>Structure</td>
<td>Name</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>14*</td>
<td><img src="image1" alt="Structure" /></td>
<td><strong>cyclo-N1′-Me-L-Trp-L-Trp DKP</strong></td>
</tr>
<tr>
<td>15*</td>
<td><img src="image2" alt="Structure" /></td>
<td><strong>cyclo-L-Trp-N1′-Me-C3′-&quot;prenyl-L-Trp DKP</strong></td>
</tr>
<tr>
<td>16*</td>
<td><img src="image3" alt="Structure" /></td>
<td><strong>cyclo-C3-Me-L-Trp-N1′-Me-L-Trp DKP</strong></td>
</tr>
<tr>
<td>17*</td>
<td><img src="image4" alt="Structure" /></td>
<td><strong>Des-N1′-Me-nocardioazine B</strong></td>
</tr>
<tr>
<td>18</td>
<td><img src="image5" alt="Structure" /></td>
<td><strong>C3-Methyl-indole</strong></td>
</tr>
<tr>
<td>19</td>
<td><img src="image6" alt="Structure" /></td>
<td><strong>Methyl 2-(benzyloxy)carbonyl-amino acrylate</strong></td>
</tr>
<tr>
<td>20a, b</td>
<td><img src="image7" alt="Structure" /></td>
<td><strong>(2S,3aR,8aR)-1-((benzyloxy)carbonyl)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carboxylic acid methyl ester</strong></td>
</tr>
<tr>
<td>20c</td>
<td><img src="image8" alt="Structure" /></td>
<td><strong>(2S,3aR,8aR)-1-((benzyloxy)carbonyl)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carboxylic acid</strong></td>
</tr>
<tr>
<td></td>
<td>Chemical Structure</td>
<td>Name</td>
</tr>
<tr>
<td>---</td>
<td>-------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>21</td>
<td><img src="image" alt="N-Me-L-Trp-Methyl Ester" /></td>
<td>N-Me-L-Trp-Methyl Ester</td>
</tr>
<tr>
<td>22</td>
<td><img src="image" alt="Benzyl-(2S,3aR,8aR)-2-((3-(1H-indol-3-yl)-1-methoxy-1-oxopropan-2-yl)carbamoyl)-3a-methyl-3,3a,8,8a-tetrahydropyrrolo[2,3-b]indole-1(2H)-carboxylate" /></td>
<td>Benzyl-(2S,3aR,8aR)-2-((3-(1H-indol-3-yl)-1-methoxy-1-oxopropan-2-yl)carbamoyl)-3a-methyl-3,3a,8,8a-tetrahydropyrrolo[2,3-b]indole-1(2H)-carboxylate</td>
</tr>
<tr>
<td>23</td>
<td><img src="image" alt="Methyl((2S,3aR,8aS)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carbonyl)tryptophanate" /></td>
<td>Methyl((2S,3aR,8aS)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carbonyl)tryptophanate</td>
</tr>
<tr>
<td>24</td>
<td><img src="image" alt="Benzyl-(2S,3aR,8aR)-2-((1-methoxy-3-(1-methyl-1H-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-3a-methyl-3,3a,8,8a-tetrahydropyrrolo[2,3-b]indole-1(2H)-carboxylate" /></td>
<td>Benzyl-(2S,3aR,8aR)-2-((1-methoxy-3-(1-methyl-1H-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-3a-methyl-3,3a,8,8a-tetrahydropyrrolo[2,3-b]indole-1(2H)-carboxylate</td>
</tr>
<tr>
<td>25</td>
<td><img src="image" alt="N-methyl-(2S,3aR,8aS)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carbonyl methyl-1-tryptophanate" /></td>
<td>N-methyl-(2S,3aR,8aS)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carbonyl methyl-1-tryptophanate</td>
</tr>
<tr>
<td>26</td>
<td><img src="image" alt="N-Me-L-Trp-COOMe ester" /></td>
<td>N-Me-L-Trp-COOMe ester</td>
</tr>
<tr>
<td>27a/b</td>
<td><img src="image" alt="C3’-‘prenyl-L-Trp-pyrroloindoline methyl ester" /></td>
<td>C3’-‘prenyl-L-Trp-pyrroloindoline methyl ester</td>
</tr>
<tr>
<td>28a/b</td>
<td><img src="image" alt="N1’-Me-C3’-‘prenyl-L-Trp-pyrroloindoline methyl ester" /></td>
<td>N1’-Me-C3’-‘prenyl-L-Trp-pyrroloindoline methyl ester</td>
</tr>
<tr>
<td></td>
<td>Chemical Structure</td>
<td>Name</td>
</tr>
<tr>
<td>---</td>
<td>-------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>29</td>
<td><img src="image" alt="Molecule 29" /></td>
<td>N1′-phthalyl-L-Trp-acid</td>
</tr>
<tr>
<td>30</td>
<td><img src="image" alt="Molecule 30" /></td>
<td>C3′-&quot;prenyl-pyrrolindoline-methyl-ester-N1-phthalyl-L-Trp-amide</td>
</tr>
<tr>
<td>31</td>
<td><img src="image" alt="Molecule 31" /></td>
<td>N1′-Me-C3′-&quot;prenyl-pyrrolindoline-methyl-ester-N1-phthalyl-L-Trp-amide</td>
</tr>
<tr>
<td>32</td>
<td><img src="image" alt="Molecule 32" /></td>
<td>L-Serine methyl ester hydrochloride</td>
</tr>
<tr>
<td>33</td>
<td><img src="image" alt="Molecule 33" /></td>
<td>N-Cbz-L-Serine methyl ester</td>
</tr>
<tr>
<td>34</td>
<td><img src="image" alt="Molecule 34" /></td>
<td>O-Boc-N-Cbz-L-Serine methyl ester</td>
</tr>
<tr>
<td>35</td>
<td><img src="image" alt="Molecule 35" /></td>
<td>Methyl (S)-2-(1,3-dioxoisodolin-2-yl)-3-(1H-indol-3-yl)propanoate</td>
</tr>
<tr>
<td>36</td>
<td><img src="image" alt="Molecule 36" /></td>
<td>N1-Me-N2-phth-L-Trp-methyl ester</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>38</td>
<td><img src="image" alt="Malonic acid" /></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td><img src="image" alt="3-(tert-butoxy)-3-oxopropanoic acid" /></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td><img src="image" alt="Tert-butyl 3-((2-acetamidoethyl) thio)-3-oxopropanoate" /></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td><img src="image" alt="3-((2-acetamidoethyl) thio)-3-oxopropanoic acid" /></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td><img src="image" alt="Methlymalonic acid" /></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td><img src="image" alt="3-(tert-butoxy)-2-methyl-3-oxopropanoic acid" /></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td><img src="image" alt="Tert-butyl 3-((2-acetamidoethyl) thio)-2-methyl-3-oxopropanoate" /></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td><img src="image" alt="3-((2-acetamidoethyl)thio)-2-methyl-3-oxopropanoic acid" /></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structure</td>
<td>Name</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>46 and ent-46</td>
<td><img src="image1.png" alt="Structure Image" /></td>
<td>D/L-Trp- N-BOC-cyanomethyl ester</td>
</tr>
<tr>
<td>47 and ent-47</td>
<td><img src="image2.png" alt="Structure Image" /></td>
<td>D/L-Trp-cyanomethyl ester</td>
</tr>
</tbody>
</table>
5.10. Synthesis of cyclo-L-Trp-L-Trp DKP and cyclo-D-Trp-D-Trp DKP

L/D-Trp-N-Cbz-carbamate (8 or ent-8)\(^{100}\)

To a clear solution of L/D-Trp (500 mg, 2.45 mmol) in 20 mL of CH\(_3\)CN-H\(_2\)O (2:3) were added NaHCO\(_3\) (308 mg, 3.68 mmol) and Na\(_2\)CO\(_3\) (390 mg, 3.68 mmol). The resulting turbid solution (pH = 10~11) was cooled to 0 °C (H\(_2\)O/ice bath) and stirred for 15 min. To this mixture was added Cbz-Cl (350 μL, 1.20 mmol) drop wise. The resulting solution was stirred for 15-20 min at 0 °C, the ice bath was removed and reaction was stirred at rt for 3h, at which time no starting material remained (TLC analysis). After acidification by drop wise addition of a 1 N HCl solution and removal of CH\(_3\)CN by rotary evaporation, the reaction mixture was transferred to a separatory funnel and washed three times with EtOAc. The combined organic phase was washed with brine, dried over anhyd. Na\(_2\)SO\(_4\), and filtered. Concentration under reduced pressure gave 813 mg (98%) of 8 or ent-8 as a colorless powder which was directly used for the next step without further purification.

mp: 125-126 °C; IR (KBr): 3413, 3020, 2934, 1702, 1596, 1519, 1415, 1345, 1218, 1137,

1067, 760, 672 cm\(^{-1}\). \(^1\)H NMR (400 MHz, CD\(_3\)OD): \(\delta\) 7.58 (d, \(J = 8.0\) Hz, 1H), 7.33-7.23 (m, 7H), 7.06 (s, 1H), 7.09-7.02 (m, 1H), 6.95 (ddd, \(J = 7.8, 7.0, 0.8\) Hz, 1H), 5.04 (AB, \(J = 12.5\) Hz, 1H), 4.98 (AB, \(J = 12.5, \) Hz, 1H), 4.42 (dd, \(J = 7.4, 4.8\) Hz, 1H), 3.36 (ABX, \(J = 14.6, 4.8\) Hz, 1H), 3.14 (ABX, \(J = 14.6, 7.5\) Hz, 1H). \(^{13}\)C NMR (100 MHz, CD\(_3\)OD, DEPT): \(\delta\) 158.1 (C), 138.3 (C), 137.9 (C), 129.4 (2 * CH), 129.3 (C), 129.2 (C), 128.8 (CH), 128.7 (CH), 124.4 (CH), 122.1 (CH), 119.6 (CH), 119.5 (2 * CH), 112.1 (CH), 111.7 (C), 67.3 (CH\(_2\)), 57.8 (CH), 29.2 (CH\(_2\)). HRMS (EI, \(M^+\)): m/z calcd. for C\(_{19}\)H\(_{18}\)O\(_4\)N\(_2\) 338.1267, found 338.1262.

L/D-Trp methyl ester hydrochloride (9 or ent-9)

Thionyl chloride (7.15 mL, 98 mmol) was added drop wise to a cold (0 °C) solution of anhydrous methanol (220 mL) under magnetic stirring. The solution was stirred at 0 °C

---

for 30 min. and then L/D-Trp (8.00 g, 39.2 mmol) was added and the resulting solution was heated at 60 °C for 18 h. After evaporation of the solvent, a white residue of hydrochloride salt was obtained, which was neutralized by a satd. Na₂CO₃ solution (25 mL) and the ester was extracted with equal volume of ethyl acetate three times. The organic layer was dried over anhyd. Na₂SO₄ and evaporated under reduced pressure yielding pale yellow oil, which solidified upon standing to a pale yellow crystalline solid.

Yield: 8.48 g (99%, 38.8 mmol) of L/D-Trp methyl ester (9 or ent-9). mp: 91-93 °C IR (KBr): 3386, 3017, 1730, 1582, 1441, 1351, 1215, 1101, 1014, 758, 665 cm⁻¹. ¹H NMR (400 MHz, CDCl₃): δ 8.27 (brs, 1H), 7.62 (d, J = 7.9 Hz, 1H), 7.35 (td, J = 8.1, 0.9 Hz, 1H), 7.20 (ddd, J = 8.2, 7.1, 1.2 Hz, 1H), 7.13 (ddd, J = 8.0, 7.1, 1.1 Hz, 1H), 7.04 (d, J = 2.3 Hz, 1H), 3.85 (dd, J = 7.7, 4.8 Hz, 1H), 3.72 (s, 3H), 3.29 (ABXY, J = 14.4, 4.8, 0.8 Hz, 1H), 3.06 (ABXY, J = 14.4, 7.7, 0.4 Hz, 1H), 1.60 (brs, 2H). ¹³C NMR (100 MHz, CDCl₃, DEPT): δ 175.9 (C), 136.4 (C), 127.6 (C), 123.1 (CH), 122.3 (CH), 119.6 (CH), 118.7 (CH), 111.4 (CH), 111.2 (C), 55.1 (CH), 52.2 (CH₃), 30.8 (CH₂). HRMS (EI, M⁺): m/z calcd. for C₁₂H₁₄O₂N₂ 218.1055, found 218.1055.

Methyl ((benzyloxy)carbonyl)-L/D-tryptophyl-L/D-tryptophanate (10 or ent-10)
To a cold (−10 °C) magnetically stirred solution of \(N\)-Cbz-acid 8 or ent-8 (400 mg, 1.18 mmol) with D/L-Trp-methyl ester (ent-9 or 9) (284.0 mg, 1.30 mmol) in dry THF (5.0 mL) was added Et\(_3\)N (0.66 mL, 4.73 mmol) followed by BOP-Cl (903 mg, 3.55 mmol) and the resulting mixture was stirred at same temperature overnight and was then quenched by addition of water (20 mL) and extracted with ethyl acetate (3 × 20 mL). The combined organic layer was washed with brine and dried (anhyd. Na\(_2\)SO\(_4\)). Evaporation of the solvent under reduced pressure and purification of the residue on a silica gel column using ethyl acetate–hexanes (1:1) as eluent furnished the coupled-product 10 or ent-10 as a colorless dense liquid in 93% yield (595 mg, 1.11 mmol). \(^1\)H NMR (400 MHz, CDCl\(_3\)): \(\delta\) 7.80 (brs, 2H), 7.67 (d, \(J = 8.2\) Hz, 1H), 7.37-7.27 (m, 7H), 7.25 (d, \(J = 8.4\) Hz, 1H), 7.20 (t, \(J = 7.3\) Hz, 1H), 7.17-7.08 (m, 2H), 6.93 (t, \(J = 7.3\) Hz, 1H), 6.89 (s, 1H), 6.55 (s, 1H), 6.15 (d, \(J = 7.3\) Hz, 1H), 5.42 (d, \(J = 7.6\) Hz, 1H), 5.08 (s, 2H), 4.79 (td, \(J = 7.8, 5.4\) Hz, 1H), 4.51 (q, \(J = 4.4\) Hz, 1H), 3.61 (s, 3H), 3.35 (ABX, \(J = 13.7, 3.0\) Hz, 1H), 3.21-3.05 (m, 3H). \(^{13}\)C NMR (100 MHz, CDCl\(_3\), DEPT): \(\delta\) 171.9 (C), 171.1 (2 × C), 136.2 (C), 136.1 (C), 128.6 (4 × CH), 128.3 (CH), 128.1 (2 × CH), 127.4 (CH), 123.7 (CH), 123.1 (C), 122.3 (C), 122.2 (2 × CH), 119.9 (CH), 119.7 (CH), 119.0 (C), 118.5 (CH), 111.4 (CH), 111.3 (CH), 109.4 (CH), 67.0 (CH\(_2\)), 55.5 (CH\(_3\)), 52.8 (CH), 52.5 (CH), 28.6 (CH\(_2\)), 27.5 (CH\(_2\)). HRMS (EI, M\(^+\)): m/z calcd. for C\(_{31}\)H\(_{30}\)O\(_5\)N\(_4\) 538.2216, found 538.2182.
Methyl L/D-tryptophyl-L/D-tryptophanate (11 or ent-11)

To a homogenous solution (under stirring with a magnetic bar) of coupling compound 10 or ent-10 (235 mg, 0.44 mmol) in MeOH and ethyl acetate (1:1) was added 10% palladium on activated charcoal (5 mg, 0.04 mmol) and the reaction mixture stirred under hydrogen at 1 atm for 7 h. The solution was then filtered over a pad of celite and washed with ethyl acetate. Evaporation of the solvent under reduced pressure yielded 11 or ent-11 (98%, 173 mg, 0.43 mmol) in sufficiently pure form as an off-white solid and was subjected to the next step. mp: 207 °C. IR (KBr): 3407, 3018, 2938, 1723, 1672, 1515, 1350, 1221, 1060, 757, 678 cm$^{-1}$. $^1$H NMR (400 MHz, CD$_3$OD): $\delta$ 7.89 (s, 2H), 7.59 (td, $J$ = 8.0, 1.0 Hz, 1H), 7.34 (td, $J$ = 8.1, 1.0 Hz, 1H), 7.29 (td, $J$ = 8.1, 1.0 Hz, 1H), 7.24 (td, $J$ = 8.0, 1.0 Hz, 1H), 7.10 (ddd, $J$ = 8.0, 7.0, 1.0 Hz, 1H), 7.07 (s, 1H), 7.06-6.99 (m, 2H), 6.89 (ddd, $J$ = 8.0, 7.0, 1.0 Hz, 1H), 6.85 (s, 1H), 4.72 (t, $J$ = 6.0 Hz, 1H), 3.61 (s, 3H), 3.59 (dd, $J$ = 6.8, 5.6 Hz, 1H), 3.13 (ABXY, $J$ = 14.6, 6.6, 0.6 Hz, 1H), 3.09 (ABXY, $J$ = 14.6, 5.6, 0.6 Hz, 1H), 3.03 (ABXY, $J$ = 14.6, 5.6, 0.6 Hz, 1H), 2.98 (ABXY, $J$ = 14.2, 6.6, 0.6 Hz, 1H). $^{13}$C NMR (100 MHz, CD$_3$OD, DEPT): $\delta$ 177.1 (C),
cyclo-L/D-Trp-L/D-Trp DKP (5 or ent-5)

A homogenous solution (under stirring with a magnetic bar) of amine 11 or ent-11 (396 mg, 0.98 mmol) was refluxed overnight in 14 M methanolic ammonia (15.0 mL). Evaporation of the solvent under reduced pressure and washing of the resulting residue with chloroform furnished the pure diketopiperazine 5 or ent-5 (345 mg, 95% yield, 0.93 mmol) as a pale yellow solid. mp: 242 °C. IR (KBr): 3409, 3326, 3018, 2926, 2481, 1659, 1536, 1453, 1336, 1225, 1088, 1018, 932, 758, 669 cm⁻¹. ¹H NMR (400 MHz, CD₃OD): δ 7.45 (td, J = 8.0, 1.0 Hz, 2H), 7.30 (td, J = 8.0, 1.0 Hz, 2H), 7.09 (ddd, J = 8.1, 7.0, 1.1 Hz, 2H), 7.01 (ddd, J = 8.1, 7.0, 1.1 Hz, 2H), 6.46 (s, 2H), 4.04 (dd, J = 6.7, 3.9 Hz, 2H), 2.92 (dd, J = 14.4, 3.8 Hz, 2H), 2.17 (dd, J = 14.4, 7.2 Hz, 2H). ¹³C NMR (100 MHz, CD₃OD, DEPT): δ 169.7 (2 × C), 138.0 (2 × C), 128.6 (2 × C), 125.9 (2 × CH), 122.5 (2 × CH), 120.1 (2 × CH), 119.7 (2 × CH), 112.4 (2 × CH), 109.4 (2 × C), 56.8 (2 × CH), 31.4 (2 × CH₂). HRMS (EI, M⁺): m/z calcd. for C_{22}H_{20}O_{2}N_{4} 372.1586,
found 372.1595. \([\alpha]^{21}_D - 52 (c 0.05, \text{MeOH})\) for cyclo-L-Trp-L-Trp DKP (5) and \([\alpha]^{21}_D + 52 (c 0.05, \text{MeOH})\) for cyclo-D-Trp-D-Trp DKP (ent-5). HPLC of individual enantiomer provided below.

