INVESTIGATION OF THE DISCOIPYRROLES AND OTHER MARINE BACTERIA

DERIVED NATURAL PRODUCTS

APPROVED BY SUPERVISORY COMMITTEE

John MacMillan, PhD

Richard Bruick, PhD

Deepak Nijhawan, M.D., PhD

Joseph Ready, PhD

INVESTIGATION OF THE DISCOIPYRROLES AND OTHER MARINE BACTERIA DERIVED NATURAL PRODUCTS

by

DOMINIC ANDREW COLOSIMO

DISSERTATION / THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2017

Copyright

by

Dominic Andrew Colosimo, 2017

All Rights Reserved

INVESTIGATION OF THE DISCOIPYRROLES AND OTHER MARINE BACTERIA DERIVED NATURAL PRODUCTS

Publication No. _____

Dominic Andrew Colosimo, PhD

The University of Texas Southwestern Medical Center at Dallas, 2017

Supervising Professor: John B. MacMillan, PhD

Natural products are a rich source of scientific innovation. These chemical compounds are canonically celebrated as biomedical tools or synthetic chemistry feats. Discovery of new chemical compounds has benefited from the study of natural product biosynthesis, or the methods by which they are constructed. Additionally, these studies have advanced the fields of ecology, industrial chemistry, and protein biophysics. This work will demonstrate how the

discoipyrrole family of natural products inspired novel biomedical and biosynthetic discovery. In particular, their biosynthesis features unique mechanisms that are independent of protein catalysts, a growing trend found in natural products. Unrelated to the discoipyrroles, the manipulation of biosynthesis pathways using the methodology known as ribosome engineering will be discussed. This work demonstrates how the discovery and development of bioactive natural products can be enhanced by biosynthetic interrogations.

TABLE OF CONTENTS

PAGE

PRIOR PUBLICATIONS ix

LIST OF FIGURES x

LIST OF TABLES xv

LIST OF SCHEMES xvi

LIST OF APPENDICES xvii

CHAPTER ONE 1

NON-ENZYMATIC BIOSYNTHESIS OF NATURAL PRODUCTS....1

1.1 Natural products from afar *1*

1.1.1 Brief history of natural products 1

1.1.2 Biomedical importance of natural products 2

1.2 Natural products biosynthesis *2*

1.2.1 Impetus to study natural products biosynthesis 2

1.2.2 Foundation of natural product biosynthesis 3

1.3 Non-enzymatic natural products biosynthesis *5*

1.3.1 Current knowledge of non-enzymatic reactions 5

1.3.2 Incomplete categorization of non-enzymatic processes in natural products 7

1.4 Categories of non-enzymatic mechanisms *9*

1.4.1 Previously discussed categories of non-enzymatic biosynthetic reactions 9

1.4.2 Intramolecular reactions 10

1.4.3 Multi-component reactions 14

1.4.4 Homologous multi-component reactions 14

1.4.5 Heterologous multi-component reactions 33

1.4.6 Tailoring reactions 41

1.4.7 Light induced reactions 47

1.5 Non-biologically relevant reactions *55*

1.6. Conclusion *56*
1.6.1 Discussion of common precursors 56
1.6.2 Final remarks 59

CHAPTER TWO 61

MECHANISTIC STUDIES OF THE DISCOIPYRROLE BIOSYNTHESIS 61
2.1 The discoipyrroles: novel marine natural products *61*
2.2 Biosynthetic studies of the marine bacteria Bacillus hunanensis *61*
2.2.1 Origins of metabolites secreted by SNA-048 63
2.2.2 Development of a biomimetic synthetic system 68
2.2.3 Establishing method for examining reaction in real time 69
2.2.4 Tracking reaction intermediates using enriched NMR-active isotopes 71
2.2.5 Examining spontaneous oxidation of pyrrole core 72
2.2.6 Determination of early reaction intermediates 74
2.2.7 Conclusion 78
2.2.8 Future directions 79
2.3 Experimental *80*

CHAPTER THREE 87

EXPANDING NATURAL PRODUCT LIBRARY CHEMICAL DIVERSITY 87
3.1 Cryptic biosynthesis *87*
3.2 Ribosomal engineering *88*
3.3 Ribosome engineering of Streptomyces rubrogriseus sp. SNC-031 *91*
3.4 Ribosome engineering optimization of the MacMillan microbe library *98*
3.4.1 Ribosomal engineering pipeline in the MacMillan lab 99
3.4.2 Testing of MacMillan laboratory ribosomal engineering pipeline 104
3.4.3 Genetic sequencing of resistant mutants 109
3.4.4 Metabolic analysis of resistant mutants 112
3.5 Conclusion *116*
3.6 Experimental *118*

CHAPTER FOUR 124

BIOLOGICAL STUDIES OF THE DISCOIPYRROLES 124

4.1 Background *124*

4.1.1 A brief history of the discoidin domain containing receptors 124

4.1.2 Targeting DDRs in the context of non-small cell lung cancer 126

4.1.3 Connection of the discoidin domain-containing receptors to the discoipyrroles 126

4.3 Exploration of discoipyrroles cytological effects in vitro *127*

4.3.1 Toxicity profile of the discoipyrroles in NSCLC cell lines 127

4.3.2 Cytostatic effect of the DPs 128

4.3.3 DPs cause irregular cell morphology and affect migration 131

4.3.4 Assessment of DPs role in collagen associated pathways 133

4.5 Exploration of the molecular target of the discoipyrroles *136*

4.5.1 Affinity based probes to attempt molecular target identification 136

4.5.2 Pipeline for preliminary target validation studies 139

4.5.3 Visualization of DP probes in lysate and fixed cells 140

4.5.4 Discussion of proteome labeling limitations 143

4.5.5 Gene regulation in response to discoipyrrole treatment 144

4.5.6 GPCR screening of the discoipyrroles 147

4.6 Conclusion *149*

4.7 Experimental *150*

BIBLIOGRAPHY 178

PRIOR PUBLICATIONS

1. Colosimo, D. A.; MacMillan, J. B., Ammosamides Unveil Novel Biosynthetic Machinery. Cell Chem Biol 2016, 23 (12), 1444-1446. PMID: 28009976

2. Colosimo, D. A.; MacMillan, J. B., Detailed Mechanistic Study of the Non-enzymatic Formation of the Discoipyrrole Family of Natural Products. Journal of the American Chemical Society 2016, 138 (7), 2383-8. PMID: 26824832

3. Oyer, J. L.; Igarashi, R. Y.; Kulikowski, A. R.; Colosimo, D. A.; Solh, M. M.; Zakari, A.; Khaled, Y. A.; Altomare, D. A.; Copik, A. J., Generation of highly cytotoxic natural killer cells for treatment of acute myelogenous leukemia using a feeder-free, particle-based approach. Biology of blood and marrow transplantation: Journal of the American Society for Blood and Marrow Transplantation 2015, 21 (4), 632-9. PMID: 25576425

4. Hu, Y.; Potts, M. B.; Colosimo, D.; Herrera-Herrera, M. L.; Legako, A. G.; Yousufuddin, M.; White, M. A.; MacMillan, J. B., Discoipyrroles A-D: isolation, structure determination, and synthesis of potent migration inhibitors from Bacillus hunanensis. Journal of the American Chemical Society 2013, 135 (36), 13387-92. PMID: 23984625

LIST OF FIGURES

Figure 1. Possible types of non-enzymatic biosyntheses 4

Figure 2. Non-enzymatic biosynthesis of the endiandric acids 11

Figure 3. Elansolid family of natural products. 12

Figure 4. Non-enzymatic biosynthesis of elansolids A1 and A2 13

Figure 5. Carpanone biosynthetic hypothesis and biomimetic validation 16

Figure 6. Non-enzymatic biosynthesis of oxazinin A 17

Figure 7. Progression of the mechanistic hypothesis of paracaseolide A 18

Figure 8. Securinega family of plant alkaloids. 21

Figure 9. Proposed biosynthetic mechanism of flueggenine. 22

Figure 10. Synthetic attempts of securinega dimerization 24

Figure 11. Biosynthetic proposal for homodimericin A. 27

Figure 12. Sorbicillinoid monomers and dimers 28

Figure 13. Symmetric dimers from various species 31

Figure 14. Proposed biosynthetic mechanism of dibohemamine A 32

Figure 15. Non-enzymatic biosynthesis of jadomycins A and B 35

Figure 16. Structure and proposed biosynthesis of the rubrolones 37

Figure 17. Biosynthetic mechanism of the rubrolones 38

Figure 18. Maximiscin and pseudomaximiscin biosynthesis 40

Figure 19. Biosynthetic mechanism of maximiscin 41

Figure 20. Non-enzymatic derivatization of the ammosamides 43

Figure 21. Biosynthetic proposal of juanlimycins A and B. 44

Figure 22. Progression of the stealthin biosynthetic hypothesis 46

Figure 23. The γ-pyrone natural products and biomimetic studies 49

Figure 24. Proposed biosynthesis of phototridachiahydropyrone 50

Figure 25. Biosynthetic hypothesis for pseudopterane and validation using bipinnatin J 52

Figure 26. Intricarene biomimetic studies 53

Figure 27. Biomimetic studies of aplydactylone 54

Figure 28. Anthranilic acid in non-enzymatic biosyntheses 57

Figure 29. Potential biological applications of non-enzymatic reactions 59

Figure 30. Compounds produced by SNA-048 62

Figure 31. Anthranilic acid biogenesis 64

Figure 32. Acyloin containing amino acid dimers 65

Figure 33. Biosynthesis of 4-hydroxysattabacin from amino acids 66

Figure 34. Quantification of 4-hydroxybenzaldehyde in SNA-048 crude extract 67

Figure 35. Model system of DP biosynthesis 69

Figure 36. Chemical shifts of ^{15}N NMR 70

Figure 37. Reaction timeline of DP formation 71

Figure 38. MS traces of reactions performed under $^{18}O_2$ or $^{16}O_2$ 74

Figure 39. Proposed biosynthetic mechanism of the discoipyrroles 75

Figure 40. Aldimine formation based on aniline substitution 76

Figure 41. Analysis of Pathway 3 77

Figure 42. DP synthesis carried out in deuterated sea water 79

Figure 43. Antibacterial effect of SNC031 wild-type and mutants 92

Figure 44. 16s rRNA sequences of SNC031 mutants and rpoβ protein comparison 93

Figure 45. Metabolomes of SNC-031 and mutants 94

Figure 46. Compounds isolated from SNC-031 mutants 95

Figure 47. Piperazimycin A structure and mammalian cytotoxicity 96

Figure 48. Comparison of piperidamycin and piperazimycin induction 97

Figure 49. Pipeline for MacMillan lab ribosome engineering 101

Figure 50. Acquired resistance to rifampicin (L-R: WT, M1, M2) 108

Figure 51. rpoβ sequences from SNF-022 wild-type and resistant clones 111

Figure 52. Crude extracts of interest in List 1 (280 nm) 113

Figure 53. Comparison of Erythrobacter metabolomes 114

Figure 54. SNC-119 crude extracts (280 nm) 115

Figure 55. Initial toxicity assays (IC_{50}) and structures of the discoipyrroles. 128

Figure 56. Structures and growth inhibition (IC_{50}) of DP analogs 128

Figure 57. Time course of DP treatment 129

Figure 58. Short term toxicity of DPs 130

Figure 59. Flow cytometry analysis of DP treated cells 131

Figure 60. Phenotypic effects of DP treatment on A549 132

Figure 61. Confocal microscopy of treated A549 cells 133

Figure 62. Collagen effects on actin skeleton and DP growth inhibition 134

Figure 63. Migration assays of A549 cells 135

Figure 64. Assessment of EMT in A549 cells treated with DPs 136

Figure 65. Conceptual overview of affinity based target identification 137

Figure 66. Possible binding modes of DP to a protein target 138

Figure 67. Various alkynylated DP analog probes 139

Figure 68. Overview of proteomic visualization 140

Figure 69. SDS-PAGE used to view proteome labeling 141

Figure 70. Fluorescence confocal microscopy of probe treated cells 142

Figure 71. Wash assays of SW212917 143

Figure 72. The ^{1}H-NMR spectrum of compound 42 158

Figure 73. The 13C-NMR spectrum of compound 42 159

Figure 74. The COSY-NMR spectrum of compound 42 160

Figure 75. The HSQC-NMR spectrum of compound 42 161

Figure 76. The HMBC-NMR spectrum of compound 42 162

Figure 77. The ^{1}H-^{15}N HMBC NMR spectrum of compound 42 163

Figure 78. SNA-076 crude extracts (210 nm) 164

Figure 79. SNA-076 crude extracts (280 nm) 164

Figure 80. SNA-079 crude extracts (210 nm) 165

Figure 81. SNC-102 crude extracts (210 nm) 165

Figure 82. SNE-016 crude extracts (210 nm) 166

Figure 83. SNA-053 crude extracts (210 nm) 166

Figure 84. SNF-022 crude extracts (210 nm) 167

Figure 85. SNA-079 WT (red) and M3 (blue) 167

Figure 86. Downfield 1H NMR of SNF-022 WT, M2, M3 (bottom to top) 168

Figure 87. 1H NMR of SNF-022 WT, M2, M3 (bottom to top) 168

Figure 88. Upfield ^{1}H NMR of SNF-022 WT, M2, M3 (bottom to top) 169

Figure 89. SNB-017 crude extracts (210 nm) 169

Figure 90. SNB-035 crude extracts (210 nm) 170

Figure 91. SNB-054 crude extract (210 nm) 170

Figure 92. SNC-041 crude extract (210 nm) 171

Figure 93. SNC-066 crude extracts (210 nm) 171

Figure 94. SNE-029 crude extracts (210 nm) 172

Figure 95. SNF-015 crude extracts (210 nm) 172

Figure 96. ^{1}H NMR spectra of SNC-119 WT and M2 173

Figure 97. SNA-073 crude extracts (210 nm) 173

Figure 98. SNC-012 crude extracts (210 nm) 174

Figure 99. SNC-090 crude extracts (210 nm) 174

Figure 100. SNE-037 crude extracts (210 nm) 175

Figure 101. SNE-038 crude extracts (210 nm) 175

Figure 102. SNE-046 crude extracts (210 nm) 176

LIST OF TABLES

Table 1. List 1 of candidate bacteria for ribosome engineering 105

Table 2. List 2 of candidate bacteria for ribosome engineering 106

Table 3. List 3 of candidate bacteria for ribosome engineering 107

Table 4. Phenotypic resistance development of the three candidate groups. 108

Table 5. Summary of 16s rRNA sequencing 110

Table 6. Pathway analysis of DP treated cells 146

Table 7. GPCR panel results 148

Table 8. Antibodies List 155

Table 9. NMR table for compound 42 157

Table 10. Primer list for mutant rpoβ sequencing 177

LIST OF SCHEMES

Scheme 1. Aldimine formation 84

Scheme 2. Pathway 2 interrogation 85

Scheme 3. Pathway 2 interrogation with base 85

Scheme 4. Pathway 3 interrogation 85

Scheme 5. Pathway 1 interrogation 86

Scheme 6. Imine trapping 86

LIST OF APPENDICES

Appendix 1. Chapter 2 157

Appendix 2. Chapter 3 164

LIST OF DEFINITIONS

NP	natural product
RNA	ribonucleic acid
BGC	biosynthetic gene cluster
YMC	yeast metabolic cycle
PKC	polyketide synthase
AT	acetyltransferase
DFT	density functional theory
RC	Rauhut-Currier
MBH	Morita-Baylis-Hillman
UV	ultraviolet
TBAF	tert-n-butylammonium fluoride
THF	tetrahydrofuran
NMR	nuclear magnetic resonance
IMDA	intramolecular Diels-Alder
SAM	S-adenosyl methionine
PFBHA	pentafluorobenzyl hydroxyl amine
RuMP	ribulose monophosphate
DP	discoipyrrole
OSMAC	one strain many conditions
NRPS	non-ribosomal peptide synthase

ROS	reactive oxygen species
RNS	reactive nitrogen species
RSS	reactive sulfur species
JACS	Journal of the American Chemical Society
FuSiOn	functional signature ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
TPP	thiamine pyrophosphate
MS	mass spectrometry
SIM	single ion monitoring
HMBC	heteronuclear multiple bond correlation
COSY	homonuclear correlation spectroscopy
HSQC	heteronuclear single quantum correlation
DMSO	dimethyl sulfoxide
LC/MS	liquid chromatography mass spectroscopy
DMAP	4-dimethylaminopyridine
HDAC	histone deacetlyase
ATP	adenosine triphosphate
GTP	guanosine triphosphate
pppGpp	guanosine pentaphosphate
ppGpp	guanosine tetraphosphate
rpoβ	RNA polymerase β subunit
RelA	GTP pyrophosphokinase

Ef-Tu	elongation factor thermos unstable
RPM	revolutions per minute
NSCLC	non-small cell lung cancer
siRNA	small interfering RNA
DDR	discoidin domain containing receptor
RTK	receptor tyrosine kinase
EGFR	epidermal growth factor receptor
SHP2	SH2 domain containing tyrosine phosphatase 2
SCC	squamous cell carcinoma
EMT	epithelial mesenchymal transition
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
DNA	deoxyribonucleic acid
GPCR	G-protein coupled receptor
IFN-α	interferon alpha
IFN-β	interferon beta
CXCR4	C-X-C Motif Chemokine Receptor 4
CXCL12	C-X-C Motif Chemokine Ligand 12
MIF	migration inhibitory factor
CALCRL-RAMP1	calcitonin receptor-like receptor
PTGIR	prostacyclin receptor

CHAPTER ONE

NON-ENZYMATIC BIOSYNTHESIS OF NATURAL PRODUCTS

1.1 Natural products from afar

1.1.1 Brief history of natural products

Over the last century, the frontline treatments for human diseases have been chemical agents of high purity, broadly described as drugs. However, before the process of discovering, testing, and ultimately prescribing drugs became commonplace, many ailments were treated by using impure mixtures of chemicals coming from rudimentary solvent extractions of biological sources, most frequently plants. Even as late as 2005, 80% of the world's inhabitants depend on plant based natural therapeutics for their healthcare, as demonstrated by the World Health Organization (Cragg Gordon & Newman David, 2005). Conservation of these remedies over time has been driven by empirically demonstrated utility, such as the analgesic property of morphine from the poppy *Papaver somniferum*. The most useful concoctions were passed on by primitive populations in written or oral form. Analytical chemistry in the 19th and 20th century provided tools such as chromatography and spectrophotometry to examine many of the complicated plant extracts passed on from earlier generations. Now we know that many of the biological effects of popular plant extracts were driven by a single molecule, or family of closely related molecules known as natural products. The chemical complexity of these molecules demonstrated the chemical power that had evolved over time by these organisms. The process by which these complex molecules were made became known as biosynthesis and the study of

biosyntheses begged a question that has since dominated the fields of biomedical focused chemistry: can we find molecules evolved by nature to modulate biological systems?

1.1.2 Biomedical importance of natural products

Natural products (NPs) have become an integral component of modern chemistry. Research of NPs commonly focus on the discovery, chemical synthesis, or biological characterization of these molecules. These combined efforts have been fruitful for the medicinal industry of the United States, with natural products, natural products based compounds, and natural product mimics together making up 36% of the FDA approved drugs from 1981-2014 (Newman & Cragg, 2016). The most common use of natural products is as anti-infective or anti-cancer therapeutics, with 73% and 83% of the small molecule approved drugs since 1981 being natural products or compounds derived from natural products, respectively. In addition to the sheer numbers of natural products used in treatment of human disease, "workhorse" therapeutics that serve as the last line of defense against certain diseases have been derived from nature. This was recently demonstrated by the awarding of the 2015 Nobel Prize for Physiology and Medicine to Drs. Omura, Campbell, and Tu for their role in discovering the anti-parasitic compounds avermectin and artemisinin, which were derived from bacteria and plants respectively. These compounds have been vital in the fight against several neglected tropical diseases, which are not traditionally profitable in the pharmaceutical world (McKerrow, 2015).

1.2 Natural products biosynthesis

1.2.1 Impetus to study natural products biosynthesis

The utility of studying biosynthesis is intuitive to natural products chemists. Recently, the broader value of natural product biosynthetic studies has been emphasized by two major factors.

First, a published data-mining effort suggested that basic science research, as biosynthetic projects may be classified, is as effective as applied science in promoting technological innovations (D. Li, Azoulay, & Sampat, 2017). This advocates that knowledge of natural product producing mechanisms can aid in the development of drugs and chemical tools. Second, through genome sequencing, researchers have demonstrated that microbes grown in the laboratory commonly do not express their full potential of biosynthetic machinery (Bentley et al., 2002; Rutledge & Challis, 2015). This suggests that understanding the intricacies of biosynthetic regulation is crucial for discovering new chemistry. These two factors create a necessity for in-depth studies of how microbes synthesize complex molecules that have the potential to be scientifically relevant and whose nature is not always obvious from superficial experimental analysis. A facet of natural products whose study has been beneficial particularly to synthetic and medicinal chemistry is the presence of non-enzymatic reactions within biosynthetic pathways, which will be discussed in this review.

1.2.2 Foundation of natural product biosynthesis

The anabolic pathways used to derive natural products can best be explained in context of the central dogma of molecular biology, first detailed by Francis Crick (Crick, 1970). Genes are transcribed to produce RNA which is translated into proteins that perform catalytic functions to covalently link basic metabolic substrates made *de novo* or taken up from the environment (Figure 1).

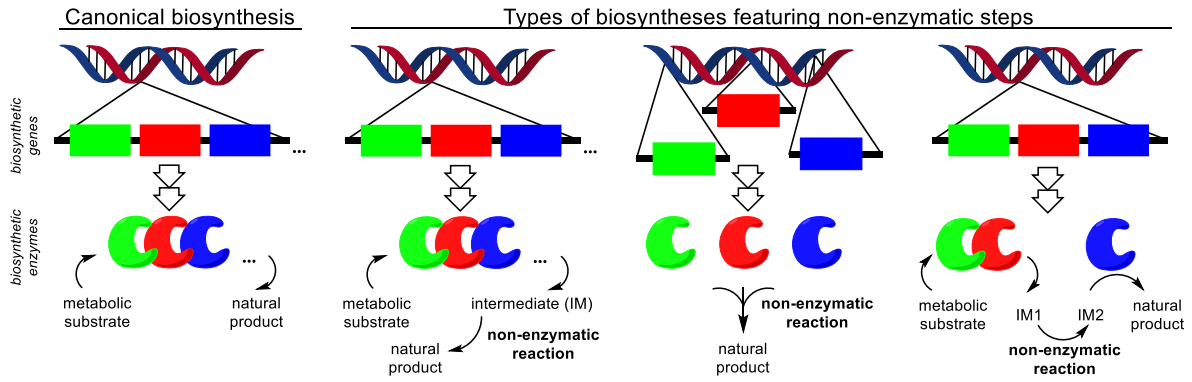


Figure 1. Possible types of non-enzymatic biosyntheses

As genes are the blueprint for functional proteins within the cell, their importance in biosynthesis does not need overstating. Early work in the 1980's by D.A. Hopwood on the actinomycete *Streptomyces coelicolor A3(2)* showed that secondary metabolite production and sporulation, both hallmarks of defense, were dependent on each other for their function (Hopwood, 1988). This was one of the earliest pieces of evidence to suggest that microbial defense tactics were part of a dedicated system intimately tied to each other, and thus genes involved in any such mechanism were likely to be spatially clustered on the genome (Hopwood, 1988). We now know that in the vast majority of studied cases, genes responsible for the production of natural products lie in contiguous sequences, known as biosynthetic gene clusters (BGCs) and frequently are accompanied by associated genes involved in active transport, resistance, and regulation (Jensen, 2016).

The functional unit of biosynthetic machinery are the proteins responsible for carrying out the catalytic steps to bond chemical substrates. Most commonly, after translation these proteins form large multimeric complexes that serve as localized NP factories. Since the advent of cloning, molecular biology has been used extensively to study these enzymes *in vivo* and *in vitro* using reconstituted systems. Studying natural products from diverse families, the scientific

community has built a strong foundation for how these enzymes perform secondary metabolism. However, unveiling a truly complete picture of microbial enzymatic machinery is a long and complicated process that will grow in complexity as researchers deviate from canonical protein scaffolds.

Certainly these protein-centered studies will cover the large majority of the global secondary metabolome, but there are exceptions to these protein-centric view that deserve attention. Specifically, the use of protein-independent mechanisms for covalent bond construction in the cell is a recently exemplified process in biosynthesis that has been well established in other areas of biology for years.

1.3 Non-enzymatic natural products biosynthesis

1.3.1 Current knowledge of non-enzymatic reactions

In the packed environment of a cell, of any type, there are chemical reactions that may occur in the absence of enzymes due to inherent chemical reactivity and physical proximity. These reactions, known as non-enzymatic reactions, have been studied particularly in the context of primary metabolism. Recently these types of chemical reactions in microbes (de Lorenzo, Sekowska, & Danchin, 2015) and humans (Keller, Piedrafita, & Ralser, 2015; L. Shi & Tu, 2015) have been thoroughly reviewed. Additionally, a biophysical explanation of how their utility is amplified by cellular compartmentalization has recently been discussed (Weiss, 2014). Finally, the Balskus laboratory recently wrote an intriguing opinion on how to manipulate non-enzymatic reactions within primary metabolism to study the cellular biology of microbes (Wallace, Schultz, & Balskus, 2015). While the details of these reviews will not be discussed

here, they establish a role of non-enzymatic chemistry as an important intermediate in the cross-talk between primary metabolic flux and cell functioning.

A single example of the biological impact of non-enzymatic chemistry is the role of acetyl-CoA, whose levels fluctuate during the yeast metabolic cycle (YMC) as a result of energy consumption. Acetyltransferases are responsible for carrying out specific acetylation during these cycles to regulate gene expression base on acetyl-CoA pools (Cai & Tu, 2011). However, non-enzymatic acetylation has been shown to be widespread across the proteome, particularly in the mitochondria and to a lesser extent the nucleus (Kim et al., 2006). These modifications have been shown to be either deleterious or advantageous for protein activity, reflecting the ability for non-enzymatic reactions to affect cellular function (Hallows, Yu, & Denu, 2012). Finally, genetic regulation of the sirtuin deacetylases is intimately tied to these pools of acetyl-CoA demonstrating a direct pathway between primary metabolism and gene expression through non-enzymatic chemistry (Wagner & Hirschey, 2014).

The production of natural products in ecological niches is intimately tied to organism health (Giordano et al., 2015). Given that non-enzymatic reactions are a known communication line between primary metabolism and cellular function as mentioned above, it is possible that secondary metabolism can utilize non-enzymatic reactions in a conserved manner to improve cell survival.

Almost 10 years ago Gravel and Poupon wrote a micro-review detailing many non-enzymatic mechanisms in natural products biosynthesis (Gravel & Poupon, 2008), particularly focused on the justification of these reactions using biomimetic total syntheses. Since then, a growing body of literature has accumulated which highlight non-enzymatic steps in secondary

metabolism. These studies allow trends in substrates and mechanisms to be identified, but also beg the question: are these reactions biologically relevant? In this review, I will continue to characterize chemical trends in non-enzymatic reactions while focusing on their potential for biological relevance. I hope to provide the reader chemical and biological details of non-enzymatic mechanisms in natural products biosyntheses so that direct approaches in the future will be taken to support or contradict the argument of physiological function of non-enzymatic biosynthesis.

1.3.2 Incomplete categorization of non-enzymatic processes in natural products

The history of non-enzymatic chemistry in natural products biosynthesis is convoluted due to inconsistent terminology and experimental details reported over the greater part of the last century. Commonly, when a new natural product is discovered a biosynthetic hypothesis is made in an attempt to support a logical path to the proposed structure. A lack of molecular biology efforts in chemo-centric natural products laboratories have historically prevented a thorough experimental process for publication. This works against the accurate definition of non-enzymatic chemistry two fold.

First, authors could describe biosynthetic processes with no precedent or obvious catalyst as being spontaneous without thorough exploration for a possible enzyme. Commonly, identification of a racemic mixture of isomers and sensible arrow pushing is used to justify such claims, as these are a necessary first step for exploring non-enzymatic reactions. Reactive intermediates are often unstable or produced in low titers (Colosimo & MacMillan, 2016; Czechowski et al., 2016), therefore the difficulty of these experiments may deter interrogation.

These technical issues can result in a lack of biosynthetic validation and thus mislabelled non-enzymatic mechanisms.

In direct contrast, mislabelling can occur if authors describe a process as being enzymatically driven without vetting of the proposed BGC. When gene clusters are identified, it is accepted to predict the role of each gene product based on sequence alignment to previously studied biosyntheses. This is a powerful practice that is the conceptual basis for the natural product community to identify BGC in microbial genomes (Monroe et al.). However, sequence similarity does not always accurately predict the output of a biosynthetic pathway, as evident by the continued discoveries of unique biosynthetic gene clusters(Jordan & Moore, 2016). Rigorous studies are needed to confirm biosynthetic gene cluster predictions. Fortunately, this process is aided by the increased sophistication of scientific tools which has lowered the technical threshold for multi-disciplinary laboratories.

In addition to the lack of experimental validation, misrepresentation has also worked against the accurate description of non-enzymatic processes. The definition of non-enzymatic biosynthesis lies on a razors edge between a purposeful function and an artifact of scientific intervention. The authors argue, as described with examples later, that there is a distinction between non-enzymatic chemistry that is functional in the context of an organism and non-enzymatic chemistry that occurs serendipitously due to biologically irrelevant conditions. While this distinction is difficult to parse, it remains an important characterization of any non-enzymatic reaction. At the same time, identification of a non-enzymatic reaction, regardless of its characterization as artificial or functional, can further scientific endeavours particularly in the

case of medicinal chemistry. There lies a responsibility to explore the biological context without ignoring the chemical potential.

Finally, a fundamental component inhibiting the growth of studying non-enzymatic mechanisms is the disjointed language used to describe said reactions. Inconsistent terms used to label these reactions include domino, spontaneous, cascade, concerted, chemoreactive, non-proteogenic, and more. Accurately and wholly identifying potential non-enzymatic mechanisms has thus been difficult. Hopefully, with this review and the growing examples of proven mechanistic studies of non-enzymatic reactions, a unified language will be used to assure future identification and consolidation.

1.4 Categories of non-enzymatic mechanisms

1.4.1 Previously discussed categories of non-enzymatic biosynthetic reactions

As the literature of this field continues to mount, it is necessary to define the types of non-enzymatic mechanisms in natural products. The authors of this review largely agree with the categorization introduced by Gravel and Poupon in their 2008 review of non-enzymatic biosyntheses which highlighted four main categories: intramolecular rearrangement, dimerization, multi-component reactions and light induced changes. We suggest a slightly expanded version in which there are 5 major categories. These categories are 1) intramolecular reactions, 2) homologous multi-component reactions (including dimerizations), 3) heterologous multi-component reactions, 4) light induced reactions and 5) tailoring reactions. These categories will be described below using examples of studied non-enzymatic reactions, with a focus on recent publications. Categorization is not necessarily mutually exclusive and exceptions will be noted where a single reaction fits multiple classifications.

Finally, this work will focus primarily on bacterial and fungal natural products, with some plant and invertebrate examples. Due to the previously mentioned limitations of identifying non-enzymatic mechanisms from the literature, we must classify this review as non-comprehensive. Rather, we hope to consolidate validated and detailed examples of non-enzymatic mechanisms to portray a clearer picture of the field and set a precedent for future investigations.

1.4.2 Intramolecular reactions

Intramolecular reactions, or those occurring between two distinct hemispheres of the same molecule are one of the hardest to characterize due to the inability to isolate the two components of the reaction before they react. However, these are one of the most important types of non-enzymatic mechanisms in natural product biosynthesis in that they have been used extensively in biomimetic total syntheses (Beaudry, Malerich, & Trauner, 2005; Nicolaou, Montagnon, & Snyder, 2003; Razzak & De Brabander, 2011).

The most well studied intramolecular reactions are electrocyclizations between electron rich π systems commonly found in conjugated polyenes. Even as early as 1980 the chemist David Black postulated that due to the racemic mixture of the isolated polyenes, the endiandric acids, they possibly arose from so-called “non-enzymic reactions” (Figure 2) (Bandaranayake, Banfield, & Black, 1980). These hypotheses were validated in K.C. Nicolaou’s synthesis of the endiandric acids and their precursors (Nicolaou, Petasis, & Zipkin, 1982; Nicolaou, Petasis, Zipkin, & Uenishi, 1982; Nicolaou, Zipkin, & Petasis, 1982). The details of the biosynthesis and biomimetic synthesis of the endiandric acids have been reviewed frequently (Beaudry et al., 2005; Gravel & Poupon, 2008; Nicolaou et al., 2003).

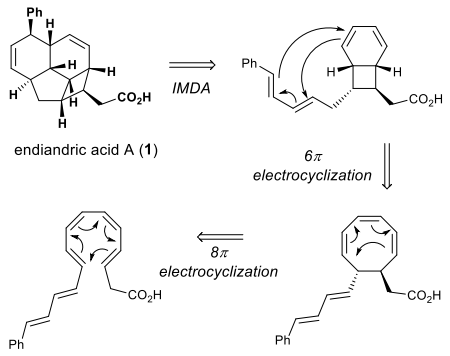


Figure 2. Non-enzymatic biosynthesis of the endiandric acids

The endiandric acids feature one of the most utilized electrocyclization reactions in total synthesis of natural products, the [4+2] Diels-Alder reaction (Juhl & Tanner, 2009). Several of the works highlighted in the Stocking and Williams review of biomimetic Diels-Alder reactions exist as clear examples of how to distinguish enzymatic processes from their non-enzymatic counterpart (Stocking & Williams, 2003). These studies commonly involve synthetic systems in which substrates are produced, then incubated with or without heat denaturation, a simple yet effective preliminary experiment. Additionally, understanding the distribution of isomers and its reflection on enzymatic involvement are commonly discussed.