5.11. Asymmetric Synthesis of cyclo-C3-Me-L-Trp-L-Trp DKP (13)

\[
\begin{align*}
\text{L-Serine} & \xrightarrow{\text{SOCl}_2, \text{MeOH}, \text{rt}} \text{L-Serine methyl ester hydrochloride} \\
& \xrightarrow{\text{petroleum ether}} \text{L-Serine methyl ester hydrochloride (32)}
\end{align*}
\]

**L-serine methyl ester hydrochloride**\(^{103}\) (32)

Thionyl chloride (3.79 mL, 52.0 mmol) was added dropwise to a cold (0 °C) solution of anhydrous methanol (50 mL) under magnetic stirring. The solution was stirred at 0 °C for 30 min and then L-Serine (5.0 g, 47.6 mmol) was added. The reaction mixture was stirred at room temperature for 24 h and TLC analysis (CHCl$_3$/CH$_3$OH, 9:1) indicated complete disappearance of L-serine. The reaction mixture was evaporated under reduced pressure and the residue was triturated with petroleum ether (~5 times) to provide 7.2 g (98%, 46.6 mmol) of L-serine methyl ester hydrochloride salt (32) as a colorless powder which was directly used in the subsequent step without further purification. mp: 161–162 °C; IR (KBr): 3402, 3024, 2951, 2691, 1739, 1625, 1524, 1448, 1244, 1072, 893 cm$^{-1}$. \(^1\)H NMR (400 MHz, CDCl$_3$): $\delta$ 4.13 (t, $J = 3.8$ Hz, 1H), 4.01 (ABX, $J = 11.9$, 4.4 Hz, 1H), 3.91 (ABX, $J = 11.9$, 3.4 Hz, 1H), 3.85 (s, 3H). \(^1\)C NMR (100 MHz, CDCl$_3$, DEPT): $\delta$ 169.3


(C), 60.6 (CH₂), 56.1 (CH), 53.8 (CH₃). HRMS (ESI, M+H⁺): m/z calcd. for C₄H₁₀O₃N 120.0655, found 120.0660.

**N-Cbz-L-serine methyl ester (33)**

L-serine methyl ester hydrochloride (4.0 g, 25.7 mmol) was dissolved in a mixture of saturated NaHCO₃ (11.0 g in 50 mL H₂O) and CH₂Cl₂ (70.0 mL). To this solution, benzyl chloroformate (3.85 mL, 27.0 mmol) was added at 0 °C and the reaction mixture was stirred for 6 h at rt. After quenching the reaction with 1.0 M aqueous HCl at 0 °C, the organic layer was washed with water and brine and dried over anhyd. Na₂SO₄. The solvent was evaporated under reduced pressure, and the resultant residue was purified by column chromatography (EtOAc/hexane 1:1) to yield N-Cbz-L-serine methyl ester (33) (6.3 g, 24.9 mmol, 97% yield) as colorless oil. IR (KBr): 3393, 3024, 2951, 1715, 1532, 1443, 1342, 1224, 1066, 893 cm⁻¹. ¹H NMR (400 MHz, CDCl₃): δ 7.40-7.27 (m, 5H), 5.75 (d, J = 6.56 Hz, 1H), 5.12 (s, 2H), 4.45 (dd, J = 7.80, 3.68 Hz, 1H), 5.05-3.85 (m, 2H), 3.78 (s, 3H), 2.35 (t, J = 5.76 Hz, 1H). ¹³C NMR (100 MHz, CDCl₃, DEPT): δ 171.1 (C), 156.4 (C), 136.1 (C), 128.6 (2 × CH), 128.4 (2 × CH), 128.2 (CH), 67.3 (CH₂), 63.3 (CH₂), 56.1 (CH), 52.9 (CH₃). HRMS (EI, M⁺): m/z calcd. for C₁₂H₁₅O₅N 253.0950, found 253.0954.
Methyl 2-(benzyloxy)carbonyl-amino acrylate (19)

To a homogenous solution (under stirring) of (6.3 g, 24.8 mmol) N-Cbz-L-serine methyl ester (33) in dry CH₃CN (50 mL) was added DMAP (0.5 g, 4.1 mmol) followed by di-tert-butyl dicarbonate (5.16 g, 23.6 mmol) under rapid stirring at room temperature. The reaction was monitored by TLC (diethyl ether/n-hexane, 1:1) until all the reactants had been consumed. Tetra-methyl guanidine (1.57 mL, 12.4 mmol) was added at room temperature and the reaction mixture was further stirred overnight. Evaporation of the solvent at reduced pressure gave a residue that was partitioned between diethyl ether (100 mL) and water. The organic phase was washed with brine and dried (anhyd. Na₂SO₄). The solvent was evaporated under reduced pressure and the residue was purified by column chromatography to yield the pure olefin product (19) (3.03 g, 12.9 mmol, 52% yield) as a colorless liquid. IR (KBr): 3402, 3024, 2951, 2691, 1719, 1640, 1522, 1448, 1320, 1210, 1072, 893 cm⁻¹. ¹H NMR (400 MHz, CDCl₃): δ 7.40–7.29 (m, 5H), 6.27 (s, 1H), 5.80 (d, J = 1.5 Hz, 1H), 5.17 (s, 2H), 3.81 (s, 3H). ¹³C NMR (100 MHz, CDCl₃, DEPT): δ 164.2 (C), 153.1 (C), 135.8 (C), 131.0 (C), 128.6 (2 × CH), 128.4 (CH), 128.3 (2 × CH), 106.1 (CH₂), 67.1 (CH₂), 53.0 (CH₃). HRMS (EI, M⁺): m/z calcd. for C₁₂H₁₃O₄N 235.0845, found 235.0845.

The intermediate O-Boc-N-Cbz-L-serine methyl ester (34) was characterized as a colorless liquid. IR (KBr): 3366, 2972, 1735, 1524, 1453, 1366, 1274, 1165, 1075, 853.
$^1$H NMR (400 MHz, CDCl$_3$): $\delta$ 7.38-7.28 (m, 5H), 5.61 (d, $J = 8.3$ Hz, 1H), 5.12 (s, 2H), 4.61 (dd, $J = 8.4$, 3.6 Hz, 1H), 4.48 (ABX, $J = 11.2$, 3.6 Hz, 1H), 4.34 (ABX, $J = 11.2$, 3.6 Hz, 1H), 3.78 (s, 3H), 1.46 (s, 9H). $^{13}$C NMR (100 MHz, CDCl$_3$, DEPT): $\delta$ 169.9 (C), 155.8(C), 153.1 (C), 136.1 (C), 128.6 (2 × CH), 128.3 (CH), 128.2 (2 × CH), 83.0 (C), 67.3 (CH$_2$), 66.2 (CH$_2$), 53.5 (CH), 53.0 (CH$_3$), 27.7 (3 × CH$_3$). HRMS (ESI, M+H$^+$): m/z calcd. for C$_{17}$H$_{24}$O$_7$N 354.1547, found 354.1552.

1-benzyl-2-methyl-(2R,3aR,8aR)-3a-methyl-3,3a,8,8a-tetrahydropyrrolo[2,3-b]indole-1,2(2H)-dicarboxylate (20a)

Enantioselective formation of 20a followed literature procedure reported by Repka et al.$^{104}$ Accordingly, to a homogenous solution (under stirring) of 3-methyl indole (0.9 g, 6.86 mmol) in dry CH$_2$Cl$_2$ (40.0 mL) was added (S)–BINOL (0.393 g, 1.37 mmol) and methyl 2-(benzyloxy)carbonyl-amino acrylate (19) (1.61 g, 6.86 mmol) followed by slow addition of SnCl$_4$ (1.2 equiv. in 1.0 M CH$_2$Cl$_2$) at room temperature over a period of 30 minutes and stirring was continued for 4 h. To the reaction mixture 1M HCl was added and the organic layers extracted with CH$_2$Cl$_2$. The organic phase was washed with brine and dried (anhyd. Na$_2$SO$_4$). The solvent was evaporated under reduced pressure and the

residue was purified by column chromatography to yield the cyclic product 20 in 61% yield (1.54 g, 4.20 mmol). The product was detected to be present as a mixture of rotational isomers in 3:2 ratio as indicated by $^1$H NMR signals. IR (KBr): 3396, 3034, 2950, 1743, 1703, 1607, 1448, 1454, 1416, 1344, 1271, 1207, 1172, 1127, 1062, 1001, 959, 917, 818, 750, 696 cm$^{-1}$. **Data for major isomer:** $^1$H NMR (400 MHz, CDCl$_3$) δ 7.43-7.28 (m, 5H), 7.12-7.03 (m, 2H), 6.81-6.74 (m, 1H), 6.64 (d, $J = 7.7$ Hz, 1H), 5.41 (brs, 1H), 5.27 (s, 1H), 5.19 (AB, $J = 12.2$ Hz, 1H), 4.93 (AB, $J = 12.2$ Hz, 1H), 4.11 (t, $J = 7.8$ Hz, 1H), 3.47 (s, 3H), 2.65 (ABX, $J = 12.9$, 7.7 Hz, 1H), 2.16 (ABX, $J = 12.9$, 8.4 Hz, 1H), 1.40 (s, 3H). $^{13}$C NMR (100 MHz, CDCl$_3$, DEPT): δ 173.1 (C), 154.3 (C), 147.9 (C), 135.8 (C), 133.3 (C), 128.8 (CH), 128.4 (2 * CH), 128.2 (CH), 128.0 (2 * CH), 122.4 (CH), 119.3 (CH), 109.9 (CH), 83.5 (CH), 67.3 (CH$_2$), 59.4 (CH), 52.2 (CH$_3$), 52.1 (C), 41.7 (CH$_2$), 24.1 (CH$_3$). HRMS (EI, M$^+$): m/z calcd. for C$_{21}$H$_{22}$O$_4$N$_2$ 366.1580, found 366.1580.

(2S,3aR,8aR)-1-((benzyloxy)carbonyl)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carboxylic acid (20c)

To a homogenous solution (under stirring with a magnetic bar) of N-Cbz-methyl ester 20a (373 mg, 1.02 mmol) in MeOH (7.0 mL) and THF (7.0 mL) was added aqueous solution of LiOH (98.0 mg, 4.0 mmol in 7.0 mL H$_2$O) and the reaction mixture stirred at
room temperature overnight. It was then quenched with 1N HCl at 0 °C till the pH is between 4-5 followed by extraction with ethyl acetate (3 × 10 mL). The combined organic layers were washed with brine and dried (anhyd. Na₂SO₄). Evaporation of the solvent under reduced pressure and purification of the residue on a silica gel column using ethyl acetate–hexanes (1:1) as eluent furnished the acid 20c in 69% yield (241 mg, 0.69 mmol).

A similar prep was executed for synthesis of 20c on a gram-scale: To a magnetically stirred solution of N-Cbz-methyl ester 20a (1.0 g, 2.73 mmol) in MeOH (10.0 mL) and THF (10.0 mL) was added excess aqueous solution of LiOH (656.0 mg, 10.0 mmol in 10.0 mL H₂O) and the reaction mixture stirred at room temperature overnight. It was then quenched with 1N HCl at 0 °C till the pH of solution become 4-5 and extracted with ethyl acetate (3 × 50 mL). The combined organic layer was washed with brine and dried (anhyd. Na₂SO₄). Evaporation of the solvent under reduced pressure and purification of the residue on a silica gel column using ethyl acetate–hexanes (1:1) as eluent furnished the acid 20c in 50% yield (480 mg, 1.36 mmol). Pale yellow liquid.

IR (KBr): 3391, 2961, 1701, 1608, 1462, 1414, 1354, 1317, 1205, 1157, 1127, 1050, 1016, 978, 917, 746 cm⁻¹. ¹H NMR (400 MHz, CDCl₃): δ 7.45-7.36 (m, 2H), 7.33-7.20 (m, 3H), 7.13-7.03 (m, 1H), 6.82-6.75 (m, 1H), 6.65 (d, J = 7.7 Hz 1H), 5.28 (s, 1H), 5.14 (AB, J = 12.4 Hz, 1H), 5.01 (AB, J = 12.4 Hz, 1H), 4.18 (t, J = 8.0 Hz, 1H), 2.72 (ABX, J = 13.0, 8.1 Hz, 1H), 2.22 (ABX, J = 13.0, 8.0 Hz, 1H), 1.41 (s, 3H). ¹³C NMR (100 MHz, CDCl₃, DEPT): δ 178.3 (C), 154.5 (C), 147.9 (C), 135.9 (C), 133.3 (C), 129.0 (CH), 128.5 (2 × CH), 128.2 (CH), 127.7 (2 × CH), 122.5 (CH), 119.4 (CH), 110.1
Benzyl-(2S,3aR,8aR)-2-((3-(1H-indol-3-yl)-1-methoxy-1-oxopropan-2-yl)carbamoyl)-3a-methyl-3,3a,8,8a-tetrahydropyrrolo[2,3-b]indole-1(2H)-carboxylate (22)

To a homogenous solution of \(N\)-Cbz-acid (20c) (218.0 mg, 0.62 mmol) and L-Trp-methyl ester hydrochloride (9) (151.0 mg, 0.69 mmol) in dry THF (4.0 mL) (under stirring with a magnetic bar) at \(-10\) °C was added Et\(_3\)N (0.38 mL, 2.75 mmol) followed by BOP-Cl (395.0 mg, 1.55 mmol) and the resulting mixture was stirred at \(-10\) °C overnight. It was then quenched with water (5.0 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic layers were washed with brine and dried (anhy. Na\(_2\)SO\(_4\)). Evaporation of the solvent under reduced pressure and purification of the residue on a silica gel column using ethyl acetate–hexanes (1:1) as eluent furnished the coupled product (22) as a colorless dense liquid in 90% yield (306 mg, 0.55 mmol). IR (KBr): 3359, 3012, 2960, 2927, 1684, 1612, 1519, 1418, 1349, 1216, 1155, 1127, 748 cm\(^{-1}\). \(^1\)H NMR (400 MHz, CDCl\(_3\)): \(\delta\) 8.54 (brs, 1H), 7.44 (d, \(J = 7.8\) Hz, 1H), 7.38 (brs, 2H), 7.32 (d, \(J = 8.0\) Hz, 1H), 7.30-6.95 (m, 8H), 6.84 (d, \(J = 2.2\) Hz, 1H), 6.76 (t, \(J = 7.4\) Hz, 1H), 6.60 (d, \(J = 7.8\) Hz, 1H), 5.40 (brs, 1H), 5.11 (s, 1H), 4.97 (AB, \(J = 12.3\) Hz, 1H), 4.88 (AB, \(J = 12.3\) Hz,
1H), 4.73 (q, \( J = 5.6 \) Hz, 1H), 3.93 (t, \( J = 7.8 \) Hz, 1H), 3.62 (s, 3H), 3.31 (dd, \( J = 5.4, 2.3 \) Hz, 1H), 3.17 (t, \( J = 6.0 \) Hz, 1H), 2.41 (dd, \( J = 13.0, 7.8 \) Hz, 1H), 2.09 (dd, \( J = 13.0, 8.0 \) Hz, 1H), 1.26 (s, 3H). \(^{13}\)C NMR (100 MHz, CDCl\(_3\), DEPT): \( \delta \) 172.0 (C), 171.9 (C), 154.6 (C), 147.8 (C), 136.2 (C), 135.9 (C), 133.7 (C), 128.9 (CH), 128.4 (2 * CH), 128.1 (CH), 127.9 (2 * CH), 127.6 (C), 123.0 (CH), 122.4 (CH), 122.3 (CH), 119.7 (CH), 119.2 (CH), 118.4 (CH), 111.5 (CH), 109.7 (CH), 109.5 (CH), 84.1 (CH), 67.2 (CH\(_2\)), 61.4 (CH), 52.9 (C), 52.3 (CH), 51.9 (CH\(_3\)), 42.2 (CH\(_2\)), 27.4 (CH\(_2\)), 23.8 (CH\(_3\)). HRMS (ESI, M\(^+\)): m/z calcd. for C\(_{32}\)H\(_{32}\)O\(_5\)N\(_4\) 552.2373, found 552.2367.