These and other classical electrocyclizations have been reviewed in a synthetic chemistry context by many others including the Trauner (2005) and Nicolaou (2003) groups (Beaudry et al., 2005; Nicolaou et al., 2003). Since these reviews, elegant syntheses of natural products have continued to be completed with electrocyclizations attributed to biomimetic mechanisms (Raju, Piggott, Conte, & Capon, 2010; Webster, Gaspar, Mayer, & Trauner, 2013). However, not all of these mechanisms have been validated in biological settings.

A thorough, modern study of non-enzymatic electrocyclizations features the elansolids, which represent the dichotomy of artifacts and natural products in non-enzymatic biosynthesis.

The elansolids are a family of polyketide antibiotics from the gliding bacteria *Chitinophaga sancti*. Initially, alongside the original tropoisomers elansolid A1 (**2**) and A2 (**3**), the derivatives elansolid B1, B2, and B3 (**4**, **5,** not shown) were identified in the bacterial fermentation extract (Figure 3). Steinmetz and co-authors suggested that these were methanol, ammonia, and water derivatives resulting from extraction artifacts (H. Steinmetz et al., 2011).

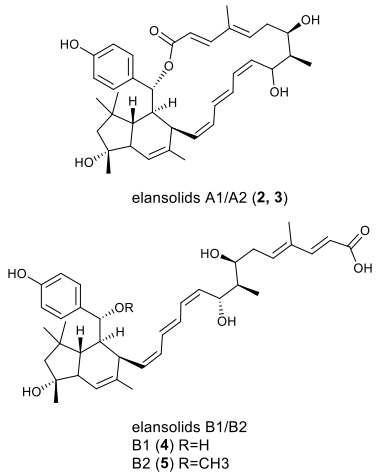


Figure 3. Elansolid family of natural products.

In the hunt for the reactive precursor of these artifacts, Jansen and co-workers carried out an extremely careful isolation to avoid contact of the unknown, sensitive intermediate with common nucleophilic agents such as water and methanol (Jansen et al., 2011). The resulting compound, known as elansolid A3 (**6**), was found in high yield (6.9 mg/L) and contained a highly electrophilic para-quinone methide functionality, giving credence to its transient nature during extraction. It was then shown that not only does this intermediate undergo non-enzymatic nucleophilic attack by water and methanol to form the **4** and **5**, but many nucleophilic amines could be used in a medicinal chemistry effort to construct a library of synthetic analogs with varying antibacterial activity (Steinmetz et al., 2012; L. L. Wang et al., 2017).

To fully elucidate the biosynthesis of the transient **6**, Dehn and co-workers constructed a cosmid gene library of the *Chitinophaga* strain and utilized known polyketide synthase (PKS) sequences to identify putative BGCs (Dehn et al., 2011). Consequently, a trans-AT PKS cluster was found whose predicted enzymes matched with the known chemical functionalities of the elansolids, particularly in the context of methylation and β-branching. Once the genes within the BGC were defined, there was no putative enzyme identified that would catalyze the final step of the biosynthesis to **6** from its precursor, **7**. Comparing the two structures, it was suggested that generation of intermediate **7** via enzymatic dehydration yields the intermediate **8**, which undergoes a non-enzymatic intramolecular Diels-Alder cyclization to form **6** (Figure 4). To interrogate this mechanism, Dehn and co-workers synthesized several hypothetical intermediates and confirmed the nature of the Diels-Alder cyclization to occur in a reaction flask. These family of compounds exemplify a non-enzymatic mechanism that is carefully installed by more canonical enzymatic machinery.

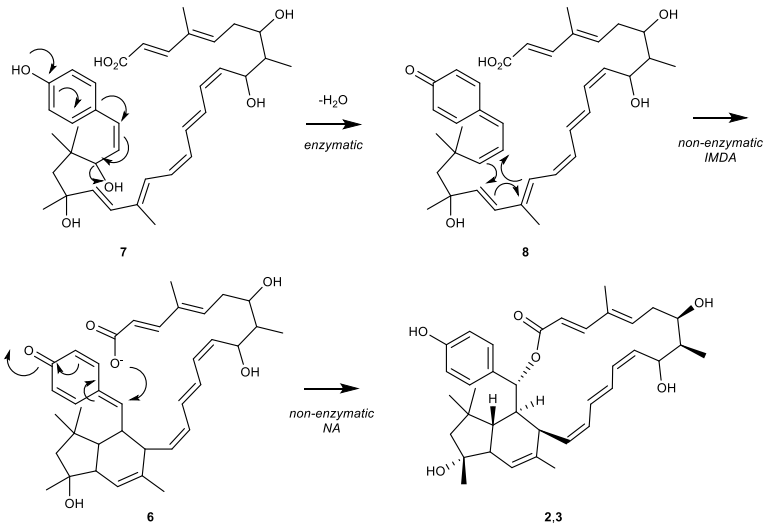


Figure 4. Non-enzymatic biosynthesis of elansolids A1 and A2

Historically, intramolecular reactions are the most well studied biomimetic non-enzymatic mechanisms. The cascades involved in these mechanisms are remarkable in their ability to construct complex structures without need for enzymatic catalysts or lengthy synthetic steps, thus drawing intense interest in the field (Razzak & De Brabander, 2011). However, many of these reactions cannot be studied extensively in biological systems due to their inherent reactivity upon formation of the immediate biosynthetic precursor. Their study relies on synthetic formation of such an intermediate, followed by measurement of its ability to progress to the following biosynthetic step in moderate conditions. Studies like the elansolids performed by Dehn and co-workers represent a holistic way to approach these types of non-enzymatic reactions that involve a heavy synthetic chemistry effort alongside BGC identification, to access transient molecules.

1.4.3 Multi-component reactions

Recently, there has been a notable increase in the number of reported natural products biosyntheses featuring multi-component non-enzymatic reactions. These reactions are characterized by two or more reasonably large components (> 50 Da) that come together through the formation of multiple bonds. This category is broadly split into two subclasses: those which involve chemically similar components (dimerizations and pseudo-dimerizations) and those that involve functionally disparate molecules.

1.4.4 Homologous multi-component reactions

Non-enzymatic homologous multicomponent reactions, or bond formation between structurally similar molecules, have been prevalent in natural products for many years. One of the earliest works in this category began in 1969 when Brophy and co-workers suggested that

novel lignans discovered from a Carpano tree bark extract, including carpanone (**9**), might be derived from dimerization of the related molecule, carpacin (**10**) (Brophy et al., 1969). The dimerization was suggested to occur via phenolic coupling of desmethyl-carpacin (**11**) followed by nucleophilic addition of enol ether into the ortho-quinonemethide. This hypothesis was later validated in a biomimetic total synthesis by the Clardy laboratory in 1971 (Chapman, Engel, Springer, & Clardy, 1971) (Figure 5). Examples of non-enzymatic dimerization that have been reviewed previously include the bisorbicillinoids (Abe, Murata, & Hirota, 1998), torreyanic acids (Julie C. Lee, Strobel, Lobkovsky, & Clardy, 1996; J. C. Lee, Yang, Schwartz, Strobel, & Clardy, 1995), carpanone (Chapman et al., 1971), stephacidin B (Artman, Grubbs, & Williams, 2007; Baran, Guerrero, Hafensteiner, & Ambhaikar, 2005; Baran, Hafensteiner, Ambhaikar, Guerrero, & Gallagher, 2006; Herzon & Myers, 2005), stylissadines (Grube, Immel, Baran, & Köck, 2007), and the complex rugulin bisanthroquinolones (Nicolaou, Lim, Piper, & Papageorgiou, 2007; Nicolaou, Papageorgiou, Piper, & Chadha, 2005). Here we will discuss modern examples of dimerizations and pseudo-dimerizations. The latter is defined as the bond formation between two monomeric units that involves additional small molecules.

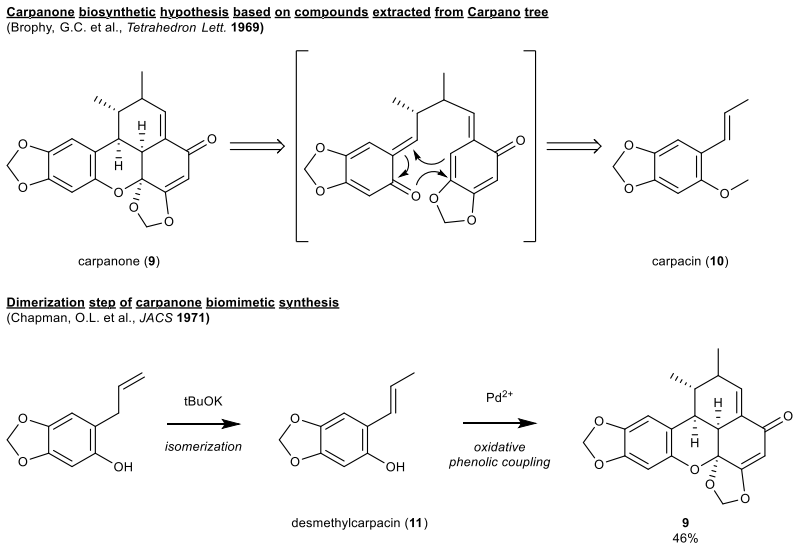


Figure 5. Carpanone biosynthetic hypothesis and biomimetic validation

In 2014, Lin et al. isolated a fungal strain known as *Lissoclinum patella,* from a sea squirt originating in Papa New Guinea whose extract possessed antibacterial activity (Z. Lin et al., 2014). The modest biological activity was due to the compound oxazinin A (**12**), a polyketide whose biosynthesis was suggested to be due to a non-enzymatic dimerization (Figure 6). A polyketide skeleton is prenylated to form **13**, which contains a reactive aldehyde. Nucleophilic addition of anthranilic acid into the aldehyde forms an electrophilic imine which undergoes nucleophilic addition by the para-substituted benzoic acid. This pushing of electrons is proposed to result in both cyclization of a benzoxazine ring and dimerization with another prenylated polyketide via an iminium intermediate. Evidence of this non-enzymatic reaction is due to the racemic mixture of oxazinin A found in the fungal extract. Although this is the first step in any investigation of non-enzymatic formation, further evidence is required to confirm the non-enzymatic mechanism of oxazinin A. As prenylated polyketides of similar structure have been isolated before (Tanaka et al., 2011; Teles et al., 2006), their reactivity should be able to be assayed in a flask to confirm the absence of protein catalysts.

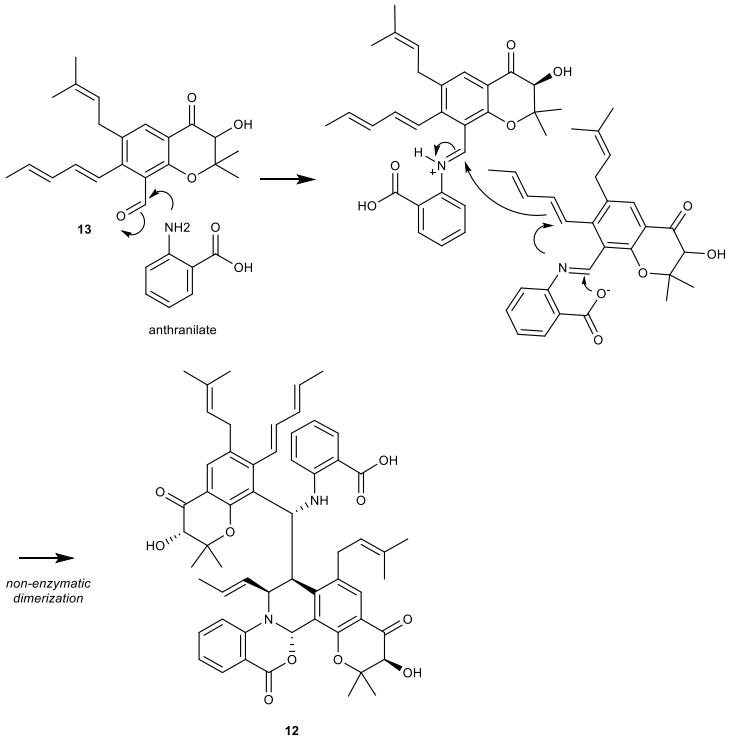


Figure 6. Non-enzymatic biosynthesis of oxazinin A

One of the newest and fastest moving discoveries of a natural product whose biosynthesis features non-enzymatic dimerization involves the study of the paracaseolides, which originate from the bark of a *Sonneratia paracaseolaris* Chinese mangrove tree (X. L. Chen, Liu, Li, Xin, & Guo, 2011). Their structure features a unique tetracyclic core composed of a tetraquinane oxa-cage bislactone, with two long chain alkyl appendages (Figure 7). Original biosynthetic hypotheses by Chen and co-workers in 2011 suggested that paracaseolide A (**14**) was derivative of a Diels-Alder [4+2] cycloaddition-based dimerization of the monomeric α-alkylbutenolide, **15**. In 2013, Noutsias and Vassilikogiannakis completed the first total synthesis of paracaseolide A using such a biomimetic Diels-Alder as the last step (Noutsias & Vassilikogiannakis, 2012). In their synthesis, monomeric **15** was heated in the absence of solvent to yield **14** in 59% yield. A year later, Yin and co-workers utilized a similar approach involving high temperature

dimerizations to yield a variety of paracaseolide analogs whose activity as phosphatase inhibitors were evaluated (L. Chen et al., 2014). These with multiple other explorations of the paracaseolides (Giera & Stark, 2013; Vasamsetty, Khan, & Mehta, 2013) indicated that the Diels-Alder [4+2] cycloaddition of butenolides was chemically tractable.

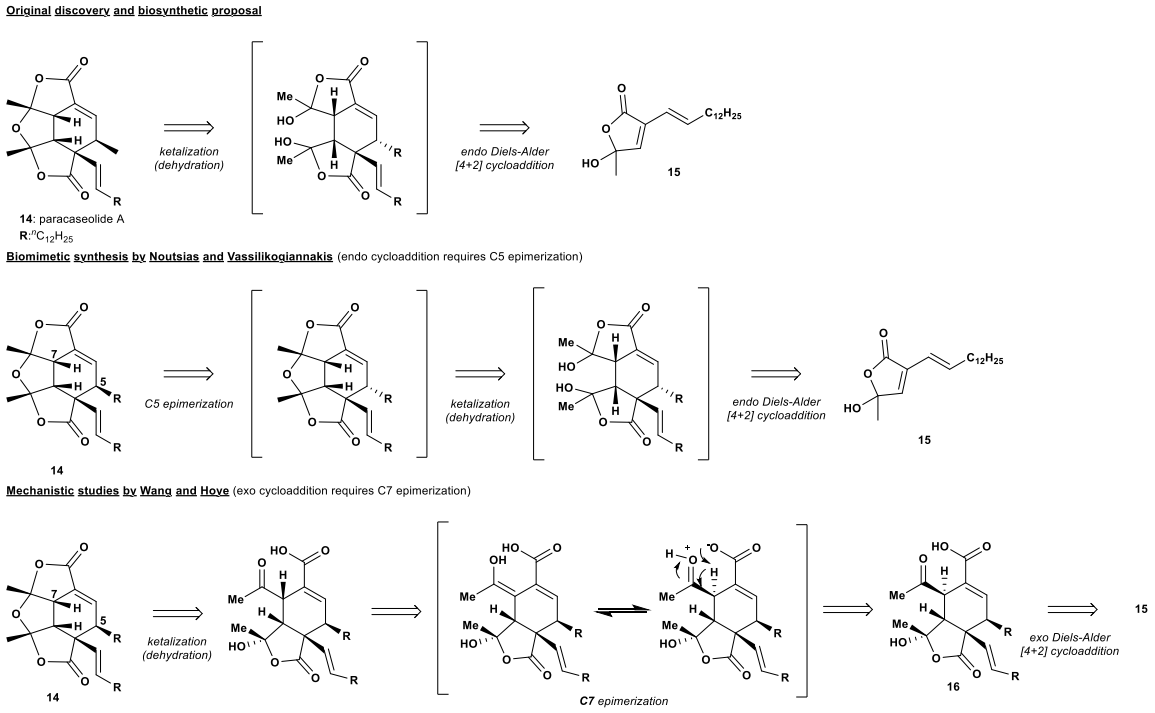


Figure 7. Progression of the mechanistic hypothesis of paracaseolide A

The bowl-like structure of **14** and the *cis* orientation of the alkyl chains has prompted detailed studies of the mechanism of this dimerization (Vasamsetty, Sahu, Ganguly, Khan, & Mehta, 2014). Specifically, it was first hypothesized that the cycloaddition was a concerted *endo*-Diels Alder [4+2] cycloaddition. However, this reaction would result in the *trans* configuration of the alkyl chains, an isomer not reported by the initial authors. While an epimerization of these stereocenters could explain their dynamic nature, several alternate groups participating in total synthetic routes reported that epimerization of the final paracaseolide product was not observed.

In their 2014 report Vasamsetty and co-workers demonstrated that the dimerization was racemic and these diastereomers could not epimerize, an experimental finding supported by DFT calculations. Further computational evaluation of the energetics of this system suggested a step-wise cycloaddition in contrast to a concerted mechanism. This prediction could explain the weighted isolation of the *cis*-paracaseolide A as a step-wise biradical mechanism would favour this diastereomer by an estimated 2.3 kcal/mol.

In 2015, Wang and Hoye further explored the non-enzymatic mechanism of paracaseolide A through various experimental and computational methods (T. Wang & Hoye, 2015). Initially, the authors were able to show that the non-enzymatic dimerization occurred at lower temperatures than previously reported, supporting its feasibility in a biological setting. One of the most important findings was the isolation of the intermediate **16**, the immediate result of the Diels-Alder cycloaddition before dehydration to yield the full paracaseolide A. This intermediate has previously been predicted, but never isolated.

Interestingly, the stereochemistry of this molecule revealed that the electrocyclization occurred with an *exo* orientation of the carbonyl, a generally unfavorable mechanism that contradicted the initial proposals that were made without experimental evidence. The *exo* configuration suggested by the isolation of intermediate **16** was supported by computational calculation of the energies for each transition state (TS_{endo} vs TS_{exo}) and an argument for a stabilized bis-pericyclic transition state that was made for increased *exo* favorability.

The details of this stereochemical orientation also shed light on the epimerization of paracaseolide. Previous reports had studied the epimerization of paracaseolide at the C5 position, due to the fact that the proposed *endo* configuration would result in the opposite orientation of

the final product. Their findings, as detailed above, declared that epimerization was not possible. However, the C5 stereochemistry of **16** assuming an *exo* [4+2] addition is set correctly, while the orientation of C7 is opposite to the final product, paracaseolide A. The authors provided evidence for this novel epimerization using deuterium labeling experiments of the intermediate **15** and a model system in which the alkyl substitution was simplified. In these experiments, the C7-H was shown to exchange with deuterium at ambient temperatures quite rapidly (50% conversion in 4 hours). The authors suggest this mechanism is due to keto-enol tautomerism that leads to formation of a zwitterionic carboxylate ion capable of abstracting the C7 proton, intramolecularly. The result of this abstraction would be an enol olefin capable of protonation before the final dehydration step. This finding was key as it demonstrated that there is a feasible non-enzymatic mechanism for generating a planar environment at the C7 position amenable to an obligate step for generating the preferred stereochemistry of **14**. Finally, the catalytic function of the carboxylate ion was supported by the decreased yields (35% conversion in 10 days) when esterified substrates were utilized.

The mechanistic studies performed by Wang and Hoye exemplify the detail needed to study multi-stepped non-enzymatic biosynthetic reactions. In addition to the findings discussed here, the authors raised more questions that require attention. Most importantly, they showed that a monomeric butenolide that was opened irreversibly, which contained a favorable dienophile with two electron withdrawing groups, could not only undergo dimerization, but could do so much more efficiently than its closed form. While the exact nature of the biological dienophile remains to be determined, the work performed by the Hoye group and other synthetic chemistry groups over the course of five years is an impressive feat that has brought a necessary level of

experimentation to a preliminary hypothesis of non-enzymatic biosynthesis. Where these studies lack is a biological exploration of the nature of these molecules in the context of the Chinese mangrove they were isolated from, which the authors of this review look forward to.

In 1956 the alkaloid natural product securinine (**17**) was isolated from the leaves of the plant, *Securinega suffructicosa* (IaA & Turova, 1956). Since its discovery, over 60 related natural products from five different genus of related Euphorbiaceae have been discovered (Weinreb, 2009). These compounds are characterized by their bridged tetracyclic core consisting of a fully saturated 5-membered lactone ring and either piperidine (e.g. securinine, virosecurinine, phyllantadine) or pyrrolidine rings (e.g. norsecurinine) (Figure 8).

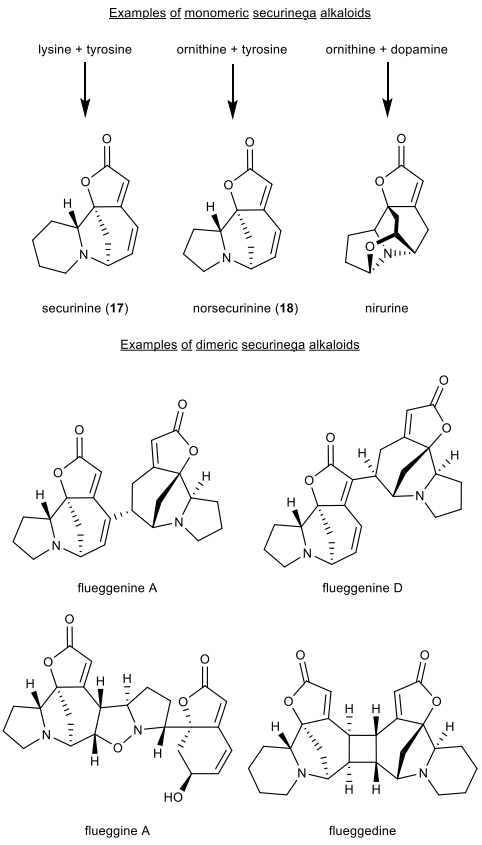


Figure 8. Securinega family of plant alkaloids.

Starting in 2006, the chemical diversity of the securinega alkaloids expanded exponentially. Dimeric, trimeric and higher order oligomeric derivatizations have been discovered (a complete collection of oligomeric securinega alkaloids can be found in work by Jeon and Han, 2017). The complicated structures of the securinega monomers with their heavy substitution, compact ring structures, and short chain bridges was confounded even further in dimeric analogs. These compounds included asymmetric orientations (e.g. flueggenines A-I and flueggines A-B) as well as traditional symmetric molecules (flueggedine).

The hypothesis that these oligomeric molecules could be constructed via non-enzymatic mechanisms began in 2006 when Gan and co-workers first isolated flueggenines A and B from the roots of a *Flueggea virosa* plant (Gan et al., 2006). The extract's most abundant metabolite, norsecurinine (**18**), was proposed to be the biosynthetic precursor to the flueggenines. The authors suggested an intramolecular vinylogous Morita-Baylis-Hillman reaction, also known as the Rauhut-Currier (RC) reaction, as a mechanistic rationale for this dimerization (Figure 9).

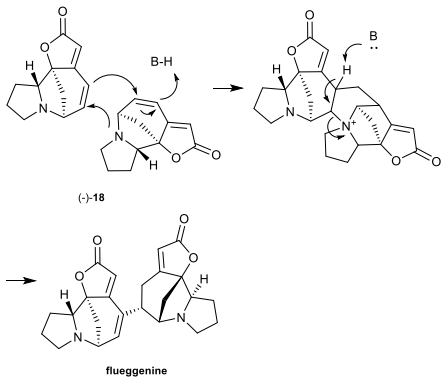


Figure 9. Proposed biosynthetic mechanism of flueggenine.

A conventional Morita-Baylis-Hillman reaction involves an aldehyde as the ultimate electrophilic acceptor of the anionic carbon, whereas in the vinylogous Morita-Baylis-Hillman reaction or Rauhut-Currier reaction a second Michael acceptor is the electrophilic carbon source.

While RC reactions typically involve phosphine catalysts, the classification of the securinega alkaloids as having a RC-like mechanism is more appropriate due to the identity of the electrophile.

In contrast to conventional MBH and RC reactions where a nucleophilic catalyst is used, the three necessary components (activated alkene, electrophilic carbon, and nucleophilic amine) are all contained on the securinega monomers. This provides a self-catalyzed system, as seen in several other non-enzymatic mechanisms discussed here. As a direct follow up, the same laboratory in 2013 proposed that due to the regeneration of the nitrogen catalyst and an abundance of unhindered enone systems after putative dimerization, there should be a possibility for continued oligomerization (H. Zhang et al., 2013). After careful isolation, three new trimeric securinega compounds, named the fluevirosines, were discovered whose oligomeric state supported the author's hypothesis regarding repetitive RC reactions. Stirring of the monomers in organic solvent did not lead to oligomerization suggesting the necessity of an enzyme or a particular chemical environment, although these studies were limited (Figure 10).

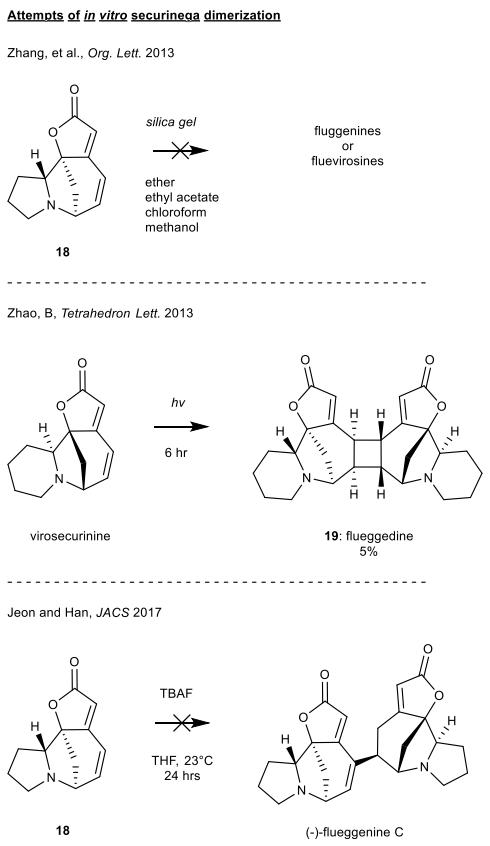


Figure 10. Synthetic attempts of securinega dimerization

A year later Zhao and co-workers reported the new securinega dimer, fluggedine (**19**), which contained a novel cyclobutane ring connecting two identical units of virosecurinine (B.-X. Zhao et al., 2013). Interestingly, the authors proposed that this ring was constructed via a [2+2] cycloaddition, a hypothesis tested by irradiating virosecurinine for several hours at room temperature under UV light (350 nm). Indeed, the authors saw formation of **19**, albeit in low yield (5% over 6 hours), suggesting that this UV- based cycloaddition could occur (Figure 9, middle). More so, the stereochemistry of **19** suggested that the cycloaddition occurred predominantly in the *exo* configuration yielding a caged structure. This was the first experimental evidence for the dimerization of securinega monomers under biological conditions. Interestingly, mass spectrometry of the leaves and roots of the *F. virosa* plant demonstrated that

the monomers were isolated from both sites, but the dimeric **19** was only isolated from the leaves. This finding suggested that the UV-induced dimerization seen *in vitro* may be relevant during conventional plant growth. While a total synthetic route to flueggine A had been completed (Wei, Qiao, Liu, Yang, & Li, 2013), there were no biomimetic studies performed with the proposed RC reaction before the recent publication by Jeon and Han in 2017. In their total synthetic effort to (-)-flueggenine C, Jeon and Han incorporated a biomimetic RC step early in their route. This is in contrast to a majority of biomimetic total syntheses, where biomimetic coupling are used as terminal steps. The authors constructed a model system of the proposed biosynthesis that differed quite drastically from endogenous substrates in the hopes of making the reaction more efficient. After optimizing the reaction conditions for their model substrate, the authors tried to dimerize **18** using TBAF as a base in THF at ambient temperatures with no success after 24 hours. This is in agreement with Gan and co-workers who found that stirring **18** in various organic solvents did not lead to dimerization.

The studies of the flueggenines and related compounds introduce many interesting questions about their biosynthesis and the role of non-enzymatic reactions within. First, Zhao and co-workers showed that **19** dimerization could occur non-enzymatically under UV light in low yields in vitro. However, the proposed [2+2] cyclization is drastically different from the suggested mechanisms of every other example of the 16 known securinega dimers so conclusions from this report must then be tempered. The value of these studies is the demonstration that the RC reaction of the model dienone system can occur in the absence of enzymes. Secondly, two groups have shown that stirring monomers in organic solvents do not lead to dimerization. This could be due to the fact that the RC mechanism requires enzymes or it

could be that they require a non-proteogenic promoter that would be endogenous to the plant environment such as UV light or an oxidative environment, neither of which were tested for the monomer, **18**. Relatedly, a proposed mechanism for the formation of the dimers flueggine A and B suggests that exogenous oxidation is required for a cyclization cascade (B. X. Zhao et al., 2011). Regardless of these conditions, it is likely that the non-enzymatic RC reaction will be inefficient as it is known to be of limited reactivity in a flask due to the transient nature of the nucleophilic trigger. The story of securinega oligomerization is still developing, however their chemical complexity and diversity raises must interest in their biosynthetic origins.

Recently, the Clardy lab isolated the complex fungal derived natural product, homodimericin A (**20**), from a *Trichoderma* strain of fungus (Mevers et al., 2016). Homodimericin A, an extremely carbon rich structure whose structural elucidation pushed the limits of modern NMR spectroscopy, was isolated as the racemate. The structure appears to be a derivative of the complex dimerization of a hydroquinone monomer, **21**. Due to the structure and racemic nature, the authors suggest that the biosynthetic dimerization is non-enzymatic under oxidative conditions. However, no appreciable amount of **21** was isolated to validate the dimerization hypothesis.

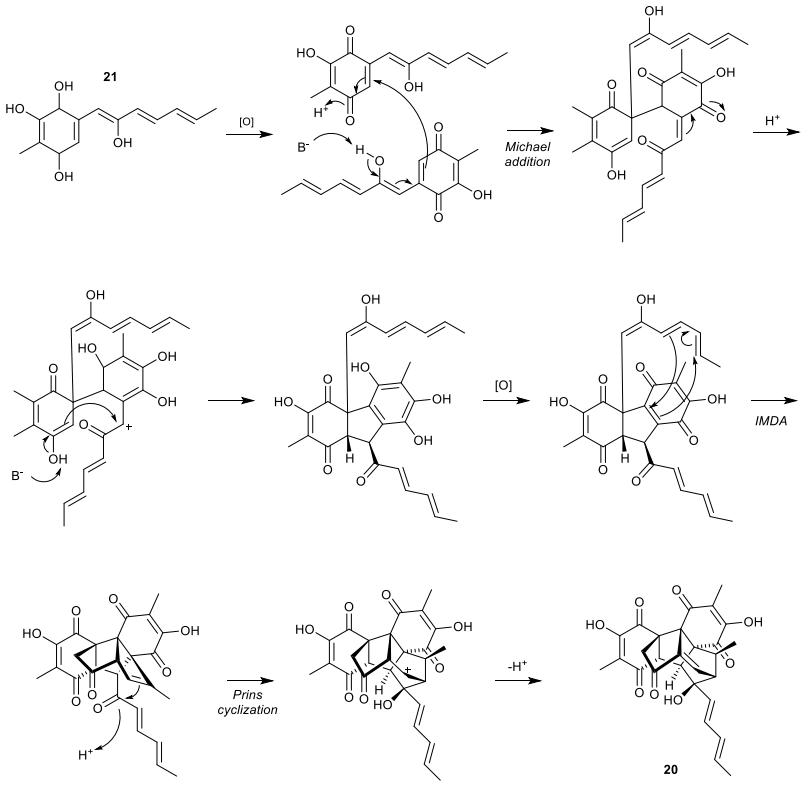


Figure 11. Biosynthetic proposal for homodimericin A.

Structurally, **20** is similar to the bisorbicillinoids, a well-studied family of fungal natural products that are derived from various dimerizations of the sorbicillinoids (Harned & Volp, 2011). Although many sorbicillinoid analogs have been discovered, they are proposed to be derived from a single common intermediate, sorbicillinol (**22**). The exact identity of **22** was unknown for 20 years after the discovery of the original dimeric compounds due to its extreme reactivity (Trifonov, Dreiding, Hoesch, & Rast, 1981). This fact may explain the Clardy group's inability to isolate the putative monomeric starter unit **21**.

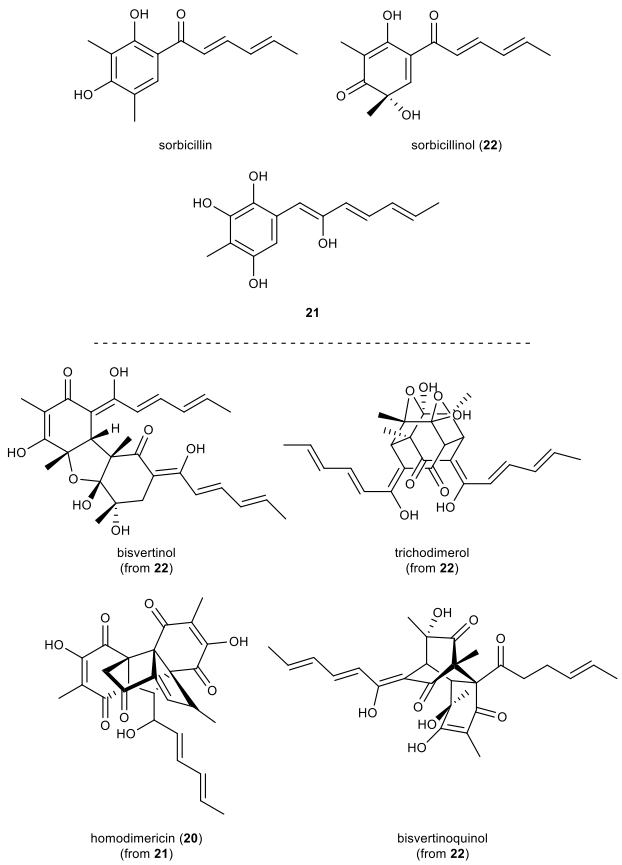


Figure 12. Sorbicillinoid monomers and dimers

As **22** is a precursor to three different bisorbicillinoid skeletons (represented by the bisvertinols, trichodimerols, and bisorbicillinol in Figure 12), there have been multiple non-enzymatic mechanisms suggested for these transformations. These proposed biosyntheses were discussed by Gravel and Poupon in the 2008 review and their mechanism will not be discussed in detail here. Interestingly, the mechanism put forth by the Clardy lab for **20** is a chimera of the mechanism suggested for bisvertinol/trichodimerol and bisorbicillinol. The early steps proposed for **20** biosynthesis feature Michael addition to form an initial C-C bond, similar to bisvertinol/trichodimerol biosynthesis. An intramolecular Diels-Alder is proposed later in the biosynthesis of **20** similar to the intermolecular Diels-Alder proposed for bisorbicillinol.