Methyl((2S,3aR,8aS)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carbonyl)tryptophanate (23)

To a homogenous solution (under stirring) of coupling compound 22 (95.0 mg, 0.188 mmol) in MeOH and ethyl acetate (1:1) was added 10% palladium on activated charcoal (18 mg, 0.017 mmol) and the reaction mixture stirred under hydrogen at 1 atm. for 7h. It was then filtered with celite pad and washed with ethyl acetate. Evaporation of the solvent under reduced pressure yielded 97% (70.0 mg, 0.167 mmol) of product 23 which was subjected for the next step without any purification. IR (KBr): 3359, 3012, 2960, 2927, 1684, 1519, 1418 cm\(^{-1}\). \(^{1}\)H NMR (400 MHz, CDCl\(_3\)): \( \delta \) 7.52 (d, \( J = 7.9 \) Hz, 1H), 7.38 (d, \( J = 8.1 \) Hz, 1H), 7.19-7.00 (m, 5H), 6.82 (t, \( J = 7.4 \) Hz, 1H), 6.68 (d, \( J = 7.9 \) Hz, 1H), 5.20 (s, 1H), 4.84 (dd, \( J = 9.6, 5.0 \) Hz, 1H), 3.84 (dd, \( J = 11.9, 5.9 \) Hz, 1H), 3.70 (s,
3H), 3.35 (ABX, J = 14.6, 4.8 Hz, 1H), 3.08 (ABX, J = 14.6, 9.6 Hz, 1H), 2.23 (dd, J = 13.2, 6.0 Hz, 1H), 1.62 (t, J = 12.8 Hz, 1H), 1.30 (s, 3H). \(^{13}\)C NMR (100 MHz, CDCl\(_3\), DEPT): \(\delta\) 173.2 (C), 168.3 (C), 149.2 (C), 137.9 (C), 133.8 (C), 130.0 (CH), 128.7 (C), 124.6 (CH), 124.0 (CH), 122.4 (CH), 121.3 (CH), 119.9 (CH), 119.2 (CH), 112.4 (CH), 110.8 (C), 110.4 (CH), 85.2 (CH), 59.5 (CH), 55.5 (C), 54.8 (CH), 52.9 (CH\(_3\)), 44.1 (CH\(_2\)), 28.5 (CH\(_2\)), 25.2 (CH\(_3\)). HRMS (ESI, M\(^+\)): m/z calcd. for C\(_{24}\)H\(_{26}\)O\(_3\)N\(_4\) 418.2005, found: 418.2008.

\(\text{Cyclo-C3-Me-L-Trp-L-Trp DKP (13)}\)

To a homogenous solution (under stirring with a magnetic bar) of amine 23 (77.0 mg, 0.184 mmol) in 14 M methanolic ammonia (4.0 mL) was refluxed for overnight. Evaporation of the solvent under reduced pressure and the residue was washed with chloroform furnished the pure diketopiperazine 13 (44.0 mg, 68% yield) as a pale yellow color solid. IR (KBr): 3431, 3297, 3025, 1666, 1631, 1528, 1447, 1340, 1286, 1072, 911, 754, 625 cm\(^{-1}\). \(^1\)H NMR (600 MHz, CD\(_3\)OD): \(\delta\) 7.90 (s, 1H), 7.51 (dd, J = 6.8, 1.5 Hz, 1H), 7.11 (dd, J = 7.1, 1.5 Hz, 1H), 7.07-7.00 (m, 2H), 6.97 (dt, J = 7.7, 0.8 Hz, 1H), 6.93 (s, 1H), 6.85 (d, J = 7.9 Hz, 1H), 6.62 (t, J = 7.4 Hz, 1H), 6.39 (d, J = 7.8 Hz, 1H), 5.09 (s, 1H), 4.59 (brs, 1H), 4.23 (t, J = 3.7 Hz, 1H), 3.42 (ABX, J = 14.7, 3.7 Hz, 1H), 3.08 (ABX, J = 14.7, 4.3 Hz, 1H), 2.38 (dd, J = 12.0, 5.8 Hz, 1H), 2.24 (dd, J = 12.4, 5.8 Hz, 1H), 1.84 (t, J = 12.0 Hz, 1H), 1.28 (s, 3H). \(^{13}\)C NMR (150 MHz, CD\(_3\)OD, DEPT): \(\delta\)
171.0 (C), 168.1 (C), 150.0 (C), 137.6 (C), 133.0 (C), 129.2 (CH), 128.7 (C), 125.6 (CH), 123.1 (CH), 122.8 (CH), 120.1 (2 × CH), 119.2 (CH), 112.5 (CH), 110.7 (CH), 108.7 (C), 82.1 (CH), 59.4 (CH), 59.1 (CH), 51.9 (C), 43.1 (CH₂), 31.0 (CH₂), 24.7 (CH₃). HRMS (ESI, M+H⁺): m/z calcd. for C₂₃H₂₃O₂N₄ 387.1816, found 387.1824.

5.12. Asymmetric Synthesis of cyclo-C₃-Me-L-Trp-N¹'-Me-L-Trp DKP (16)

**Methyl (S)-2-(1,3-dioxoisodolin-2-yl)-3-(1H-indol-3-yl)propanoate or N-phth-L-Trp-methyl ester (35)**

To a refluxing solution of L-Trp-methyl ester (9) (1.0 g, 4.58 mmol) and phthalic anhydride (0.747 g, 5.04 mmol) in toluene (35 mL) was added triethylamine (0.702 mL, 5.04 mmol) and the reflux was continued overnight. Evaporation of the solvent under reduced pressure yielded 98% (1.57 g, 4.51 mmol) of product 35 which was subjected for the next step without any purification as an orange fluffy solid. mp: 80 °C. IR (KBr): 3611, 3417, 2941, 1853, 1717, 1635, 1585, 1524, 1455, 1385, 1254, 1187, 1093, 1018, 880.6, 733.1 cm⁻¹. ¹H NMR (400 MHz, CDCl₃): δ 7.84 (brs, 1H), 7.65 (dd, J = 5.5, 3.0 Hz, 2H), 7.30 (dd, J = 5.5, 3.0 Hz, 2H), 7.49 (ddd, J = 7.9, 6.8, 0.72 Hz, 1H), 7.16 (ddd, J = 8.0, 6.6, 0.96 Hz, 1H), 7.02 (ddd, J = 8.1, 6.8, 1.1 Hz, 1H), 6.95 (ddd, J = 8.1, 6.9, 1.2
Hz, 1H), 6.89 (d, $J = 2.4$ Hz, 1H), 5.17 (dd, $J = 9.5$, 6.4 Hz, 1H), 3.69 (s, 3H), 3.65 (dd, $J = 4.3$, 0.92 Hz, 1H), 3.63 (dd, $J = 2.0$, 0.88 Hz, 1H). $^{13}$C NMR (100 MHz, CDCl$_3$, DEPT): $\delta$ 169.7 (C), 167.6 (2 $\times$ C), 136.7 (C), 134.0 (2 $\times$ CH), 131.7 (2 $\times$ C), 127.6 (C), 127.3 (CH), 123.4 (2 $\times$ CH), 121.6 (CH), 119.0 (CH), 118.6 (CH), 109.5 (C), 109.2 (CH), 52.8 (CH), 32.6 (CH$_3$), 24.8 (CH$_2$). HRMS (EI, M$^+$): m/z calcd. for C$_{20}$H$_{16}$N$_2$O$_4$ 348.1110, found 348.1109.

Methyl (S)-2-(1,3-dioxoisooindolin-2-yl)-3-(1-methyl-1H-indol-3-yl)propanoate or N$^1$-Me-$N^2$-phth-L-Trp-methyl ester (36)

To a cold (0 °C) solution of $N$-phth-L-Trp-methyl ester (35) (800 mg, 2.30 mmol) in dry DMF (8.0 mL) was added NaH (101 mg, 2.53 mmol) followed by slow addition of methyl iodide (215 $\mu$L, 3.45 mmol) and the resulting mixture was stirred at same temperature for 8h (reaction monitored by TLC). It was then quenched with water (10 mL) and extracted with ethyl acetate (3 $\times$ 15 mL). The combined organic layer was washed with brine and dried (anhyd. Na$_2$SO$_4$). Evaporation of the solvent under reduced pressure and purification of the residue on a silica gel column using ethyl acetate–hexanes (1:20) as eluent furnished the product 36 in 58% yield (480 mg, 1.32 mmol). Crystalline solid. mp: 124-125 °C; IR (KBr): 3026, 2952, 1745, 1714, 1614, 1553, 1470,
1435, 1390, 1328, 1256, 1210, 1127, 1017, 967, 917, 880, 750, 720, 662 cm$^{-1}$. $^1$H NMR (400 MHz, CDCl$_3$): $\delta$ 7.77 (dd, $J = 5.5$, 3.0 Hz, 2H), 7.67 (dd, $J = 5.5$, 3.0 Hz, 2H), 7.59 (d, $J = 8.0$ Hz, 1H), 7.21 (d, $J = 8.2$ Hz, 1H), 7.15 (ddd, $J = 8.0$, 6.8, 1.0 Hz, 1H), 7.04 (ddd, 8.0, 6.9, 1.2 Hz, 1H), 6.87 (s, 1H), 5.26 (dd, $J = 8.8$, 7.1 Hz, 1H), 3.79 (s, 3H), 3.74 (d, $J = 1.2$ Hz, 1H) 3.72 (s, 1H), 3.65 (s, 3H). $^{13}$C NMR (100 MHz, CDCl$_3$, DEPT): $\delta$ 169.8 (C), 167.7 (2 × C), 136.9 (C), 134.1 (2 × CH), 131.8 (2 × C), 127.7 (C), 127.4 (CH), 123.5 (2 × CH), 121.7 (CH), 119.0 (CH), 118.7 (CH), 109.6 (C), 109.2 (CH), 52.9 (CH), 52.8 (CH$_3$), 32.7 (CH$_3$), 24.9 (CH$_2$). HRMS (EI, M$^+$): m/z calcd. for C$_{21}$H$_{18}$N$_2$O$_4$ 362.1267, found 362.1253.

$N$-Me-L-Trp-Methyl Ester (21)

To a solution of phthalyl amine (36) (525 mg, 1.45 mmol) in MeOH (7.0 mL) and CH$_2$Cl$_2$ (7.0 mL) was added hydrazine hydrate (78 μL, 1.59 mmol) and the reaction mixture stirred for 24h. It was then filtered on a celite pad and washed with ethyl acetate. Evaporation of the solvent under reduced pressure and purification of the residue on silica gel (pre-neutralized with Et$_3$N) column using MeOH–CHCl$_3$ (1:20) as eluent furnished the amine (21) in (180 mg, 0.77 mmol) 54% yield. IR (KBr): 3369, 2950, 2922, 1736, 1608, 1546, 1473, 1439, 1373, 1325, 1249, 1207, 1173, 1128, 1062, 1014, 837, 742 cm$^{-1}$. $^1$H NMR (400 MHz, CDCl$_3$): $\delta$ 7.61 (td, $J = 8.0$, 1.0 Hz, 1H), 7.30 (td, $J = 8.2$, 1.0 Hz, 2H), 7.59 (d, $J = 8.0$ Hz, 1H), 7.21 (d, $J = 8.2$ Hz, 1H), 7.59 (d, $J = 8.0$ Hz, 2H), 7.15 (ddd, $J = 8.0$, 6.8, 1.0 Hz, 1H), 7.04 (ddd, 8.0, 6.9, 1.2 Hz, 1H), 6.87 (s, 1H), 5.26 (dd, $J = 8.8$, 7.1 Hz, 1H), 3.79 (s, 3H), 3.74 (d, $J = 1.2$ Hz, 1H) 3.72 (s, 1H), 3.65 (s, 3H). $^{13}$C NMR (100 MHz, CDCl$_3$, DEPT): $\delta$ 169.8 (C), 167.7 (2 × C), 136.9 (C), 134.1 (2 × CH), 131.8 (2 × C), 127.7 (C), 127.4 (CH), 123.5 (2 × CH), 121.7 (CH), 119.0 (CH), 118.7 (CH), 109.6 (C), 109.2 (CH), 52.9 (CH), 52.8 (CH$_3$), 32.7 (CH$_3$), 24.9 (CH$_2$). HRMS (EI, M$^+$): m/z calcd. for C$_{21}$H$_{18}$N$_2$O$_4$ 362.1267, found 362.1253.
Hz, 1H), 7.23 (ddd, J = 8.1, 6.9, 1.1 Hz, 1H), 7.12 (ddd, J = 8.0, 6.9, 1.1 Hz, 1H), 6.93 (s, 1H), 3.82 (dd, 7.7, 4.8 Hz, 1H), 3.75 (s, 3H), 3.72 (s, 3H), 3.28 (ABXY, J = 14.4, 4.8, 0.8 Hz, 1H), 3.04 (ABXY, J = 14.4, 7.7, 0.6 Hz, 1H), 1.58 (brs, 2H). ¹³C NMR (100 MHz, CDCl₃, DEPT): δ 175.9 (C), 137.1 (C), 128.0 (C), 127.8 (CH), 121.8 (CH), 119.1 (CH), 119.0 (CH), 109.7 (C), 109.4 (CH), 55.2 (CH), 52.1 (CH₃), 32.8 (CH₃), 30.7 (CH₂).

HRMS (EI, M⁺): m/z calcd. for C₁₃H₁₆O₂N₂ 232.1212, found 232.1208.

**Benzyl-(2S,3aR,8aR)-2-((1-methoxy-3-(1-methyl-1H-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-3a-methyl-3,3a,8,8a-tetrahydropyrrolo[2,3-b]indole-1(2H)-carboxylate (24)**

To a cold (0 °C) magnetically stirred solution of N-Cbz-acid 20c (135.0 mg, 0.38 mmol) with N-Me-L-Trp-methyl ester 21 (98.0 mg, 0.42 mmol) in dry THF (3 mL) was added Et₃N (213.0 μL, 1.53 mmol) followed by BOP-Cl (254.0 mg, 1.00 mmol) and the resulting mixture was stirred at room temperature for 12 h (reaction monitored by TLC). It was then quenched with water (3 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic layer was washed with brine and dried over anhyd. Na₂SO₄.
Evaporation of the solvent under reduced pressure and purification of the residue on silica gel column using ethyl acetate–hexanes (1:3) as eluent furnished the coupled amide \( 24 \) in 81% yield (175 mg, 0.31 mmol) as a mixture of two rotational isomers. IR (KBr): 3381, 3019, 2954, 2927, 1740, 1692, 1610, 1515, 1481, 1465, 1443, 1416, 1355, 1328, 1252, 1214, 1157, 1126, 1054, 1012, 986, 749, 696 cm\(^{-1}\). \(^1\)H NMR (400 MHz, CDCl\(_3\)): \( \delta \) 7.51 (d, \( J = 7.9 \) Hz, 0.5H), 7.40 (d, \( J = 8.1 \) Hz, 1H), 7.37 (s, 2H), 7.30-7.20 (m, 5H), 7.20-7.05 (m, 5H), 6.99 (d, \( J = 7.5 \) Hz, 1H), 6.89 (s, 1H), 6.78-6.71 (m, 1.5H), 6.70 (s, 1H), 6.59 (d, \( J = 7.8 \) Hz, 1H), 6.55 (d, \( J = 7.8 \) Hz, 0.5H), 6.43 (d, \( J = 8.0 \) Hz, 0.5H), 6.12 (d, \( J = 7.6 \) Hz, 1H), 5.41 (brs, 1H), 5.31 (AB, \( J = 12.4 \) Hz, 0.5H), 5.20 (s, 0.5H), 5.12 (s, 1H), 5.09 (AB, \( J = 12.4 \) Hz, 0.5H), 4.96 (AB, \( J = 12.4 \) Hz, 1H), 4.89 (AB, \( J = 12.4 \) Hz, 1H), 4.67 (dd, \( J = 12.8 \), 5.4 Hz, 1H), 4.07 (t, \( J = 7.9 \) Hz, 0.5H), 3.90 (t, \( J = 7.9 \) Hz, 1H), 3.73 (s, 1.5H), 3.72 (s, 3H), 3.68 (s, 1.5H), 3.64 (s, 3H), 3.68-3.62 (m, 0.5H), 3.32 (t, \( J = 4.8 \) Hz, 0.5H), 3.17 (ABX, \( J = 14.8 \), 5.5 Hz, 1H), 3.11 (ABX, \( J = 14.8 \), 5.5, Hz, 1H), 2.42 (ABX, \( J = 13.0 \), 7.8, Hz, 1H), 2.32 (ABX, \( J = 13.0 \), 7.8, Hz, 0.5H), 2.21 (ABX, \( J = 13.0 \), 7.0, Hz, 0.5H), 2.10 (ABX, \( J = 13.0 \), 8.2, Hz, 1H), 1.34 (s, 1.5H), 1.30 (s, 3.0H). \(^1\)C NMR (100 MHz, CDCl\(_3\), DEPT): \( \delta \) 172.2 (C), 172.0 (C), 171.7 (C), 171.2 (C), 154.8 (C), 154.6 (C), 147.8 (C), 147.2 (C), 136.9 (2 * C), 136.1 (C), 136 (C), 133.8 (C), 133.6 (C), 128.9 (2 * CH), 128.6 (C), 128.5 (4 * CH), 128.4 (C), 128.1 (4 * CH), 128.0 (CH), 127.9 (2 * CH), 127.4 (CH), 122.4 (CH), 122.3 (CH), 122.0 (CH), 121.8 (CH), 119.4 (CH), 119.3 (CH), 119.2 (CH), 119.1 (CH), 118.7 (CH), 118.6 (CH), 109.8 (CH), 109.5 (CH), 109.4 (CH), 109.3 (CH), 108.3 (C), 108.2 (C), 84.1 (CH), 83.8 (CH), 67.6 (CH\(_2\)), 67.1 (CH\(_2\)), 61.8 (CH), 61.5 (CH), 53.3 (C), 52.8 (C), 52.7 (CH), 52.4 (CH\(_3\)), 52.3 (CH), 51.9
(CH$_3$), 42.1 (CH$_2$), 41.4 (CH$_2$), 32.8 (2 * CH$_3$), 27.5 (CH$_2$), 27.3 (CH$_2$), 23.8 (CH$_3$), 23.6 (CH$_3$). HRMS (EI, M$^+$): m/z calcd for C$_{33}$H$_{34}$O$_5$N$_4$: 566.2529 found: 566.2529.