However this electrocyclizations is distinct with respect to the location of the diene. In **20**, the conjugated polyene tail is involved in the [4+2] cycloaddition, whereas the two quinone-based monomers are responsible for the proposed electrocyclization in bisorbicillinol. Additionally, a Prins cyclization is suggested as the final ring-forming step, a mechanism not proposed for the bisorbicillinoids. The structural complexity of homodimericin A makes the biosynthetic hypothesis more intricate. Therefore, further detailed studies of these steps would be an enlightening discovery for a large family of related natural products.

Homodimericin A was found in several fungi, but in one particular strain it was only produced when co-cultured with a *Streptomyces* bacterium that secretes the bafilomycin family of anti-fungals. Additionally, pure bafilomycin C1 induced production of **20** in this strain. Most intriguing is the fact that **20** does not possess anti-fungal or antibacterial activity in the least, but is strictly induced in the *Trichoderma* fungus by the presence of bafilomycin C1. Mevers and co-workers suggest that this might be due to the mechanism of action of the bafilomycins which are known eukaryotic inhibitors of vacuolar ATPase, an effect linked to oxidative stress (Nishikawa et al., 2016; Yokomakura et al., 2012). A highly oxidative environment has been shown to be crucial for bisorbicillinoid dimerization. Additionally, a concerted effort by Abe and co-workers in three separate publications have suggested that the bisorbicillinoids are effective radical scavengers (Abe & Hirota, 2002; Abe et al., 1998; Abe, Murata, Yamamoto, & Hirota, 1999). However, why other members of the bafilomycin family, which have the same effect on eukaryotic machinery, do not induce the production of homodimericin remains a mystery. Perhaps this difference is explained by varying cell permeability or solubility. Whether these homodimericin products are the result of oxidative stress caused by the anti-fungal effect of

bafilomycins remains to be determined, however the experimental induction is convincing that these compounds are produced as a biological response to bacterial challenge.

The bohemamines are a family of pyrrolizidine alkaloids, a large and structurally diverse array of over 650 molecules with similar heterocyclic structures (J. Robertson & Stevens, 2017). In 2016, our laboratory published the first known example of bohemamine dimers isolated from a *Streptomyces spinoverrucosus* species (Fu, Legako, La, & MacMillan, 2016) (, top). Interestingly, these were not true dimers as an extra methylene carbon was found to connect the two monomeric structures. These types of dimers where two symmetric units are connected by aryl-aryl bonds via a single methylene bridge have been reported in various sources, primarily in plants (Arisawa et al., 1990; L. Chen et al., 2014; Cheng, Xiang, Izumikawa, Meluzzi, & Moore, 2007; Kalaitzis, 2013; Kalaitzis, Cheng, Thomas, Kelleher, & Moore, 2009; Kalaitzis, Izumikawa, Xiang, Hertweck, & Moore, 2003; Piel et al., 2000; Shao et al., 2011) (Figure 13). Several of these monomers have been explored biosynthetically, but the dimerization process remains speculative. Several groups have utilized synthetic mechanisms under mild conditions to demonstrate that the reaction can occur (Chauthe et al., 2010; Eaton, Dalal, Cassera, Zhao, & Kingston, 2016; Minassi et al., 2012; Penttila, Kapadia, & Fales, 1965), giving credence to this non-enzymatic mechanism in a biological setting.

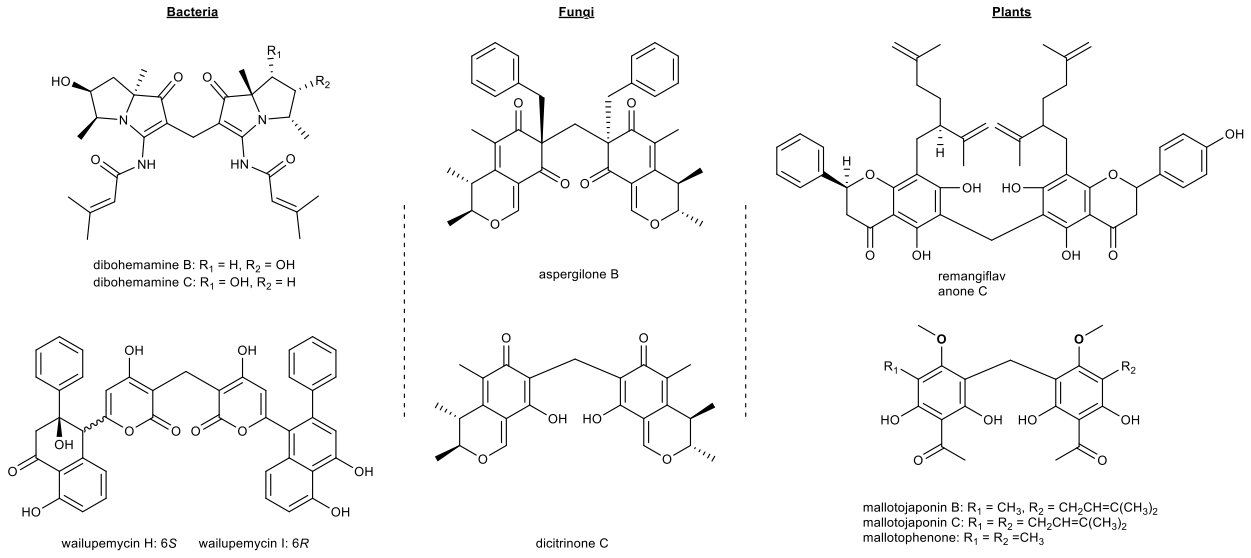


Figure 13. Symmetric dimers from various species

The utilization of a C1 unit in secondary metabolism is rather limited, as these building blocks are commonly incorporated as terminal modifications via SAM-based methylation (Dewick, 2001). Formaldehyde is a common metabolic by-product and it was hypothesized that excess formaldehyde could be responsible for this dimerization. Indeed, reaction of bohemamine A with formaldehyde in THF led to the formation of two products, the hydroxyl derivative of bohemamine A, **23**, and dibohemamine A. The enol from a resonance structure of bohemaine A could participate in a nucleophilic addition with formaldehyde to generate **23**. Then under acidic conditions, **23** could be protonated to generate a reasonable electrophile for a second intermolecular nucleophilic attack from a bohemamine A monomer (Figure 14). Isolation of **23** and exposure to acid generated the dimeric dibohemamine A in 1 hour, suggesting that acidification is essential and that the original formaldehyde condensation is reversible. Utilizing different aryl and alkyl aldehydes, it was shown that the dimerization was robust to aryl substitution under normal atmosphere, but alkyl aldehydes led to facile oxidation unless

performed under nitrogen. These experiments suggest that the dimerization mechanism is tuned to C1 metabolism.

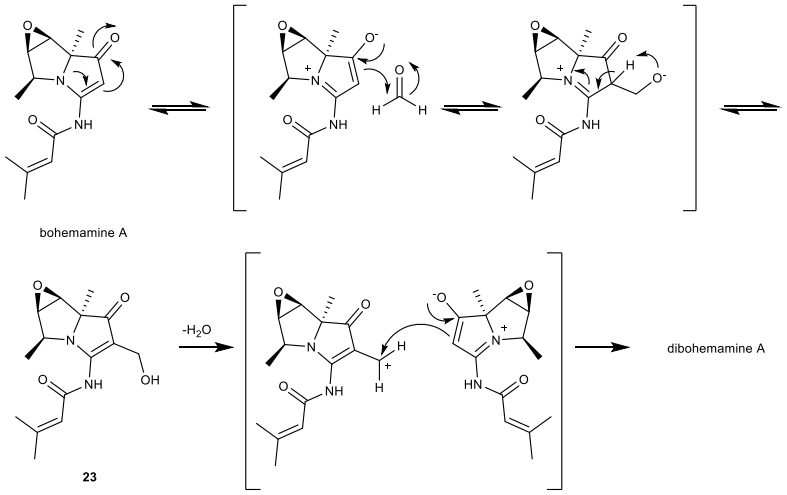


Figure 14. Proposed biosynthetic mechanism of dibohemamine A

Using the PFBHA assay of quantification (Perkin Elmer), the levels of extracellular formaldehyde in dibohemamine producing bacteria were found to be four times higher than in a similar non-dibohemamine producing *Streptomyces* strain. As formaldehyde is highly diffusible across the biological membrane, this result suggests that internal levels of formaldehyde were also elevated. Formaldehyde is toxic due to its highly electrophilic nature resulting in cross-linking of macromolecular structures within the cell, particularly after combining with proteogenic thiols and amines to form thiazolidine and electrophilic imines, respectively (N. H. Chen, Djoko, Veyrier, & McEwan, 2016). Furthermore, formaldehyde levels are known to be dependent on metabolic alterations, particularly as recent studies have shown that flux through the pentose phosphate pathway results in excess accumulation of the toxic C1 metabolite. This has challenged canonical views that mainly methanotrophs produce formaldehyde as a product of methanol fixation. However, the metabolic data is sensible as genetic evidence of formaldehyde

neutralization techniques has been found in a variety of bacteria that cannot fix methanol. Detoxification of formaldehyde in bacteria is known to occur via four major mechanisms in order of abundance: 1) thiol-based systems, 2) the ribose monophosphate or RuMP system, 3) the pterin-dependent system, and 4) direct oxidation system. Interestingly, in the thiol and pterin systems non-enzymatic reactions between sulfur or nitrogen nucleophiles and formaldehyde have been shown to trigger the downstream reactions. It is enticing to ascribe the use of these highly reactive monomeric subunits found in bacteria, fungi, and plants as a detoxification mechanism of endogenously high levels of formaldehyde.

In conclusion, the examples provided here demonstrate that homologous multi-component reactions are an important sect of non-enzymatic reactions in biosynthesis. In addition to the classical examples of electrocyclizations, these new studies represent new discoveries in carbon and nitrogen based nucleophiles catalyzing cascades of varying intramolecular non-enzymatic reactions. Most importantly, the last three dimerization stories of fuggedine, homodimericin A, and the dibohemamines have generated many questions about the potential biological function of these reactive molecules. An area of interest which will continue to be in focus throughout this review.

1.4.5 Heterologous multi-component reactions

An important addition to the field of non-enzymatic biosynthesis over the last five years is the discovery of more than five families of compounds who are postulated or proven to be the result of heterologous multi-component reactions. These reactions are characterized by the formation of multiple covalent bonds between distinct substrates. There will be a heavy focus on

this section of the review as these studies have greatly expanded the known complexity and chemical diversity of non-enzymatic mechanisms.

The jadomycins are a family of polyketide antibiotics produced by a *Streptomyces venezuelae* strain under stress conditions such as organic solvent exposure and heat (Doull, Ayer, Singh, & Thibault, 1993; Doull, Singh, Hoare, & Ayer, 1994). Previously, it has been suggested via both genetic sequencing of the BGC of the jadomycins and biomimetic synthesis (Sharif & O'Doherty, 2012) that the biosynthesis of jadomycin A involves a canonical polyketide extension of malonate subunits to form the intermediate UWM6 in Figure 15, that can be oxidatively cleaved to form the aldehyde **24** which can non-enzymatically combine with isoleucine via nucleophilic addition to afford intermediate **25**. The imine formed can then react with the nucleophilic carboxylate of the amino acid substrate to trigger a Michael addition into the attached quinone to cyclize oxidative decarboxylation. Biosynthetic studies of this molecule have utilized homologous recombination and mutagenesis to show that no enzyme involvement is needed for the incorporation of isoleucine.

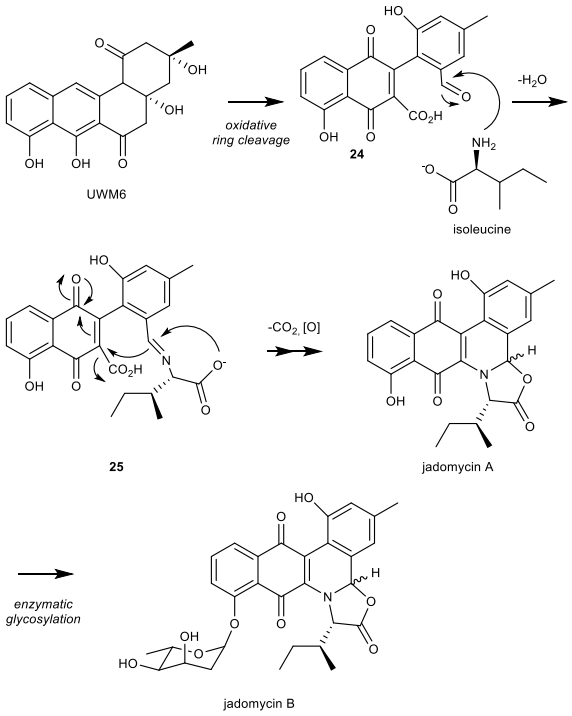


Figure 15. Non-enzymatic biosynthesis of jadomycins A and B

Several groups have utilized the reactivity of intermediate **24** to incorporate a slew of amino acids whose common primary amine serves as a potent nucleophile (Dupuis et al., 2011; Forget, Robertson, Overy, Kerr, & Jakeman, 2017; Jakeman et al., 2009; Jakeman, Graham, & Reid, 2005; Martinez-Farina & Jakeman, 2015; Martinez-Farina et al., 2015; Rix et al., 2004; A. W. Robertson et al., 2015). For these reactions, amines that lack the spatial proximity of a carboxylic acid found in amino acids would hypothetically not result in the cyclization of the full jadomycins. While experiments such as these would further support the non-enzymatic mechanism, they have not been performed. Additionally, intercepting the reactive iminium group of intermediate **25** with alternate nucleophiles would help determine how the proximity of the carboxylic acid affects the reactivity of the imine. That is, do proximally located aryl/alkyl primary amines and carboxylic acids have privileged reactivity compared to other nucleophiles.

This question is intriguing as the examples in which these substrates (anthranilic acid and amino acids) constitute a majority of the known non-enzymatic biosynthetic mechanisms, a topic that will be discussed at the conclusion of this review.

Interestingly, only one of the known analogs of jadomycins is not glycosylated and it is suggested that this glycosylation occurs after the non-enzymatic incorporation of isoleucine. This is a rarity in non-enzymatic biosynthesis, where it is more common for these non-enzymatic transformations to occur on fully matured natural products, or those that no longer require enzymatic involvement. Furthermore, aglycone jadomycins produced naturally or unnaturally, do not retain the same level of bactericidal activity suggesting that the non-enzymatic incorporation of amino acids which is necessary for said glycosylation is required for environmental fitness.

Continuing the theme of nucleophilic amines, rubrolone A and B are pyridine containing type II polyketides isolated from different *Streptomyces* strains of bacteria in 1971 and 2016, respectively (Palleroni et al., 1978; Schuep, Blount, Williams, & Stempel, 1978) (Figure 18). While these compounds represent interesting chemical complexity, their lack of bioactivity has potentially stymied biosynthetic studies until recently. In 2016, Yan and co-workers utilized the unique dehydratase gene, previously determined to be effective for genome mining of BGC, to search for the putative rubrolone BGC within the sequenced genome of *Streptomyces* sp. KIB-H033 (Yan et al., 2016). After defining the biosynthetic pathway needed for synthesis of the late stage intermediate **26**, it was postulated that this molecule reacted non-enzymatically with the different endogenous nucleophilic nitrogen species ammonia and anthranilic acid, in the bacterium to product the final rubrolones A and B, respectively. Interestingly, in their heterologous expression of the BGC in the model bacterium *Streptomyces albus*, which has low

endogenous titers of anthranilate, the yields of rubrolone B were diminished unless supplemented with exogenous ^{15}N-anthranilate, which was readily incorporated into rubrolone B, but not A. These data and the lack of any appropriate aminotransferase in the BGC provided biological evidence of this non-enzymatic mechanism which was tested empirically by reacting isolated substrates in an aqueous phosphate buffer at the appropriate fermentation conditions *in vitro*.

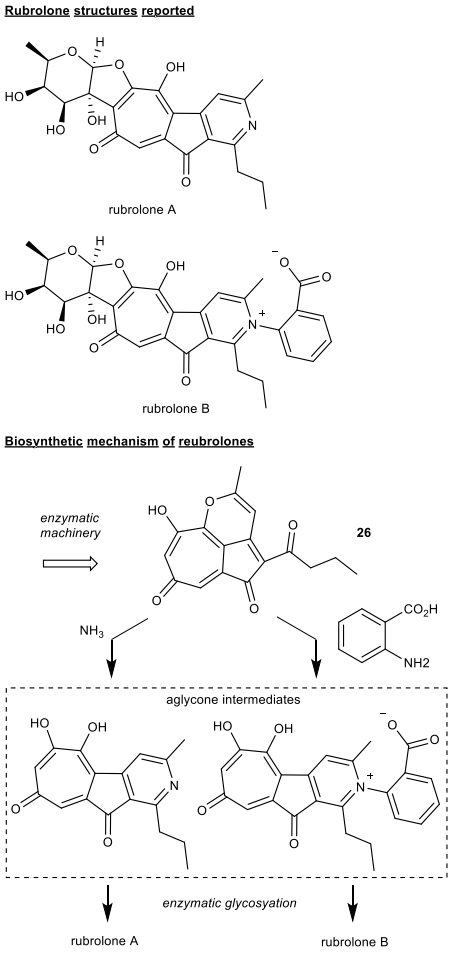


Figure 16. Structure and proposed biosynthesis of the rubrolones

The structures of **26** and the rubrolones, the non-enzymatic nucleophilic addition is not the final step of the biosynthesis. Rather, the initial nucleophilic attack causes a secondary

reaction to generate the pyridine ring of rubrolone. Specifically, **26** is proposed to exist in equilibrium with **27**, due to a hydrolytic ring opening demonstrated by deuterium exchange experiments. Nucleophilic attack of **27** with an amine substrate occurs to form the intermediate **28**. Intermediate **28**, which features a reactive 1,5-dione, can trap the non-enzymatically formed imine nitrogen resulting in the subsequent formation of the pyridine ring, a mechanism never before reported in natural product biosynthesis. Reactions between amines and 1,5-diones have been reported for the formation of pyridine rings in synthetic chemistry efforts under high temperatures. This suggests that the inherent properties of the intermediate **28** could provide a uniquely selective advantage for cyclization.

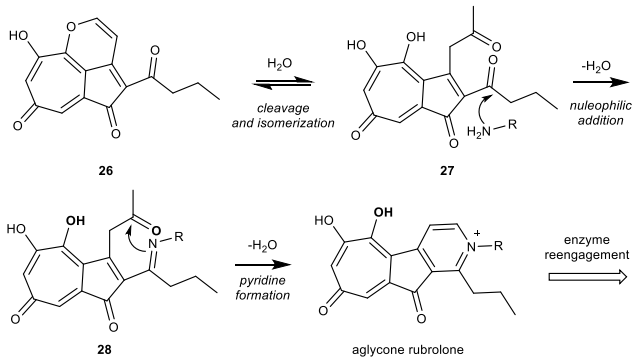


Figure 17. Biosynthetic mechanism of the rubrolones

An intriguing component of the multi-step non-enzymatic pyridine formation is that it occurs prior to enzyme engagement downstream in the biosynthetic pathway. This was demonstrated when glycosylated rubrolone levels directly correlated with the pool of pyridine formed intermediates, but not acyclic intermediates. This is similar to the case of the jadomycins in which non-enzymatic reactions occur before the natural product is reengaged with catalytic proteins. Speaking to the substrate specificity of said enzymes in the rubrolone biosynthesis, a variety of functionalized anthranilic acid analogs were able to be successfully incorporated into

mature rubrolones *in vitro*. This surprising finding could be explained by a function of the rubrolones that would allow promiscuity in the substrate binding pocket if driven by an environmental pressure, albeit there is no known biological function attributed to these natural products.

The last multi-component non-enzymatic biosynthetic mechanism discussed in this section is in stark contrast in that these reactions have demonstrated a unique biological function. The fungal based maximiscins are a family of natural products which have biological intrigue due to their cytotoxicity to various types of cancer *in vitro* and *in vivo* (Lin Du et al., 2014; Robles, Du, Cichewicz, & Mooberry, 2016). These molecules have been studied extensively by the Cichewicz group since their 2014 discovery after a series of OSMAC-like experiments successfully elicited altered metabolite profiles of a natural product producing *Tolypocladium* species of fungus (Lin Du et al., 2014).

In their 2014 report, Du et al. utilized thorough isotope labeling experiments to derive the biosynthetic origins of every carbon atom in the molecule which features a unique amalgam of precursors, hence their name meaning “from many sources”. The final step of the biosynthesis was proposed to be a bond formation between a shikimate-derived sugar and a polyketide-NRPS derived subunit. Interestingly, this reaction was reversible as demonstrated by the slow isomerization of maximiscin to the isomer isomaximiscin. This data alongside careful spectrometry observations in co-culture experiments of the same fungal strain led the Cichewicz group to study the ability of the electrophilic precursors, pericoxide and (+)-pericosine A, to participate in non-enzymatic nucleophilic additions to form the maximiscins, via pyridoxatin,

and various related molecules through exogenously added nucleophiles (L. Du, You, Nicholas, & Cichewicz, 2016) (Figure 19).

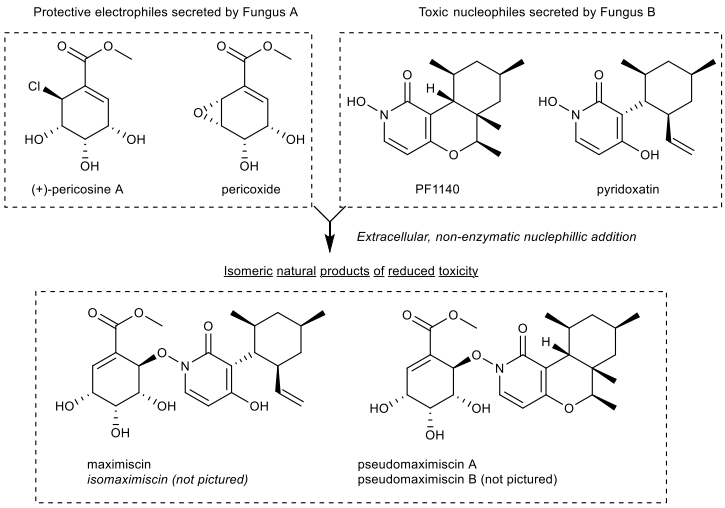


Figure 18. Maximiscin and pseudomaximiscin biosynthesis

The formation of maximiscin and analogs from hydroxamic acids were found to be stereoselective, a rare phenomenon in non-enzymatic biosynthesis. Du and co-workers used their substrate diversification studies alongside computational DFT calculations to examine the stereoselectivity of different nucleophiles. They were able to characterize the mechanism by which hydroxamic acids such as pyridoxatin undergo nucleophilic addition to pericoxide and pericosine, which proceeds via anti-SN2' and syn-SN2', respectively to form the non-racemic maximiscin (Figure 19). The authors of this review encourage readers to consult the work of Du and co-workers as a guide for a high quality demonstration of non-enzymatic mechanism.

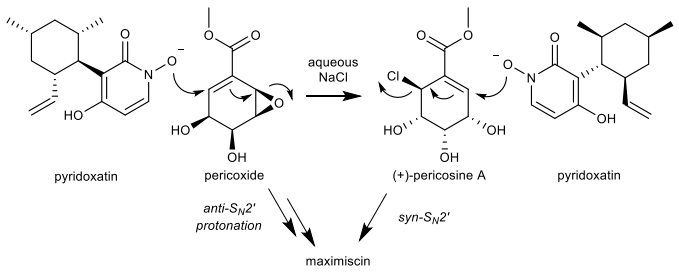


Figure 19. Biosynthetic mechanism of maximiscin

Most interestingly, the story of the electrophilic natural products pericoxide and pericosine A are a prime example of how biological function can be attributed to a non-enzymatic reaction. While studying the *Tolypocladium* fungal strain T1, it was found that many of the nucleophiles that combined non-enzymatically with pericoxide and pericosine A possessed anti-fungal activity. Once combined, the new molecules were non-toxic. More so, supplementation of spent media from T1 fungus, containing the electrophilic molecules, suppressed the effect of exogenously added nucleophilic antifungals. These results suggest that the T1 fungal strain has developed a non-enzymatic detoxification system to protect against competing fungi. While this system of protection remains to be studied in depth on a genetic level, it suggests an exciting, non-enzymatic relationship between secondary metabolism and basic microbial function.

1.4.6 Tailoring reactions

The multi-component reactions discussed above involve the formation of chemical bonds which often trigger subsequent intramolecular reactions. However, there also exists non-enzymatic reactions in which the reaction occurs on mature natural products, no subsequent reactions occur, and the transformation results in a modest change to the structure of the molecule. Based on the complexity of the substrate reacted with the mature natural product, we

can split this category into two main classes: reactions involving simple molecular species (O_2, SH_2, NH_3, etc.) and those involving higher molecular weight species.

The simplest non-enzymatic tailoring reaction is between a secondary metabolite and reactive oxygen, nitrogen, and sulfur species (ROS, RNS, and RSS). These molecules are common by-products of primary metabolism. Their ability to participate in non-enzymatic reactions is demonstrated by their documented reactivity with the proteome, which has been reviewed elsewhere (Keller et al., 2015).

Canonical enzymatic natural product biosynthesis utilize simple building blocks, such as O_2 for oxidation, therefore the delineation between non-enzymatic processes and enzymatic processes in complex pathways is difficult to make. What stands out to the authors is reactions that occur between these reactive species and mature natural products. In this context a mature natural product is one who has undergone a majority of its biosynthetic pathway. In many cases, the identification of these reactions is assisted by the co-isolation of the mature natural products and their non-enzymatic derivative.

The biosynthesis of the ammosamides is one of the most well studied cases of a pathway featuring a non-enzymatic tailoring reaction. After the original description of the ammosamide family in the extracts of a *Streptomyces* strain isolated from the Bahamian Islands (Hughes, MacMillan, Gaudencio, Jensen, & Fenical, 2009), it was found that the biosynthetic gene cluster for the ammosamides in *Salinispora* and *Streptomyces* did not contain a gene for an oxidoreductase that would catalyze the oxidation of ammosamide C to ammosamide B (Jordan & Moore, 2016). Relatedly, our group found that ammosamide C could easily react with several nucleophilic amines to form non-biogenic derivatives (Pan et al., 2013). These two independent

reports suggest that ammosamide C, which contains a reactive iminium functional group, was a precursor for a non-enzymatic derivatization to form the remaining ammosamides.

Ammosamides A, B, and E are single atom derivatives of ammosamide C (Figure 21), suggesting that these originated from reactive oxygen, sulfur, and nitrogen species produced by the bacteria. While the idea of having an iminium-based electrophilic sink to either neutralize toxins or respond to metabolic by-products is alluring, further biological studies are needed to suggest that the ammosamides are a functional tool. Function aside, the ammosamides are an ideal example of how identifying non-enzymatic mechanisms can be used to quickly and efficiently perform medicinal chemistry on biologically active compounds (Fenical, Jensen, Macmillan, Hughes, & Laclair, 2012; Hughes, MacMillan, Gaudencio, Fenical, & La Clair, 2009) as over 20 analogs were able to be synthesized using benign conditions and a variety of nucleophilic amine substrates (Pan et al., 2013).

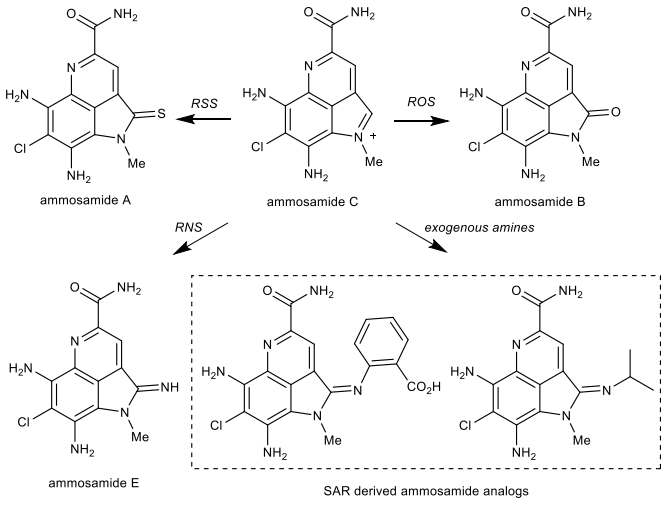


Figure 20. Non-enzymatic derivatization of the ammosamides

The second class of tailoring reactions involves the modification of mature natural products with higher molecular weight substrates. These substrates are commonly metabolic by-

products. Specifically, instead of reactions involving O_2, NH_2, or SH_2, the following reactions involve amino acids or their derivatives. While similar to heterologous multi-component reactions, tailoring reactions occur on the periphery of the molecule and commonly do not result in subsequent reactions.

The ansamycins are a well characterized and biomedically important family of macrolactams which include rifamycin and geldanamycin. Recently, Zhang and co-workers utilized the unique gene structure of the BGC for these family of molecules to search for the BGC in a collection of *Streptomyces* (Zhang et al., 2014). The pentaketides known as the juanlimycins were discovered in a bacterium associated with a Chinese plant of the *Kandelia candel* species. The final structures of juanlimycins A and B were postulated to be derived from a reaction between the thioesterification product intermediate **28** and aryl amines, anthranilate and 3-hydroxy-5-aminobenzoic acid (Figure 21). While the authors did not postulate or experimentally interrogate this reaction, the known availability and reactivity of such aryl amines in previous non-enzymatic mechanisms is intriguing.

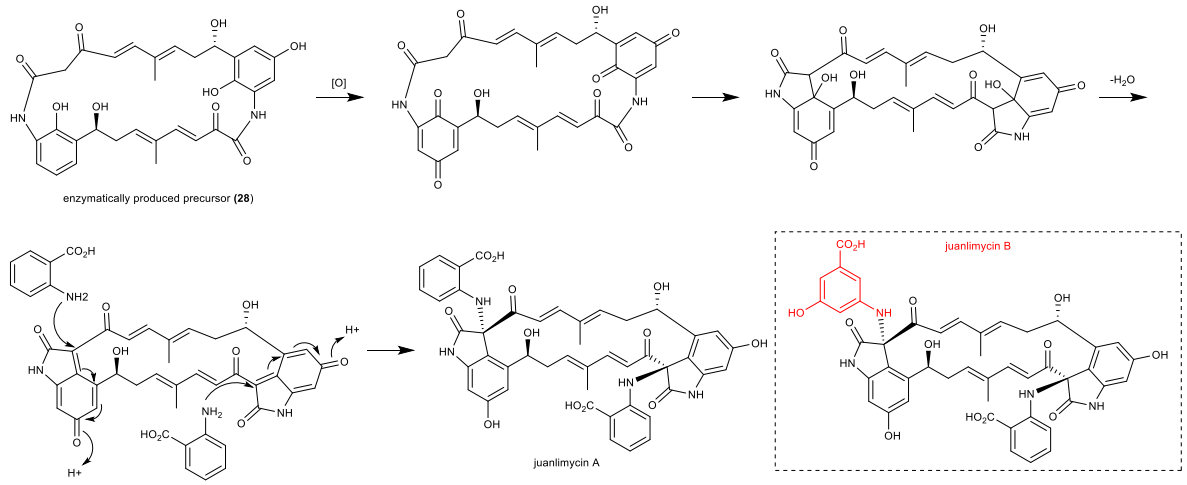


Figure 21. Biosynthetic proposal of juanlimycins A and B.

Recently, the Balskus laboratory published a biosynthetic interrogation of kanamycin-like molecules in *Streptomyces murayamaensis*, which featured several non-enzymatic steps (P. Wang, Hong, Wilson, & Balskus, 2017). Their studies focused on the construction of seongomycin and stealthin C from the enzymatically produced precursor for many angucycline polyketides, dehydrorabelomycin (Figure 22). During *in vitro* experiments, Wang and co-workers discovered that dehydrorabelomycin in the presence of pure enzymes AlpJ and Fre and L-cysteine formed stealthin C and the product, **29**. The authors proposed that the AlpJ and Fre oxidases catalysed a Baeyer-Villiger based C-C bond cleavage/ring contraction and subsequent oxidation, respectively, to produce an intermediate subject to non-enzymatic nucleophilic addition by thiols such as L-cysteine. The result of this non-enzymatic addition was proposed to be N-deacetylseongomycin, which could undergo an intramolecular S-N-type Smiles arrangement to form **29** that is subject to non-enzymatic oxidation and hydrolysis forming stealthin C. The non-enzymatic thiol addition was validated by incorporation of various thiols into new dehydrorabelomycin derivatives, demonstrating a lack of enzyme-like specificity. The non-enzymatic oxidation and hydrolysis of **29** to stealthin C was validated by boiling solutions of **29**, to degrade any present enzymes, before observing conversion to stealthin C.