N-methyl-(2S,3aR,8aS)-3a-methyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-b]indole-2-carbonyl) methyl-1-tryptophanate (25)

To a magnetically stirred solution of 24 (75 mg, 0.13 mmol,) in MeOH (2 mL) and ethyl acetate (1 mL) was added 10% palladium on activated charcoal (14 mg, 0.1 equiv.) and the reaction mixture was stirred under hydrogen at 1 atm. Pressure for 6h. It was then filtered with Celite pad and washed with ethyl acetate. Evaporation of the solvent under reduced pressure furnished the amine 25 (57 mg, 0.13 mmol) in quantitative yield. The identity of the compound was confirmed by HRMS and subjected for the next step without any purification. HRMS (EI, M$^+$): m/z calcd. for C$_{25}$H$_{28}$N$_4$O$_3$ 432.2161; found, 432.2164.

C3-Me-L-Trp- N1′-Me-L-Trp DKP (16)
To a magnetically stirred solution of amine (57 mg, 0.13 mmol) in 10M methanolic ammonia (4.0 mL) was refluxed for overnight. Evaporation of the solvent under reduced pressure and the residue was washed with chloroform furnished the pure diketopiperazine 16 (48 mg, 0.12 mmol) in 91% yield. IR (KBr): 3431, 3297, 3025, 1666, 1631, 1528, 1447, 1340, 1286, 1072, 911, 754, 625 cm⁻¹. ¹H NMR (600 MHz, CD₃OD): δ 7.51 (d, J = 5.2 Hz, 1H), 7.10-7.07 (m, 2H), 7.06-6.98 (m, 2H), 6.91 (d, J = 4.9 Hz, 1H), 6.86 (s, 1H), 6.65 (t, J = 4.9 Hz, 1H), 6.51 (d, J = 5.2 Hz, 1H), 5.15 (s, 1H), 4.59 (brs, 2H), 4.22 (s, 1H), 3.40 (s, 3H), 3.35 (dd, J = 9.8, 2.0 Hz, 1H), 3.12 (dd, J = 9.8, 3.0 Hz, 1H), 2.36 (dd, J = 7.9, 3.9 Hz, 1H), 2.25 (dd, J = 8.3, 3.9 Hz, 1H), 1.84 (t, J = 8.1 Hz, 1H), 1.29 (s, 3H). ¹³C NMR (150 MHz, CD₃OD, DEPT): δ 170.8 (C), 168.2 (C), 150.2 (C), 138.0 (C), 133.2 (C), 129.8 (CH), 129.5 (C), 129.4 (CH), 123.4 (CH), 122.7 (CH), 120.3 (CH), 120.1 (CH), 119.7 (CH), 110.5 (CH), 110.3 (CH), 108.3 (C), 82.3 (CH), 59.6 (CH), 59.1 (CH), 52.0 (C), 42.7 (CH₃), 32.9 (CH₂), 30.7 (CH₂), 24.9 (CH₃). HRMS (EI, M⁺): m/z calcd. for C₂₄H₂₄O₂N₄ 400.1899, found 400.1897.

5.13. Synthesis of cyclo-L-Trp-C₃⁺⁻prenyl-L-Trp DKP (12) and cyclo-L-Trp-N¹⁻ Me-C₃⁺⁻prenyl-L-Trp DKP (15)
To a magnetically stirred solution of L-Trp-methyl ester (9) (400 mg, 1.83 mmol) in sodium acetate-acetic acid solution (pH = 2.7) (30 mL) was added prenyl bromide (635 μL, 5.50 mmol) over a period for 45-50 minutes at room temperature. The resulting mixture was stirred at same temperature for overnight. Evaporation of acetic acid under reduced pressure resulted in a solid residue which was dissolved in ethyl acetate. Solution was neutralized by addition of sodium carbonate solution and the ester was extracted with ethyl acetate three times. The combined organic layers were dried over anhyd. Na₂SO₄ and evaporated under reduced pressure yielding a diastereomeric mixture of the cyclic product 27a and 27b (in a 4:1 ratio) (358 mg, 1.25 mmol) in 67% overall yield (based on recovered starting material). Data for major diastereomer 27a: IR (KBr): 3366, 3044, 2933, 1823, 1737, 1672, 1605, 1477, 1365, 1225, 1110, 1024, 833, 747 cm⁻¹. ¹H NMR (400 MHz, CDCl₃): δ 7.06-7.01 (m, 2H), 6.72 (dt, J = 7.4, 0.9 Hz, 1H), 6.56 (d, J = 7.9 Hz, 1H), 5.09 (t, J = 7.3 Hz, 1H), 4.91 (s, 1H), 3.71 (dd, J = 10.5, 5.8 Hz, 1H), 3.70 (s, 3H), 3.42 (br’s, 2H), 2.47-2.41 (m, 2H), 2.38 (dd, J = 12.0, 5.8 Hz, 1H), 2.00 (dd, J = 12.0, 10.7 Hz, 1H), 1.68 (s, 3H), 1.55 (s, 3H). ¹³C NMR (100 MHz, CDCl₃, DEPT): δ 174.4 (C), 150.0 (C), 134.7 (C), 133.1 (C), 128.2 (CH), 123.6 (CH), 119.6 (CH), 118.8 (CH), 109.1 (CH), 82.2 (CH), 59.4 (CH), 58.8 (CH₃), 52.2 (CH₂), 44.2 (C), 36.9 (CH₂), 26.0 (CH₃), 18.2 (CH₃). HRMS (ESI, M+H⁺): m/z calcd. for C₁₇H₂₅O₂N₂ 287.1681, found 287.1758.
C3'-"prenyl-pyrroloindoline-methyl-ester-N1-phthalyl-L-Trp-amide (30)

To a cold (0 °C) magnetically stirred solution of C3'-"prenyl-pyrroloindole methyl ester 27a (42.0 mg, 0.15 mmol) with L-Trp-N-phth-acid 29 (44.0 mg, 0.13 mmol) in dry THF (1.5 mL) was added Et₃N (101 μL, 0.73 mmol) followed by BOP-Cl (93 mg, 0.37 mmol) and the resulting mixture was stirred at room temperature for overnight. It was then quenched with water (10 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic layer was washed with brine and dried (anhyd. Na₂SO₄). Evaporation of the solvent under reduced pressure and purification of the residue on a silica gel column using ethyl acetate–hexanes (4:6) as eluent furnished the coupling compound as a mixture of rotamers in 59 mg, 67% yield. The identity of the compound was confirmed by HRMS and subjected for the next step without any purification. IR (KBr): 3372, 3057, 2937, 1717, 1646, 1449, 1371, 1101, 913, 730 cm⁻¹. HRMS (ESI, M+Na⁺): m/z calcd. for C₃₆H₃₄O₅N₄Na 625.2427, found 625.2441.
Cyclo-L-Trp-C3′-nprenyl-L-Trp DKP (12)

To a magnetically stirred solution of 30 (56 mg, 0.09 mmol) in MeOH (1.0 mL) and CH₂Cl₂ (1.0 mL) was added hydrazine hydrate (45 μL, 0.93 mmol) and the reaction mixture stirred under nitrogen atmosphere for 24h. It was then quenched with water (10 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic layer was washed with brine and dried (anhyd. Na₂SO₄). Evaporation of the solvent under reduced pressure and purification of the residue on a silica gel column chromatography using EtOAc as eluent furnished the cyclo-L-Trp-C3′-nprenyl-L-Trp DKP (12) in 70 % yield (29.0 mg, 0.06 mmol) as a single diasteromer. IR (KBr): 3297, 3058, 1663, 1452, 1324, 1192, 1088, 922, 814, 740, 608 cm⁻¹. ¹H NMR (400 MHz, CDCl₃): δ 7.74 (brs, 1H), 7.59 (d, J = 7.3 Hz, 1H), 7.20 (m, 3H), 7.05 (dt, J = 7.7, 1.2 Hz, 1H), 6.89 (d, J = 6.9 Hz, 1H), 6.86 (d, J = 2.4 Hz, 1H), 6.70 (t, J = 7.4 Hz, 1H), 6.40 (d, J = 7.8 Hz, 1H), 6.04 (brs, 1H), 5.22 (s, 1H), 5.04 (t, J = 7.3 Hz, 1H), 4.44 (brs, 1H), 4.26 (dd, J = 9.2, 4.0 Hz, 1H), 3.37 (ABX, J = 14.6, 5.4 Hz, 1H), 3.12 (ABX, J = 14.6, 3.8 Hz, 1H), 2.72 (dd, J = 12.0, 5.5 Hz, 1H), 2.36-2.26 (m, 3H), 2.00 (t, J = 12.0 Hz, 1H), 1.65 (s, 3H), 1.49 (s, 3H).

¹³C NMR (100 MHz, CDCl₃, DEPT): δ 169.3 (C), 165.9 (C), 149.2 (C), 136.0 (C), 135.8 (C), 131.0 (C), 128.3 (CH), 127.3 (C), 123.7 (CH), 123.3 (CH), 122.8 (CH), 120.0 (CH), 119.0 (CH), 118.7 (CH), 118.5 (CH), 111.5 (CH), 109.3 (CH), 108.9 (C), 79.1 (CH), 58.5
(CH), 57.7 (CH), 55.1 (C), 39.2 (CH₂), 35.5 (CH₂), 30.6 (CH₂), 26.1 (CH₃), 18.1 (CH₃).

HRMS (ESI, M+Na⁺): m/z calcd. for C_{27}H_{28}O_{2}N_{4}Na 463.2110, found 463.2104.


\[
\text{N1’-Me-C3’-”prenyl-L-Trp-pyrroloindoline methyl ester (28)}
\]

To a magnetically stirred solution of N-Me-L-Trp-methyl ester 26 (680 mg, 2.93 mmol) in acetate sloution (pH = 2.7) (30 mL) was added prenyl bromide (1.29 ml, 8.79 mmol) over a period for 45-50 minutes at room temperature. The resulting mixture was stirred at same temperature overnight. After evaporation of acetic acid under reduced pressure the resulting solid residue was dissolved in ethyl acetate. Then the mixture was neutralized through addition of sodium carbonate solution and the aqueous layer was extracted with ethyl acetate three times. The collected organic layer was dried over anhyd. Na₂SO₄ and evaporated under reduced pressure yielding diastereomeric mixture of the cyclic product in 11% yield for 28b (99 mg, 0.33 mmol) and 21% yield for 26a (185 mg, 0.62 mmol) along with a 40% recovery of 26 amounting to a 53% overall yield based on recovered starting material. Data for diastereoisomer 28a:

- IR (KBr): 3366, 3044, 2933, 1823, 1737, 1672, 1605, 1477, 1365, 1225, 1110, 1024, 942, 833, 747, 662 cm⁻¹.
- ^{1}H NMR (400 MHz, CDCl₃): δ 7.08 (dt, J = 7.7, 1.2 Hz, 1H), 7.00 (dd, J = 7.3, 1.2 Hz, 1H), 6.63 (dt, J
= 7.4, 0.9 Hz, 1H), 6.34 (d, J = 7.8 Hz, 1H), 5.09 (qt, J = 6.7, 1.3 Hz, 1H), 4.67 (s, 1H), 3.71 (s, 3H), 3.66 (dd, J = 10.3, 6.2 Hz, 1H), 3.08 (brs, 1H), 2.83 (s, 3H), 2.45-2.38 (m, 2H), 2.35 (dd, J = 12.2, 6.2 Hz, 1H), 2.00 (t, J = 11.2 Hz, 1H), 1.68 (s, 3H), 1.56 (s, 3H). 

13C NMR (100 MHz, CDCl3, DEPT): δ 174.6 (C), 151.20 (C), 134.6 (C), 133.4 (C), 128.3 (CH), 123.0 (CH), 119.7 (CH), 117.0 (CH), 105.6 (CH), 88.8 (CH), 59.5 (CH), 57.1 (C), 52.2 (CH3), 44.0 (CH2), 36.7 (CH2), 31.7 (CH3), 26.0 (CH3), 18.2 (CH3). HRMS (EI, M+): m/z calcd. for C18H24O2N2 300.1838, found 300.1838.

Data for diastereoisomer 28b: IR (KBr): 3366, 3044, 2933, 1823, 1737, 1672, 1605, 1477, 1365, 1225, 1110, 1024, 942, 833, 747, 662 cm−1. 1H NMR (400 MHz, CDCl3): δ 7.05 (dt, J = 7.6, 1.2 Hz, 1H), 6.98 (dd, J = 7.3, 0.9 Hz, 1H), 6.60 (dt, J = 7.4, 0.8 Hz, 1H), 6.29 (d, J = 7.8 Hz, 1H), 5.13 (qt, J = 6.7, 1.2 Hz, 1H), 4.63 (s, 1H), 3.91 (dd, J = 7.8, 3.1 Hz, 1H), 3.50 (brs, 1H), 3.33 (s, 3H), 2.86 (s, 3H), 2.49-2.36 (m, 3H), 2.33 (ABX, J = 12.8, 7.9 Hz, 1H), 1.69 (s, 3H), 1.58 (s, 3H). 13C NMR (100 MHz, CDCl3, DEPT): δ 174.6 (C), 150.8 (C), 134.5 (C), 133.6 (C), 128.3 (CH), 123.2 (CH), 119.9 (CH), 117.0 (CH), 105.6 (CH), 88.7 (CH), 60.0 (CH), 56.0 (CH), 52.0 (CH3), 41.0 (CH2), 36.2 (CH2), 31.4 (CH3), 26.1 (CH3), 18.2 (CH3). HRMS (ESI, M+): m/z calcd. for C18H24O2N2 300.1838, found 300.18375.
A cold (−10 °C) solution of N1’-Me-C3’-prenyl-L-Trp-pyrroloindoline methyl ester 28a (100 mg, 0.33 mmol) was mixed with N-Phth-L-Trp acid 29 (101, 0.30 mmol) under stirring in dry THF (8 mL) and Et₃N (168 μL, 1.21 mmol) was added followed by BOP-Cl (192 mg, 0.75 mmol) and the resulting mixture was stirred at same temperature for 8 h. TLC was used for monitoring the progress. The reaction was then quenched with water (10 mL) and extracted with ethyl acetate (3 × 15 mL). The combined organic layers were washed with brine and dried over anhyd. Na₂SO₄. Evaporation of the solvent under reduced pressure and purification of the residue on a silica gel column using ethyl acetate–hexanes (1:20) as eluent furnished the coupled product 31 as a mixture of rotational isomers in 97% yield (180 mg, 0.29 mmol). This material was subjected to the DKP-forming step directly. HRMS (EI, M⁺): m/z calcd. for C₂₉H₃₂N₃O₃ 470.2444, found 470.2446.

Cyclo-L-Trp-N1’-Me-C3’-prenyl-L-Trp DKP (15)

To a magnetically stirred solution of 31 (180 mg, 0.29 mmol) in MeOH (5.0 mL) and CH₂Cl₂ (5.0 mL) was added hydrazine hydrate (152 μL, 0.31 mmol) and the reaction mixture stirred under nitrogen atmosphere for 24 h. It was then quenched with water (10 mL) and extracted with ethyl acetate (3 × 15 mL). The combined organic layers were washed with brine and dried (anhyd. Na₂SO₄). Evaporation of the solvent under reduced
pressure and purification of the residue on preparative TLC using MeOH–CHCl₃ (1:40) as eluent furnished the Cyclo-L-Trp-N¹'-Me-C3''-prenyl-L-Trp DKP (15) in 74% yield (99.0 mg, 0.22 mmol). IR (KBr): 3308, 3243, 3059, 2927, 2864, 1788, 1670, 1501, 1442, 1359, 1305, 1233, 1172, 1104, 984, 919, 733 cm⁻¹. ¹H NMR (600 MHz, CDCl₃): δ 8.21 (brs, 1H), 7.61 (d, J = 8.1 Hz, 1H), 7.39 (d, J = 8.1 Hz, 1H), 7.22 (t, J = 7.2 Hz, 1H), 7.21 (t, J = 7.4 Hz, 2H), 7.18 (brs, 1H), 7.00 (d, J = 7.3 Hz, 1H), 6.74 (t, J = 7.2 Hz, 1H), 6.48 (d, J = 7.7 Hz, 1H), 5.98 (brs, 1H), 5.41 (s, 1H), 4.97 (t, J = 7.2 Hz, 1H), 4.43 (d, J = 7.6 Hz, 1H), 4.02 (dd, J = 11.0, 6.2 Hz, 1H), 3.68 (dd, J = 14.6, 3.1 Hz, 1H), 3.09 (s, 3H), 3.07 (dd, J = 14.6, 10.0 Hz, 1H), 2.52 (dd, J = 12.0, 6.0 Hz, 1H), 2.36-2.24 (m, 2H), 1.90 (t, J = 12.0 Hz, 1H), 1.67 (s, 3H), 1.52 (s, 3H). ¹³C NMR (150 MHz, CDCl₃, DEPT): δ 168.8 (C), 165.4 (C), 136.6 (2 × C), 135.6 (C), 131.8 (C), 129.0 (CH), 126.9 (CH), 124.0 (CH), 123.3 (CH), 122.9 (CH), 120.2 (CH), 120.1 (C), 118.9 (CH), 118.7 (CH), 111.6 (CH), 109.5 (CH), 109.0 (C), 85.6 (CH), 58.7 (CH), 55.0 (CH), 54.5 (C), 41.0 (CH₃), 36.7 (CH₂), 29.8 (CH₂), 28.5 (CH₂), 26.1 (CH₃), 18.2 (CH₃). HRMS (EI, M⁺): m/z calcd. for C₂₈H₃₀N₄O₂ 454.2369, found 454.2368.
5.15. Conditions for HRMS, HPLC, LC-MS and MS2 Characterization and Identification of signature peaks.

A 2.1x50 mm column packed with BEH C18 1.7 μm particles (Waters) was held at 45 °C throughout the separation; mobile phase A was 5% v/v Omnisolve grade CH₃CN (EMD Millipore, Billerica, MA), 0.1% v/v formic acid (Sigma Aldrich) in Omnisolve grade water (EMD). Mobile phase B was 5% v/v Omnisolve water, 0.1% v/v formic acid in acetonitrile and the flow rate was maintained at 0.3 mL/min. The gradient profile was: Start at 10% B, linear gradient to 100% B over 30 minutes, hold 3 minutes at 100% B, and a linear gradient to 0% B over two minutes followed by a 3 minute re-equilibration period between injections. All effluent was directed into the ESI source of the G2 (3.0 kV on capillary, 120 °C source temperature, 850 L/h of nitrogen desolvation gas @ 600 °C, 20 L/h of cone gas, 40 V on sample cone, 4 V on extraction cone) which was used in resolution mode (20,000 resolving power). Separate chromatograms were recorded simultaneously with a 0.2 sec MS¹ scan from m/z 275-600 for the LC eluent with no collision energy, along with nine 0.2 second MS¹ and MS² scans directing the quadrupole to sequentially pass 205.10 (L/D-tryptophan, [M+H]⁺); 373.17 (cyclo-L-Trp-L/D-Trp DKP (5), Rₜ for [M-H]⁻, 5.05 min and Rₜ for [M+H]⁺, 7.05 min.); m/z 387.18 (cyclo-C3-Me-L-Trp-L-Trp DKP (13), M+H⁺, Rₜ 10.32 min.); m/z 401.19 (cyclo-C3-Me-L-Trp-N1′-Me-L-Trp DKP (16), [M+H]⁺, Rₜ 12.15 min.); m/z 441.23 (cyclo-L-Trp-C3′-prenyl-L-Trp DKP (12), [M+H]⁺, Rₜ 12.12 min.); m/z 455.24 (cyclo-L-Trp-N1′-Me-C3′-prenyl-L-Trp DKP (15), [M+H]⁺, Rₜ 18.56 min. (17.25-19.25 min.); m/z 483.24 (nocardioazine A 167
(3), [M+H]^+, R_t 8.8 min.), 469.26 (nocardioazine B (4), [M+H]^+, R_t 18.40 min.), and m/z 574.37 (nocardiopsin A (1), M+H^+).

During the nine MS^2 experiments, the trap voltage (collision energy) was scanned from 20-50 V over 0.2 seconds; the trap was filled with 7.9*10^{-3} mbar of UHP Ar. The 9th chromatogram was another MS1 scan of the lock spray nozzle which had 5 μL/min of a 2 mg/L solution of Leucine-enkephalin (Sigma) in 50% v/v methanol, 0.1% v/v formic acid in water using the m/z 556.2765 (M+H)^+ ion to dynamically correct the mass axis calibration throughout the experiment. After acquisition, extracted ion chromatograms for fragment ions were generated using MassLynx 4.1 software (Waters).

cyclo-L-Trp-C3'-prenyl-L-Trp DKP (12) was identified at m/z = 441.2292 (Figure 4C). Of all the daughter ions identifiable that are characteristic of the 2,5-DKP structure, presence of a C3'-prenyl group gave rise to a unique signal with 100% relative abundance among the MS^2 peaks at m/z = 198.129 (Figure 4B and Table S5). For the molecular ion, loss of a neutral prenyl-derived C4-containing side chain was observed resulting in a m/z = 385.166. The minor diastereomer of 12 (formed from the corresponding minor isomer 27b) showed MS^2 fragmentation pattern identical to the major isomer, except for the LC-MS retention time (Figure S2). Figure 4D shows the presence of [M-H]^− for 13 under negative ion mode MS^2 fragmentation.

Characteristically, the fragment ion at 256.108 corresponds to a half of the DKP with the C3-methyl portion intact. Presence of the N1'-methyl group was confirmed through fragment ions m/z = 184.0761, 156.0815 and 144.0813 assigned to structures as depicted in Figure 4E. Next we subjected cyclo-C3-Me-L-Trp-N-Me-L-Trp DKP (16) synthesized earlier, to LC-MS^1 and MS^2 fragmentation analyses. As shown in Figure S1 (SI) and
listed in Table 1, 16 was identified through its LC retention time of 12.15 min and further confirmed through its [M+H]^+ ion m/z = 401.1972. MS^2 fragmentation revealed ions at m/z = 384.1707 (loss of NH₃ at 17 Da); 256.1089 (loss of neutral Trp fragment at 129 Da); 201.1023; 184.0761; 156.0813 and 144.0813 (Figure 4C). The fact that 13 and 16 differ from each other through a single N1'-Me group was evident in the daughter ion populated by 256.109 instead of 242.093 (an increase of a CH₂ group, 14 Da) and the exact position of the CH₂ group was confirmed through ESI-(−)MS^1 and MS^2 analyses as discussed below. Like DKP 5, 13 fragmented to give daughter ions at m/z = 242.0930; 184.0761; 159.0923; 156.0815; 144.0813; 130.0656 and 103.0547 as listed in Table S3 (SI). Ions with m/z = 184.0761 and 156.0815 appear newly in comparison to those fragmented out of 5, and the ion at m/z = 214.0980 and 169.0759 disappear from the MS^2 distribution of fragments.

For the two diastereomers of cyclo-L-Trp-N1'-Me-C3'-"prenyl-L-Trp DKP (15), that differed from 12 by possession of a single additional N1'-Me group, the LC retention time was 18.32 and 18.56 min for the major and minor diastereomers respectively. The protonated molecular ion [M+H]^+ was identified at m/z = 455.2435 and their tandem mass spectra are shown in Figures 4E and 4F. Fragmentation pattern through MS^2 analysis revealed uniquely C3'-prenylated and N1'-methylated ion at m/z = 212.144 that was not present in the non-prenylated DKPs such as 5, 13 or 16. Between 12 and 15, the ion at m/z = 198.129 (C_{14}H_{16}N^+) observed only in 7 and a corresponding ion seen for 15 in addition to the ion at m/z = 212.144 (C_{15}H_{18}N^+) fingerprinted the presence of N1'-methyl group. Overall, with the synthetic standards whose structures were established through detailed NMR spectroscopy along with their LC-MS^1 and MS^2 data on hand, we
established the ground work for addressing specifically the biosynthetic logic and order of enzymatic events that included the C3'-prenyltransferase (NozC activity) and two methyl transferase enzymatic steps (NozD and NozE activity) downstream of DKP biosynthesizing NozA-catalyzed event, leading to the biosynthesis of 4. 

*Cyclo-C3-Me-L-Trp-L-Trp* DKP (13) was identified through its LC retention time (10.32 min) (Figure S1), further through its [M+H]+ ion (at 387.1825 Da) and its [M-H]- ion (at m/z = 385.1711).
Two diastereomers of cyclo-L-Trp-C3′-prenyl-L-Trp DKP (12) were synthesized and their LC traces were observed as distinct peaks at Rt 9.37 (minor) and 12.11 (major) min respectively.

**Figure 25.** LC-ESI-TOF MS spectra for biosynthetic intermediates in their purely synthesized state. A: LC-ESI (+) ions from mixture of synthesized standards 5, 13, 15 and 16; B: LC-ESI (-) ions from mixture of synthesized standards 5, 13, 15 and 16. C. Extracted ion chromatograms for positive ion tandem mass spectra of 5, 8, 10, and 11 from biological extract.

**Figure 26.** Two diastereomers of cyclo-L-Trp-C3′-prenyl-L-Trp DKP (12) were synthesized and their LC traces were observed as distinct peaks at Rt 9.37 (minor) and 12.11 (major) min respectively.
Figure 27. MS1 and MS2 for two diastereomers of cyclo-L-Trp-C3′-"prenyl-L-Trp DKP (12).
Table 5. MS² fragments for [M+H]⁺ peak at m/z = 373.1659 corresponding to cyclo-L-Trp-L-Trp DKP (5).

<table>
<thead>
<tr>
<th>Observed Mass</th>
<th>% Rel. Int.</th>
<th>Predicted Mass</th>
<th>Mass Diff. (ppm)</th>
<th>Ion Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>373.1662</td>
<td>2.8</td>
<td>373.1659</td>
<td>0.80</td>
<td>C₂₂H₂₁N₄O₂⁺</td>
</tr>
<tr>
<td>242.0930</td>
<td>2.7</td>
<td>242.0924</td>
<td>2.48</td>
<td>C₁₃H₁₁N₃O₂⁺</td>
</tr>
<tr>
<td>214.0980</td>
<td>1.6</td>
<td>214.0975</td>
<td>2.34</td>
<td>C₁₂H₁₂N₁O⁺</td>
</tr>
<tr>
<td>169.0759</td>
<td>5.7</td>
<td>169.0760</td>
<td>-0.59</td>
<td>C₁₁H₈N₂⁺</td>
</tr>
<tr>
<td>159.0924</td>
<td>2.5</td>
<td>159.0917</td>
<td>4.40</td>
<td>C₁₀H₁₁N₂⁺</td>
</tr>
<tr>
<td>144.0809</td>
<td>1.3</td>
<td>144.0808</td>
<td>0.69</td>
<td>C₁₀H₁₀N⁺</td>
</tr>
<tr>
<td>130.0654</td>
<td>100</td>
<td>130.0651</td>
<td>2.31</td>
<td>C₉H₆N⁺</td>
</tr>
<tr>
<td>103.0544</td>
<td>2.8</td>
<td>103.0542</td>
<td>1.94</td>
<td>C₉H⁺</td>
</tr>
</tbody>
</table>
Table 6. MS² fragments for [M+H]⁺ peak at m/z = 387.1816 corresponding to cyclo-C3-Me-L-Trp-L-Trp DKP (13).

<table>
<thead>
<tr>
<th>Observed Mass</th>
<th>% Rel. Int.</th>
<th>Predicted Mass</th>
<th>Mass Diff. (ppm)</th>
<th>Ion Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>387.1825</td>
<td>14.7</td>
<td>387.1816</td>
<td>2.32</td>
<td>C₂₂H₂₁N₄O₂⁺</td>
</tr>
<tr>
<td>242.0936</td>
<td>3.0</td>
<td>242.0924</td>
<td>4.96</td>
<td>C₁₃H₁₂N₃O₂⁺</td>
</tr>
<tr>
<td>184.0761</td>
<td>10.1</td>
<td>184.0757</td>
<td>2.17</td>
<td>C₁₂H₁₀NO⁺</td>
</tr>
<tr>
<td>159.0923</td>
<td>3.2</td>
<td>159.0917</td>
<td>3.77</td>
<td>C₁₀H₁₁N₂⁺</td>
</tr>
<tr>
<td>156.0815</td>
<td>3.9</td>
<td>156.0808</td>
<td>4.48</td>
<td>C₁₁H₁₀N⁺</td>
</tr>
<tr>
<td>144.0813</td>
<td>100</td>
<td>144.0808</td>
<td>3.47</td>
<td>C₁₀H₁₀N⁺</td>
</tr>
<tr>
<td>130.0656</td>
<td>24.5</td>
<td>130.0651</td>
<td>3.84</td>
<td>C₉H₈N⁺</td>
</tr>
<tr>
<td>103.0547</td>
<td>1.0</td>
<td>103.0542</td>
<td>4.85</td>
<td>C₈H₇⁺</td>
</tr>
</tbody>
</table>
Table 7. MS² fragments for [M+H]+ peak at m/z = 401.1972 corresponding to cyclo-C3-Me-L-Trp-N1′-Me-L-Trp DKP (16).

<table>
<thead>
<tr>
<th>Observed Mass</th>
<th>% Rel. Int.</th>
<th>Predicted Mass</th>
<th>Mass Diff. (ppm)</th>
<th>Ion Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>401.1981</td>
<td>8.8</td>
<td>401.1972</td>
<td>2.24</td>
<td>C₂₅H₂₅N₄O₂⁺</td>
</tr>
<tr>
<td>384.1707</td>
<td>0.5</td>
<td>384.1707</td>
<td>0.00</td>
<td>C₂₄H₂₂N₃O₂⁺</td>
</tr>
<tr>
<td>256.1089</td>
<td>1.0</td>
<td>256.1081</td>
<td>3.12</td>
<td>C₁₄H₁₄N₃O₂⁺</td>
</tr>
<tr>
<td>201.1023</td>
<td>0.5</td>
<td>201.1022</td>
<td>0.50</td>
<td>C₁₂H₁₃N₂O⁺</td>
</tr>
<tr>
<td>184.0761</td>
<td>2.6</td>
<td>184.0757</td>
<td>2.17</td>
<td>C₁₂H₁₀NO⁺</td>
</tr>
<tr>
<td>173.1078</td>
<td>0.5</td>
<td>173.1079</td>
<td>-0.58</td>
<td>C₁₁H₁₃N₂⁺</td>
</tr>
<tr>
<td>156.0813</td>
<td>0.8</td>
<td>156.0808</td>
<td>3.20</td>
<td>C₁₂H₁₄N⁺</td>
</tr>
<tr>
<td>144.0813</td>
<td>100</td>
<td>144.0808</td>
<td>3.47</td>
<td>C₁₃H₁₄N⁺</td>
</tr>
</tbody>
</table>
Table 8. MS² fragments for [M+H]⁺ peak at m/z = 441.2285 corresponding to cyclo-L-Trp-C3’-prenyl-L-Trp DKP (12)

<table>
<thead>
<tr>
<th>Observed Mass</th>
<th>% Rel. Int.</th>
<th>Predicted Mass</th>
<th>Mass Diff. (ppm)</th>
<th>Ion Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>441.2292</td>
<td>1.5</td>
<td>441.2285</td>
<td>1.59</td>
<td>C₂₇H₂₉N₄O₂⁺</td>
</tr>
<tr>
<td>385.1664</td>
<td>2.6</td>
<td>385.1659</td>
<td>1.30</td>
<td>C₂₃H₂₁N₄O₂⁺</td>
</tr>
<tr>
<td>373.1665</td>
<td>3.6</td>
<td>373.1659</td>
<td>1.61</td>
<td>C₂₂H₂₁N₄O₂⁺</td>
</tr>
<tr>
<td>271.1235</td>
<td>1.2</td>
<td>271.1230</td>
<td>1.84</td>
<td>see below</td>
</tr>
<tr>
<td>242.0930</td>
<td>2.3</td>
<td>242.0924</td>
<td>2.48</td>
<td>C₁₉H₁₂N₃O₂⁺</td>
</tr>
<tr>
<td>198.1287</td>
<td>100</td>
<td>198.1277</td>
<td>5.05</td>
<td>C₁₄H₁₆N⁺</td>
</tr>
<tr>
<td>183.1044</td>
<td>8.0</td>
<td>183.1043</td>
<td>0.55</td>
<td>C₁₃H₁₃N⁺</td>
</tr>
<tr>
<td>168.0815</td>
<td>6.3</td>
<td>168.0808</td>
<td>4.16</td>
<td>C₁₂H₁₀N⁺</td>
</tr>
<tr>
<td>156.0809</td>
<td>5.6</td>
<td>156.0808</td>
<td>0.64</td>
<td>C₁₁H₁₀N⁺</td>
</tr>
<tr>
<td>144.0810</td>
<td>7.0</td>
<td>144.0808</td>
<td>1.39</td>
<td>C₁₀H₁₀N⁺</td>
</tr>
<tr>
<td>130.0656</td>
<td>25.0</td>
<td>130.0651</td>
<td>3.84</td>
<td>C₉H₈N⁺</td>
</tr>
</tbody>
</table>
Table 9. MS² fragments for [M+H]+ peak at m/z = 455.2442 corresponding to Cyclo-L-Trp-N1’-Me-C3’-nprenyl-L-Trp DKP (15).