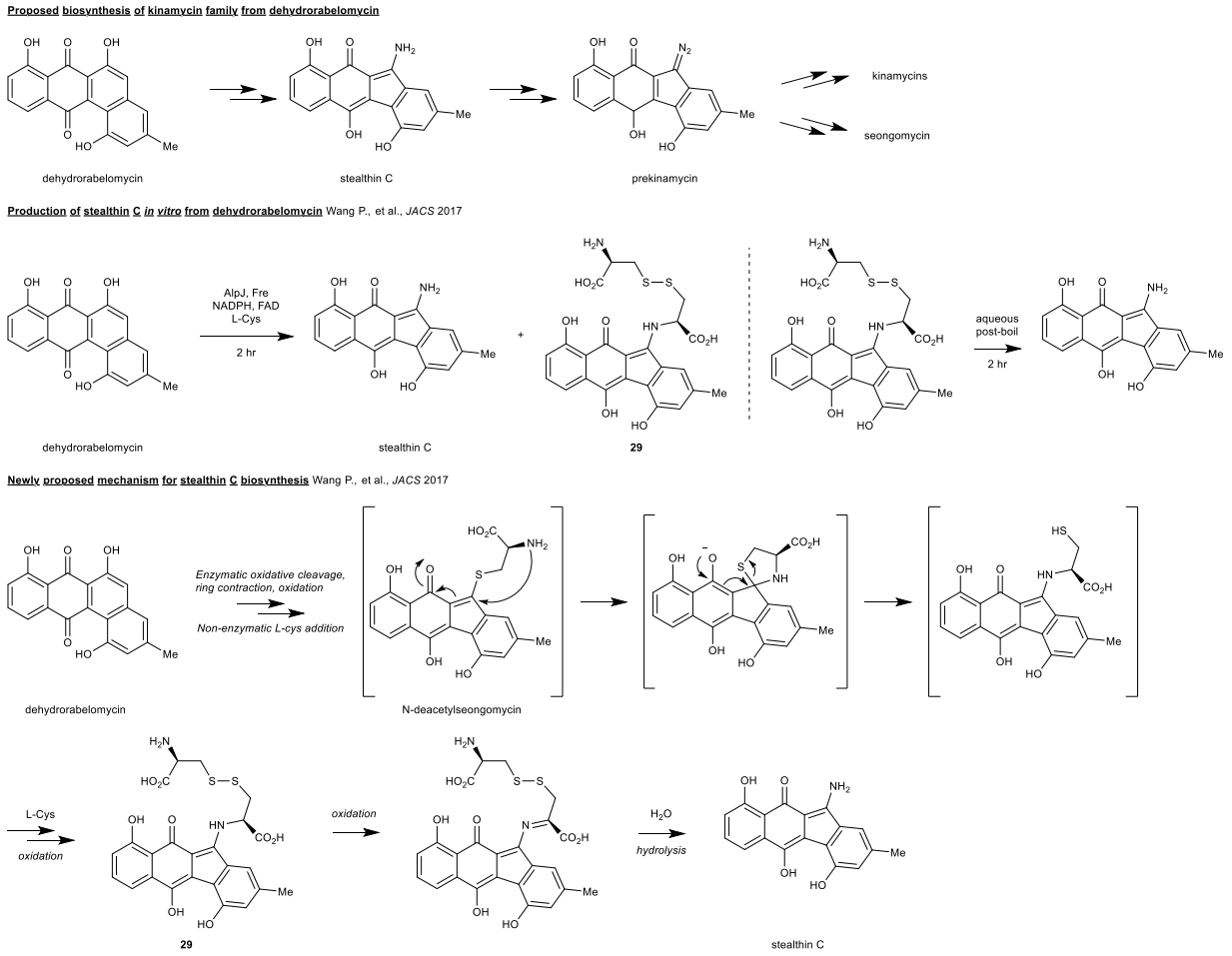


Figure 22. Progression of the stealthin biosynthetic hypothesis

The discovery that stealthin C is derived from non-enzymatic mechanisms called into question an important component of angucycline polyketide biosynthesis. It has been suggested that stealthin C is a biosynthetic intermediate of the kinamycins, based on ^{13}C based isotope feeding studies and the accumulation of stealthin C in heterologous expression of the partial kanamycin biosynthetic gene cluster (Steven J. Gould et al., 1996; S. J. Gould, Hong, & Carney, 1998; Steven J. Gould, Melville, Cone, Chen, & Carney, 1997). However, these experiments did not completely conclude the direct conversion of stealthin C to kinamycins. Furthermore, Wang et al. hypothesized that the reactivity of the non-enzymatic mechanism would make stealthin C

unstable as an intermediate in the kanamycin pathway. They validated this alternative hypothesis using ^{15}N-isotope labelling, suggesting that stealthin C is not an intermediate to the kinamycins, but rather a diverted pathway from the common dehydrorabelomycin that features non-enzymatic diversification.

The study of the stealthins by the Balskus lab highlight how studies of non-enzymatic steps can lead to broader scientific discovery. Identification and experimental validation of these non-enzymatic mechanisms provided insight to a biosynthetic pathway central to a large family of natural products. Also, the stealthins join the jadomycins as examples of angucycline polyketides that participate in non-enzymatic reactions. While it remains to be seen if other members of this family are constructed with non-enzymatic steps, the angucycline polyketides may represent a privileged scaffold for chemical reactivity.

1.4.7 Light induced reactions

The utility of UV induced transformations in synthetic chemistry is widely known (Bach & Hehn, 2011). However, its utilization in natural product biosynthesis is questionable as many times the ecology of the natural product source is not explored in the biomimetic synthesis from the aforementioned review on synthetic photochemistry. Additionally, a truly biomimetic source of light is rarely used, but rather a high intensity UV irradiation source is traditionally utilized to conduct these reactions. That being said, there are cases of ecologically relevant UV induced biosynthesis including the brevianamides which feature a Norrish-type cleavage, the biyouyanagins which undergo [2+2] cycloaddition, and most importantly Vitamin D3, which is formed after a UV induced rearrangement of 7-deoxycholesterol. These are formally discussed in the previous review of non-enzymatic biosynthesis by Gravel and Poupon (Gravel & Poupon,

2008). Here we will speak about several modern examples of UV induced non-enzymatic biosynthesis.

The aforementioned fuggedine, classified primarily as a homologous multi-component reaction or a dimerization, has been shown to be synthesized from its monomeric precursor by a UV light induced [2+2] cycloaddition, albeit in low yields (5%). Additionally, this dimer was only found on the leaves of the plant, not the roots suggesting this could be an ecologically relevant non-enzymatic reaction. However, because fuggedine differs from the rest of the 15 other securinega dimers in its construction mechanism, the whole family is mainly classified as non-enzymatic dimerizations with this single example classified secondarily as a light induced reaction.

When considering UV induced non-enzymatic biosyntheses, it is obvious that not all ecological niches are created equal. In the case of fuggedine the sequestering of the reactive monomer within the leaves is intuitive given the role of foliage in carbon fixation. Likewise, UV induced reactions have been implicated in marine invertebrates that are subject to continuous sunlight.

In the late 1970's Ireland and Scheuer were investigating the Hawaiian based sacoglossan mollusk *Placobranchus ocellatus* and their ability to assimilate chloroplasts from marine algae, when they discovered that a family of molecules featuring a γ-pyrone ring, which had previously been described from *Tridachia* and *Tridachiella* mollusks (Chris Ireland, Faulkner, Finer, & Clardy, 1979; Chris Ireland, Faulkner, Solheim, & Clardy, 1978), could undergo photochemical isomerization (C. Ireland & Scheuer, 1979) (Figure 23). Specifically, 9,10-deoxytridachione could isomerize to photodeoxytridachione (Figure 23). Most importantly, the authors found that

this could happen *in vitro*, as well as *in vivo* as detailed by radioisotope labeling experiments in which light exposed mollusks showed increased flux of ^{14}C from carbon dioxide to 9,10-deoxytridachione to photodeoxytridachione compared to mollusks kept in the dark. Their rationale was that sacoglossan mollusks, which lack a protective shell, could use byproducts of algae photosynthesis to generate γ-pyrones which could serve as sunscreen against harmful irradiation, a direct reflection of their environment.

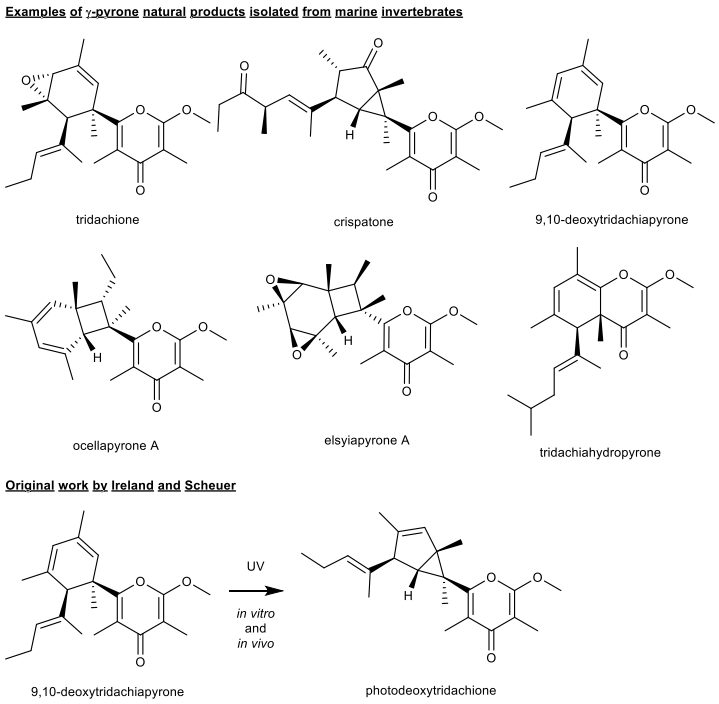


Figure 23. The γ-pyrone natural products and biomimetic studies

Over the last 40 years, more of these γ-pyrones have been discovered, expanding the chemical diversity of the family. However, in 2008 Eade and co-workers suggested that regardless of their diversity, most γ-pyrones could be derivatives of a singular precursor consisting of a conserved γ-pyrone heterocycle linked to serial (*E*)-polyenes (Eade et al., 2008). They hypothesized that this all-(*E*)-polyene could undergo a series of *Z*/*E* isomerizations, thermal or photochemical electrocyclizations, and [4+2] cycloadditions or [2+2] concerted

rearrangements to yield a variety of the γ-pyrone family members. The authors were able to demonstrate this with several of the family members, albeit not consistently under truly biologically relevant conditions (no exogenous catalysts or organic solvents). Interestingly, the very next year the same laboratory undertook a biomimetic synthesis of the complete set of tridachiahydropyrones and demonstrated that each could be derived photochemically from the precursor **30** (Sharma, Lygo, Lewis, & Moses, 2009) (Figure 24). Important for the biological context of these molecules, the authors showed that exposing the precursor **30** to "UK sunlight in the month of March for 3 consecutive, partially sunny, days" led to a 75% conversion to the three products, intermediate **31**, tridachiahydropyrone A, and phototridachiahydropyrone (ratio of 20:7:1). Furthermore, tridachiahydropyrone A irradiated for 24 hours led to quantitative conversion to phototridachiahydropyrone presumably via a light induced 1,3-sigmatropic alkyl migration (Figure 24).

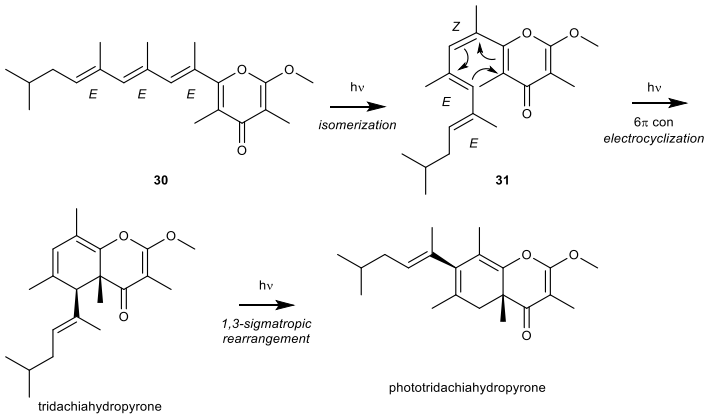


Figure 24. Proposed biosynthesis of phototridachiahydropyrone

While the extent to which these isomerizations, cycloadditions, and rearrangements occur non-enzymatically in the context of the marine invertebrates has yet to be seen, they remain an open inquiry on the matter. The authors of this review predict that many of the initial steps in

these biosynthetic routes occur enzymatically with isomerases and related proteins, but that the final steps in which UV light can be absorbed by γ-pyrones, particularly in the exposed dorsal body of sea slugs as originally demonstrated by the elegant isotope labeling experiments of Ireland and Scheuer can occur non-enzymatically.

Early work in the Fenical laboratory in the 1980's uncovered the family of macrocyclic diterpenes known as the pseudopteranes (Figure 25, top) (Bandurraga, Fenical, Donovan, & Clardy, 1982). In their original publication, the authors suggest that the structure could be derivative of a ring contraction of a related family of natural products referred to as the cembranes. Giving credence to this hypothesis, in 1998 Rodriquez and Shi isolated the known pseudopterane kallolide A and the novel furanocembranolide bipinnatin J from the Eastern Caribbean Sea plume *Pseudopterogorgia bipinnata* (Rodriguez & Shi, 1998*)*. Due to the presence of both a cembrane and pseudopterane in the same sea plume, the authors set out to determine if bipinnatin J, with a 14 carbon ring, could undergo a [1,3]-sigmatropic rearrangement as Fenical had suggested to yield the 12 carbon kallolide A (Figure 25, bottom). Most importantly for this review, the authors contemplated that UV light could provide the energy for the rearrangement. Indeed, incubation of the bipinnatin J molecule in UV light for several hours led to a respectable 40% stereoselective conversion to kallolide A. While the amount of ecological studies on the cembranes does not allow for comparison to the y-pyrones, it could serve a similar sunscreen-like function for these coral-like invertebrates.

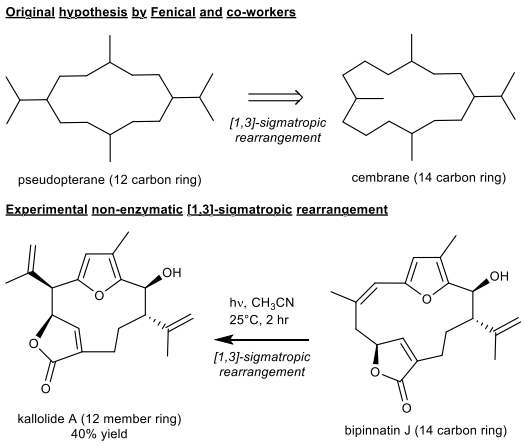


Figure 25. Biosynthetic hypothesis for pseudopterane and validation using bipinnatin J

The Trauner lab and others have utilized UV-induced photochemistry to convert cembrane molecules to a family member of the pseudopterane molecules known as intricarene by a 1,3-dipolar cycloaddition (Roethle, Hernandez, & Trauner, 2006; Tang, Bray, & Pattenden, 2006). Before 2014, the conditions to carry out this photochemistry were non-biological as they featured extremely high temperatures and organic solvents. However, in 2014 Stichnoth and co-workers undertook an elegant and serendipitous biomimetic synthetic effort to intricarene from the cembrane bipinnatin J, the same precursor used previously for the sigmatropic rearrangement photoconversion to kallolide A (Stichnoth et al., 2014) (Figure 26). The researchers decided to use a UV lamp commonly used in reptile terrariums as the light source and an aqueous acetone solvent system, an altogether more mild setting that previously established for bipinnatin-based synthetic efforts. The photochemistry was not performed directly on bipinnatin J, but rather an oxidized derivative (**32**), to yield intricarene in 25% yield. These experiments suggest that an enzymatically derived intermediate similar to **32** could undergo non-enzymatic 1,3-dipolar cycloaddition, albeit these findings remain to be justified *in vivo*.

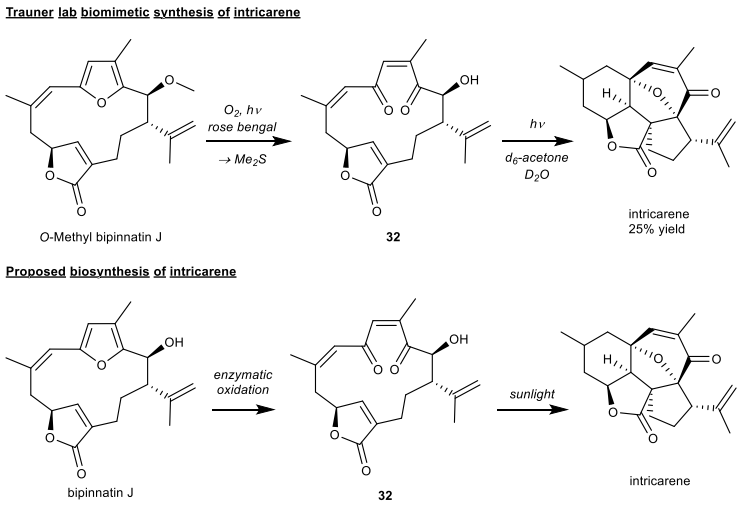


Figure 26. Intricarene biomimetic studies

The chamigrenes are a unique family of brominated sesquiterpenes of which there have been more than 50 molecules isolated from terrestrial and marine plants, as well as marine invertebrates. Dactylone (**33**) and aplydactylone (**34**) are two members of this family discovered in the sea hare *Aplysia dactylomela* in the early 2000's (Sergey N. Fedorov et al., 2001; Sergei N. Fedorov, Shubina, Kalinovsky, Lyakhova, & Stonik, 2000). In their *JACS* Communication, Fedorov and co-workers hypothesize that **34** could be derived from **33** via an enzymatic [2+2] cycloaddition. Attempts to cyclize these molecules under UV light were unsuccessful (Figure 28). Then, in 2016 the Burns lab of Stanford took on the challenge of coming up with a general approach to the enantioselective synthesis of the family of brominated chamigrenes and took this opportunity to examine the relationship between **33** and **34** (Burckle, Vasilev, & Burns, 2016). Conversion of (-)-**33** to (+)-**34** occurred under a variety of UV irradiation conditions, including an impressive 100% conversion after 15 minutes in THF irradiated at 254 nm. However, the more biologically relevant conditions, letting the compound sit in chloroform for 8 days in "Californian sunlight" lead to a modest 15% yield. These experiments indicated that the

cyclobutane present in **34** could indeed be formed non-enzymatically during exposure to UV light.

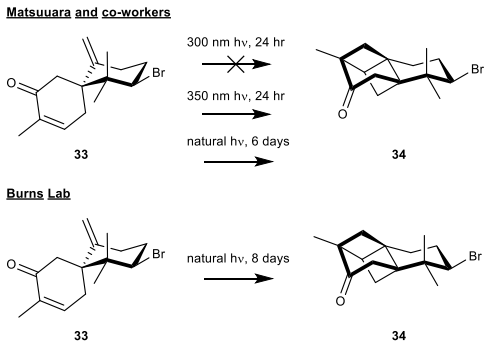


Figure 27. Biomimetic studies of aplydactylone

In late 2016 Matsuura and co-workers defined the mechanistic characteristics of the non-enzymatic UV driven [2+2] cycloaddition reaction of **33** to **34** as well as the potential electrocyclization of the diastereomer 10-epi-dactylone (Matsuura, Kolle, Trauner, de Vivie-Riedle, & Meier, 2017). After synthesizing their respective precursors, the researchers took great care in assessing the type of UV light utilized in their photochemical reactions, selecting ranges of light that would be available in natural sunlight. Subjecting **33** and 10-epi-dactylone to 350 nm UV light resulted in a complete conversion over 24 hours, while 300 nm light led to degradation of the reagents. Additionally, 6 days of "Munich sunlight" led to the conversion of **33** to **34**, further demonstrating that the [2+2] cycloaddition could occur non-enzymatically in response to biological solar exposure. An important component to these studies was the justification for cyclobutane ring formation over a more relaxed cyclopentane structure and the authors utilized detailed quantum calculations for these determinations that will not be discussed here. Nevertheless, these works add to the growing examples of mechanistic studies being utilized to justify exactly how these non-enzymatic biosynthetic reactions are occurring.

1.5 Non-biologically relevant reactions

Generation of new metabolites via exploitation of non-enzymatic reactions is a potential approach to increase the chemical diversity of natural products. However, there lies an important distinction between processes such as these and natural biosynthetic approaches to produce biologically functional molecules non-enzymatically. There is a dichotomy of utility here: a medicinal chemistry perspective suggests that the circumstances of nature do not matter in the push to expand a chemical repertoire. A biosynthetic perspective suggests that the biological reconciliation of a non-enzymatic reaction is relevant. It is our belief that both perspectives are pipelines to impactful science.

Validation is imperative to claims of non-enzymatic biosynthesis, as highlighted by the case of the elansolids. Elansolids A1 and A2 do not require exogenous sources for formation and were isolated in high yields from the fermentation. Discovery of the isolation artifacts, elansolid B1-B3, provided evidence that a non-enzymatic reaction could occur with various substrates. This metabolome suggested that an intermediate existed that predominantly reacts with endogenous substrates, but could also form artificial derivatives during isolation. Elansolid A1 and A2 provide a survival function to their host bacteria as antibiotics therefore the non-enzymatic formation is beneficial, but can still result in artifacts under lab conditions. The authors experimentally validated this hypothesis by characterizing the intermediate, elansolid A3. Knowledge of elansolid A3 was then exploited in a medicinal chemistry effort and a biosynthetic investigation in the producing organism. Therefore, this case exemplified how d*etermining the presence these reactions, endogenous or not, is imperative to further exploration.*

At the core of this story is the proper validation of non-enzymatic reactions. Specifically, the two major constituents of any claim are mechanism and function. It is paramount that one demonstrate analytically that a reaction can occur in solvent free of protein catalysts. To the second point, an effort must be made to determine that the reaction serves a biological benefit to the organism. This pursuit is much more difficult and is commonly attempted with indirect qualitative data. An example of these types of indirect validation is the rubrolones. A non-enzymatic reaction is required to generate the substrate for downstream glycosylation. As glycosylation is needed for antibiotic activity, the non-enzymatic pyridine formation is necessary for the biological function of the rubrolones. This suggests that this non-enzymatic reaction can truly be considered a component of the biosynthesis, rather than a side product. Regardless of their complexity, mechanistic and functional studies should be pursued for all proposed non-enzymatic biosynthetic reactions.

1.6. Conclusion

1.6.1 Discussion of common precursors

Just as enzymes have evolved to improve the efficiency of biologically important reactions, it is possible that chemical reactivity is inherent in particular substrates. An attractive hypothesis for such design is the re-purposing of primary metabolites for secondary metabolite function that would eliminate the need for a new set of anabolic genes. Evidence for this claim is shown by the re-occurrence of several substrates in diverse biosynthetic pathways.

Anthranilic acid appears in the jadomycins, discoipyrroles, and rubrolones. Additionally, this compound has been shown to be reactive with the elansolids and the ammosamides albeit in the formation of non-endogenous metabolites (Figure 29). The presence of anthranilic acid is not

rare as it is a common byproduct of primary metabolism via the kynurinase pathway and the tryptophanase pathway, however it might be possible that upregulation of anthranilate which results in an environmental fitness could allow for said upregulation to be conserved. The power of anthranilate as well as other nucleophilic amines lies in their dual function as 1) initiators of new bonds via nucleophilic addition into carbonyls (discoipyrroles, jadomycins, juanlimycins, rubrolones, and oxazinins) and 2) generators of secondary reactive sites including electrophilic moieties (imine and iminium carbons as seen in discoipyrroles and oxazinins) and secondary nucleophiles (secondary amines as seen in discoipyrroles, oxazinins, and rubrolones).

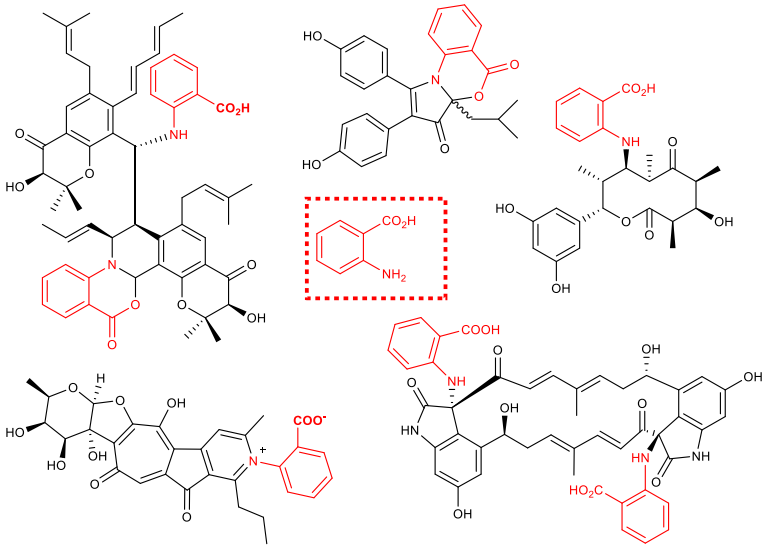


Figure 28. Anthranilic acid in non-enzymatic biosyntheses

Another common reactive functionality found in non-enzymatic natural product biosynthesis are polyenes, particularly in the context of electrocyclization reactions as discussed in the intramolecular, dimerizations, and light induced categories. Obviously, the power of olefins in synthetic chemistry as a whole is well known and the amount of enzymatic processes involving conjugated systems of double bonds is overwhelming. Therefore it is inaccurate to say that this functional group is a signature of non-enzymatic processes. However when mature

natural products with multiple double bonds are found alongside heavily cyclized natural products in extracts of ecologically sensible organisms, it is important to interrogate the relationships between those reactions and their biological function. The authors of this review argue that any light based non-enzymatic reaction shown to occur in a biological setting reveals a potential functional role for these natural products as molecular sunscreens or radical scavengers, a relationship that would bolster an argument for conserved non-enzymatic biosynthesis.

A third fundamental functional group found in both non-enzymatic biosynthesis and organic synthesis are the α,β-unsaturated ketones. These functional groups are encountered as both nucleophilic components and electrophilic components commonly as Michael acceptors. As with the polyenes, these functional groups are commonly derivative of polyketide biosynthesis suggesting that most reactions should be enzymatically driven. However, conjugated cyclic systems especially are indicative of mature or late stage natural products that can engage in non-enzymatic reactions.

The last two instances of functional groups subject to non-enzymatic chemistry demonstrate the intimate relationship between synthetic chemistry and natural products biochemistry. It is no surprise to any chemist that reactions that develop in biological systems in the absence of protein catalysts are the same as those that are integral to performing basic organic chemistry. This is the most important reasoning for the amount of non-enzymatic, concerted reactions that are discovered through the eyes of fantastic synthetic chemists. Beyond the simple classification of reactions as enzymatic and non-enzymatic, it is important to utilize these synthetic tools to identify reactions that can be applied in total synthesis or medicinal chemistry, two fields that directly benefit biomedicine.

In addition to the clinical aspect of identifying non-enzymatic reactions, biological insights can be gained from studying these mechanisms. This leads us to the discussion of the common functional group that is harder to define, that of the biological toxin. This was exemplified by four major stories within this review: the dibohemamines, homodimericin A, the tridachiahydropyrones, and the maximiscins. Each of these natural products are derived from a presumed biological detoxification of common biological toxins whether they be formaldehyde, reactive oxygen species, UV light, or general toxins (Figure 29). The nuances of their biological functions require further mechanistic studies, although the preliminary data for each of these natural products is certainly intriguing. These also are informative for future studies. When examining natural products biosynthesis, it is imperative to consider the incorporation of these common toxins amongst others in the absence of genetic information regarding their source.

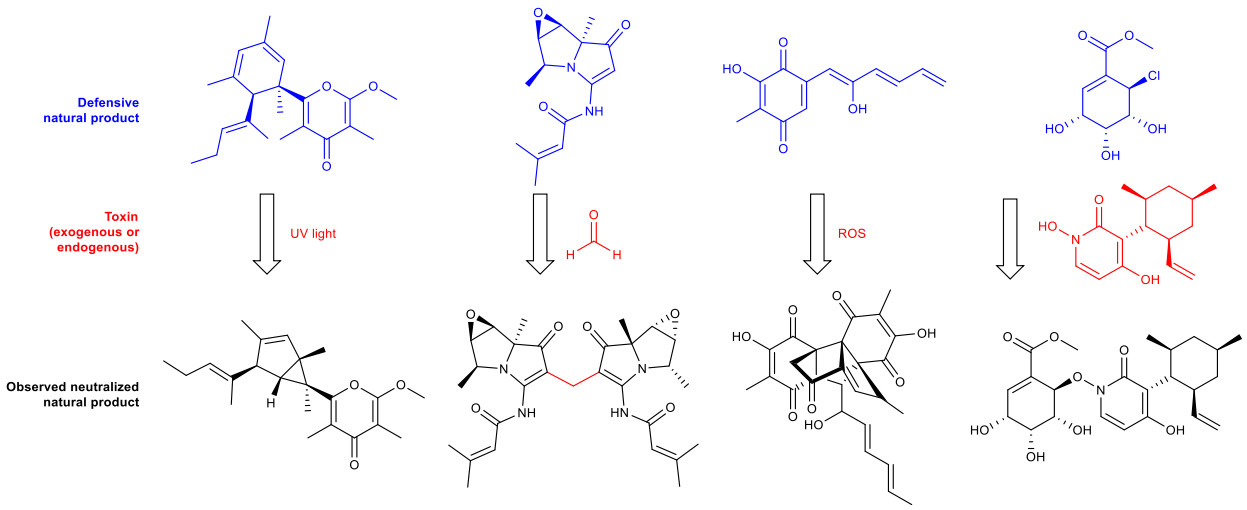


Figure 29. Potential biological applications of non-enzymatic reactions

1.6.2 Final remarks

Although non-enzymatic mechanisms have been relevant historically, the number of occurrences in recent years demonstrates the need for a conscious effort for the accurate and

mechanistic assessment of these reactions in natural products biosynthesis. Reactions ranging from synthetically important reactions such as intramolecular electrocyclizations to novel multi-component reactions that participate in biological functions all can be arranged within the concept of non-enzymatic chemistry.

In conclusion, the authors believe it is most important to understand the scope of these reactions. *Non-enzymatic biosynthetic steps are the vast minority of construction mechanisms for natural products*. The superior efficiency, regulation, and complexity of protein machinery cannot be overstated. Clearly, enzymatic processes have evolved over time to be responsible for the chemistry we isolate on a daily basis. However, for a holistic understanding of how organisms use simple building blocks to construct complex molecules every detail, as minute or rare as it may be, must be considered. Furthermore, as is seen in many eukaryotic systems such as the non-enzymatic acetylation of proteins (Kim et al., 2006), detoxification system of glutathione (BAEZ, SEGURA-AGUILAR, WIDERSTEN, JOHANSSON, & MANNERVIK, 1997; Kolm, Danielson, Zhang, Talalay, & Mannervik, 1995; Satoh, 1995), and oxidative regulation of the HIF system (Tarhonskaya et al., 2014), there exists the possibility that the inherent reactivity of the biological milieu is used to convey important signals for cell survival.

CHAPTER TWO

MECHANISTIC STUDIES OF THE DISCOIPYRROLE BIOSYNTHESIS

2.1 The discoipyrroles: novel marine natural products

As part of an ongoing effort to find biologically active natural products from the MacMillan collection of marine bacteria, our lab recently carried out a screen known as FuSiOn that will be discussed in Chapter 4. Bioactivity guided isolation resulted in the isolation of the novel compounds, discoipyrrole A-D. These compounds are characterized by their 3-pyrrolidone based core which is tetra substituted with various aryl and alkyl groups. After the isolation of these compounds we began a dual pronged approach to studying both their biosynthetic origins and their biological activity in the context of non-small cell lung cancer, which will be the subject of Chapter 4.

2.2 Biosynthetic studies of the marine bacteria *Bacillus hunanensis*

The MacMillan laboratory microbe collection is derived from the sediment of various marine environments including shallow and deep salt-water sources. Locations include the Bahamas, Tonga, and the North American Gulf Coast. Originally, the discoipyrroles were isolated from the organic solvent extract of a 7-day bacterial fermentation. The producing organism was isolated from an inland lagoon of the isle of Little San Salvador, Bahamas. This bacteria belonged to the species *Bacillus hunanensis* and was given the classification of SNA-048 as dictated by our canonical lab naming protocols.

The compounds 4-hydroxysattabacin (**35**), anthranilic acid (**36**), and 4-hydroxybenzaldehyde (**37**) were also found in the excreted metabolome of SNA-048 (Figure 30). The structural similarities between **35**-**37** and the discoipyrroles suggested that these compounds

could serve as substrates for discoipyrrole biosynthesis. Interestingly, the relative concentrations of these potential substrates in the media were higher than the discoipyrroles, as evident by UV absorbance and yields after isolation. Therefore, we hypothesized that **35-37** were secreted or diffused into the extracellular environment at which point they could participate in discoipyrrole biosynthesis.

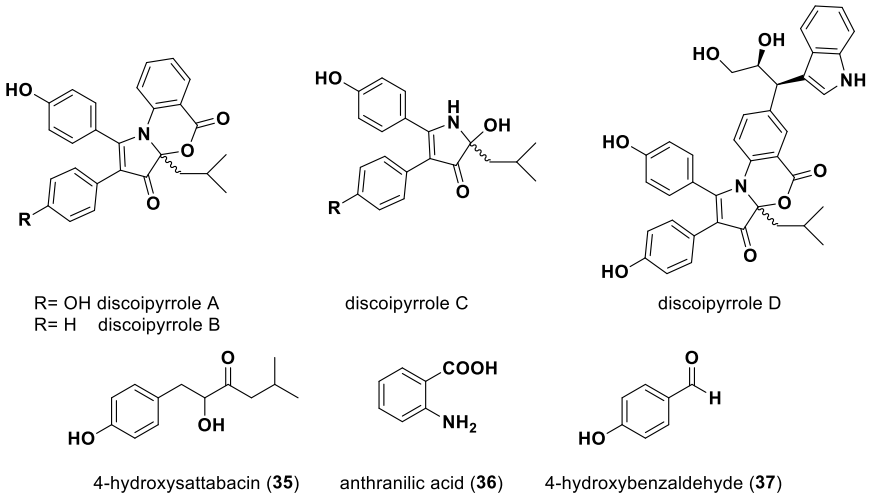


Figure 30. Compounds produced by SNA-048

We first tested the involvement of factors in the fermentation media to promote a potential reaction between the three compounds (Y. Hu et al., 2013). Two factors proposed to be involved were secreted proteins and the unique chemical environment induced by fermentation. Therefore we sought to emulate a post-fermentation condition that was depleted of proteins to individually assess these factors. This was accomplished by collecting the media from a 7-day fermentation of a related *Bacillus* species from our microbe collection. Large macromolecules were removed via filtration through a 0.2 um filter. To degrade potential proteins, the media was boiled for 10 minutes resulting in so-called spent media. Theoretically, this media did not contain macromolecules or functional proteins while maintaining similar small molecules and trace

elements. The compounds **36** and **37** were purchased from a commercial source (Sigma), while **35** was purified from SNA-048 crude extract. These molecules were incubated in spent media and product accumulation was measured via single-ion monitoring (SIM) of the molecular mass of discoipyrrole A. Surprisingly, discoipyrrole A was formed from **35-37** in the spent media. This indicated that the presence of proteins was not necessary for the biosynthetic reaction, but that the chemical environment could be specialized for allowing such a non-enzymatic reaction.