<table>
<thead>
<tr>
<th>Observed Mass</th>
<th>% Rel. Int.</th>
<th>Predicted Mass</th>
<th>Mass Diff. (ppm)</th>
<th>Ion Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>455.2435</td>
<td>0.8</td>
<td>455.2442</td>
<td>-1.54</td>
<td>C_{28}H_{31}N_{4}O_{2}^{+}</td>
</tr>
<tr>
<td>399.1806</td>
<td>2.1</td>
<td>399.1816</td>
<td>-2.51</td>
<td>C_{24}H_{23}N_{4}O_{2}^{+}</td>
</tr>
<tr>
<td>387.1823</td>
<td>2.0</td>
<td>387.1816</td>
<td>1.81</td>
<td>C_{23}H_{23}N_{4}O_{2}^{+}</td>
</tr>
<tr>
<td>285.1396</td>
<td>3.8</td>
<td>285.1387</td>
<td>3.16</td>
<td>C_{20}H_{17}N^{2+}</td>
</tr>
<tr>
<td>268.1089</td>
<td>2.1</td>
<td>268.1081</td>
<td>2.98</td>
<td>C_{15}H_{14}N_{3}O_{2}^{2+}</td>
</tr>
<tr>
<td>212.1441</td>
<td>100</td>
<td>212.1434</td>
<td>3.30</td>
<td>C_{15}H_{18}N^{+}</td>
</tr>
<tr>
<td>197.1207</td>
<td>9.7</td>
<td>197.1199</td>
<td>4.06</td>
<td>C_{14}H_{15}N^{2+}</td>
</tr>
<tr>
<td>182.0972</td>
<td>8.6</td>
<td>182.0964</td>
<td>4.39</td>
<td>C_{13}H_{12}N^{+}</td>
</tr>
<tr>
<td>158.0969</td>
<td>4.1</td>
<td>158.0964</td>
<td>3.16</td>
<td>C_{11}H_{11}N^{+}</td>
</tr>
<tr>
<td>144.0814</td>
<td>52.1</td>
<td>144.0808</td>
<td>4.16</td>
<td>C_{9}H_{8}N^{+}</td>
</tr>
<tr>
<td>130.0657</td>
<td>6.5</td>
<td>130.0651</td>
<td>4.61</td>
<td>C_{8}H_{7}^{+}</td>
</tr>
</tbody>
</table>
Figure 28: cyclo-L-Trp-L-Trp DKP (5). A. LC-ESI (-)-TOF- MS$^1$ and MS$^2$ spectrum from Nocardiosis sp. CMB-M0232. B. ESI (-) MS$^2$ for 5 extracted from Nocardiosis sp. CMB-M0232. C. ESI(-) MS$^2$ for synthetic 5.
Figure 29: LC-ESI (+)-TOF-MS$^1$ and MS$^2$ traces from *Nocardiopsis* sp. CMB-M0232 confirming presence of tabulated metabolites as given in Table 1 for m/z = 373.2 (5); 387.2 (13) and 574.4 (nocardiopsin A, * - pending full confirmation).
Figure 30: LC-ESI (+)-TOF- MS¹ and MS² traces from Nocardiopsis sp. CMB-M0232 confirming presence of tabulated metabolites as given in Table 1 for m/z = 401.2 (16); 455.3 (des-N1'-Me-nocardioazine B (17)) and 469.3 (nocardioazine B (4)).
Figure 31. ESI (+)-TOF MS$^1$ and MS$^2$ spectra for m/z = 373.1665 (5) extracted from from *Nocardiopsis* sp. CMB-M0232.
Figure 32: ESI (+)-TOF MS$^1$ and MS$^2$ spectra for m/z = 387.1791 (13) extracted from *Nocardiopsis* sp. CMB-M0232. Cyclo-C3-Me-L-Trp-L-Trp DKP (13) (7.7 ppm error) is shown with blue arrow, 387.2384 (148 ppm error) is not 13, and is a different unidentified metabolite. Also, 387.2361 not fragmented during MS$^1$ and MS$^2$ where Cyclo-C3-Me-L-Trp-L-Trp DKP (13) was totally fragmented in standard sample. Presence of 144.0811 fragment (2.1 ppm error) indicates cyclo-C3-Me-L-Trp-L-Trp DKP (13).
Figure 33. LC-ESI+ TOF MS$^1$ and MS$^2$ spectra for m/z = 469.27 (nocardioazine B, 4) extracted from from Nocardiopsis sp. CMB-M0232.
Figure 34. ESI (+)-TOF MS\(^1\) and MS\(^2\) spectra for m/z = 469.27; noccardiozine B (4) from *Nocardiopsis* sp. CMB-M0232.
Figure 35. TLC images of intermediates extracted from *Nocardiopsis* sp. CMB M0322 corresponding to alkaloidal fractions. A1, B1 and C1 are synthetic standards of Cyclo-C3-Me-L-Trp-L-Trp DKP (13). A2 and A3 are extracts of *Nocardiopsis* sp. CMB M0322 at 7 days from the time of inoculation. B2 and B3 are extracts of *Nocardiopsis* sp. CMB M0322 at 14 days from the time of inoculation. C2 and C3 are extracts of *Nocardiopsis* sp. CMB M0322 at 21 days from the time of inoculation. A4, B4 and C4 are co-spots between respective extracts and synthetic standard of 13.
5.16. NMR for 13 from *Nocardiopsis* extract:

![NMR spectrum](image)

**Figure 36.** $^1$H NMR spectrum for 13 extracted from *Nocardiopsis* sp. CMB-M0232 after 21 days of culture followed by fractionation and purification. Purification methods followed procedures adopted for the synthesis of 13 except culture extracts were used instead of crude reaction mixtures.
Figure 37. $^{13}$C NMR spectrum for 13 extracted from *Nocardiopsis* sp. CMB-M0232 – 6.9-7.15 ppm.
Cyclo-L-Trp-L-Trp-DKP (5): and Cyclo-D-Trp-D-Trp-DKP (ent-5) HPLC analysis

A considerable asymmetry is noticeable in the HPLC traces and we attribute this to the fact that 5 and ent-5 are polar entities and show significant “tailing” effect on a normal phase chiral AS column under the conditions observed.

Figure 38: HPLC analyses of cyclo-L-Trp-L-Trp-DKP (5).

Figure 39: HPLC analyses of cyclo-D-Trp-D-Trp-DKP (ent-5).

105 A considerable asymmetry is noticeable in the HPLC traces and we attribute this to the fact that 5 and ent-5 are polar entities and show significant “tailing” effect on a normal phase chiral AS column under the conditions observed.
5.17. Synthesis of malonyl SNAc Thioester

![Chemical Structure](image)

3-(tert-butoxy)-3-oxopropanoic acid (39)\(^{106,107}\)

Malonic acid (2.00 g, 19.21 mmol) was dissolved in dry THF (25 mL). Pyridine (3.40 mL, 42.2 mmol) and tert-butanol (4.00 mL, 42.26 mmol) were added with stirring. After cooling to 0 °C, methanesulfonyl chloride (3.71 ml, 48.02 mmol) was added over 10 min. The mixture was stirred at room temperature for 3 h and then filtered to remove the pyridine hydrochloride salt. The filtrate was diluted in water (20 mL). This solution was brought to pH ~11 using 4 N NaOH and then washed with CH\(_2\)Cl\(_2\) (3 x 20 mL). The aqueous layer was acidified to pH ~3 with 1.0 M aqueous HCl and extracted with CH\(_2\)Cl\(_2\) (4 x 20 mL). Evaporation of the solvent afforded 39 as a colorless oil (2.44 g, 79 %).

**Physical appearance:** colorless oil. \(^1\)H NMR (400 MHz, CDCl\(_3\)): δ 10.97 (brs, 1H), 3.26 (brs, 2 H), 1.38 (brs, 9H); \(^13\)CNMR (100 MHz, CDCl\(_3\), DEPT): δ 171.9 (C), 166.1 (2 × C), 42.1 (CH\(_2\)), 27.7 (3 × CH3). HRMS (ESI, M+): m/z calcd. for C\(_7\)H\(_{12}\)O\(_4\) 160.0736, found 160.0736.


Tert-butyl 3-((2-acetamidoethyl) thio)-3-oxopropanoate (40)

To a solution of 39 (460 mg, 27.68 mmol) in CH$_2$Cl$_2$ (6 mL) were successively added EDCI (1.00 g, 65.44 mmol) and DMAP (6.0 mg, 0.50 mmol). The mixture was stirred at 0 °C for 15 min and then HSNAC (3.00 mL, 25.17 mmol) was added. The reaction mixture was stirred overnight at room temperature, then concentrated in vacuo. The residue was dissolved in 25mL of chloroform and washed successively with 1.0 M aqueous HCl (3 x 10 mL) and brine, then dried over MgSO$_4$. After concentration in vacuo, the clear oil residue was purified by silica gel chromatography eluted with ethyl acetate: hexane (6:4) to yield 40 (635 mg, 87%). Physical appearance: a colorless oil. $^1$H NMR (400 MHz, CDCl$_3$): $\delta$ 6.79 (brs, 1H), 3.37 (s, 2H), 3.29-3.25 (m, 2H), 2.94 (t, J = 6.52 Hz, 2H), 1.83 (s, 3H), 1.32 (s, 9H). $^{13}$C NMR (100 MHz, CDCl$_3$, DEPT): $\delta$ 191.6 (C), 170.6 (C), 165.0 (C), 82.3 (C), 50.5 (CH$_3$), 38.9 (CH$_2$), 28.7 (CH$_2$), 27.7 (3 × CH$_3$), 22.8 (CH$_2$). HRMS (ESI, M$^+$): m/z calcd. for C$_{11}$H$_{19}$NO$_4$S 261.1035, found 261.1035.

3-((2-acetamidoethyl) thio)-3-oxopropanoic acid (41)

3-((2-acetamidoethyl) thio)-3-oxopropanoic acid (41)
In an oven-dried, nitrogen-purged 15-mL round-bottom flask equipped with a stir bar, compound 40 (338 mg, 1.29 mmol) was dissolved in 3 mL of CH$_2$Cl$_2$ and cooled to 0 °C, then 3 mL of TFA was added and the mixture was stirred for 2 h at 0 °C. After evaporation of TFA in vacuo, the residue was repeatedly dissolved in benzene and concentrated in vacuo to afford 238 mg of 41 (90%). Physical appearance: a yellow liquid. $^1$H NMR (400 MHz, CD$_3$OD): $\delta$ 3.62 (brs, 2 H), 3.35 (t, $J = 6.64$ Hz, 2H), 3.06 (t, $J = 6.60$ Hz, 2H), 1.92 (s, 3H). $^1$3C NMR (100 MHz, CD$_3$OD, DEPT): $\delta$ 193.0 (C), 173.5 (C), 169.6 (C), 50.1 (CH$_2$), 39.9 (CH$_3$), 29.6 (CH$_2$), 22.5 (CH$_2$). HRMS (ESI, M$^+$): m/z calcd. for C$_7$H$_{11}$NO$_4$S 205.0409, found 205.0409.

5.18. Synthesis of methyl malonyl SNAC Thioester

\[
\begin{align*}
\text{HO-COO} & \quad \text{SO-CI} \\
\text{t-BuOH, Pyridine} & \quad \text{THF, 0 °C-rt, 3h} \\
\text{HO-COO} & \quad \text{HO-COO} \\
42 & \quad 43
\end{align*}
\]

3-(tert-butoxy)-2-methyl-3-oxopropanoic acid (43)

Methylmalonic acid (2.00 g, 16.93 mmol) was dissolved in dry THF (25 mL). Pyridine (3.00 mL, 37.2 mmol) and tert-butanol (4.00 mL, 42.3 mmol) were added with stirring. After cooling to 0 °C, methanesulfonyl chloride (2.62 mL, 33.8 mmol) was added over
10 min. The mixture was stirred at room temperature for 3 h and then filtered to remove the pyridine hydrochloride salt. The filtrate was diluted in water (20 mL). This solution was brought to pH ~11 using 4 N NaOH and then washed with CH₂Cl₂ (3 x 20 mL). The aqueous layer was acidified to pH ~3 with 1.0 M aqueous HCl and extracted with CH₂Cl₂ (4 x 20 mL). Evaporation of the solvent afforded 43 as a colorless oil (2.15 g, 73 %).

**Physical appearance:** a colorless oil. **¹H NMR (400 MHz, CDCl₃):** δ 11.25 (brs, 1H), 3.26 (dd, J = 14.6, 7.3 Hz, 1H), 1.33 (s, 9 H), 1.25 (d, J = 7.3 Hz, 3H). **¹³CNMR (100 MHz, CDCl₃, DEPT):** δ 176.0 (C), 169.0 (C), 82.0 (C), 46.8 (CH), 27.6 (3 × CH₃), 13.3 (CH₃). HRMS (ESI, M⁺): m/z calcd. for C₈H₁₄O₄ 174.6892, found 174.6892.

**Tert-butyl 3-((2-acetamidoethyl) thio)-2-methyl-3-oxopropanoate (44)**

To a solution of 43 (482 mg, 27.68 mmol) in CH₂Cl₂ (6 mL) were successively added EDCI (970 mg, 62.9 mmol) and DMAP (6.7 mg, 0.55 mmol). The mixture was stirred at 0 °C for 15 min and then HSNAC (3.00 mL, 25.16 mmol) was added. The reaction mixture was stirred overnight at room temperature, then concentrated *in vacuo*. The residue was dissolved in 25ml of chloroform and washed successively with 1.0 M aqueous HCl (3 x 10 mL) and brine, then dried over MgSO₄. After concentration *in vacuo*, the clear oil residue was purified by silica gel chromatography eluted with ethyl acetate: hexane (6:4) to yield 44 (428 mg 56%). **Physical appearance:** a colorless oil. **¹H NMR (400 MHz, CDCl₃):** δ 6.71 (brs, 1H), 3.43 (dd, J = 14.3, 7.2 Hz, 1H), 3.33-3.22 (m,
2H), 2.94 (ddd, J = 8.3, 6.6, 1.7 Hz, 2H), 1.84 (s, 3H), 1.31 (s, 3H), 1.24 (d, J = 7.2 Hz, 3H); 13C NMR (100 MHz, CDCl3, DEPT): δ 196.3 (C), 170.6 (C), 168.4 (C), 82.0 (C), 54.8 (CH), 39.1 (CH2), 28.4 (CH2), 27.6 (3 × CH3), 22.9 (CH3), 13.9(CH3). HRMS (ESI, M+): m/z calcd. for C12H21NO4S 275.1191, found 275.1191.

3-((2-acetamidoethyl)thio)-2-methyl-3-oxopropanoic acid (45)

In an oven-dried, nitrogen-purged 15-mL round-bottom flask equipped with a stir bar, compound 44 (224 mg, 0.81 mmol) was dissolved in 3 mL of CH2Cl2 and cooled to 0 °C, then 3 mL of TFA was added and the mixture was stirred for 2 h at 0 °C. After evaporation of TFA in vacuo, the residue was repeatedly dissolved in benzene and concentrated in vacuo to afford 164 mg of 45 (92%). Physical appearance: brown oil. 1H NMR (400 MHz, CD3OD): δ 5.60 (brs, 1H), 3.71 (dd, J = 14.3, 7.2 Hz, 1H), 3.36-3.31 (m, 2H), 3.05 (t, J = 6.6 Hz, 2H), 1.91(s, 3H), 1.38 (d, J = 7.1 Hz, 3H); 13C NMR (100 MHz, CD3OD, DEPT): δ 197.6 (C), 173.5 (C), 172.5 (C), 55.0 (CH), 39.9 (CH2), 29.4 (CH2), 22.5 (CH3), 14.5 (CH3). HRMS (ESI, M+): m/z calcd. for C8H13NO4S 219.0565, found 219.0565.
5.19. Synthesis of D/L-Trp CME ester (47)

\[
\begin{align*}
N\text{-BOC-D-Trp} & \quad \xrightarrow{\text{CICH}_2\text{CN, DMF, Et}_3\text{N}} \quad N\text{-BOC-L-Trp} \\
\text{rt, overnight} & \quad \text{rt, overnight}
\end{align*}
\]

D/L-Trp- N-BOC- cyanomethyl ester (46 or *ent*-46)\(^{108}\)

\(N\)-Boc-L/D-tryptophan (235 mg, 0.77 mmol) was mixed with triethylamine (0.323 mL, 2.31 mmol) and chloroacetonitrile (0.146 mL, 2.31 mmol) in 1.00 mL of DMF, and the reaction mixture was stirred for overnight at room temperature. Excess amount of chloroacetonitrile was removed under reduced pressure and AcOEt (10 mL) was added.

The resulting solution was washed with 1 M HCl (3 mL x 3), sat. NaHCO3 (3 mL x 3) and brine (5 mL), and the organic layer was dried over MgSO4 and concentrated under reduced pressure. The crude residue was purified by silica-gel flash chromatography (hexane/AcOEt = 6:1) to yield 46 or ent-46 (226 mg 85%). **Physical appearance:** pink oil.  \(^1\text{H NMR (400 MHz, CDCl}_3\): \(\delta\) 8.45 (brs, 1H), 7.56 (d, \(J = 7.7\) Hz, 1H), 7.35 (d, \(J = 8.1\) Hz, 1H), 7.21 (ddd, \(J = 9.1, 7.2, 1.1\) Hz, 1H), 7.15 (ddd, \(J = 8.6, 7.0, 1.0\) Hz, 1H), 7.01 (brs, 1H), 5.13 (d, \(J = 8.0\) Hz, 1H), 4.72- 4.67 (m, 1H), 4.63- 4.49 (m, 2H), 3.30 (d, \(J = 5.7\) Hz, 2H), 1.44 (s, 9H). ; \(^{13}\text{C NMR (100 MHz, CDCl}_3, \text{DEPT)}: \delta 171.2\) (C), 155.3(C), 136.2 (C), 127.4 (C), 123.3 (CH), 122.4 (CH), 119.8 (CH), 118.4 (CH), 114.1 (C), 111.5 (CH), 109.0 (C), 80.5 (C), 54.3 (CH), 48.8 (CH), 28.3 (3 ×CH3), 27.8 (CH2). HRMS (ESI, M+): m/z calcd. for C\(_{18}\)H\(_{21}\)N\(_3\)O\(_4\) 343.1532, found 343.1532.

---

**D/L-Trp-cyanomethyl ester (47 or ent- 47)**

The obtained product D/L-Trp- \(N\)-BOC- cyanomethyl ester (230 mg, 0.82 mmol) was added to a mixture of 3 mL of CH\(_2\)Cl\(_2\) and 3 mL of trifluoroacetic acid at 0 °C. The mixture was stirred for 2 h. The resulting mixture was concentrated under reduced
pressure and the remained solvent was removed by repeating the addition of benzene and concentrated in vacuo to afford 154 mg of 47 or ent-47 (77%). Physical appearance: pink oil. ¹H NMR (400 MHz, CD₃OD): δ 7.56 (t, J = 7.8 Hz, 1H), 7.40 (d, J = 8.1 Hz, 1H), 7.23 (d, J = 9.1 Hz, 1H), 7.16-7.12 (m, 1H), 7.10-7.04 (m, 1H), 4.91 (s, 1H), 4.57 (dd, J = 32.0, 14.5 Hz, 1H), 4.45-4.39 (m, 1H), 3.55-3.37 (m, 2H). ¹³C NMR (100 MHz, CD₃OD, DEPT): δ 169.6 (C), 138.2 (C), 128.0 (C), 125.8 (CH), 123.0 (CH), 120.4 (CH), 118.7 (CH), 115.3 (C), 112.7 (CH), 107.0 (C), 54.4 (CH), 50.9 (CH₂), 28.7 (CH₂). HRMS (ESI, M⁺): m/z calcd. for C₁₃H₁₃N₃O₂ 243.1008, found 243.1008.
APPENDIX
Appendix 2: Spectral Data
100 MHz, CD$_3$OD
400 MHz, CD$_2$OD
$\text{400 MHz, CD}_3\text{OD}$
400 MHz, CD$_3$OD
HN
NH
O
NH
HN
H
H
O

400 MHz, CD$_3$OD
$\text{CO}_2\text{Me}$

$\text{NH}\text{Cbz}$

100 MHz, CDCl$_3$
$100 \text{ MHz, CDCl}_3$
600 MHz, CD$_2$OD
150 MHz, C$_3$D$_2$
600 MHz, CDCl₃
100 MHz, CDCl₃
400 MHz, CDCl₃
Appendix 2:

An Overview of the History and Current Relevance of Natural Products
A2.1.1. Why Natural Products are important?

Natural Products (NPs) are compounds that are derived from natural sources such as plants, animals and microorganisms. Their innate pharmacological activities have been used for pharmaceutical drug design and discovery.\textsuperscript{109,110} NPs have been used to produce crude therapeutic formulations for more than thousands of years. For example medicines such as aspirin, digitoxin, quinine, morphine, and pilocarpine, originally were developed as end products derived from plant extracts.\textsuperscript{111} Through much of the last century, the diversity of natural product scaffolds as well as that of their associated functional groups has served as major inspirations (at the core of pharmaceutical discovery) for chemical diversity of starting materials.\textsuperscript{112} Natural products continue to play a vital role in therapeutic discovery against infectious diseases; however, pharmaceutical companies have dramatically decreased activities in natural product discovery over the past few decades. This is mostly due to the complexity of many natural products; which limits the possibilities of making chemical modifications to these scaffolds.\textsuperscript{113} With biotechnology companies having limited success pursuing projects in the fields of combinatorial biosynthesis, genetic engineering, and metagenomic approaches to identify renewable active substances from natural products, the use of natural products in therapeutic


discovery has prompted questions about the relevance of natural products in a day and
time when advances in synthetic procedures have made the production of any molecule a
realistic target.112

Evidence of the deep historical role of natural products dates as far back as 2600
B.C. when the first medical texts with records of natural products were depicted on
hundreds of clay tablets in cuneiform.114,115 These documents are filled with descriptions
of plants and plant derived substances such as oils from Cupressus sempervirens
(Cypress), the oils of Cedrus species (cedar), and those from the Commiphora species
(myrrh), as well as juice of the poppy seed Papaver somniferum.114, 115 Many of these
substances are still used today in the treatments of cough, cold, and inflammations.114, 115
Another early pharmaceutical record of natural products, the Egyptian Ebers Papyrus
which dates back to about 1550 B.C., contains approximately 800 complex medicines
and 700 natural (plant) based drugs ranging from pills to ointments like Aloe vera (aloe),
frankincense (Boswellia carteri), and castor oil (Ricinus communis).114, 115 The Chinese
Materia Medica (circa 40-90 A.D.) compiled by Pendanius Dioscorides, acted as the
basis of pharmacology in Europe through the collection, storage and documentation of
almost 600 plant based therapeutic solutions.114

With the passage of time, and the development of analytical and structural
chemistry, scientists developed the tools to isolate compounds, determine their structures

115 Dias, A. D.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery.
and their effects on the human body, which facilitated the synthesis of molecules, rather than their isolation from natural products, the latter which was economically unproductive.\textsuperscript{114} The first pure drug to be naturally derived was morphine, from opium by German pharmacist Friedrich Wilhelm in 1805, and this ushered in an era where large numbers of well-known natural products were located, broken down and analyzed and then synthesized for pharmaceutical purposes, and many of these are still used as drugs even today. These products included but were not limited to Salicin from \textit{Salix alba}, quinine from \textit{Cinchona ledgeriana}, caffeine from \textit{Coffea arabica}, nicotine from \textit{Nicotiana tabacum}, and cocaine from \textit{Erythroxylum coca}.\textsuperscript{114} With a rich history of the discovery and usage of natural products in the pursuit of pharmaceutical solutions against diseases, perhaps the golden era of natural products came during the twentieth century, specifically around World War II, with the discovery of the antibacterial properties of the derivate of \textit{Penicillium notatum}, penicillin, \textsuperscript{110} which paved the way for the discovery of new antibiotics against infectious diseases. The discovery of penicillin was complimented by the discoveries of streptomycin, gentamicin and tetracycline, among other antibiotics, that triggered the development of extensive research and development sectors in the pharmaceutical industry centered on the discovery of new natural products with possible pharmaceutical solutions for not just anti-bacterial, and anti-fungal targets, but also other infectious diseases.\textsuperscript{110} Notwithstanding the massive growth and success of drug discovery and engineering programs based on natural products, pharmaceutical companies switched focus from natural products in the 1990’s and 2000’s, due to the inability of natural products to meet the sheer numbers necessitated for successful automated high throughput screening.\textsuperscript{110}
Despite the negative trend in the use of natural products in the pursuit of pharmaceutical solutions to infectious diseases among other things, there is data that shows that natural products still matter in the production of drugs and clinical alternatives. Between 1981 and 2002, natural products consisted of 5% of all chemical constructs approved as drugs by the US Food and Drug Administration (FDA), with a further 23% of those drugs being natural product derived molecules.\textsuperscript{113} Between 1981 and 2006, 34% of small molecule new chemical entities were natural products or their semisynthetic derivatives, and these molecules made up about two-thirds (68%) of the anti-bacterial new chemical entities, and 54% of the anticancer new chemical entities.\textsuperscript{111} In another study, it was reported by Proudfoot that of the 29 small molecule drugs launched in the year 2000, 8 were derived from natural products or hormones.\textsuperscript{111} The decline in discovery efforts for natural products has seemingly not had much of an impact on the contribution of natural products and their semisynthetic derivatives to new chemical entities, as this has remained relatively stagnant over the previous two decades. In the year 2000, the percentage of natural product derived drugs was 40% but went to 24% in 2001, and then 26% in 2002, showing that natural products not only still remain relevant for their contributions in therapeutic solutions, but they also contribute significantly to the profit margins of companies.\textsuperscript{111}

Several of theories that attempt to offer an explanation as to why natural products are still a vital source of pharmaceutical solutions has to do with the origin of beneficial natural products. One of these; the coevolution theory, asserts that natural products, driven by natural selection and evolution have developed profound specificity which continues to make them vital in the pharmaceutical business, as this ability to specifically
affect a single node in a cellular network makes them valuable biological probes. An example of the invaluable function of natural products as biological probes is rapamycin, which contributed to the discovery of serine/threonine protein kinase mTOR (mammalian target of rapamycin), and helped identify the role of mTOR in signaling pathways that lead to cell proliferation and protein synthesis. An alternative hypothesis that has recently been proposed to explain the origin of beneficial natural products, which is one of the main reasons why natural products remain relevant in the 21st century, is dubbed xenohormesis. This theory declares that the common ancestor of plants and animals must have had the ability to produce a profusion of stress induced secondary metabolites, an ability which got lost to plants and fungi down the evolutionary path, but was retained in plants to allow them to timely sense changes in environmental circumstances.

Transitioning to a drug candidate from a screening hit is an exhaustive process that requires both experience and expertise. Natural products while widely recognized as sources for therapeutic solutions have faced a decline in discovery efforts over the years for several aforementioned reasons. Nevertheless, there has been renewed interest in the contribution of natural products to drugs, and this is most likely due to their structural diversity, their biodiversity, and the invention of new technologies which have facilitated the screening of natural products in developing new drugs. Natural products contain larger scale structural diversity than synthetic products and are major resources for bioactive agents, which have been at the core of drug discovery. The enormous structural and chemical diversity of natural products gives them about 40% more

chemical scaffolds than are present in today’s medicinal chemistry, which makes natural
products highly complementary to today’s synthetically produced molecules. Furthermore, natural products remain vital because there are still many of them to discover. Bacterial genome sequences have shown that a single strain can lead to the production of 25-30 distinct molecules, leading to renewed efforts to discover the abundant remaining natural products that remain hidden. The renewed interest in natural products has also been spurred by the failures of alternative drug discovery methods, which have attempted to deliver compounds to therapeutic fields such as immunosuppression among others. A glaring example of the failures of other methods of drug discovery is highlighted by a study conducted by Newman and Cragg. The authors report that over the 30 years documented by their literature review, combinatorial chemistry has been used as a method of drug discovery for approximately 70% of that time, and yet the review can find only one de novo new chemical entity in public domain engineered through this method, and approved for drug use anywhere. Sorafineb (Nexavar, 1), as this antitumor chemical is known, was declared as an appropriate treatment for renal cell carcinoma by the FDA back in 2005, and then in 2007 as an appropriate treatment for hepatocellular carcinoma. The compound has gotten approval for usage in Switzerland, the European Union, and the People’s Republic of China, and is awaiting approval in other countries. The developmental potential of combinatorial chemistry as a pathway for structural optimization for therapeutic purposes is unparalleled. However, this potential has thus far failed to materialize as was expected. In light of the shortcomings of other drug discovery methods, natural products have

\[\text{Newman, D. J.; Cragg, M. G. Natural Products as Sources of New Drugs over the 30 Years from 1981 to 2010. } J. Nat. Prod. \textbf{2012, 75}, 311–335.\]
continued to inspire research geared at exploring a variety of lead structures that have the potential of being used as templates for the development of new drugs by pharmaceutical companies.116

A2.1.2. Are Marine Microbiota Sources for Novel Drugs?

Oceans are fundamental to our planet for the sheer area that they cover, and for the ecosystems that they provide.118 Of all animal phyla that have been identified and described so far, 15% in habitat the world’s oceans exclusively and only phylum Onychopora has not been identified in a marine environment.118 Oceans make up about approximately 70% of the earth’s surface and claim habitat to more than 300,000 described plant and animal life species. Therefore, to claim that oceans are home to huge diversity and microbial populations would be quite an understatement.119,120 According to previous documentation, the first marine living organisms date back some 3500 million years, ensuring that evolutionary forces have had enough time to confer these microbes with adaptive properties against some of the harsh characteristics of marine environments such as extreme temperatures, changes in salinity and pressure as well as overcoming the adversities that can be posed by viral pathogens, mutations, and mutations.120 Life on earth can be traced back to origins in the sea, and in certain marine ecosystems such as coral reefs or deep sea floors, the species diversity is extraordinarily rich and trumps the

118 Leal, C. M.; Puga, J.; Serodio, J.; Gomez, C. M. N.; Calado, R. Trends in the Discovery of New Marine Natural Products from Invertebrates over the Last Two Decades – Where and What Are We Bioprospecting?. *Plos One*. **2012**, *7*.


biological diversity of even the tropical rain forest.\textsuperscript{121} Most marine organisms are inactive and soft bodied individuals and require chemical means of defense. Also over the years they have developed exquisitely complex biological mechanisms such as the capability of synthesizing poisonous organisms or obtaining them from their environment in order to fend off predators, immobilize prey, or keep competitors under control.\textsuperscript{121} Much like terrestrial natural products, interest in marine natural products and their role as therapeutic solutions is evidenced as far back as the ancient maritime people (Chinese and Japanese), who ate a variety of seaweeds rich in iodine to help prevent goiter.\textsuperscript{120} In Japan, China, Taiwan, and India, products of corals have customarily been used to create various treatments.\textsuperscript{120} It had been known for centuries that sponges contained bioactive compounds with great pharmaceutical potential; however, it was not until the 20\textsuperscript{th} century that scientists began to methodically investigate the oceans for medicines.\textsuperscript{120} In this regard, the field of marine natural product chemistry is a very novel pursuit, which finds its roots just in the 1960’s.\textsuperscript{122}

Early studies in this field, spurred on by the marvels of the structural diversity of marine natural products, concentrated on the most noticeable and collectable organisms, such as intertidal and shallow sub-tidal microalgae and invertebrates.\textsuperscript{122} These studies were however driven by a need to identify and comprehend the chemistry behind many toxins that people encountered in order to describe the adaptations of marine life.\textsuperscript{123} These toxins included the likes of, but were not limited to saxitoxin, tetrodotoxin and the


With the arrival of SCUBA diving and the application of deep sea diving vessels to the cause of marine natural product chemistry, deeper water organisms have been collected, as well as smaller and more cryptic organisms. The focus of studies in this field has also shifted to biomedical screening programs, with several drugs being engineered using clues from marine natural products. From the efforts of biomedical screening of deeper water organism for pharmacologically significant secondary metabolites, over 22,000 distinct marine metabolites have been identified, isolated and characterized by structure. A substantial number of these metabolites have been evaluated for some level of pharmacological activity such as mammalian cell toxicity. More recently, the enzymatic and genetic foundations of these metabolites have been eye-marked with the vision of isolating these biochemical facilitators in order to promote molecular diversity that can be utilized in biomedical and agrochemical settings.