The evidence for a non-enzymatic biosynthesis of the DPs from **35-37** led us to investigate the finer details of the biosynthetic mechanism. Mass spectrometry unveiled no obvious intermediates, suggesting a concerted reaction mechanism. While non-enzymatic reactions are not uncommon in natural products biosynthesis (reviewed in Chapter 1), detailed studies of complex multi-component mechanisms, such as the one implicated in DP formation, are limited. By studying this mechanism, we hoped to gain insight into why these individual chemical moieties were reactive and how these reactivity's led to the structural complexity of the DPs without the influence of enzymes.

2.2.1 Origins of metabolites secreted by SNA-048

To gain a more complete understanding of DP biosynthesis, we studied the origin of the substrates **35**-**37**. Potentially, the source of these compounds could inform us to the function of the non-enzymatic extracellular reaction they took part in.

Anthranilic acid (**36**) is ubiquitous in primary metabolism, namely amino acid catabolism and anabolism (KEGG Compound C00108) (Kanehisa & Goto, 2000; Kanehisa et al., 2014). Specifically, **36** can be produced by the enzymatic breakdown of tryptophan through various mechanisms, most of which involve indole intermediates (Kurnasov et al., 2003; Sharma et al.,

2009). Feeding into the fermentation of SNA-048 with L-tryptophan (1 g/L) provided a 22 fold increase in excreted **36** (Figure 31) suggesting that indeed tryptophan was a source of **36** in this bacteria. Based on the observation that SNA-048 produced higher quantities of **36** compared to alternate *Bacillus* bacteria in our soil microbe library, we hypothesized that a supplemental mechanism might be driving the anthranilic acid production. A second, less common mechanism for production of **36** from L-Trp is directly via the oxidative kynurenine pathway in three steps, similar to mammalian systems (Aziz et al., 2008). Analysis of our annotated SNA-048 genome indicated that the kynurenine gene cluster was present, a potential explanation for the increased source of excreted **36**.

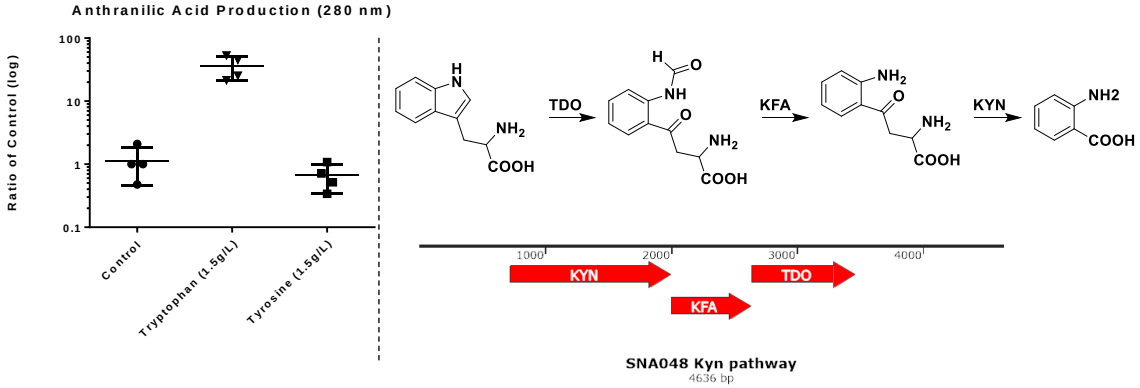


Figure 31. Anthranilic acid biogenesis

The compound 4-hydroxysattabacin (**35**) is part of a growing family of acyloin-containing molecules that are products of condensation between two amino acids. This family includes the circumcins, kurasoins, soraphinol C, and xenocycloins (Figure 32) (Lampis et al., 1995; X. Li et al., 2008; Lin et al., 2013; Park et al., 2014; Proschak et al., 2014; Uchida, Shiomi, Inokoshi, et al., 1996; Uchida, Shiomi, Sunazuka, et al., 1996). The construction of these simple

molecules occurs via decarboxylative condensation of amino acid pyruvates. Acyloin dimers have also been implicated as key intermediates in complex natural product biosynthesis as in the case of the UV-protectant scytonemin A by cyanobacteria (Balskus & Walsh, 2008).

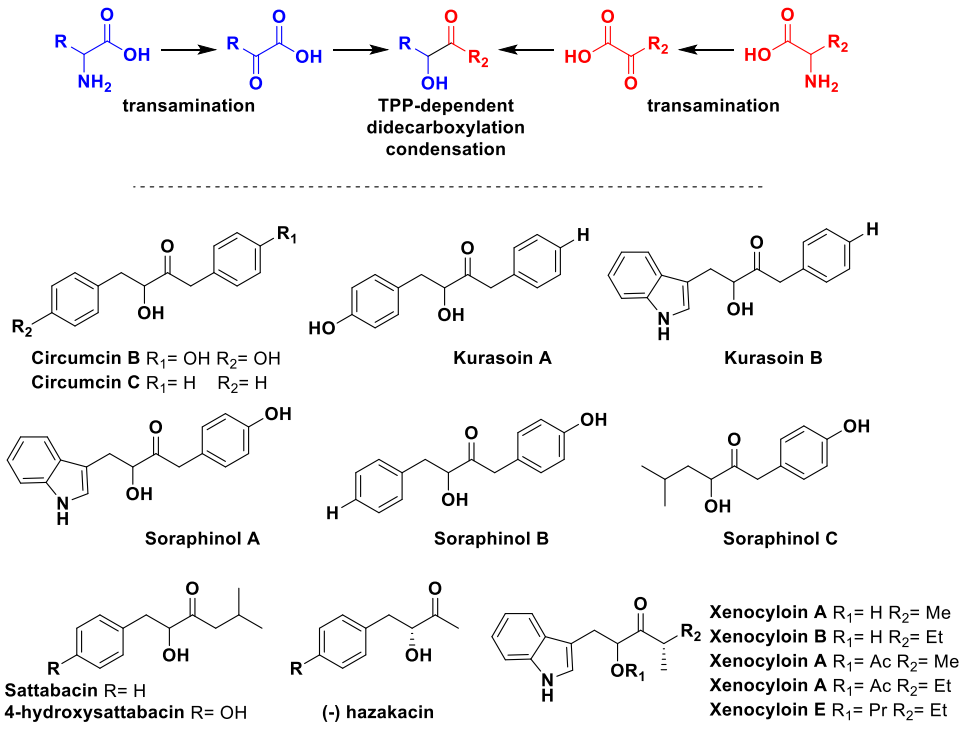


Figure 32. Acyloin containing amino acid dimers

The TPP dependent decarboxylase, or acyloin synthase, capable of performing these condensations has been cloned, purified, and tested *in vitro* (Balskus & Walsh, 2008; Gocke et al., 2007; Park et al., 2014). Recently, Park et al. thoroughly characterized the protein Thzk1050 from the thermophile *Thermosporothrix hazakensis* that produces sattabacin from phenylalanine and leucine *in vitro (Park et al., 2014)*. Sattabacin differs from **35** by a single aromatic hydroxyl substitution. We hypothesized that a homologous enzyme in SNA-048 could carry out this

reaction using tyrosine instead of phenylalanine to produce **35**. Using the published gene sequences of TPP dependent decarboxylases (kdcA, scyA, and thzk1050)(Balskus & Walsh, 2008; Gocke et al., 2007; Park et al., 2014; Smit et al., 2005) we performed PSI-BLAST analysis of the annotated genome of *B. hunanensis* SNA-048. Three genes showed high sequence homology to TPP dependent decarboxylases. Based on the presence of these putative biosynthetic genes in SNA-048, feeding studies were performed with [2-^{13}C]-L-tyrosine and [2-^{13}C]-L-leucine. Mass spectrometry confirmed their incorporation into **35** (Figure 33). These studies validated that *B. hunanensis* SNA-048 could produce 4-hydroxysattabacin from L-tyrosine and L-leucine, likely through the acyloin synthases found in our genomic analysis. Recombinant expression and *in vitro* study of the three gene products is required to confirm the identity of the **35** acyloin synthase.

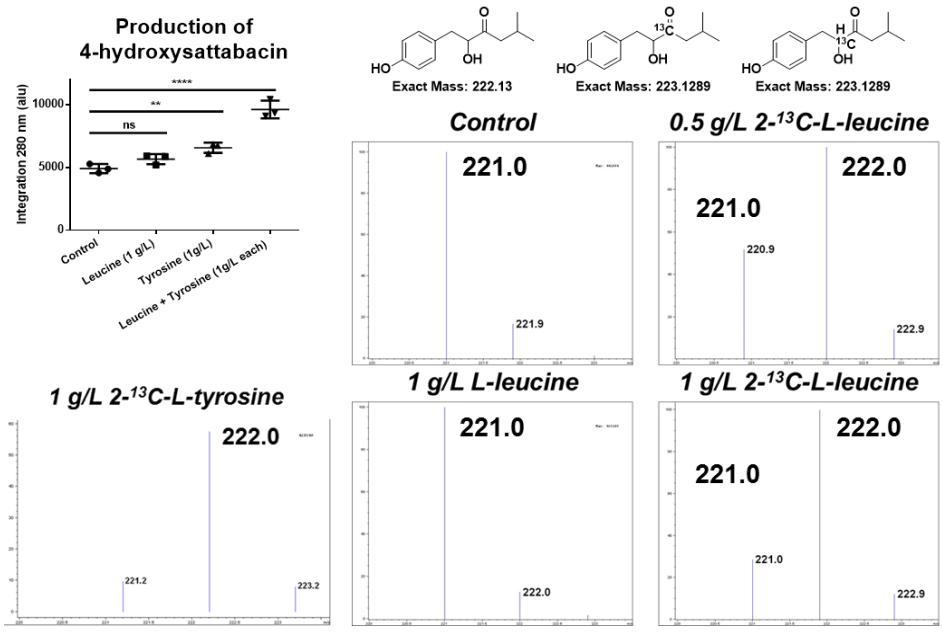


Figure 33. Biosynthesis of 4-hydroxysattabacin from amino acids

The final substrate, 4-hydroxybenzaldehyde (**37**) is derived from oxidative degradation of metabolites such as L-tyrosine and toluene (KEGG Compound C00633) (Kanehisa & Goto,

2000; Kanehisa et al., 2014). It has been identified in bacteria, fungus, and algae as a byproduct of primary and secondary metabolism (Bhat & Vaidyanathan, 1976; Fenical & McConnell, 1976; Jamaluddin, Rao, & Vaidyanathan, 1970). In addition **37** can be produced from degradation of oxidized 4-hydroxysattabacin when left at ambient temperature for a period of time. In a dose dependent manner, feeding of L-tyrosine increased levels of **37** in SNA-048 fermentations (Figure 34). This suggests that **37** is derived from tyrosine via a degradation pathway. However, as **35** also is derived from tyrosine there is a possibility that increased tyrosine leads to increased **35** production which provides a larger pool of **35** to degrade into **37**.

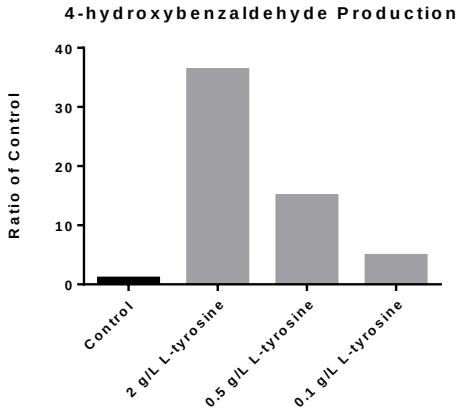


Figure 34. Quantification of 4-hydroxybenzaldehyde in SNA-048 crude extract

Study of the three substrates **35-37** did not reveal an obvious connection between their individual biosyntheses. The genes for the putative acyloin synthases, kynurinase or tryptophanase gene clusters, and tyrosine catabolic pathways were not oriented together in the SNA-048 genome. This suggested that the non-enzymatic formation of the DPs is not due to a canonically organized BGC. The biological role of the DPs, if any, to bacterial systems is unknown. Therefore, it is reasonable to suggest that their formation is nonfunctional. The inherent reactivity of **35-37** could be used for toxin neutralization in the case of the maximiscin

biosynthetic substrates (L. Du et al., 2016). In this scenario, the laboratory fermentation might lead to the non-endogenous formation of the DPs due to the lack of an ecological threat. However, it remains to be seen whether these compounds have a role in eukaryotic cells. If a role does exist it could serve a protective function to *Bacillus* bacteria in their natural habitat.

2.2.2 Development of a biomimetic synthetic system

While their biological function to SNA-048 remained enigmatic, the mechanism of the non-enzymatic reaction between substrates **35**-**37** was of great interest. To construct the core pyrrole of DPA, these substrates were thought to undergo 4 covalent bonding events and multiple changes of oxidation state. Conventional non-enzymatic, multi-component reactions occur on late stage intermediates of a biosynthetic pathway and involve periphery functionalization (Rix et al., 2004). Therefore, studying DP formation with mechanistic detail was of great significance to our understanding of non-enzymatic reactions in biosynthesis.

To investigate the formation of the DP scaffold we utilized our previously established synthetic model reaction system in organic solvent. This model takes advantage of starting with the oxidized 4-methoxysattabacin (**38**) (Figure 35), the formation of which we have previously found to be the rate-limiting step in both the fermentation media and organic solvent. The resulting analog, **39**, differs from DPA by a single methylation of the 4-hydroxysattabacin-derived aromatic ring. The model system removes analytical barriers associated with complex fermentation media. Past successes using NMR active isotope labels for studying reactions through use of kinetic isotopes (Cleland, 2005; Gomez-Gallego & Sierra, 2011), to elucidate biosynthetic pathways (Esquenazi, Jones, Byrum, Dorrestein, & Gerwick, 2011; Mahmud, 2007; Rinkel & Dickschat, 2015), and in the identification of natural products (Bode et al., 2012;

Kwon, Park, Shin, & Oh, 2014; Vizcaino & Crawford, 2015) led us to utilize ^{13}C and ^{15}N labeled substrates to probe the multi-component reaction by NMR and MS.

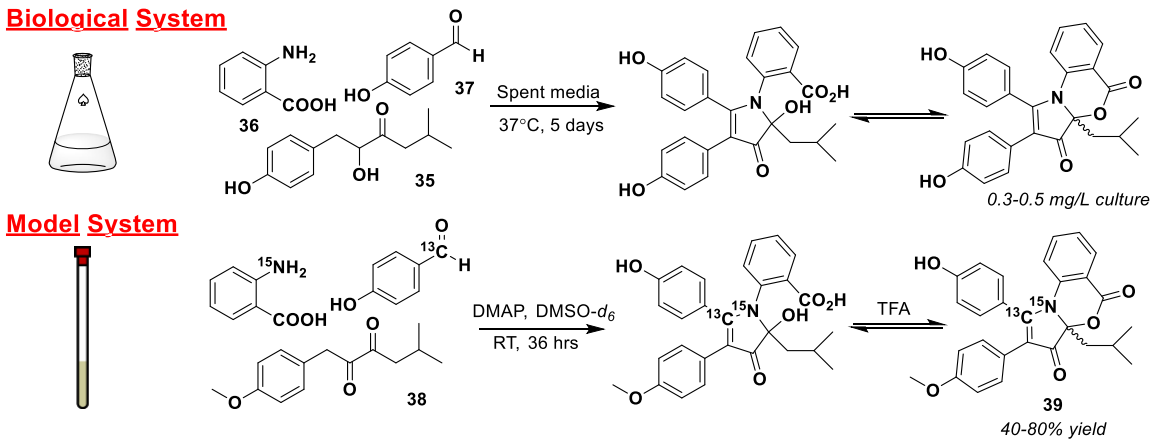


Figure 35. Model system of DP biosynthesis

2.2.3 Establishing method for examining reaction in real time

To track the formation of reaction intermediates, the chemical shifts and heteronuclear correlations of key carbon and nitrogen atoms around the pyrrole core were observed in real time. Utilizing [1-^{13}C]-*p*-hydroxybenzaldehyde and ^{15}N-anthranilic acid allowed for observation of heteronuclear NMR correlations from these isotope labels during each of the key bond forming steps.

Experimentally we began by carrying out the multi-component reaction with ^{15}N-anthranilic acid and conducted continuous reaction monitoring in 30 minute intervals for 36 hours using ^{1}H-^{15}N HMBC. This methodology provided evidence of C-N bond formation through developing ^{1}H-^{15}N correlations, specifically to downfield ^{1}H signals of the isobutyl chain of **38**. Due to the spectral range of ^{15}N NMR, the ^{15}N chemical shift of anthranilate-derived intermediates provided key structural information on the nature of the C-N bond of an intermediate (Figure 36)

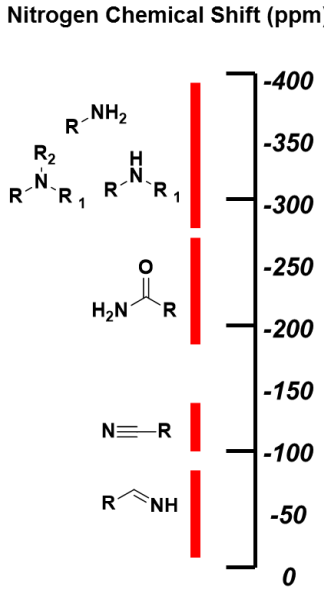


Figure 36. Chemical shifts of ^{15}N NMR

The first experiment performed provided a time frame for key ^{15}N containing intermediates (Figure 37). For example, it was determined that the first major intermediate (red box in Figure 37) appeared two hours into the reaction and was not present by 12 hours. The second major intermediate (blue box in Figure 37) appeared at 4 hours and was absent by 14 hours. From these types of information, additional 2D NMR experiments, such as ^{1}H – ^{1}H COSY, ^{1}H – ^{13}C HSQC, and ^{1}H-^{13}C HMBC were used to elucidate the structural identity of the time sensitive intermediates (Figure 37).

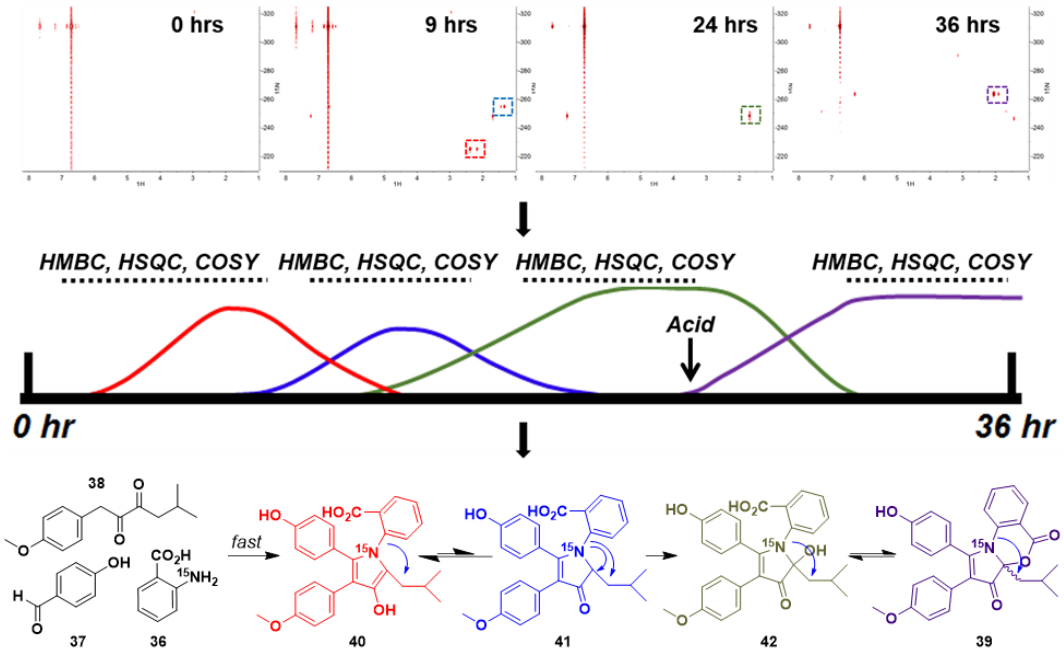


Figure 37. Reaction timeline of DP formation

2.2.4 Tracking reaction intermediates using enriched NMR-active isotopes

The starting material, ^{15}N-anthranilic acid, had a ^{15}N shift of -311.1 ppm and a strong correlation to aromatic protons at 6.71 and 7.67 ppm. Immediately, after the three reagents were added to the NMR in DMSO-d_6 we saw ^{1}H-^{15}N HMBC correlations from a ^{15}N shift of -225.0 ppm to protons at 6.96, 2.17, and 2.38 ppm appear, suggesting bond formation between **38** and **36**. This was validated by further COSY and HSQC correlations that established the ^{1}H's at 2.17 and 2.38 ppm to be the methylene of the isobutyl side-chain of the discoipyrroles. The key COSY correlations were 2.17 and 2.38 ppm to a septet at 1.28 ppm that was further coupled to methyl doublets at 0.62 and 0.70 ppm. The slightly downfield chemical shift of the diastereotopic methylene and the ^{15}N chemical shift indicated an electron withdrawing environment, potentially suggesting a pyrrole ring. This was confirmed by ^{1}H-^{13}C HMBC correlations from methylene protons to carbons at 127.0 and 138.5 ppm consistent with structure

of **40**. The absence of ^{1}H-^{13}C HMBC correlations between these carbons and exchangeable NH ^{1}H signals further suggested the formation of the pyrrole ring.

A weaker ^{15}N signal of -254.9 ppm appeared with correlations to an aromatic proton at 6.95 ppm and aliphatic signals at 4.40 and 1.34 ppm. Using similar strategies as described for **40**, correlations between these signals and to the isobutyl chain led to structure **41**, the ketone tautomer of **40**.

After 16 hours, the ^{1}H-^{15}N HMBC correlations of **40** and **41** decrease, while a third ^{15}N signal at -248.3 ppm appears with correlations to protons at 7.24 and 1.69 ppm, representing compound **42**. Up to 24 hours into the reaction, this signal existed independently and continued to build in intensity. Due to the apparent stability and abundance of the intermediate an aliquot of the reaction was removed and subjected to LC/MS analysis. The retention time and mass indicated that compound **42** was the open form of discoipyrrole that exists in pH dependent equilibrium with **39**. This reaction was repeated at a later time to yield larger quantities of **42**, which was isolated and characterized in full (Table 9).

After 24 hours, 1% TFA was added to the NMR tube and monitoring via ^{1}H-^{15}N HMBC was continued. In the first NMR scan after acid addition, two ^{15}N species at -251.3 ppm, with ^{1}H correlations to 7.29 and 1.67 ppm, and -246.4 ppm, with proton correlations to 7.61, 1.75, and 1.44 ppm, appeared. Within minutes, a major ^{15}N species of -263.8 ppm with correlations to ^{1}H of 6.29, 2.06, and 1.91 ppm appeared. This was assigned as the final product, **39**.

2.2.5 Examining spontaneous oxidation of pyrrole core

These experiments demonstrated that the formation of the pyrrole core of the discoipyrroles occurs rapidly to produce the first major intermediate, **40**, which persists in

equilibrium with the tautomer **41**. The loss of oxygen indicated that formation of the C-N bond in **40** induces an elimination of water. Subsequent oxidation of the pyrrolone ring to yield **42**, which under exposure to acid undergoes cyclization to **39**. To confirm this mechanism, we carried out the formation of **42** using the same conditions under an atmosphere of $^{18}O_2$. Monitoring of the reaction by LC-MS led to observation of a peak with an *m/z* $[M - H]^-$ of 475.2, a +2 shift from canonical **42** as evident by the control reaction run with $^{16}O_2$ (Figure 38). Acid free isolation and high resolution mass spectrometry of the ^{18}O-labeled product gave an experimental value of an *m/z* $[M + H]^+$ of 477.1923 (expected $[M + H]^+$ *m/z* = 477.1924). The isolated ^{16}O control product gave an experimental value *m/z* $[M + H]^+$ of 475.1881 (expected $[M + H]^+$ *m/z* = 475.1881). This oxidation could be facilitated by the hydroxyl group or carboxylic acid of **37** or **36**, respectively. However the syntheses of a diverse amount of discoipyrrole analogs with various alkylated benzaldehydes and anthranilate substrates proceeded with high yields, arguing against this hypothesis.

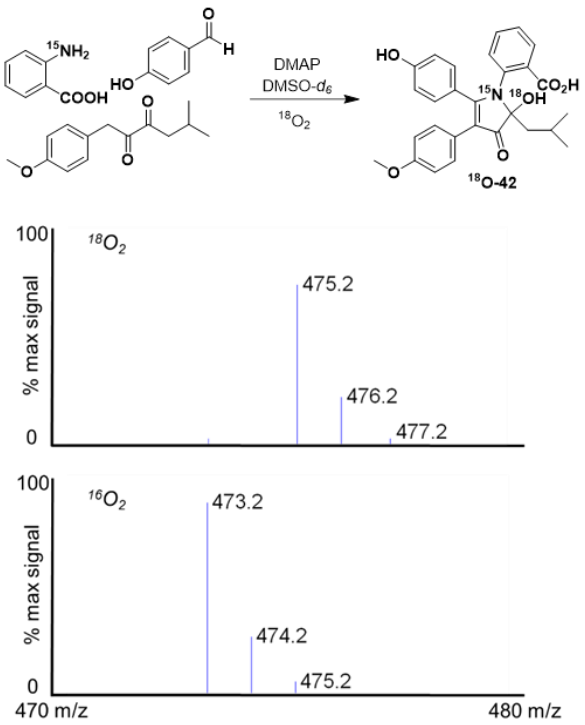


Figure 38. MS traces of reactions performed under $^{18}O_2$ or $^{16}O_2$

Isolation of the ^{18}O-**42** allowed us to explore the acid-induced ring closure. Surprisingly, exposure to TFA led to an isolated product with *m/z* $[M + H]^+$ of 457.1774 (expected $[M + H]^+$ *m/z* = 457.1776) that indicated the loss of ^{18}O label upon ring closure. This suggested that under these anhydrous, acidic conditions the elimination of H_2O led to iminium formation, followed by rapid cyclization. During analysis of the ring closure via ^{15}N-^{1}H HMBC, we did not observe any intermediates, albeit this was attributed to the temporal limitations of the method.

2.2.6 Determination of early reaction intermediates

While monitoring ^{1}H-^{15}N HMBC, we detected no intermediates prior to pyrrole formation, i.e. no combination of **36** with either **38** or with **37** alone. This indicated that formation of intermediate **40** was irreversible and the multi-component reaction occurred too rapidly to be detected by isotope enhanced NMR. We hypothesized that there were three

potential pathways to form **40** (Figure 39, *pathways 1 – 3*). To test these individually, we performed three independent reactions, each with one of the substrates omitted. These reactions were performed using the described model conditions, except when noted.

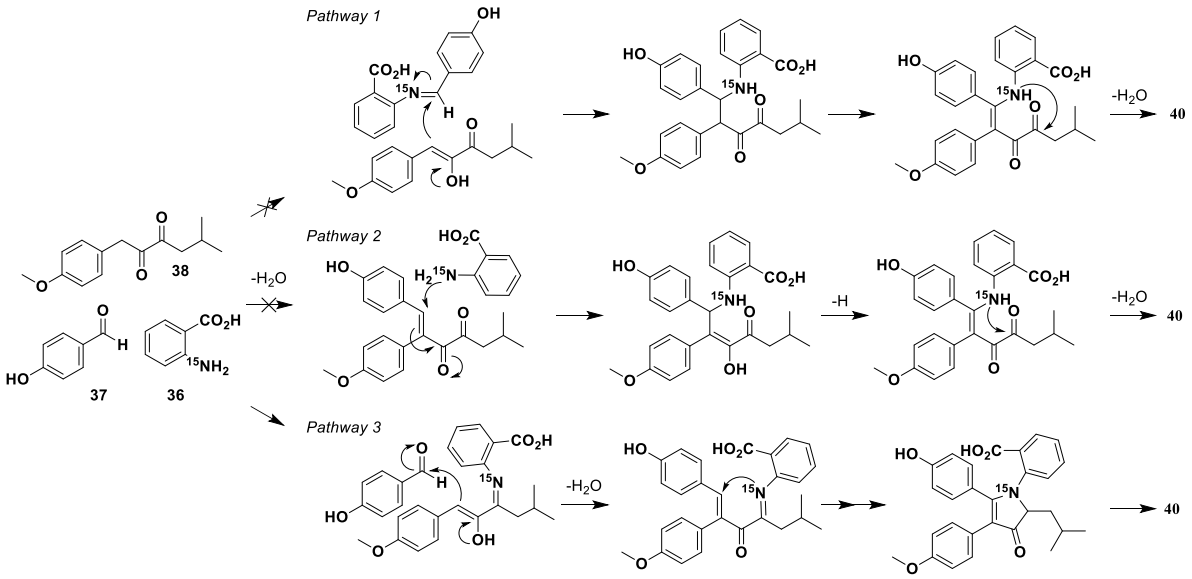


Figure 39. Proposed biosynthetic mechanism of the discoipyrroles

Pathway 1 relies on initial formation of an aldimine between **36** and **37**, setting up for a Mannich reaction. We monitored imine formation using ^{1}H–^{15}N HMBC with an expected imine ^{15}N shift in the 0 – 100 ppm range. Under model conditions, after 24 hours there was no formation of **43** (Figure 40). This was further confirmed by LC-MS analysis. As a control reaction we were readily able to form aldimine **44** between ^{15}N aniline and **37** under the model conditions to give a product with a ^{15}N at -65.0 ppm. This suggested that the mechanism of discoipyrrole formation does not involve a Mannich reaction. To further test this, we took commercially available **43**, **38** and DMAP in our model conditions and monitored for formation of **42** by LC-MS. Product was not detected after 48 hours, well within the timeframe expected for product formation to occur.

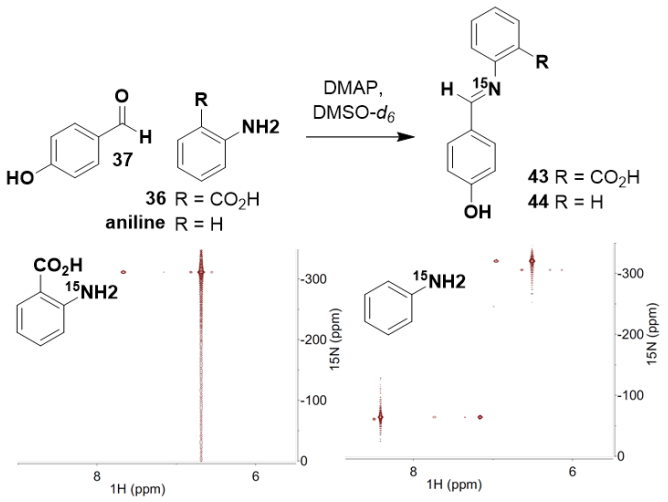


Figure 40. Aldimine formation based on aniline substitution

To test pathway 2, [1-^{13}C]-**37** and **38** were treated with DMAP and monitored by NMR and LC-MS for the Claisen-Schmidt product. After 48 hours we only observed the presence of [1-^{13}C]-**37** and **38**. Increasing reaction time, temperature, or strength of the base did not facilitate condensation. It is possible that **36** could act as a catalyst for the Claisen-Schmidt reaction, however the complete lack of product formation under all of the conditions attempted is strong evidence against pathway 2.

Pathway 3, in which **38** undergoes imine formation was tested by omitting **37** in the reaction (Figure 41). However to our surprise, after 24 hours the main product was a di-methoxy discoipyrrole analog. Utilizing NMR and LC/MS we observed 4-methoxybenzaldehyde in the reaction, demonstrating that **38** could be degraded to form 4- methoxybenzaldehyde. Therefore, aldehyde contamination of reactions involving oxidized sattabacin analogs prevented further analysis due to the shuttling of intermediates to pyrrole formation.

To circumvent this issue, we utilized **37** in lieu of **38**, removing the possibility for condensation with aldehyde coming from **38**. No discernable amount of the desired product was

detected, however slightly increased temperatures induced minor product formation. We reasoned that an equilibrium between the acyloin group and any imine might favor the former, compared to the diketone of **38**. Therefore, we set out to trap potential amination products using reductive conditions. Indeed, reacting **35** and **36** in the presence of sodium triacetoxyborohydride in dichloroethane with acetic acid yielded a reductive amination product (31% yield) within 16 hours. Reduction of **35** to the diol side product was prevalent, potentially obscuring the calculated reaction efficiency.

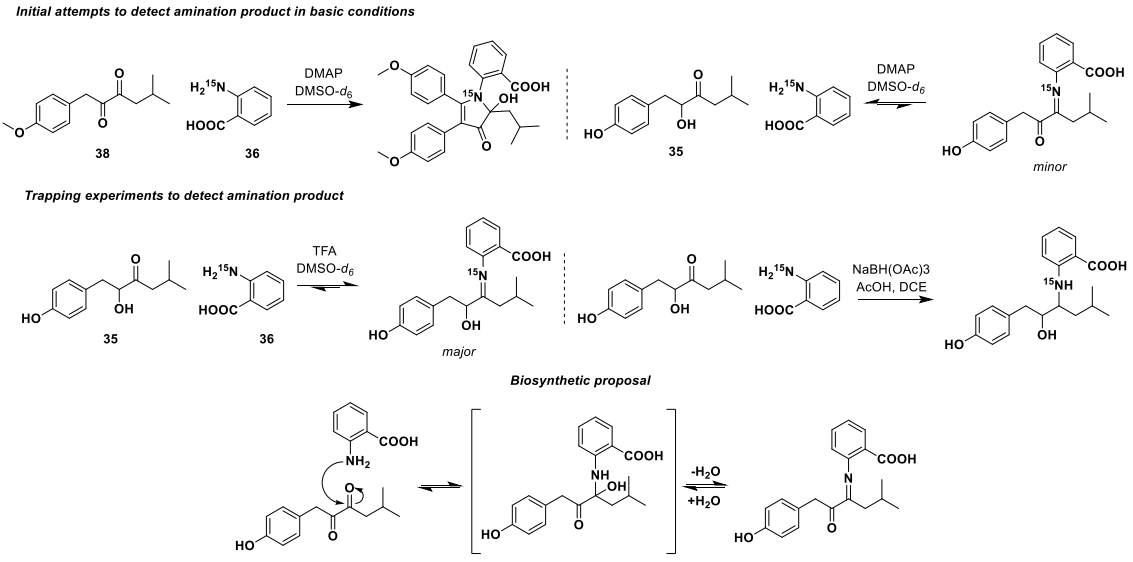


Figure 41. Analysis of Pathway 3

The efficiency of these omission experiments compared to the complete reaction containing all three reactants indicated that the mechanism was enhanced by a concerted flow of reactivity. Our current hypothesis is that pyrrole formation can proceed through amination of **38** by **36** to an imine intermediate that can be trapped by a Claisen-Schmidt condensation between the C1 position of **38** and **37**. We believe that the amination increases the acidity of the C1 methylene to cause the "snapping" together of the three substituents. The resulting conjugated olefin can then be attacked by the nitrogen of **36** to form the five membered core.

2.2.7 Conclusion

The DPs presented a unique biosynthetic case, in that they can be produced by a specific and concerted mechanism in aqueous media from diverse excreted metabolites. To study how **35-37** were able to undergo such a reaction we constructed a model system in organic solvent and utilized various analytical techniques, including the novel application of ^{1}H-^{15}N HMBC as a tool for monitoring reactions *in situ*. Building on the foundation made possible by this technique, we performed various isotope labeling experiments with ^{13}C, ^{15}N, and ^{18}O to elucidate the mechanism of non-enzymatic discoipyrrole formation. This multi-component reaction was characterized by the initial, rapid steps to induce pyrrole formation and the following changes in oxidation state that dictate the final closure of the lactone ring. Particularly useful was the ^{18}O labeling experiments which allowed us to clearly demonstrate the atmospheric oxidation of the pyrrolone using careful isolation and high resolution mass spectrometry. Furthermore, the subsequent loss of this isotope was integral in establishing the dehydration dependent lactone ring closure.

Importantly, once the details of the mechanism were uncovered, we constructed fermentation media with deuterated water and used our ^{1}H-^{15}N HMBC monitoring method to validate discoipyrrole formation in biologically equivalent conditions without organic solvent or added DMAP (Figure 42). We observe identical intermediates and a similar reaction time course under both conditions.

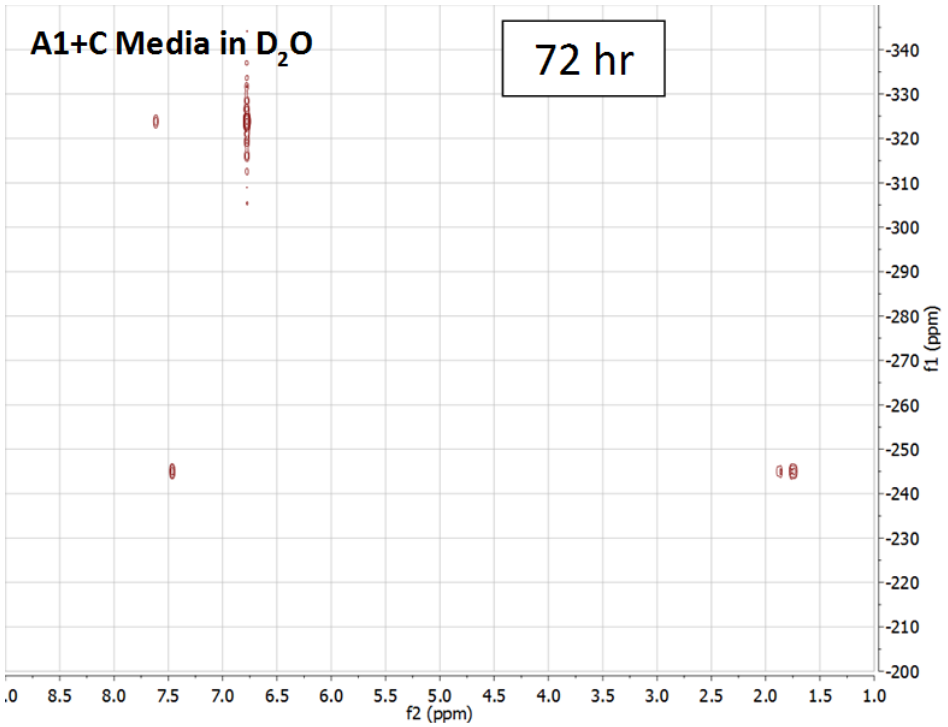


Figure 42. DP synthesis carried out in deuterated sea water

2.2.8 Future directions

Aryl amines have been found to participate in many of the recent examples of natural products amenable to non-enzymatic incorporation including the ammosamides, elansolids, and oxazinin A (Zhenjian Lin et al., 2014; Pan et al., 2013; Heinrich Steinmetz et al., 2011; Steinmetz et al., 2012). Through our mechanistic studies of the DPs, we have shown that **36** acts as the initiating factor through imine formation with **35** and that the *ortho* carboxylic acid substitution of **36** drives the selectivity as seen in the aldimine formation experiments. Our work and others suggest that aryl amines play a significant role in non-enzymatic reactions and that aryl substitution patterns, the nature of the amination sites, and the stability of intermediates all contribute to the complexity of these reactions. Using these types of nucleophiles could be a potentially cost effective and simple way to screen bacterial libraries for natural products

amenable to non-enzymatic perturbations. Concordantly, detection methods using ^{15}N labels such as the ^{1}H-^{15}N HMBC demonstrated here could provide quick and high sensitivity analysis

2.3 Experimental

General Procedures. ^{1}H and 2D NMR spectral data were recorded at 600 MHz in DMSO-d_6 on a Varian System spectrometer, and chemical shifts were referenced to the corresponding residual solvent signals (δ_H 2.50/δ_C 39.51). ^{13}C NMR spectra were acquired at 400 MHz on a Varian System spectrometer. High resolution ESI-TOF mass spectra were provided by The Scripps Research Institute, La Jolla, CA. Low-resolution LC/ESI-MS data were measured using an Agilent 1200 series LC/MS system with a reversed-phase C_{18} column (Phenomenex Kinetix, 150 mm × 4.6 mm, 2.5 μm) at a flow rate of 1.0 mL/min. Preparative HPLC was performed on an Agilent 1200 series instrument with a DAD detector, using a reversed-phase C_{18} column (Phenomenex Luna, 250 × 10.0 mm, 5 μm) or a phenyl-hexyl column (Phenomenex Luna, 250 × 10.0 mm, 5 μm).

Collection and Phylogenetic Analysis of Strain SNA-048. *Bacillus hunanensis* strain SNA-048 was isolated from Little San Salvador, Bahamas. The sediment was desiccated and pasted onto agar plates using gauze 1 acidic media (10 g starch, 1 g $NaNO_3$, 0.5 g K_2HPO_4, 0.5 g $MgSO_4$, 0.5 g NaCl, 0.01 g $FeSO_4$, 1 L seawater, 15 g agar, adjust pH to 5.3 with phosphate buffer). Bacterial colonies were selected and streaked to purity using the same agar media. Analysis of the strain by 16S rRNA revealed 99.9% identity to *Bacillus hunanensis*. The sequence is deposited in GenBank under accession no. KC247801.

Cultivation and Extraction of SNA-048 Cultures. Bacterium SNA-048 was cultured in 10 × 2.8 L Fernbach flasks each containing 1 L of a seawater-based medium (10 g starch, 4 g yeast

extract, 2 g peptone, 1 g $CaCO_3$, 40mg $Fe_2(SO_4)_3 \cdot 4H_2O$, 100 mg KBr) and shaken at 200 rpm at 27 °C. After seven days of cultivation, sterilized XAD-7-HP resin (20 g/L) was added to adsorb the organic products, and the culture and resin were shaken at 200 rpm for 2 h. The resin was filtered through cheesecloth, washed with deionized water, and eluted with acetone. The acetone-soluble fraction was dried *in vacuo*.

Precursor Feeding Studies. Feeding studies with various amino acids (Cambridge Isotopes, Sigma) at indicated concentrations were carried out in triplicate in 125 mL flasks containing 50 mL of the seawater-based medium. Cultures were each inoculated with 5 mL turbid SNA-048. After 7 days, cultures were spun down at 4,800 rpm (Beckman JS5.2 Rotor) and the supernatant was removed, extracted with ethyl acetate and dried *in vacuo*. The cell pellets were weighed and used to normalize between fermentation cultures.

Quantification of metabolites secreted by SNA-048. Dried extracts were resuspended in 2 mLs methanol and spun down. Supernatant was analyzed using an Agilent 1200 series LC/MS system with a reversed-phase C_{18} column (Phenomenex Luna, 150 mm × 4.6 mm, 5 μm) at a flow rate of 0.7 mL/min.. **35**: Integration of AUC UV absorbance, λ = 280 nm (t_R = 11.5 min). **36**: Integration of AUC UV absorbance, λ = 330 nm (t_R = 8.3 min). **37**: Integration of AUC $[M - H]^-$ *m/z* 121.1 (t_R = 7.8 min). Values were normalized to cell pellet mass and compiled using Graphpad Prism.

General synthetic model of discoipyrrole formation. 1-(4-methoxyphenyl)-5-methylhexane-2,3-dione, **38**, (1 mol equivalent) was dissolved in 0.25 mL DMSO at room temperature. To this solution was added anthranilic acid, **36**, (2 mol equivalent) in 0.25 mL DMSO and 4-hydroxybenzaldehyde, **37**, (2 molar equivalent) in 0.25 mL DMSO. Then, 4-

(dimethylamino)pyridine (3 mol equivalents) in 0.25 mL DMSO was added to the reaction flask. The reaction was stirred at room temperature for indicated amount of time. If formation of **39** was desired, 1% TFA was added after 24 hours and stirred for 12 hours. Scale of reaction varied from 5-100 μmol **38**. Generally, reaction products were purified by adding NH_4Cl (*aq*) and direct loading of crude reaction onto reversed phase HPLC (RediSep Rf Gold® 30 g C18, 35 ml/min) using a gradient system from 10% to 100% H_2O:CH_3CN with monitoring at UV λ = 254, 330, and 280 nm.

Investigation of biosynthetic genes. . Sequence analyses of the annotated SNA-048 genome with the amino acid sequence of Thz1050(Park et al., 2014) were performed with the NCBI (Altschul, Gish, Miller, Myers, & Lipman, 1990) and SEED(Overbeek et al., 2005) databases and the RAST(Aziz et al., 2008) server, using PSI-BLAST. Multiple sequence alignments were performed with CLUSTALW(Thompson, Higgins, & Gibson, 1994). Genome sequencing of B. hunanensis SNA048 was performed using an Illumina MiSeq. Reads were assembled using SPAdes 3.5(Nurk et al., 2013) with read-pair merging using FLASH(Magoc & Salzberg, 2011).

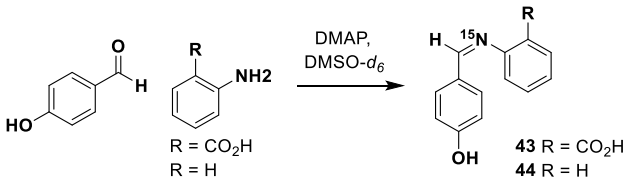
General ^{1}H-^{15}N HMBC Monitoring. ^{15}N-anthranilic acid, **36**, (Cambridge Isotopes) and 4-(dimethylamino)pyridine was added to 700 μL of DMSO-d_6 in a 5 mm thin wall, 8 inch NMR tube (Wilmad) and inserted into 600 MHz magnet (Varian). Instrument was tuned, locked, and shimmed. Then, the sample was removed and used to solubilize **38** and **37**, before being inserted into the magnet. NMR experiments were then queued to run in the following order: 1X ^{1}H NMR experiment (32 scans taking 50 seconds) followed by 4X ^{1}H-^{15}N HMBC (4 scans of 128 increments taking 30 minutes). Therefore, a single 1H NMR spectrum was collected once every 2 hours with 4 HMBC spectra collected in between. This was run constantly for 24 hours (until

42 appeared in isolation), at which point to the sample 1% TFA was added, followed by continued NMR data collection. For subsequent experiments in which alternative 2D experiments were run, a similar method was used albeit with intermittent COSY, HSQC, and HMBC experiments performed.

Isolation and characterization of 42. 1-(4-methoxyphenyl)-5-methylhexane-2,3-dione, **38**, (9 mg, 38 μmol) was dissolved in 0.25 mL DMSO-d_6 at room temperature. To this solution was added anthranilic acid, **36**, (10.6 mg, 77 μmol) in 0.25 mL DMSO and 4-hydroxybenzaldehyde, **37**, (9 mg, 76 μmol) in 0.25 mL DMSO. Then, 4-(dimethylamino)pyridine (14 mg, 116 μmol) in 0.25 mL DMSO was added to the reaction flask. The reaction was stirred for 48 hours. The reaction was then subject to preparative HPLC (Phenomenex Luna, C_{18}, 250 × 10.0 mm, 5 μm, 2.5 mL/min) using a isocratic solvent system at 17% CH_3CN:H_2O over 25 min to afford compound **42** (10.4 mg, t_R = 18.0 min) **31:** yellow oil, UV (MeOH) λ_{max} 225, 240, 265, 330, 390 nm; ^{1}H and ^{13}C NMR, see Table 9; HRESIMS $[M + H]^+$ *m/z* 475.1881 (calcd for $C_{28}H_{27}{}^{15}NO_6$, [M + H]+ m/z = 475.1881).

^{18}O Labeling Experiment. DMSO-*d6* was degassed and charged with $^{18}O_2$ gas (Sigma) by purging 10 mL DMSO-d_6 with N_2 for 10 minutes, followed by purging with $^{18}O_2$ for several minutes. 1-(4-methoxyphenyl)-5-methylhexane-2,3-dione, **38**, (45 mg, 192 μmol) was dissolved in 0.5 mL DMSO-d_6 at room temperature. To this solution was added anthranilic acid, **36**, (53 mg, 387 μmol) in 0.5 mL DMSO and 4-hydroxybenzaldehyde, **37**, (46 mg, 380 μmol) in 0.5 mL DMSO. Then, 4-(dimethylamino)pyridine (70 mg, 578 μmol) in 0.5 mL DMSO was added to the reaction flask. The reaction flask was purged with $^{18}O_2$ and stirred at room temperature under an atmosphere of $^{18}O_2$ for 24 hours. Analysis of sample was performed by subjecting aliquots to

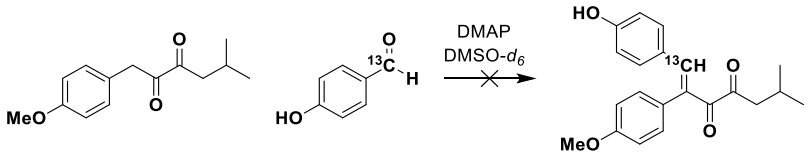
LC/MS. After completion, an aliquot of the reaction was purified directly (Phenomenex Luna, C_{18}, 250 × 10.0 mm, 5 μm, 2.5 mL/min) using an isocratic solvent system at 17% CH_3CN:H_2O over 25 min to afford compound ^{18}O-**42** (6.5, t_R = 18.0 min). As a control, this experimental design was used albeit with $^{16}O_2$. To induce closure of ^{18}O-**42**, 1% TFA was added and the reaction stirred for 12 hours. After completion, an aliquot of the reaction was subject to reversed phase chromatography (RediSep Rf Gold® 30 g C18, 35 ml/min) using a gradient solvent system from 10-100% CH_3CN:H_2O over 35 min (2.5 mg, t_R = 22 min).



Scheme 1. Aldimine formation

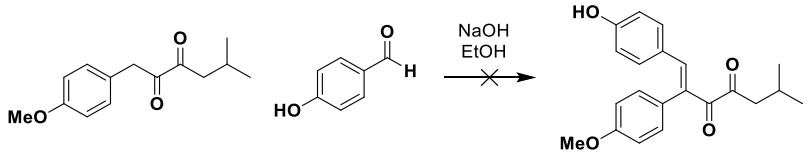
Synthesis of 43. ^{15}N-anthranilic acid (2.5 mg, 18 μmol) was dissolved in 0.2 mL DMSO-d_6 at room temperature. To this solution was added 4-hydroxybenzaldehyde, **37**, (2.4 mg, 20 μmol) in 0.2 mL DMSO. Then, 4-(dimethylamino)pyridine (4.2 mg, 34 μmol) in 0.3 mL DMSO was added to the NMR tube. The reaction was stirred for 48 hours with constant ^{1}H-^{15}N HMBC monitoring as described above.

Synthesis of 44. ^{15}N-aniline (1.8 mg, 20 μmol) was dissolved in 0.2 mL DMSO-d_6 at room temperature. To this solution was added 4-hydroxybenzaldehyde, **37**, (2.3 mg, 19 μmol) in 0.2 mL DMSO. Then, 4-(dimethylamino)pyridine (4.6 mg, 40 μmol) in 0.3 mL DMSO was added to the NMR tube. The reaction was stirred for 48 hours with constant ^{1}H-^{15}N HMBC monitoring as described above.



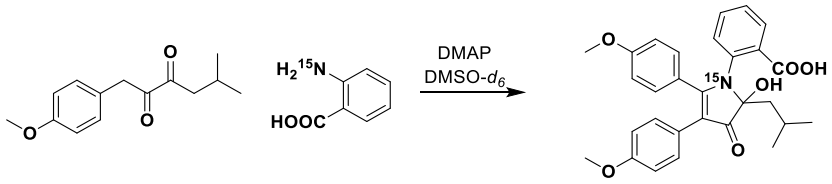
Scheme 2. Pathway 2 interrogation

[1-^{13}C]-**37** (3.6 mg, 30 μmol) was dissolved in 0.2 mL DMSO-d_6 at room temperature. To this solution was added **38** (3.5 mg, 15 μmol) in 0.2 mL DMSO-d_6. Then, 4-(dimethylamino)pyridine (5.6 mg, 45 μmol) in 0.3 mL DMSO-d_6 was added to the NMR tube. Reaction was monitored by constant ^{1}H-^{13}C HMBC for 48 hours. Aliquots were subjected to LC/MS analysis. After 72 hours, the reaction was heated to 60° C and monitored by LC/MS for several days.



Scheme 3. Pathway 2 interrogation with base

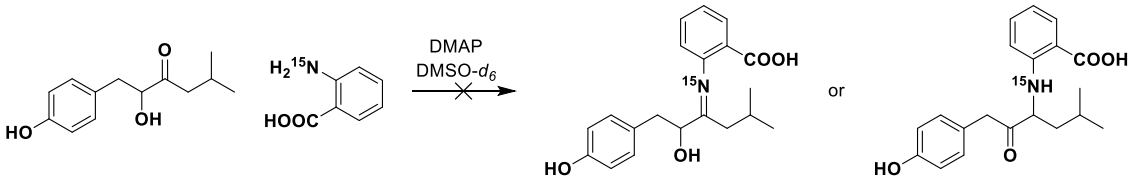
38 (13 mg, 55 μmol) was dissolved in 1 mL EtOH at room temperature. To this solution was added **37** (10 mg, 83 μmol) in 0.5 mL EtOH. Then, NaOH (25 mg, 625 μmol) was added to the round bottom flask and the reaction was stirred at room temperature for several days. Aliquots were removed and subjected to LC/MS analysis.



Scheme 4. Pathway 3 interrogation

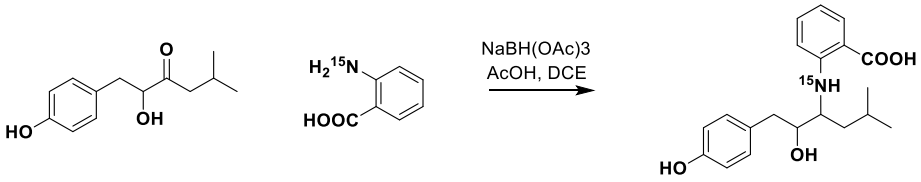
38, (12 mg, 51 μmol) was dissolved in 0.33 mL DMSO-d_6 at room temperature. To this solution was added ^{15}N-anthranilic acid, **36**, (15 mg, 109 μmol) in 0.33 mL DMSO-d_6 and 4-

(dimethylamino)pyridine (25 mg, 205 μmol) in 0.33 mL DMSO-d_6. The reaction was stirred at room temperature for 48 hours.



Scheme 5. Pathway 1 interrogation

35, (17 mg, 77 μmol) was dissolved in 0.33 mL DMSO-d_6 at room temperature. To this solution was added ^{15}N-anthranilic acid, **36**, (21 mg, 154 μmol) in 0.33 mL DMSO-d_6 and 4-(dimethylamino)pyridine (30 mg, 245 μmol) in 0.33 mL DMSO-d_6. The reaction was stirred at room temperature for 48 hours. After 48 hours, the reaction was heated to 60°C for several days.



Scheme 6. Imine trapping

35, (27 mg, 122 μmol) was dissolved in 1 mL anhydrous dichloroethane (DCE) at room temperature. To this solution was added ^{15}N-anthranilic acid, **36**, (8.4 mg, 61 μmol) in 1 mL anhydrous DCE, sodium triacetoxyborohydride (25 mg, 118 μmol) and glacial acetic acid (AcOH, 7 uL). The reaction was stirred at room temperature for 16 hours. (Figure S10) After completion, the reaction was quenched with NH_4Cl (*aq*), extracted three times with EtOAc, and the organic layer was subject to reversed phase chromatography (RediSep Rf Gold® 30 g C18, 35 ml/min) using a gradient solvent system from 10-100% CH_3CN:H_2O over 35 min (13 mg, t_R = 10 min).

CHAPTER THREE

EXPANDING NATURAL PRODUCT LIBRARY CHEMICAL DIVERSITY

The screening of natural products for biomedical application relies on the generation of a diverse compound library with chemically disparate scaffolds. There are many different ways natural products chemists can accomplish this feat. Inspired by our discovery of the non-enzymatic formation of the discoipyrroles as described in Chapter Two, our lab has recently began a program to take advantage of chemically reactive natural products to derive new molecules. Additionally, eliciting expression of silent biosynthetic gene clusters has been an integral methodology in recent years (Zarins-Tutt et al., 2016). Our lab has recently developed an approach to perform such induction experiments for our bacterial collection. In Chapter One, the concept of cryptic biosynthetic genes was introduced as a supplementary description of natural product biosynthesis. This subject will be expanded upon here before discussion of our specific cryptic biosynthesis induction approach.

3.1 Cryptic biosynthesis

Microorganisms are an occluded reservoir of natural products in that their metabolic profiles under standard laboratory conditions do not always encapsulate the entirety of their biosynthetic potential. Since pioneering work with *Streptomyces* (Bentley et al., 2002; Ikeda et al., 2003), it has been established that biosynthetic gene clusters (BGCs) can commonly be constitutively silent or expressed in low abundance (Pettit, 2011). As sequencing technology and the associated fields of gene annotation and biosynthetic gene clustering, have improved in the last decade, there has been a large push to identify and turn on these so called cryptic

biosynthetic gene clusters. The driving force behind this effort is to access different and potentially novel natural product scaffolds in existing microbial collections whose resources have not fully been elucidated.

3.2 Ribosomal engineering

In bacteria, many of the techniques to induce transcriptional reprogramming towards cryptic BGCs rely on disturbing cellular homeostasis, as stimulating a defensive state triggers natural products that permit survival (Rutledge & Challis, 2015). Stochastic techniques, or those that do not rely on targeted genetic modification, have ranged from culturing in altered media types to chemical treatment to co-culture. One of the most promising approaches has been the use of chemicals that modulate transcriptional machinery (Okada & Seyedsayamdost, 2017). These chemo-forward methods include modulation of fungal chromatin with histone deacetylase (HDAC) inhibitors (Palmer & Keller, 2010), redirection of bacterial metabolic substrates using triclosan or ARC2 compounds (Craney, Ozimok, Pimentel-Elardo, Capretta, & Nodwell, 2012), and a hypothesized stress response-based modulation using trimethoprim (Okada, Wu, Mao, Bushin, & Seyedsayamdost, 2016). These published studies represent modern approaches whose discoveries featured genomic sequencing and high-throughput screening efforts. Surely, these exciting new findings will be expounded upon in molecular detail and utility in the near future. In contrast, the classic example of stochastic chemo-centric methodologies for cryptic BGC expression is known as ribosome engineering, pioneered by Kozo Ochi of the Food Institute of Japan (Kozo Ochi & Hosaka, 2013), which was discovered based on fundamental microbiological studies.

Ribosome engineering originated from studies of the stringent response in bacteria, a process intimately tied to secondary metabolism (Kelly, Ochi, & Jones, 1991; K. Ochi, 1987; K. Ochi & Ohsawa, 1984; Strauch, Takano, Baylis, & Bibb, 1991). The stringent response is a defense mechanism triggered by limited amino acid availability sensed by unloaded tRNA in the ribosome. This key event causes the activation of an enzyme known as RelA, which is physically associated with the ribosome and catalyzes the production of pppGpp from ATP and GTP. The result of this enzymatic process is that pppGpp increases and GTP decreases. The changes to these two metabolic pools then act independently to induce different stringent response phenotypes.

First, pppGpp is dephosphorylated to form the molecule ppGpp, an alarmone that binds the β-subunit of RNA polymerase (rpoβ) directly (Artsimovitch et al., 2004) to alter the promoter preference of transcriptional machinery (Hosaka et al., 2009). Specifically, this increases natural product BGC transcription and redistributes metabolic flux from primary metabolism to enter a defensive, stationary phase (Gaca, Colomer-Winter, & Lemos, 2015). Independently, the decrease in GTP from RelA action drives sporulation and morphological differentiation, events that are integral for defensive protection from exogenous stimuli (Lopez, Dromerick, & Freese, 1981; K. Ochi, 1986; K. Ochi, Kandala, & Freese, 1981). Overall, the alteration of these two metabolic pools of GTP and ppGpp cause metabolic and growth deficiencies in addition to the switch to secondary metabolism. Therefore, from a tool perspective, manipulations of RelA create dirty systems in which a clear path to increase natural product anabolism is improbable as the positive effects of increasing RelA function are outweighed by the negative actions of the various signaling pathways involved downstream due to GTP decrease and ppGpp increase.

However in 1996 Ochi reported a mutant *Streptomyces lividians* which produced the pigment molecule known as actinorhodin, though the wild-type parent strain did not (Shima, Hesketh, Okamoto, Kawamoto, & Ochi, 1996). This mutant turned out to be resistant to streptomycin and contained a K88E point mutation in the *rpsL* gene which encodes for the ribosomal S12 protein, the known target of streptomycin. The relationship of this point mutation to actinorhodin biosynthesis was then confirmed using directed mutagenesis studies in which actinorhodin production was induced in a *Streptomyces coelicolor rpsL* K88E mutant of wild-type background (Hesketh & Ochi, 1997). Later studies demonstrated that the altered *rpsL* gene product led to increased ribosome stability causing increased translational efficiency during late stages of growth when secondary metabolism was highly active (Hosaka, Xu, & Ochi, 2006; Okamoto-Hosoya, Hosaka, & Ochi, 2003). These were the first discoveries of a ribosome-based upregulation of natural product biosynthesis that did not revolve around RelA based induction of ppGpp, which produces deleterious effects.

Later, Ochi discovered another way to avoid the side effects of ppGpp accumulation by inducing pleiotropic mutations in its target rpoβ directly by treating with known inhibitors of rpoβ, including rifampicin (Hosaka et al., 2006; H. Hu, Zhang, & Ochi, 2002). By inducing resistance, mutant rpoβ displayed altered transcription signatures favoring natural production biosynthesis without effecting the pools of GTP and ppGpp that ultimately restrict natural product biosynthesis. Since these discoveries, many labs have utilized the approach of ribosome engineering by using compounds to target rpsL (e.g. streptomycin) or rpoβ (e.g. rifampicin) and generate mutants whose secondary metabolism is upregulated (Kozo Ochi & Hosaka, 2013). One of the most important findings of these applications is the efficiency of chemically induced

mutagenesis. That is, bacteria that develop resistance to rifampicin and streptomycin are mutated in the same gene locus consistently (Hosaka et al., 2009).

3.3 Ribosome engineering of Streptomyces *rubrogriseus* sp. SNC-031

Utilizing such a simple, yet effective phenotypic approach to expand the chemical diversity of an existing microbial library was of interest to our laboratory. Previously, we applied such techniques to a subset of 20 *Streptomyces* strains from our mixed library of marine and terrestrial bacteria to assess the personal applicability of ribosome engineering (Fu, Jamison, La, & MacMillan, 2014). From one of the mutated *Streptomyces* strains, a new family of chlorinated alkaloids, the inducamides A-C, were discovered recapitulating the utility of ribosome engineering demonstrated by others.

In an attempt to expand the utility of this technique within *Streptomyces,* 14 unique actinobacteria from our collection were selected for ribosome engineering with rifampicin. Bacteria were originally plated on an artificial sea water based media with rifampicin (20 μM) and those who successfully developed resistance were eventually plated on higher concentrations of rifampicin over several weeks (up to 160 μM). Of the 15 bacteria, 4 were able to develop resistance at this dose of rifampicin.

We utilized antibiotic disc diffusion assays with the crude fermentation extracts of these 4 resistant mutants in the hopes of identifying phenotypic alterations in their secretion of bioactive secondary metabolites. Two of these mutants, derived from a species of *Streptomyces rubrogriseus,* displayed strong antibiotic activity against *Bacillus subtilis* (Figure 43). In comparison, neither the parent strain, denoted SNC-031 nor the other rifampicin mutant species (SNC-031-M1 and SNC-031-M2) showed antibiotic activity.

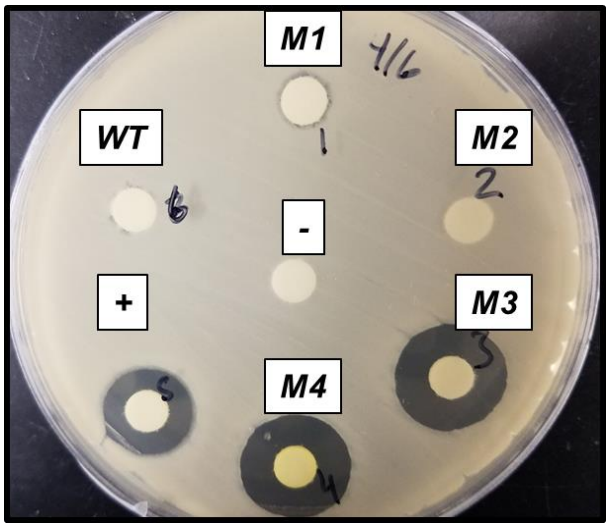


Figure 43. Antibacterial effect of SNC031 wild-type and mutants

Sequencing of the 16S subunit from the parent SNC-031 species and each of the generated mutants M1-M4 confirmed that all were derivatives of *S. rubrogriseus*, indicating that no contamination had occurred during resistance development (Figure 44). To confirm that the rifampicin target, RNA polymerase β subunit (rpoβ), was mutated in the resistant species, the gene was amplified and sequenced (Figure 44). Sequencing results indicated that rpoβ was mutated at the position H437 and S442 in M4, the former of which is characteristic of rifampicin resistant mutants (Fu et al., 2014; Kozo Ochi & Hosaka, 2013). These mutations were not found in the parent SNC-031 or other mutants.

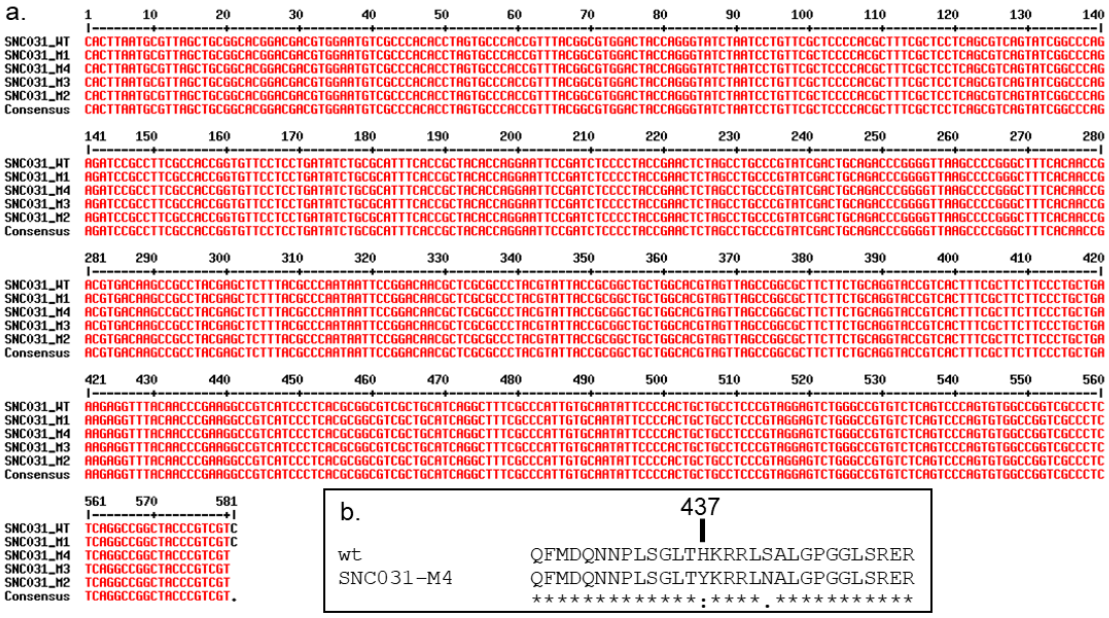


Figure 44. 16s rRNA sequences of SNC031 mutants and rpoβ protein comparison

To explore the metabolic alterations between the parent strain and mutants, the extracted metabolome of a 7 day fermentation was analyzed for each via LC/MS. The chromatogram of M3 and M4 differed drastically from that their parent strain as well as the other mutants examined (Figure 45). While all crude extracts demonstrated the same signature of peaks eluting between 3.0 and 5.0 minutes, there were marked differences in the compounds eluted between 6.5 and 7.5 minutes. Most significantly, there were major increases in the peaks eluted at 6.7 and 7.3 minutes, which shared matching UV spectra. Additionally, surrounding minor peaks with similar UV traces were observed in mutants M3 and M4 compared to the parent strains, suggesting that a family of compounds were upregulated in these cultures.