While the focus on smaller creatures such as cyanobacteria, and marine fungi that had previously been ignored in collection and examination efforts have been very productive, the new discoveries have only scratched the surface of the vast richness of microbial life in the sea. As a matter of fact, it is estimated that only 1% of bacteria present in marine settings have been isolated and cultured. A comparative study of marine and terrestrial natural products showed that marine natural products are far superior to their terrestrial counterparts in terms of chemical originality. Despite this finding, almost all of the current therapeutic discoveries stemming from natural products have terrestrial origins, thus leaving a massive potential pharmaceutical gold mine in

maritime environments if used as novel mining sources for biological novelties.\textsuperscript{124} The comparative analysis by Kong and co-workers compared molecular scaffold in the Dictionary of Natural Products to those in the marine equivalent source that is the Dictionary of Marine Natural Products, and found that roughly 71\% of molecular scaffolds reported in the Dictionary of Marine Natural Products were exclusive to marine organisms.\textsuperscript{124} Furthermore, when compared to their terrestrial counterparts, marine organisms showed higher incidences of significant bioactivity, results that were confirmed by a National Cancer Institute pre-clinical cytotoxicity screen, in which marine samples showed about 10 times more anti-tumor potential (1\% of samples tested), compared to terrestrial samples (0.1\%).\textsuperscript{124}

Besides the chemical novelty associated with bioactive secondary metabolites from the marine habitat, Montaser and Luesch (2011) show that these compounds are also very promising in a different manner, as they offer alternative mechanisms of action. They assert that the recently released marine drugs trabectedin, eribulin mesylate and ziconotide provide noteworthy examples of the unique and complex mechanisms of action of marine. According to their study, trabectedin, an anticancer drug, shows novel mechanism of action through it’s binding to DNA minor groove (leading to damage) and disrupts the transcription-coupled nucleotide repair system. This mechanism has the end effect of causing cell death of cancer cells that show highly accelerated gene expression relative to normal cells.\textsuperscript{124} Much like the anticancer drug trabectedin, eribulin mesylate, another anticancer drug, also exhibits a novel microtubule targeting mechanism of action whereby it collects tubulin and specifically interferes with microtubule growth in a manner, which is different from other drugs like it.\textsuperscript{124} Another example of the novel
mechanisms of action exhibited by marine compounds resides in the potent painkiller ziconotide, which is 1000 times more powerful than morphine, even though it does not cause tolerance like most opiate type treatments. It is the first conotoxin drug to receive approval from the FDA, and it is also the first N-type calcium channel blocker, and it acts through binding as an antagonist to N-type voltage-gated calcium channels to inhibit the activity of pain-sensing primary nociceptors.

According to Gerwick and Moore, as of 2012, there were 7 therapeutic agents which were derived from marine sources, with 13 more agents in either stages I, II, or III of clinical trials. The field of marine natural product chemistry might be a novel pursuit, but it has been at the center of some massive drug discovery stories as discussed below. One of the big success stories of marine drug discoveries resides in the isolation and engineering of ecteinascidin from the tunicate Ecteinascidia turbinata. It took 30 years, and many laboratories to go from the initial discovery of potent anticancer activity in the tunicate, to the isolation of the active compound ecteinascidin. The drug was commercialized by Spanish company Pharmamar, but met with problems due to the inability of natural stocks of tunicate to provide the required supply. This problem was later eradicated when it was found that the fermentation of the bacterium Pseudomonas flourescens, could be used to synthesize a portion of the secondary metabolite known as cyanosafracin B. On cells, ecteinascidin interacts with the minor groove of DNA to interfere with the role of the Nucleotide Excision Repair system to bring about a cytotoxic effect, which makes it useful in the treatment of soft tissue sarcomas and relapsed ovarian cancer, even though it is being explored as a solution to other types of cancer. Another significant member from marine products drug
discovery efforts is of the polyether metabolite halichondrin A from the Japan Sea sponge *Halichondria okadai* by the Uemura laboratory.\textsuperscript{124, 123,122} Halichondrins are very complex structures and powerful cancer cell cytotoxins that function via an antitubulin mechanism. Total chemical synthesis efforts at Harvard, and Kishi laboratory found that only a simplified structure of the halichondrin is required for the compound’s anticancer activity, which is reflected in the final drug Halaven which is being used in the treatment of drug refractory breast cancer.\textsuperscript{123,122} As far as anticancer drug discovery success stories from marine products go, it is impossible to overlook the amazing work of the Pettit group at Arizona State University on one of the more recent additions to anticancer agents derived from marine natural products.\textsuperscript{123, 122} Bob Pettit, and co-workers laid the groundwork for the eventual chemical synthesis of brentuximab vendotin (marketed as Adcetris), which was approved by the FDA for use in Hodgkin’s lymphoma and anaplastic large cell lymphoma in 2011.\textsuperscript{123, 122} Brentuximab vendotin is an antibody that is attached to the powerful antitubulin agent dolastatin 10 via protease-cleavable linkages.\textsuperscript{123}

The antiproliferative properties of dolastatin 10 were evaluated in several clinical studies that all ended up in failures since the substance was too toxic, however, a change of approach in 2010 was finally able to successfully drive dolastatin 10 to selective tubulin targets to unleash the powerful cytotoxic effects found in the conjugated drug Ascentris.\textsuperscript{123, 122} Finally the last example of big drug discoveries from marine products is that of the proteasome inhibitor salinosporamide A from the marine bacterium *Salinispora tropica*.\textsuperscript{123} This case is very unique, since in all 3 examples of drugs derived from marine products above, the metabolic sources of the pharmacologically active
marine natural products were microbes living with macro-organisms or being preyed upon but in this case, the metabolic source is an actual cultured marine microbe.\textsuperscript{122} Fermentation of the marine bacterium was found to produce the highly potent proteasome inhibitor salinosporamide A.\textsuperscript{122} Salinosporamide A acts by binding to the 20S proteasome to inhibit its functions, thereby ultimately leading to cell apoptosis in bortezomib-resistant multiple myeloma cells.\textsuperscript{123, 122} Since the actions of salinosporamide A resemble those of the currently approved drug Velcade, the proteasome inhibitor had been cleared for stage II clinical trials for the treatment of multiple myeloma by Nereus Pharmaceuticals.\textsuperscript{123, 122}

### A2.1.3. Bioactive metabolites from marine derived actinomycetes

Efforts in the search of novel metabolites and antibiotics from terrestrial microorganisms have regressed since the 1980’s because there is the general consensus that there have been extensive and exhaustive studies covering the terrestrial environment as a source for novel actinomycetes.\textsuperscript{125} With that in mind, the search for new sources for bioactive novel metabolites from novel sources for pharmaceutical purposes has been at the core of drug discovery over the last two decades. In that time, the search for potential drug sources has turned to the oceans, which have been demonstrated as novel sources for unique and interesting forms of actinomycetes, some of which require seawater to grow.\textsuperscript{126} This unique adaptation of some marine actinomycetes, along with other distinct qualities of marine actinomycetes has spurred the undertaking of interesting new research


for new species in the marine environment; an effort that has led to the discovery of several vital drugs to date (section 1.3). Many marine actinomycetes can be used to isolate novel new types of secondary metabolites,\(^{127}\) which possess unique biological activities, and a great potential to be developed into therapeutic solutions. This is of great significance, especially since new therapeutic agents are urgently required to fill the niche left by the slightly less than 10,000 diseases that have been clinically described, but cannot be treated symptomatically, or cured for that matter.\(^{125}\) The utilization of marine actinomycetes as a source for biologically or pharmacologically active secondary metabolites may be a novel pursuit, but numerous bioactive metabolites have been isolated from this source over the last decade. Table 1 contains a list of bioactive metabolites that have been isolated from marine actinomycetes from 2003-2010. The list is not exhaustive, but is representative of the structural and biological diversity of bioactive metabolites from marine actinomycetes.

**Table 10:** Novel metabolites produced by marine actinomycetes from 2003-2010.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Source</th>
<th>Biological activity</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinikomycins</td>
<td><em>Streptomyces sp.</em></td>
<td>Anticancer</td>
</tr>
<tr>
<td>Chloro-dihydroquinones</td>
<td>Novel actinomycete</td>
<td>Antibacterial; anticancer</td>
</tr>
<tr>
<td>Glaciapyrroles</td>
<td><em>Streptomyces sp.</em></td>
<td>Antibacterial</td>
</tr>
<tr>
<td>Frigocyclinone</td>
<td><em>Streptomyces griseus</em></td>
<td>Antibacterial</td>
</tr>
<tr>
<td>Lajollamycin</td>
<td><em>Streptomyces nodosus</em></td>
<td>Antibacterial</td>
</tr>
<tr>
<td>Mechercharmycins</td>
<td><em>Thermoactinomyces sp.</em></td>
<td>Anticancer</td>
</tr>
<tr>
<td>Salinosporamide A</td>
<td><em>Salinispora tropica</em></td>
<td>Anticancer; antimalarial</td>
</tr>
<tr>
<td>Salinosporamide B &amp; C</td>
<td><em>Salinispora tropica</em></td>
<td>Cytotoxicity</td>
</tr>
<tr>
<td>2-Allyloxyphenol</td>
<td><em>Streptomyces sp.</em></td>
<td>Antimicrobial; food preservative; oral disinfectant</td>
</tr>
<tr>
<td>Saliniketal</td>
<td><em>Salinispora arenicola</em></td>
<td>Cancer chemoprevention</td>
</tr>
<tr>
<td>Marinomycins A – D</td>
<td><em>Marinispora</em></td>
<td>Antimicrobial; anticancer</td>
</tr>
<tr>
<td>Lodopyridone</td>
<td><em>Saccharomonospora sp.</em></td>
<td>Anticancer</td>
</tr>
<tr>
<td>Arenimycin</td>
<td><em>Salinospora arenicola</em></td>
<td>Antibacterial; anticancer</td>
</tr>
<tr>
<td>Salinispyrones A &amp; B</td>
<td><em>Salinospora pacifica</em></td>
<td>Mild cytotoxicity</td>
</tr>
<tr>
<td>Pacificanones A &amp; B</td>
<td><em>Salinospora pacifica</em></td>
<td>Antibacterial</td>
</tr>
<tr>
<td>Arenicolides A-C</td>
<td><em>Salinospora arenicola</em></td>
<td>Mild cytotoxicity</td>
</tr>
<tr>
<td>Resistoflavimethyl ether</td>
<td><em>Streptomyces sp.</em></td>
<td>Antibacterial; anti-oxidative</td>
</tr>
<tr>
<td>Staurosporinone</td>
<td><em>Streptomyces sp.</em></td>
<td>Antitumor; phycotoxicity</td>
</tr>
<tr>
<td>1,8-Dihydroxy-2-ethyl-3-Methylantraquinone</td>
<td><em>Streptomyces sp.</em></td>
<td>Antitumor Methylantraquinone</td>
</tr>
<tr>
<td>Caboxamycin</td>
<td><em>Streptomyces sp.</em></td>
<td>Antibacterial; anticancer</td>
</tr>
<tr>
<td>Compound</td>
<td>Source</td>
<td>Biological activity</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Butenolides</td>
<td><em>Streptoverticillium luteoverticillatum</em></td>
<td>Antitumor</td>
</tr>
<tr>
<td>Daryamides</td>
<td><em>Streptomyces sp.</em></td>
<td>Antifungal ; anticancer</td>
</tr>
<tr>
<td>Piericidins</td>
<td><em>Streptomyces sp.</em></td>
<td>Antitumor</td>
</tr>
<tr>
<td>Proximicins</td>
<td><em>Verrucosispora sp.</em></td>
<td>Antibacterial ; anticancer</td>
</tr>
<tr>
<td>Streptokordin</td>
<td><em>Streptomyces sp.</em></td>
<td>Antitumor</td>
</tr>
<tr>
<td>ZHD-0501</td>
<td><em>Actinomadura sp.</em></td>
<td>Anticancer</td>
</tr>
<tr>
<td>Tirandamycins</td>
<td><em>Streptomyces sp.</em></td>
<td>Antibacterial</td>
</tr>
<tr>
<td>Abyssomicins</td>
<td><em>Verrucosispora sp.</em></td>
<td>Antibacterial</td>
</tr>
<tr>
<td>Aureoverticillactum</td>
<td><em>Streptomyces aureoverticillactus</em></td>
<td>Anticancer</td>
</tr>
<tr>
<td>Bonactin</td>
<td><em>Streptomyces sp.</em></td>
<td>Antibacterial; antifungal</td>
</tr>
<tr>
<td>Caprolactones</td>
<td><em>Streptomyces sp.</em></td>
<td>Anticancer</td>
</tr>
<tr>
<td>Chandranamimycins</td>
<td><em>Actinomadura sp.</em></td>
<td>Antialga; anticancer;anti-inflammatory</td>
</tr>
<tr>
<td>3,6-disubstituted indoles</td>
<td><em>Streptomyces sp.</em></td>
<td>Anticancer</td>
</tr>
<tr>
<td>IB-00208</td>
<td><em>Actinomadura sp.</em></td>
<td>Anticancer</td>
</tr>
<tr>
<td>Komodoquinone A</td>
<td><em>Streptomyces sp.</em></td>
<td>Neuritogenic activity</td>
</tr>
<tr>
<td>Gutingimycin</td>
<td><em>Streptomyces sp.</em></td>
<td>Antibacterial</td>
</tr>
<tr>
<td>Himalomycins</td>
<td><em>Streptomyces sp.</em></td>
<td>Antibacterial</td>
</tr>
<tr>
<td>L-glutaminase</td>
<td><em>Streptomyces olivochromogenes</em></td>
<td>Enzymatic activity</td>
</tr>
<tr>
<td>Alkaline Potease</td>
<td><em>Streptomyces sp.</em></td>
<td>Enzymatic activity</td>
</tr>
<tr>
<td>Xiamycin</td>
<td><em>Streptomyces sp.</em></td>
<td>Anti-HIV activity</td>
</tr>
</tbody>
</table>
From Lam (2006), Subramani and Aalbersberg (2012), and Amrita et al., (2014)\textsuperscript{125, 127, 128}


S.; Bandeira, N.; Palsson, B. O.; Pogliano, K.; Gutierrez, M.; Dorrestein, P. C.
MS/MS Networking Guided Analysis of Molecule and Gene Cluster Families.

46. Bocchi, V.; Casnati, G.; Marchelli, R. Insertion of Isoprene Units into Indole and

47. Wollinsky, B.; Ludwig, L.; Xie, X.; Li, S.-M. Breaking The Regioselectivity of
Indole Prenyltransferases: Identification of Regular C3-Prenylated
Hexahydropyrrolo 2,3-b Indoles as Side Products of The Regular C2-

48. Magnuson, R.; Solomon, J.; Grossman, A. D. Biochemical and Genetic-

49. Okada, M.; Sato, I.; Cho, S. J.; Dubnau, D.; Sakagami, Y. Chemical Synthesis of
ComX Pheromone and Related Peptides Containing Isoprenoidal Tryptophan

50. Gondry, M.; Sauguet, L.; Belin, P.; Thai, R.; Amouroux, R.; Tellier, C.; Tuphile,
Lecoq, A.; Hashimoto, S.-I.; Genet, R.; Pernodet, J.-L. Cyclodipeptide Synthases

51. (a) Aravind, L.; de Souza, R. F.; Iyer, L. M. Predicted Class-I Aminoacyl tRNA
Synthetase-Like Proteins in Non-Ribosomal Peptide Synthesis. *Biology Direct*


62. James E.; Alqahtani, N.; Viswanathan R.; Lane A. *Manuscript on prepare*


Activity of FkbP, the FK520 Pipecolate-Incorporating Enzyme.

*Biochemistry, 2005, 44, 5993–6002.*


77. Chang-Hong, P.; Matsuzaki, K.; Ikeda, H.; Tanaka, H.; Omura, S. Production of 6, 8a-seco-6, 8a-deoxy Derivatives of Avermectins by a Mutant Strain of *Streptomyces avermitilis.* *J. Antibiot. 1995, 48, 92-94.*


95. Pangborn, A. B.; Giardello, M. A.; Grubbs, R. H.; Rosen, R. K. Timmers, F. J. 
Safe and Convenient Procedure for Solvent Purification. *Organometallics*, 1996, 
15, 1518-1520.

96. Armarego, W. L. F.; Chai, C. L. L.; *Purification of Laboratory Chemicals*, 5th Ed. 

97. Still, W. C.; Kahn, M.; Mitra, A. Rapid chromatographic technique for 

98. (a). Gottileb, H. E.; Kotlyar, V.; Nudelman, A. NMR Chemical Shifts of Common 
Laboratory Solvents as Trace Impurities *J. Org. Chem.* 1997, 62, 7512-7515 and 
(b). Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottileb, H. E.; Nudelman, 
A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. NMR Chemical Shifts of Trace 
Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated 
Solvents Relevant to the Organometallic Chemist *Organometallics*, 2010, 29, 
2176-2179.

99. Note: Synthetic targets assembled in this study as putative intermediates in 
nocardioazine biosynthetic pathway are denoted by *.

100. Shao, Y.-M.; Yang, W.-B.; Peng, H.-P.; Hsu, M.-F.; Tsai, K.-C.; Kuo, T.-H.; 

101. (a) Isaacs, N. S. and Coulson, M. Effect of pressure on processes 
D.; Enzensperger, C.; AbulAzm, S. E.; Hefnawy, M. M.; El-Subbagh, H. I.; Wani,


105. A considerable asymmetry is noticeable in the HPLC traces and we attribute this to the fact that 5 and ent-5 are polar entities and show significant “tailing” effect on a normal phase chiral AS column under the conditions observed.


118. Leal, C. M.; Puga, J.; Serodio, J.; Gomez, C. M. N.; Calado, R. Trends in the Discovery of New Marine Natural Products from Invertebrates over the Last Two Decades – Where and What Are We Bioprospecting?. *Plos One.* 2012, 7.