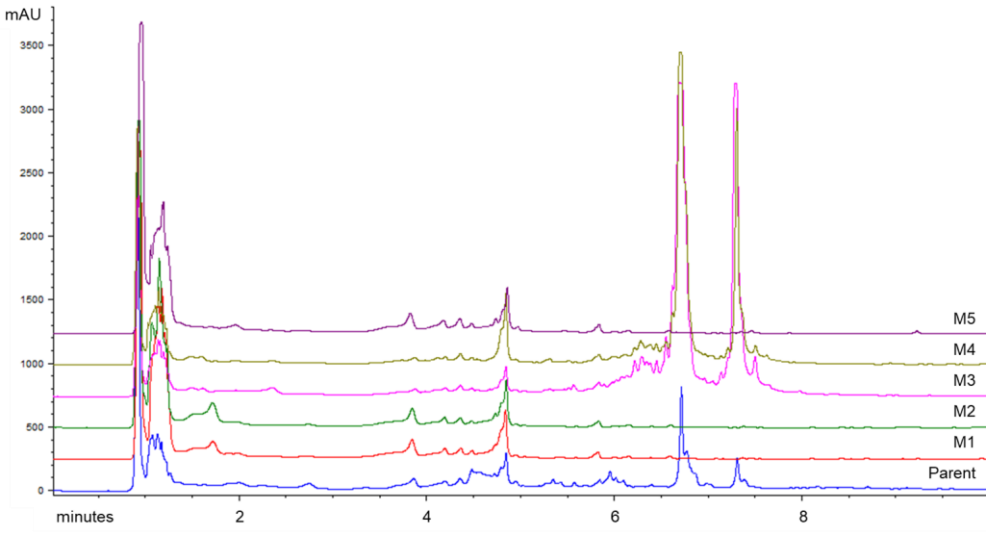


Figure 45. Metabolomes of SNC-031 and mutants

In an effort to identify these metabolites of interest, 10 liters of the mutant SNC-031 M4 was fermented for 7 days and the crude extract was chromatographed with a reverse-phase C18 resin into 9 UV absorbing fractions. Antibiotic disc diffusion assays revealed 4 of the fractions were highly cytotoxic towards *B. subtilis*, indicating that multiple antibiotics were being produced by M4. The most abundant fraction contained one main compound corresponding to the major peak in Figure 45 (t_R= 6.7') and had a very distinctive *m/z* 795.2 $[M-H]^-$. Structure elucidation led to the identification of kirromycin, or mocimycin, a hybrid polyketide-NRP with a unique pyridone ring, known for its potent antibiotic activity (Figure 46) (Wolf & Zahner, 1972). Kirromycin belongs to the family of elfamycins, whose unique mechanism of action, binding the bacterial elongation factor Ef-Tu, has been an important area of study particularly for structural biological investigations of translation machinery (Fischer et al., 2015; Parmeggiani & Swart, 1985). The biosynthesis of kirromycin has also been extensively studied as it involves an uncommon trans-acetyltransferase mechanism for substrate loading (Musiol et al., 2013) and the

conversion of aspartate into β-alanine which is used to derive the pyrridone ring (Laiple, Hartner, Fiedler, Wohlleben, & Weber, 2009; Weber et al., 2008).

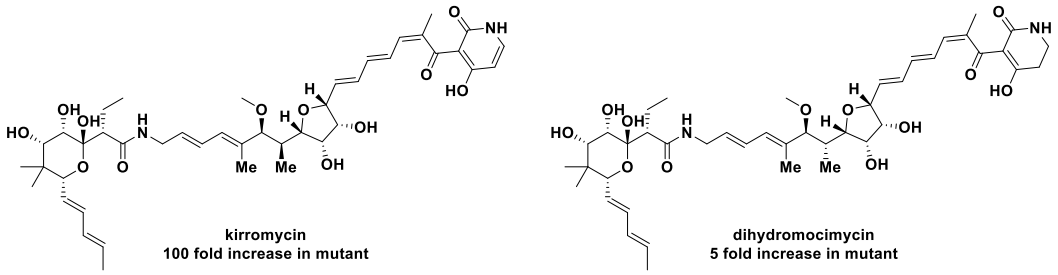
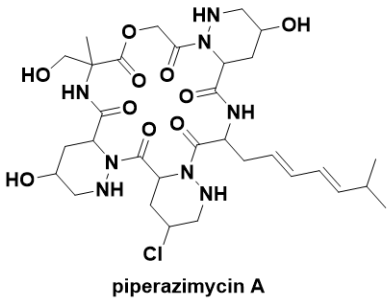


Figure 46. Compounds isolated from SNC-031 mutants

Additional members of the elfamycin family were also found, including dihydromocimycin, and aurodox, which have hydrogenated pyrridone rings and N-methylated pyrridone rings, respectively (Figure 46). Previous reports of elfamycins have cited low yields in liquid fermentation, however M4 fermentations yielded 15-20 mg/L of kirromycin. These types of titer increases have been documented previously for ribosome engineering and represent an efficient way of performing strain improvement on microbial libraries.

Originally believing that alternate minor constituents of the crude extract were elfamycin analogs, we pursued additional chromatographic fractions. Surprisingly, mass spectrometry of the major component of these fractions showed a characteristic 3:1 ratio of the +2 shifted [M+H]+ peak, indicating a chlorinated molecule. Fractionation guided by antibiotic activity and the unique chlorinated mass spectrum led to the isolation of the hexadepsipeptide, piperazimycin A, originally described by the Fenical laboratory (Miller, Kauffman, Jensen, & Fenical, 2007) (Figure 47, left). Piperazimycin A previously demonstrated pan-toxicity against the NCI panel of 60 cell lines in the low nM range (Miller et al., 2007). Submission of purified piperazimycin A

from *S. rubrogriseus* SNC-031-M4 to a small panel of non-small cell lung cancer (NSCLC) cell lines displayed similar cytotoxicity (Figure Figure 47, right)



piperazimycin A

Lung Cancer Cell Line	IC_{50} (μM)
H1395	0.529
H1993	0.364
H2087	0.191
HCC15	0.191
HCC44	0.169
HCC95	0.176
HCC827	0.227
HCC4017	0.227

Figure 47. Piperazimycin A structure and mammalian cytotoxicity

Interestingly, the discovery of piperazimycin in the *rpoβ* mutant mirrored the documented discovery of the molecule piperadimycin by Ochi and co-workers in a *rpoβ* mutant strain of *Streptomyces mauvecolor* (Hosaka et al., 2009). In their work, Hosaka et al. carried out a mutagenesis screen of terrestrial actinomycetes from their collection in which they cultured microbes with rifampicin, gentamicin, or streptomycin and assayed for increased or induced antibacterial activity. They identified a rifampicin resistant mutant strain of *S. mauvecolor*, carrying the H437L point mutation of *rpoβ*, whose extract was toxic to *Staphylococcus aureus*, even though the wild-type was innocuous. The authors then identified 8 compounds which they named the piperidamycins A-H. Interestingly, the point mutation of rpoβ matches our SNC-031 M4 mutant and the discovered piperidamycins structurally resemble the piperizamycins discovered in our SNC-031 M4 (Figure 48).

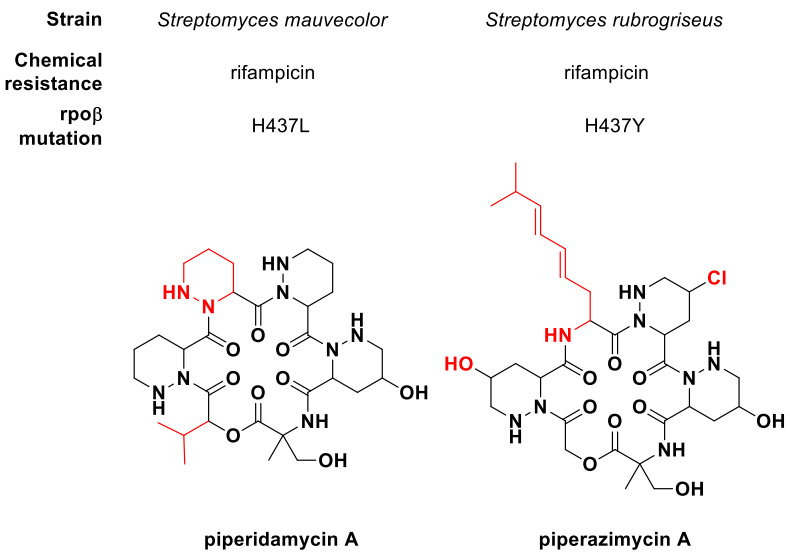


Figure 48. Comparison of piperidamycin and piperazimycin induction

Two points can be made from this comparison, one regarding the biological function of these two molecules and one regarding their biosynthesis. The similar hexadepsipeptide structure featuring contiguous piperidine rings suggest that similar biosynthetic gene clusters are responsible for these two compounds. The fact that they are both turned on in response to rifampicin resistance begs the question of whether or not certain BGCs are more prone to ribosome engineering than others. Genetic experiments in which the promoters of these BGCs are necessary to divulge this type of information. However, it might suggest that using multiple antibiotics to generate resistant mutants might cast a wider net for pleiotropic cryptic biosynthesis induction if they each are privileged for distinct BGCs. A second interesting component of this dichotomy is the fact that the piperazimycins, in our hands and in the literature, have been demonstrated to not be anti-bacterial, while the piperidamycins, as reported by Ochi, can be toxic to bacteria. Piperazimycins' inability to kill bacteria is in direct contrast to their potent ability to kill a wide range of lung cancer cell lines. For having similar structural features, their unique activity could help inspire medicinal chemistry efforts. Previous ribosomal

engineering has exclusively focused on the induction of antibiotics, as the stringent response is used by bacteria attempting to kill competing organisms. Our work with the piperazimycins suggests that ribosomal engineering may be useful in the induction of secondary metabolism to produce eukaryotic effecting natural products, a point of great interest to biomedical research.

3.4 Ribosome engineering optimization of the MacMillan microbe library

These targeted studies demonstrated that ribosome engineering by generating rifampicin resistance was easily applicable in marine *Streptomyces* from our collection as suggested by the many studies in terrestrial *Streptomyces* by Ochi and others. However, most ribosome engineering studies have been performed in this genus with some exceptions (44 of the 47 actinomycetes studied being *Streptomyces*) (Kozo Ochi & Hosaka, 2013). We postulated if non-*Streptomyces* strains from our microbial collection would be amenable to ribosome engineering with the same efficiency. Particularly, we sought to identify whether rifampicin resistance in particular clusters of genera from our collection (*Erythrobacter* and *Bacillus*) could inform us about trends in genus susceptibility. Ideally, having 10 or more wild microbes of the same genus would allow for us to determine trends within ribosomal engineering such as chemotype of natural products induced, mutations caused, and most importantly ability to be induced. Additionally, we sought to define a work-flow for generating these mutants from pure microbial strains that could be utilized in the future with general effectiveness and ease. The work by the Ochi lab suggested that these types of large scale engineering projects were viable, albeit in terrestrial actinomycetes (Hosaka et al., 2009). Therefore, we undertook a large scale ribosomal engineering effort of non-*Streptomyces* bacteria with the major goals of 1) determining the

effectiveness of the system in certain bacterial genera and 2) designing a pipeline that could be applied to diverse groups of microbial strains.

3.4.1 Ribosomal engineering pipeline in the MacMillan lab

The design of a straightforward methodology applicable to a large assortment of bacteria required many considerations. One of the main details was the antibiotic used to induce mutations. Several antibiotics have been utilized by the broader field of ribosome engineering. Most commonly, streptomycin and rifampicin have been used to generate rpsL and rpoβ mutations respectively. As our previous efforts had shown rifampicin resistance to be quite effective, the use of this compound was maintained.

The second consideration was the dosing regimen of rifampicin to trigger resistance. For a bacteria to develop resistance, they must be cultured under a toxin concentration strong enough to generate the evolutionary pressure, but weak enough to not overwhelm these mechanisms and cause debilitating toxicity. Previously, MIC values of rifampicin for each individual species were determined with dose response curves to determine what concentration the mutagenesis would be carried out under. However, performing these assays on a large scale seemed intractable. Therefore, it was decided that each microbial strain be cultured under the same 4 concentrations of rifampicin: 5 μg/mL, 50 μg/mL, 250 μg/mL, and 500 μg/mL. This range was chosen based on literature values and previous experience with marine bacteria from our collection. The benefit of using multiple values is three-fold. First, it does not require timely MIC generation, while still permitting variability in antibiotic susceptibility. Second, it allows for inherent purity controls in that a single microbe should show a range of susceptibility (i.e. the amount of colonies should be directly proportional to concentration for a single bacteria, while

contamination would likely lead to discrepancies in this relationship). Third, generally higher concentrations of antibiotics generate heavier evolutionary pressure, resulting in a less-reversible mutagenesis event. Testing the growth of microbes over a dosing range allows for selection of growth at the highest possible concentration.

Once the antibiotic and corresponding dosing regimen was determined, the fermentation conditions had to be considered. Mainly the decision to use liquid or solid culture had to be addressed. Previous laboratory attempts were inconsistently performed. Originally, work by our group on the inducamides was performed in liquid culture using multi-welled plates mainly under a consideration of throughput. However, due to the aggregation of these wild bacteria and the inherent design of multi-welled plates, the uniformity of the cultures was an issue. Additionally, these cultures were more easily contaminated due to the violent shaking conditions required. Later work with the elfamycin producing SNC-031 M4 strain was performed with solid agar plates, which were not subject to as much contamination and also allowed for easy manipulation. Therefore, we decided to move to solid media for the generation of mutants and use liquid cultures in 8-mL snap-cap tubes for quick generation of turbid cultures, as explained below.

The designed pipeline as illustrated in Figure 49 was as follows. First freezer stocks of targeted microbes were utilized to construct an initial purity check plate, using the three streak method, and a rifampicin susceptibility plate to assess resistance development later. After growth at 30°C in a humidified incubator, these plates were stored at 4°C for a short time with photo documentation.

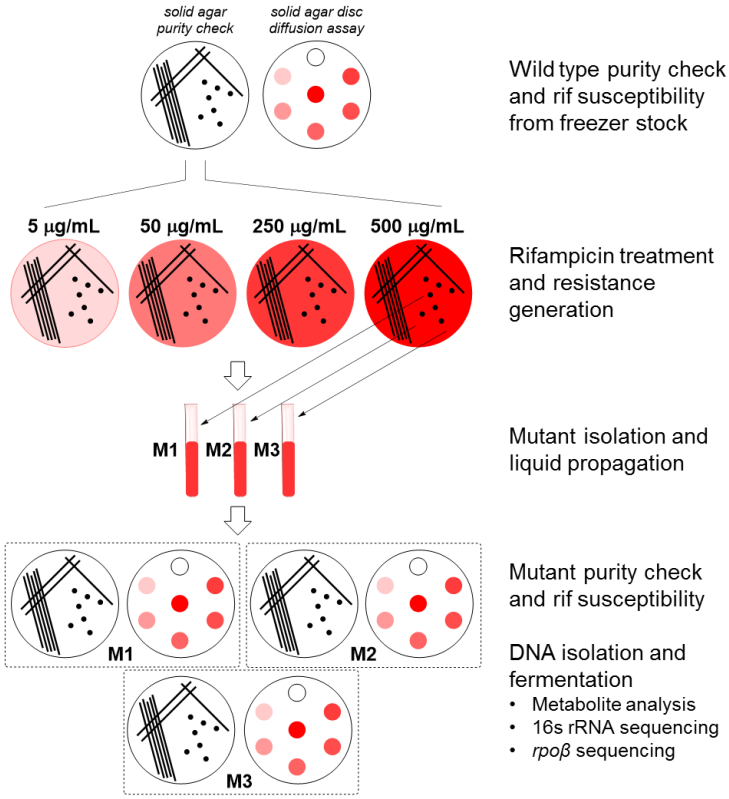


Figure 49. Pipeline for MacMillan lab ribosome engineering

The purity check served as a comparison tool against our documented microbe library for preliminary phenotypic similarity, being supplemented by 16S rRNA sequencing if needed. This plate was also a short-term source for future cultures of the wild-type strain of each bacterium. Importantly, the initial susceptibility plate was the first indication of the ribosomal engineering success. The optimal outcome was a full dynamic range of susceptibility was depicted, in that low concentrations (5 µg/mL) would permit growth and high concentrations (1.5 mg/mL) did not. The former predicted that the bacteria was not overly sensitive to rifampicin, which might have made resistance development more difficult. The latter situation in which high concentrations caused toxicity, suggested that the bacteria did not have endogenous resistance to rifampicin (e.g. efflux pumps). This toxicity was an indication that there would be evolutionary

pressure to develop resistance, as desired for effective cryptic biosynthesis induction. If this methodology was used for a particular microbe of interest, in lieu of a large scale screen, this was where alterations to the rifampicin treatment that follows could have been made.

After passing the original purity check, the bacterial plate was used to inoculate 5 mL liquid cultures of A1+C media in 8-mL snap-cap tubes. These cultures were fermented at 30°C with shaking (250 rpm) overnight or until growth appeared, as certain species were slow-growing. These cultures provided the inoculation broth to streak 4 rifampicin-containing A1+C solid agar plates ranging from 5 μg/mL to 500 μg/mL. These plates would were then incubated at 30°C in a humidified incubator until colonies formed.

At this point in the pipeline, microbial variation was demonstrated as different bacteria developed resistance at different rates or not at all. Although a range of over 100-fold rifampicin was used, certain bacteria would not grow on the lowest concentration and others replicated extensively at the highest concentration. From these rifampicin containing plates, three different colonies of resistant bacteria were selected for further consideration.

Three mutants were selected to increase the probability that the desired mutation of rpoβ had occurred, as different mechanisms could be responsible for mutagenesis. Previous large scale ribosomal engineering by Ochi indicated that 29% of type-strain Streptomyces and 43% of soil-isolated Streptomyces showed altered antibiotic activity against. However, no statistics were shown for the percentage of microbes who developed rpoβ mutagenesis or resistance. If it is assumed that the bacteria that showed novel bioactivity developed rpoβ-based resistance to rifampicin, one can assume that the correct mechanism of resistance was obtained in 1/3 of the microbes treated with rifampicin. While this is merely an estimate, Ochi’s success in generating

phenotypic responses gave us confidence in selecting three mutants. The amount of resistant bacterium selected during this step remained one of the important considerations that needed to be verified in pilot studies of this pipeline. Potentially, more mutants should be selected if the percentage of successful mutants is low, an issue discussed in the conclusion.

Another consideration for mutant selection was from which plates colonies should be picked. As explained previously, optimal mutants were taken from the highest concentration of rifampicin due to a larger evolutionary pressure and therefore more stable mutation. If there were three available colonies on the 500 μg/mL rifampicin plate, these colonies were selected. If there were not three viable colonies on this plate, colonies were selected from the next lowest concentration. Smears of colonies were avoided at this stage to avoid a mixture of potential mutants, which would obscure later sequencing results.

After the three individual mutant colonies of each bacteria were selected, they were used to inoculate liquid cultures containing the same concentration of rifampicin as the plate they were selected from. This maintained the evolutionary pressure of rifampicin and validated the resistance of the chosen strain. After fermentation at 30°C and shaking at 250 rpm, the turbid broth was used to swab a three-streak purity check plate and create a lawn for a disc diffusion rifampicin susceptibility assay. The purity check was used to compare phenotypes to the wild-type and as a bacterial source for colony-PCR for 16S rRNA and rpoβ sequencing. These two important events were used to establish that the bacteria had not been contaminated with another species and identify potential modifications in the rpoβ gene sequence. Imperative to these experiments was the inclusion of wild-type bacteria from the original purity check for a direct comparison.

Mutants free of contamination were then grown in 50 mL liquid cultures of A1+C media until turbid. Cells were removed via centrifugation and the secreted aqueous layer was extracted with organic solvent to afford the secreted metabolome of these bacteria. Fermentation and extraction processes were carried out with the wild-type bacterial strain concurrently to avoid variation. These extracts were analyzed using HPLC-MS. Specifically, the excreted metabolites of wild-type and mutant strains were examined for fluctuation in eluting metabolites or appearance of new metabolites. In the future, these extracts could be used in a range of bioassays, as some of the changes in secondary metabolism could escape detection due to lack of UV absorbance or low titers.

3.4.2 Testing of MacMillan laboratory ribosomal engineering pipeline

To carry out our first pilot screen of the ribosomal engineering, we first had to select bacteria from our collection. As explained above, the main goal was to select atypical actinomycetes, as well as non-actinomycete genera that are enriched in our collection. Three lists of bacteria were made with these criteria in mind. List 1 consisted of 32 primarily non-*Streptomyces* actinobacteria with additional unique bacteria supplemented. The goal of this List was to maintain a connection to our work and the work by Ochi by including actinomycetes, but to differentiate outside of the family of *Streptomyces* genus. We felt that this would increase the expected success while answering new questions about ribosome engineering.

Table 1. List 1 of candidate bacteria for ribosome engineering

List 1 - Non-streptomyces			
Strain	**Genus**	**Species**	**Phylum**
SNA-098	*Actinomycetospora-99*	*straminea-99, chibensis-99, chlora-99, lutea-99*	Actinobacteria
SNC-032	*Actinomycetospora-99*	*chlora-99*	Actinobacteria
SNA-053	*Agrococcus-99*	*terreus-99*	Actinobacteria
SNA-015	*Cellulosimicrobium-99*	*cellulans-99*	Actinobacteria
SNC-067	*Dietzia-94*	*maris-94*	Actinobacteria
SNA-076	*Isoptericola-100*	*variabilis-98*	Actinobacteria
SNB-036	*Isoptericola-100*	*variabilis-98*	Actinobacteria
SNA-079	*Isoptericola-99*	*nanjingensis-99*	Actinobacteria
SNC-102	*Isotericola-97*	*variabilis-97*	Actinobacteria
SNC-057	*Nonomuraea-99*	*ferruginea-99*	Actinobacteria
SNC-061	*Nonomuraea-99*	*ferruginea-99, salmonea-99*	Actinobacteria
SNC-082	*Nonomuraea-99*	*ferruginea-99*	Actinobacteria
SNC-110	*Nonomuraea-99*	*ferruginea-99*	Actinobacteria
SNB-031	*Ornithinimicrobium-98*	*pekingense-98*	Actinobacteria
SNE-016	*Promicromonospora-99*	*sukumoe-99*	Actinobacteria
SNC-040	*Pseudoncardia-99*	*oroxyli-99*	Actinobacteria
SNA-077	*Pseudonocardia-99*	*zijingensis-99, cypriaca-99, hierapolitana-99*	Actinobacteria
SNA-099	*Pseudonocardia-99*	*zijingensis-99*	Actinobacteria
SNC-017	*Pseudonocardia-99*	*halophobica-99*	Actinobacteria
SNC-030	*Pseudonocardia-99*	*ammonioxydans-99*	Actinobacteria
SNB-061	*Saccharomonospora-99*	*halophila-99*	Actinobacteria
SNC-029	*Stackebrandtia-98*	*albiflava-98*	Actinobacteria
SNF-022	*Euzebyella-99*	*saccharophila-99*	Bacteroidetes
SNE-057	*Pontibacillus-99*	*chungwhensis-98*	Bacteroidetes
SNE-027	*Pontibacter-99*	*korlensis-98, akesuensis-97*	Bacteroidetes
SNC-043	*Porifericola-91*	*rhodea-91*	Bacteroidetes
SNA-004	*Halobacillus-96*	*trueperi-96*	Fermicutes
SNE-044	*Halobacillus-99/Marinococcus-99*	*trueperi-99*	Fermicutes
SNE-045	*Pontibacillus-99*	*chungwhensis-98*	Fermicutes
SNF-021	*Lutibaculum-92*	*baratengense-92*	Proteobacteria
SNA-095	*Nitratireductor-100*	*aquimarinus-95*	Proteobacteria
SNC-044	*Phenylobacterium-95*		Proteobacteria

List Two was a shortened list which contained only bacteria from our collection that belong to the genus *Erythrobacter* (Table 2). This list was constructed due to the relative dearth of literature regarding natural products produced by *Erythrobacter* compared to other genera, while our personal studies including published work (Y. Hu, Legako, Espindola, & MacMillan,

2012; Y. Hu & MacMillan, 2011) indicate these species have the potential to be a rich source of metabolites.

Table 2. List 2 of candidate bacteria for ribosome engineering

List 2 - Erythrobacter			
Strain	**Genus**	**Species**	**Phylum**
SNB-017	*Erythrobacter-99*	*flavus-99*	Proteobacteria
SNB-020	*Erythrobacter-99*	*citreus-99*	Proteobacteria
SNB-035	*Erythrobacter-99/Halomonas-99/Bacillus-98/Citromicrobium-98*	*citreus-99, aquimaris-99*	Proteobacteria
SNB-054	*Erythrobacter-100*	*citreus-100*	Proteobacteria
SNC-041	*Erythrobacter-100*	*citreus-100*	Proteobacteria
SNC-066	*Erythrobacter-99*	*vulgaris-99*	Proteobacteria
SNC-119	*Erythrobacter-100*	*flavus-100*	Proteobacteria
SNE-029	*Altererythrobacter-99*	*epoxidivorans-97*	Proteobacteria
SNF-015	*Erythrobacter-100*	*citreus-100*	Proteobacteria

List 3, was also a single-genus assortment of microbes, albeit of *Bacillus* bacteria (Table 3). Studies indicate that *Bacillus* bacteria contain cryptic BGC (Dormeyer et al., 2015; Tojo, Tanaka, & Ochi, 2015). However, most cryptic BGC projects study actinomycetes, leaving questions about alternative bacteria including *Bacillus*. Our extensive list of Bacillus strains provide a unique opportunity to study the secondary metabolome potential of this genus using ribosome engineering.

Table 3. List 3 of candidate bacteria for ribosome engineering

List 3 - *Bacillus*			
Strain	**Genus**	**Species**	**Phylum**
SNA-012	*Bacillus-98*	*cereus-98*	Fermicutes
SNA-019	*Bacillus-99*	*endophyticus-99*	Fermicutes
SNA-034	*Bacillus-96*	*Sonorensis-96*	Fermicutes
SNA-037	*Bacillus-100*	*megaterium-100*	Fermicutes
SNA-045	*Bacillus-100/Paucisalibacillus-99/Ornithinibacillus-98*	*djibelorensis-100/globulus-99*	Fermicutes
SNA-046	*Bacillus-98*	*licheniformis-98*	Fermicutes
SNA-048	*Bacillus-99*	*hunanensis-98, lehensis-98*	Fermicutes
SNA-049	*Bacillus-99*	*cereus-99*	Fermicutes
SNA-055	*Bacillus-99*	*thuringiensis-99*	Fermicutes
SNA-064	*Bacillus-100/Brevibacillus-99*	*thuringiensis-100/brevis-99*	Fermicutes
SNA-073	*Bacillus-100*	*endophyticus-99*	Fermicutes
SNA-074	*Bacillus-99*	*licheniformis-98, niabensis-98, humi-98, herbersteinensis-98,*	Fermicutes
SNA-096	*Bacillus-99*	*sonorensis-99*	Fermicutes
SNB-008	*Bacillus-99*	*thuringiensis-99, cereus-99*	Fermicutes
SNB-028	*Bacillus-100*	*sonorensis-99*	Fermicutes
SNB-030	*Bacillus-99/Virgibacillus-99*	*sonorensis-99, licheniformis-99, aerius-99*	Fermicutes
SNB-034	*Bacillus-98*	*sonorensis-98*	Fermicutes
SNB-037	*Bacillus-95/Rheinheimera-95/Paenibacillus-95*	*lincheniformis-95/aquimaris-95/alvei-95*	Fermicutes
SNC-012	*Bacillus-99*	*vietnamensis-99*	Fermicutes
SNC-023	*Bacillus-99*	*endophyticus-99*	Fermicutes
SNC-045	*Bacillus-99*	*decolorationis-98*	Fermicutes
SNC-046	*Bacillus-99*	*methylotrophicus-99*	Fermicutes
SNC-080	*Bacillus-99*	*boroniphilus-99*	Fermicutes
SNC-087	*Bacillus-100*	*magaterium-100*	Fermicutes
SNC-090	*Bacillus-95*	*indicus-95*	Fermicutes
SNE-037	*Bacillus-99*	*algicola-99*	Fermicutes
SNE-038	*Bacillus-99*	*aquimaris-99, veitnamensis-99, alcalophilus-99, subtilis-99,*	Fermicutes
SNE-046	*Bacillus-99*	*baekryungensis-99*	Fermicutes

With these bacteria in hand, we set out to test our designed pipeline for ribosome engineering. Initial growth from freezer stock was the first major hurdle as these cultures had varying growth rates. All in all, out of the 69 bacteria selected, only 4 could not be cultured from freezer stock or consistently failed purity checks.

The ability to develop resistance was variable among the three lists as demonstrated in Table 4. Resistance generation for List 1 and 3 bacteria were similarly around 20%, whereas

Erythrobacter were highly likely to develop resistance against rifampicin as 8 of 9 were successful. However, List 2 only consisted of 9 bacterial species, while Lists 1 and 3 contained 32 and 28, respectively.

Table 4. Phenotypic resistance development of the three candidate groups.

	Rare Actinomycetes	Erythrobacter	Bacillus
Resistance development	22.9%	88.9%	20.0%
5 µg/mL	23.8%	18.2%	58.3%
50 µg/mL	9.5%	36.4%	41.7%
250 µg/mL	28.6%	22.7%	0.0%
500 µg/mL	38.1%	22.7%	0.0%

Most interestingly, the concentration of rifampicin in the plate from which resistant clones were chosen did not affect the ability of the isolated mutant to resist rifampicin. That is, every resistant bacteria resisted the highest concentration of rifampicin, which far exceeded the concentration in the resistance inducing plate, regardless of the source. This is exemplified by SNB-035, an *Erythrobacter* line incapable of forming colonies on rifampicin plates above 5 µg/mL and who showed heavy susceptibility to rifampicin in the wild-type strain (Figure 50). However, the mutant selected from the 5 µg/mL plate was able to tolerate rifampicin up to 1.5 mg/mL. This type of consistency was seen across all mutants; resistant clones, regardless of initial susceptibility, developed complete resistance of rifampicin up to 1.5 mg/mL.

Figure 50. Acquired resistance to rifampicin (L-R: WT, M1, M2)

There was some variability in the rifampicin concentration withstood by the varying group of bacteria. List 1 bacteria as a whole tolerated high concentrations of rifampicin, while the complete opposite was seen in *Bacillus* bacteria in which no colonization occurred above 50 μg/mL. The differences in these resistance development patterns will be discussed later alongside sequencing information to construct a more complete picture of how these bacteria evade toxicity.

After generating the mutants, it remained to be seen whether they 1) were the same bacteria as wild-type, 2) possess mutations of the rpoβ gene as a potential mechanism for resistance, and 3) produce different secondary metabolites than their wild-type derivative strain.

3.4.3 Genetic sequencing of resistant mutants

Contamination of cultures was an initial concern for the ribosomal engineering pipeline due to the source of the bacterial cultures and the length of time in which they are manipulated. While microbes in the MacMillan collection are painstakingly purified from collected sediments, it remains a possibility that certain strains can have contaminated spores that persist with no obvious indication. An environmental shock such as antibiotic treatment could provide opportunity for dormant, resistant bacteria to proliferate. To verify that no contamination had occurred, single colonies from each mutant were subject to DNA isolation and 16s rRNA sequencing. Of the resistant bacteria, no single mutant showed significant variation, indicating no contamination (Table 5).

Table 5. Summary of 16s rRNA sequencing

Sequence Analysis of 16S rRNA						
Strain	**Mutant**	**Identities (%)**	**Identities**	**Gaps (%)**	**Gaps**	**Score**
SNA053	SNA053_M3	100	552	0	0	1020
	SNA053_M2	100	552	0	0	1020
	SNA053_M1	100	551	0	0	1018
SNA073	SNA073_M2	100	548	0	0	1013
	SNA073_M1	100	547	0	0	1011
SNA076	SNA076_M3	100	552	0	0	1020
	SNA076_M2	100	552	0	0	1020
	SNA076_M1	100	552	0	0	1020
SNA079	SNA079_M3	100	555	0	0	1026
	SNA079_M2	100	552	0	0	1020
SNB017	SNB017_M3	100	554	0	0	1024
	SNB017_M2	100	554	0	0	1024
	SNB017_M1	100	553	0	0	1022
SNB035	SNB035_M2	100	549	0	0	1014
	SNB035_M1	100	548	0	0	1013
SNB054	SNB054_M3	100	554	0	0	1024
	SNB054_M2	100	554	0	0	1024
	SNB054_M1	100	554	0	0	1024
SNC041	SNC041_M3	100	554	0	0	1024
	SNC041_M1	100	554	0	0	1024
	SNC041_M2	99.819	554	1	0	1020
SNC066	SNC066_M2	100	548	0	0	1009
	SNC066_M1	99.635	548	2	0	1005
SNC090	SNC090_M2	100	545	0	0	1007
	SNC090_M1	100	545	0	0	1007
SNC102	SNC102_M2	99.638	552	2	0	1013
	SNC102_M1	100	552	0	0	1013
	SNC102_M3	99.637	551	2	0	1011
SNC119	SNC119_M2	100	555	0	0	1026
	SNC119_M1	100	555	0	0	1026
	SNC119_M3	100	554	0	0	1024
SNE016	SNE016_M2	100	551	0	0	1018
	SNE016_M1	100	551	0	0	1018
SNE029	SNE029_M2	100	554	0	0	1024
	SNE029_M1	100	554	0	0	1024
SNE037	SNE037_M2	95.833	552	21	2	891
SNE038	SNE038_M2	95.471	552	23	2	881
	SNE038_M1	95.471	552	23	2	881
SNF015	SNF015_M3	100	559	0	0	1033
	SNF015_M2	99.822	561	1	0	1033
SNF022	SNF022_M3	100	561	0	0	1037
	SNF022_M2	100	560	0	0	1035
	SNF022_M1	100	560	0	0	1035

Sequencing of the rpoβ gene for individual mutants was a more complicated task, due to the lack of reported sequences in contrast to 16s rRNA. The genus and species of each bacteria that developed resistance was searched in the NCBI database for reported rpoβ sequences. Related bacteria were used to generate degenerative primers that would amplify the key locations of rpoβ in which rifampicin resistance has been reported (Adekambi, Drancourt, & Raoult, 2009; Alifano, Palumbo, Pasanisi, & Tala, 2015). Particularly, the H437 location was of interest, as is the dominant location for mutation in reported ribosomal engineering studies. Successful sequencing of the rifampicin resistance domain of rpoβ was accomplished with 22 of the 66 bacterial strains. Of the 22 rpoβ sequences characterized, only a single mutant, SNF-022 M2, contained a point mutation at the H437 locus known to be responsible for rifampicin resistance after ribosomal engineering. The remaining mutants (7 strains) whose rpoβ sequencing was successful had complete homology to their wild-type origin.

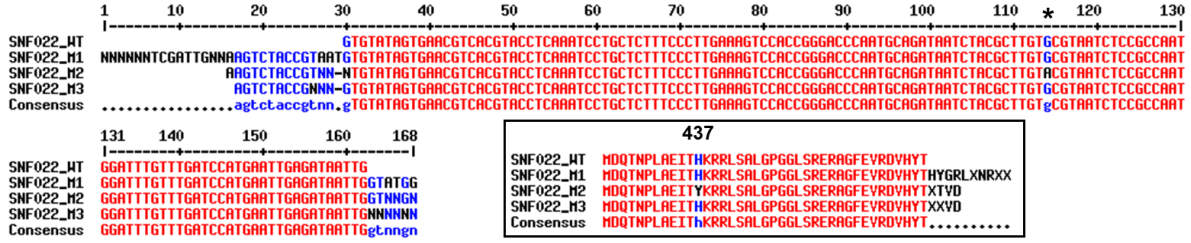


Figure 51. rpoβ sequences from SNF-022 wild-type and resistant clones

These results could be explained by several circumstances. The most likely of which is the development of resistance through alternate mechanisms. Efflux pumps and gene duplication could be the mechanism by which the bacteria developed resistance. Additionally, the region of rpoβ that was sequenced in these bacteria may not have been the region in which a mutation arose. Full sequencing of the rpoβ gene for these unique bacteria would permit study of this possibility.

3.4.4 Metabolic analysis of resistant mutants

While the genetic analysis of these resistant bacteria was underway, a chemical analysis of their secreted metabolome was carried out. This parallel analysis allowed a relationship between mutation status and natural product production without bias. Most importantly, we wanted to investigate whether alterations in the metabolome could occur without specific mutations of rpoβ in the canonical gene locations. Such a differentiation from conventional ribosome engineering was particularly of interest due to the unconventional selection of bacteria. We set out to determine if any bacteria from our pilot study fit this unique profile.

Resistant clones and their wild-type counterparts were fermented in 50 mL A1+C cultures at 30°C for 10 days. Their organic extract was examined via LC/MS and ^{1}H NMR. For LC/MS analysis, UV absorbance at λ=210 nm, 254 nm, or 280 nm was primarily examined for each strain. Alterations in the UV absorbance of each strain was then examined in more detail using PDA and mass spectrometry. To assess reproducibility of these analyses, 50 mL cultures of SNA-076 WT and M1 were grown in triplicate and analyzed (Figure 79). Replicate signals were near identical when volumetric normalization was used (i.e. all cultures extracted, resuspended, and analyzed in equivalent volumes). Cell pellet normalization gave similar results, while measurement of crude extract mass performed poorly. This phenomenon is due to the high salt content of the sea water used in A1+C media, which permits heavy salt contamination of true crude extract mass. Representative chromatograms for the complete series of crude extracts are found in the Chapter 3 Appendix.

Chromatographic analysis of the List 1 bacterial crude extracts at a wavelength of 210 nm did not reveal any obvious changes in metabolic profiles between the wild-type and rifampicin

resistant strains (Chapter 3 Appendix). As more compounds absorb light at this wavelength, it is common for significant noise to obscure smaller signals. Concordantly, comparison of UV absorbance at 280 nm by wild-type and resistant clones led to several interesting deviations in the strains, SNA-076 M1, SNA-079 M3, SNF-022 M2, and SNF-022 M3 (Figure 52). The peak magnitudes in each case were qualitatively smaller than those observed in previous ribosome engineering of *Streptomyces* strains. Therefore, validation with ^{1}H NMR was used to more precisely compare wild-type and resistant strains. The ^{1}H NMR of SNA-076, SNA-079, and SNF-022 indicated little differences between their respective wild-type and resistant derivatives (Figure 85). These results suggested that ribosomal engineering in these subset of diverse bacteria did not lead to observable changes in their metabolome.

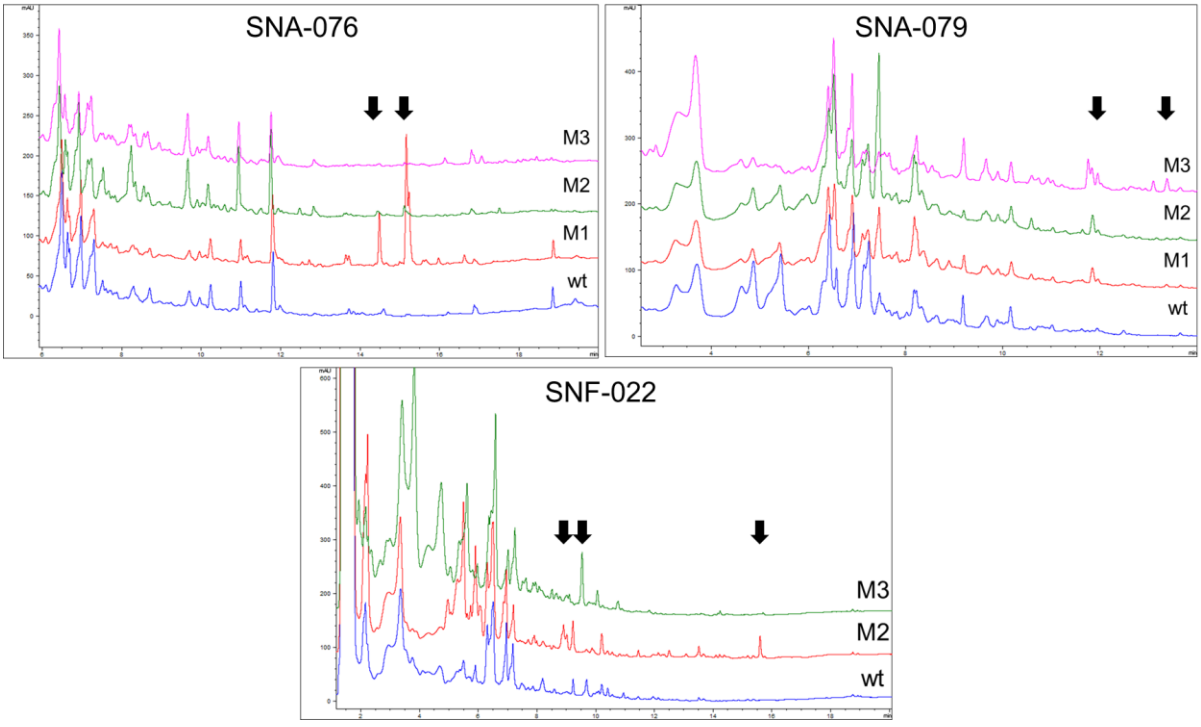


Figure 52. Crude extracts of interest in List 1 (280 nm)

Our lab has found a diverse array of natural products in the *Erythrobacter citreus* strain SNB-035 (Y. Hu et al., 2012; Y. Hu & MacMillan, 2011). However, no comparison of this strain to the remaining eight *Erythrobacter* strains from our collection has been performed. Due to their inclusion in the ribosomal engineering program, we took the opportunity to analyze these strains. Interestingly, these bacteria share a similar, metabolite rich profile via LC/MS analysis (UV absorbance λ=280 nm shown in Figure 53). A similar pattern of UV absorbance is seen between the t_R = 5.0 and 10.0 minutes. Of importance are the peaks eluting with a t_R of 8.0-16.0 minutes, as these elution times match most non-polar natural products isolated from our laboratory. No other *Erythrobacter* strain duplicated the rich chemical signature eluting between 10-14 minutes from the crude extract of SNB-035 (bottom trace). However, their species homology indicated by 16S rRNA sequencing suggested that these bacteria could contain strong biosynthetic potential.

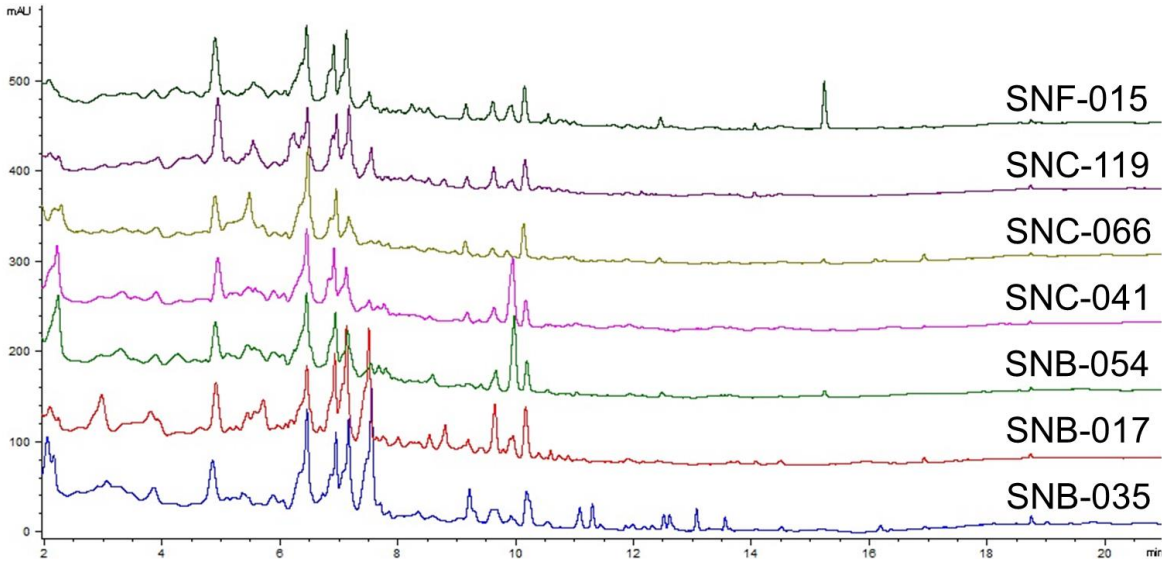


Figure 53. Comparison of Erythrobacter metabolomes

No obvious changes were observed in the chromatograms monitoring UV absorbance for all but one of the bacteria, strain SNC-119 M2 (Figure 54). A variety of peaks eluting after 8.5

minutes were of particular interest. However, follow-up studies using ^{1}H NMR yielded no stark differences between SNC-119 WT and SNC-119 M2 (Figure 96). There were small differences, such as two sets of doublets at =7.80 ppm and 7.56 ppm, present. However, the magnitude of these signals compared to the alterations seen in previous studies with *Streptomyces* bacteria suggest that any alterations to *Erythrobacter* from our collection may not be as robust. The lack of metabolic changes was also dominant in the crude extracts of the *Bacillus* strains which developed resistance (Chapter 3 Appendix). Metabolically, *Erythrobacter* and *Bacillus* strains remained the same after development of resistance to rifampicin.

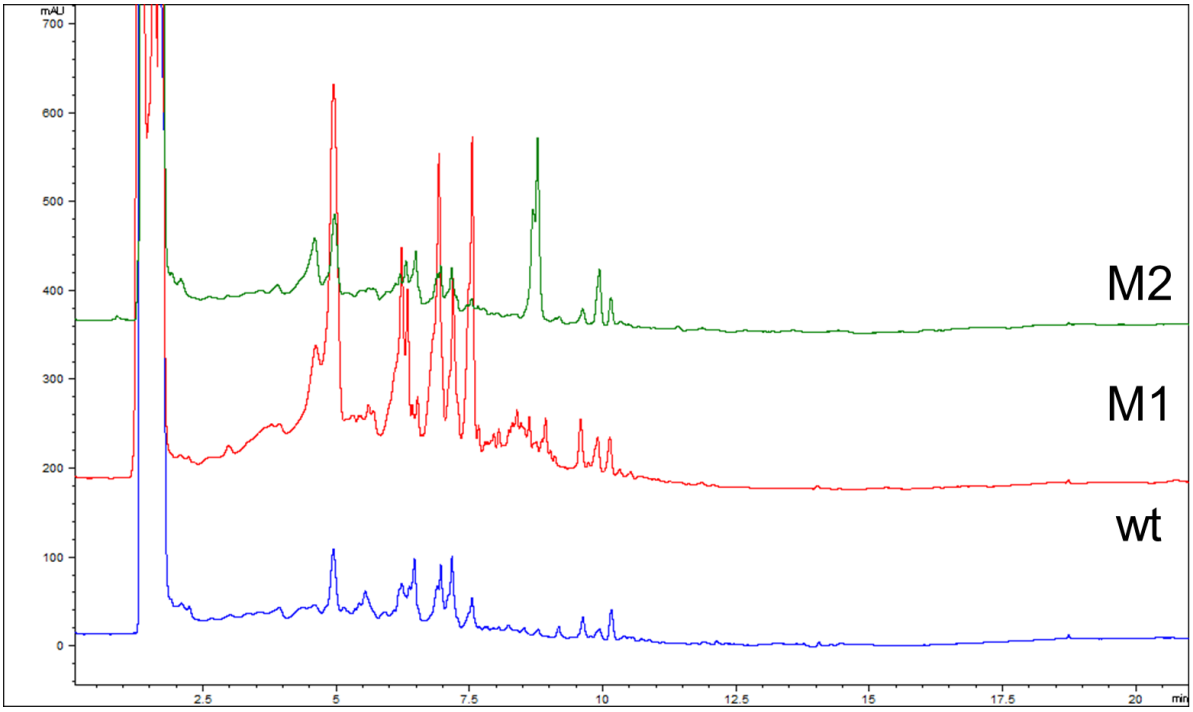


Figure 54. SNC-119 crude extracts (280 nm)

Metabolic analysis of the complete set of bacterial strains in the pilot study revealed few preliminary hits via LC/MS and no validated hits via ^{1}H NMR. Two potentially contrasting conclusions are drawn from this result. First, development of rifampicin resistance by these genera of bacteria does not correlate with altered natural product production. This could have

negative implications for the general use of rifampicin as a cryptic BGC induction tool. Potentially, other known tools such as streptomycin might be more efficient. Oppositely, the ability for rifampicin resistance to shift metabolic output might not be as robust in non-*Streptomyces* bacteria. Minute changes between wild-type strains and their rifampicin resistant clones were demonstrated with the first-pass analytical tools of LC/MS and ^{1}H NMR. Higher resolution tools such as MS/MS fragmentation or MS fingerprinting could provide a higher signal/noise ratio for weak signals in complex mixtures. Additionally, biological assays might be able to distinguish metabolic alterations of biologically active natural products. Finally, this problem could be addressed with alteration of the ribosome engineering pipeline. Using modified media and stress conditions have been reported to boost the effects of natural product BGC induction (Kozo Ochi & Hosaka, 2013). While *Streptomyces* might be better equipped for the previously established ribosomal engineering methodology, variations may be needed for widespread application in microbe libraries.

3.5 Conclusion

Induction of cryptic biosynthetic gene clusters has become a task of great interest to natural products chemists. Reports suggest that developing new techniques to access these hidden molecules will solve the problem of low chemical diversity seen in many natural product libraries. While this sentiment remains to be validated, there is evidence that different methodologies can cause production of metabolites new to a specific bacteria (Jiang et al., 2016; Shentu et al., 2016; Thanapipatsiri et al., 2016).

Studies from our own lab, including this work, have demonstrated that ribosome engineering is a powerful induction tool in *Streptomyces* bacteria (Fu et al., 2014). In a short

time, we were able to increase the titers of antibiotics (elfamycins) and anti-cancer compounds (piperazimycins) by generating rifampicin resistant clones. These clones possessed point mutations at the predicted locus of the rpoβ gene, consistent with the literature. While these compounds were known, their increased yield provided easy identification and efficient isolation. Performing strain improvement through ribosome engineering could be an efficient way to streamline validation of high-throughput screening efforts. This step is a common bottleneck in natural products screens that has only begun to be addressed (Kurita, Glassey, & Linington, 2015; J. Yang et al., 2014).

Facile application of ribosome engineering to non-*Streptomyces* bacteria remains to be demonstrated. In this work, we showed that the typical rpoβ mutation locus was not altered in most rifampicin resistant bacteria. This could have been due to several events. Primarily, a thorough rpoβ sequencing effort is required to confirm no alternative sites of the gene developed mutations. The rare and novel nature of these wild bacteria require an arduous sequencing effort, particularly for species-diverse screens. A second explanation could be the acquisition of different resistance mechanisms. Potentially, *Streptomyces* are more prone to develop rifampicin resistance than alternative genera like *Erythrobacter* or *Bacillus*. Overall, the bacterial strain in question needs to be considered before taking on this protocol. Without sufficient rpoβ sequence information, as is common for the rare extremophilic bacteria in natural product collections, mutant identification can be stymied.

Ribosome engineering has been shown to be an effective approach for large scale screens of *Streptomyces*, however more detailed studies of this methodology are needed to expand its bacterial repertoire. It would behoove the field to take on in depth case studies of different phyla

to report on their propensity to mutate rpoβ, or other applicable proteins in ribosome engineering. The success of ribosome engineering in actinomycetes suggests that these types of studies would be beneficial and should be a priority for small molecule-based cryptic BGC activation.

3.6 Experimental

Small screen mutagenesis of bacteria including SNC-031 M4. Bacterial strains were selected from freezer stock and inoculated onto A1+C agar plates. Colonies of a single strain from these plates were used to inoculate 6 well dishes with 2 mLs of media with varying concentrations of rifampicin (5 μM, 10 μM, 20 μM, 40 μM, 80 μM, 160 μM) in DMSO (1% final concentration). The cultures were shook for 72 hours at 30°C with shaking (250 rpm). Turbid cultures at 72 hours were used to streak onto two agar plates of A1+C with rifampicin at 25 μM and 50 μM to ensure resistance. Priority selection of cultures was given to those which proliferated under higher concentrations of rifampicin. Colonies from a 7 day fermentation of these plates were used to inoculate A1+C plates for a stock and 50 mL cultures for metabolome analysis. The A1+C plates were used for DNA isolation and PCR, as described below.

Disc-diffusion assay. Whatman discs were loaded with 15 uL rifampicin, specified extract, or specified chemical solubilized in methanol. The solvent was allowed to evaporate for 15 minutes, then discs were placed on freshly plated lawns of cells (*B. subtillis* for antibiotic assays or applicable microbe strain for rifampicin susceptibility) from liquid broth on to solid plates (A1+C media +2% agar). Plates were incubated at 30°C in humidified incubator until lawn was confluent. Zone of inhibition calculated as radius from center of disc until bacterial growth.

Isolation of elfamycins. SNC-031 M4 cultures (10 L) were fermented at 24°C for 10 days. Amberlite® XAD7HP resin was added (~20 mL 80% slurry in dH_2O) to each 1L Fernbach flask

and the cultures were incubated with resin for 4 hours. Then, the resin was filtered from the liquid culture and extracted with acetone overnight. This crude extract was dried *in vacuo*, resuspended in 10 mL methanol, and moved into 20 mL scintillation vials (4.5g). Crude extract was dry-loaded onto Celite® and loaded into a plastic Teledyne ISCO cartridge. Crude extract was fractionated on RediSep Rf Gold® C18 column (100 g) in two identical runs using H_2O/MeOH solvent system: 10% B hold for 5 min, up to 99%B over 30 minutes, 99% B hold for 10 minutes. Nine continuous, major peaks with multiple UV absorbance signals at 280 nm and 330 nm were collected (~100-200 mLs each). Each of these intervals was analyzed on LC/MS for peaks of interest from original comparison to wild-type SNC-031. Low-resolution LC/ESI-MS data were measured using an Agilent 1200 series LC/MS system with a reversed-phase C_{18} column (Phenomenex Kinetix, 150 mm × 4.6 mm, 2.5 μm) at a flow rate of 0.7 mL/min. The majority of these peaks were maintained in middle interval eluting from 6.2-20.1 minutes in ISCO C18 purification. Preparative HPLC was then performed on an Agilent 1200 series instrument with a DAD detector, using a reversed-phase C_8 column (Phenomenex Luna, 250 × 10.0 mm, 5 μm). The following method was used to purify kirromycin (HRMS calculated: 797.4211, HRMS measured: 797.4224, t_R = 18.5 min) and dihydromocimicin (HRMS calculated: 799.4336, HRMS measured: 799.4372, t_R = 17.5 min) using H_2O/MeOH solvent system: 50% B hold for 20 minutes, up to 99% B over 2 minutes, 99% B hold for 5 minutes.

Isolation of piperazimycin. SNC-031 M4 cultures (10 L) were fermented at 24°C for 10 days. Amberlite® XAD7HP resin was added (~20 mL 80% slurry in dH_2O) to each 1L Fernbach flask and the cultures were incubated with resin for 4 hours. Then, the resin was filtered from the liquid culture and extracted with acetone overnight. This crude extract was dried *in vacuo*,

resuspended in 10 mL methanol, and moved into 20 mL scintillation vials (4.5g). Crude extract was dry-loaded onto Celite® and loaded into a plastic Teledyne ISCO cartridge. Crude extract was fractionated on RediSep Rf Gold® C18 column (100 g) in two identical runs using H_2O/MeOH solvent system: 10% B hold for 5 min, up to 99%B over 30 minutes, 99% B hold for 10 minutes. Nine continuous, major peaks with multiple UV absorbance signals at 280 nm and 330 nm were collected (~100-200 mLs each). Each of these intervals was analyzed on LC/MS for peaks of interest from original comparison to wild-type SNC-031. Low-resolution LC/ESI-MS data were measured using an Agilent 1200 series LC/MS system with a reversed-phase C_{18} column (Phenomenex Kinetix, 150 mm × 4.6 mm, 2.5 μm) at a flow rate of 0.7 mL/min. The majority of these peaks were maintained in middle interval eluting from 6.2-20.1 minutes in ISCO C18 purification. Preparative HPLC was then performed on an Agilent 1200 series instrument with a DAD detector, using a reversed-phase C_8 column (Phenomenex Luna, 250 × 10.0 mm, 5 μm). The following method was used to purify piperazimycin (HRMS calculated: 727.3137, HRMS measured: 797.3179, t_R = 25.0 min) using H_2O/MeOH solvent system: 50% B hold for 20 minutes, up to 99% B over 2 minutes, 99% B hold for 5 minutes.

Biological testing of piperazimycin. Pure piperazimycin from isolation above was resuspended in DMSO to 10 mM. Biological testing was performed with CytoGlo assay (Promega) as previously reported (Y. Hu et al., 2013). In summary, cell lines are plated and allowed to adhere to 384-well plate, then cells are treated for 72 hrs, lysed, and the ATP quantified by CytoGlo assay. Controls for vehicle, plate, and pan-toxicity were included.

Pilot study of Ribosomal Engineering Pipeline: 1-mL freezer stock aliquots were removed from -80°C freezer and a cotton swab was used to streak a heavy 3-streak plate on solid agar

plates (A1+C + 2% agar). After growth and comparison to microbial database to preliminarily confirm purity, this plate is used to inoculate 5 mL A1+C culture which is fermented at 30°C shaking at 250 rpm until turbid (1-2 days depending on species). This liquid broth is then used to construct 3-streak purity check plate (10 uL inoculum), initial rifampicin disc diffusion assay as described generally above (conditions: methanol vehicle, 5 ug/mL, 25 ug/mL, 250 ug/mL, 500 ug/mL, 1000 ug/mL, 1500 ug/mL rifampicin), and rifampicin resistance plates. Purity check is again performed with this plate to ensure no contamination during liquid fermentation. For the rifampicin resistance plates, a 10 uL inoculum is used to streak 4 plates containing rifampicin at either 5 ug/mL, 50 ug/mL, 250 ug/mL, or 500 ug/mL. These plates are constructed as normal A1+C plates except the rifampicin is added after cooling the media after autoclave to ~45°C before pouring the plates. These rifampicin resistance plates are incubated at 30°C in a humidified incubator for 5-30 days until growth is formed. For some bacteria, no colonies were ever formed. As bacteria grew on the rifampicin resistance plates, individual colonies were selected and used to inoculate 5-mL A1+C cultures in 8 mL snap-cap tubes (which contained the same concentration of rifampicin that was used in the individual mutant's source plate) which were incubated at 30°C shaking at 250 rpm until turbid. Three colonies were selected per species and were given titles of M1, M2, and M3 with the order preference given to the higher concentrations of rifampicin (e.g. if only two colonies were present on the 500 ug/mL plate, the third mutant would be chosen from a colony on the 250 μg/mL plate and given the name M3). Once these cultures were turbid, the broth would be used to inoculate a 3-streak purity check plate (10 uL) and swab a lawn for a rifampicin susceptibility disc diffusion assay as specified earlier. This purity check was compared to the original wild-type plate for a preliminary purity

check. Also, the rifampicin susceptibility was compared to wild-type. Every colony selected from the rifampicin resistance plates showed resistance to rifampicin as evident by these disc diffusion assays. Somewhat more commonly, there were 4 specific mutants (of varying microbial species) selected that did not grow in liquid culture supplemented with rifampicin. For these mutants, new colonies were selected and grown in their case without issue. Purity check plates were used for construction of freezer stocks (in which a new 5 mL A1+C liquid culture was inoculated and grown to turbidity then 500 uL of this broth was added to glycerol to bring the final concentration to 50% glycerol, which was then frozen at -80°C) and DNA isolation (see below).

Extraction of wild-type and mutant 50-mL cultures. Once mutant cultures were established, the purity check plate was used to inoculate 50 mL cultures of A1+C in 125 mL Erlenmeyer flasks which were fermented for 7 days at 30°C shaking at 250 rpm. For each microbial species (e.g. SNX-###), the successful mutants were inoculated alongside wild-type bacteria so that all cultures (optimally 4: WT, M1, M2, M3) are grown in identical conditions and extracted together. After 7 days the contents of each 125 mL Erlenmeyer flask were poured into sterile 50 mL conical tubes and spun down at 4,000 rpm in a J5.2 Beckman Rotor for 15 minutes. The supernatant was poured into a clean 125 mL Erlenmeyer flask and this aqueous solution was extracted with an equivalent (~40 mLs for most) volume of ethyl acetate 3X in a 125 mL separatory funnel. This organic layer (~120 mL) was then dried down in 250 mL rbfs on a rotary evaporator for around 30 minutes until dry. Then, 10 mL methanol was added to the rbf and sonicated for 3-5 seconds in a bath sonicator before being transferred via Pasteur pipette to a 20 mL scintillation vial. These vials were then dried down together in a vacuum concentrator

overnight at room temperature. Dried extracts were then weighed and placed in 4°C refrigerator until analysis. Due to the varying amounts of salts in each 50 mL A1+C media (due to the saturated nature of the formula), it is personally customary to use volumetric equivalents instead of mass equivalents. A further methanol extraction to remove salts does not seem effective given the already low amount of material contained within 50 mLs of culture. Previously, this volumetric theory has been validated using standard curves from a separate project. Thus, from this point forward, these extracts are treated as equivalent in "true" mass of the secondary metabolome. To each of these extracts within a species, at one time, is added 2 mLs methanol and the extract is sonicated for 2-3 seconds in a bath sonicator. Then, 100 uL of this solution is added to a 1.5 mL Eppendorf tube and spun down at max speed ($14.5*10^3$ rpm) on a countertop micro-spin centrifuge. The supernatant (~90 uL) was removed from this microtube and added to a conical LC/MS vial. For many of these samples, insoluble salts remain in the pellet. These samples were then analyzed on an Agilent 1200 series LC/MS system with a standard gradient (H_2O:CH_3CN: 10% hold for 1 minutes, up to 99% over 8 minutes, hold at 99% for 5 minutes) reversed-phase C_{18} column (Phenomenex Kinetix, 150 mm × 4.6 mm, 2.5 μm) at a flow rate of 0.7 mL/min. Detection of analytes was allowed via UV absorption at 210 nm, 254 nm, 280 nm, and 330 nm. Positive ion observation via electron spray ionization (ESI) was used mass spectrometry analysis.

Sequencing of mutant bacteria and primer list. Colonies were selected from purity check plates and DNA was isolated and sequenced as previously described (Fu et al., 2014) with the primers indicated (Table 10). Sequencing was carried out by Sanger Sequencing Core (UTSWMC).

CHAPTER FOUR

BIOLOGICAL STUDIES OF THE DISCOIPYRROLES

4.1 Background

The discoipyrroles (DPs), whose biosynthesis was discussed in detail in Chapter 2, were originally identified due to their anti-proliferative effects in the context of non-small cell lung cancer (NSCLC). Specifically, we became interested in these family of molecules as they demonstrated a mimetic effect to siRNA knockdown of the oncogene discoidin domain containing receptor 2, or DDR2 using the FuSiOn screening platform (Y. Hu et al., 2013; Potts et al., 2013). Preliminary validation of the FuSiOn prediction propelled us to develop these molecules as cancer therapeutics. The role of DDR2 in cell biology, and more specifically in lung cancer, will be discussed briefly as a background to the drug development of the DPs

4.1.1 A brief history of the discoidin domain containing receptors

Receptor tyrosine kinases (RTKs) are transmembrane proteins whose intracellular domains possess catalytic activity capable of initiating diverse signaling pathways in the cell. Extracellularly, RTKs are known to bind growth factors and hormones (Lemmon & Schlessinger, 2010). Thus, RTKs in general function to sample the extracellular environment and relay signals to control proliferation and development. In oncology, aberrant signaling of RTKs are associated with malignant growth. Specifically, dysregulation of the RTKs EGFR, MET, and insulin-like receptors have been shown to be prevalent in lung cancer and targeting these receptors has represented a potential patient-specific therapy (Takeuchi & Ito, 2011).

The transmembrane proteins, discoidin domain containing receptor 1 and 2 (DDR1 and DDR2) are a subfamily of RTKs. These receptors were first identified and characterized by the homology of their extracellular domain to soluble lectins found in the slime mold *Dictyostelium discoideum* used for aggregation and migration (Johnson, Edman, & Rutter, 1993; Springer, Cooper, & Barondes, 1984). Over the years, studies have shown that the DDRs are quite different from canonical RTKs. Particularly, they are distinguished by their ligands and signaling kinetics. While canonical RTKs bind growth factors, the DDRs bind fibrillar collagen of varying specificity for each DDR (Shrivastava et al., 1997; Vogel, Gish, Alves, & Pawson, 1997). Extracellular matrix proteins like collagen traditionally bind integrin receptors, which are not self-catalyzing. These receptors signal by associating with catalytic RTKs (Streuli & Akhtar, 2009). Thus, the DDRs bind the ligands of integrins, while possessing self-catalytic function similar to RTKs. However, their signaling differs from traditional RTKs by virtue of its time to complete activation and duration at this level. After DDR dimers bind collagen, monomers of this pair phosphorylate each other to begin the signaling cascade. However, phosphorylation of downstream messengers and genetic regulation develop slowly and these effects are active for extended periods of time after ligand binding (Carafoli & Hohenester, 2013).

Study of the DDRs signaling pathways has recently risen in popularity. This has been spurred by the discovered role of DDR mutations in different diseases driven by accessible genome sequencing (Egan et al., 2014; Guo et al., 2015; Kenmotsu et al., 2014; Kurashige et al., 2016; X. Shi et al., 2016; Stein et al., 2012; Terashima et al., 2016). While a discussion of the diverse, putative downstream partners of DDR2 is not applicable in this text, a quick discussion on the phenotypic effects of DDR2 signaling are more appropriate. Once the discoidin-like

extracellular domains of the DDRs bind triple helical fibrillar collagen, activation is maximized 12 hours later. This has been demonstrated by the phosphorylation patterns of immediate downstream components such as SHP-2 (Iwai et al., 2013). DDR2 activation has been linked to matrix metalloproteinase secretion, epithelial-mesenchymal transition, cell migration, cell invasion, and other processes involved in the category of extracellular matrix modification and cellular movement. As these pathways suggest, the DDRs have been shown to be important in bone development, wound healing, and mesenchymal tissue function (Leitinger, 2014; Rammal et al., 2016).

4.1.2 Targeting DDRs in the context of non-small cell lung cancer

Interest in the role of DDR mutations in lung oncology began in 2011. Hammerman and co-workers utilized a sequencing effort of squamous cell carcinoma (SCC) tumors and cell lines to identify common tyrosine kinase mutations (Hammerman et al., 2011). The researchers found that 3.8% of SCC patients contained a mutation in DDR2. Cells containing this mutation were highly sensitive to DDR2 knockdown or RTK inhibition with dasatinib compared to their wild-type counterpart. Groups across the globe searched for matching DDR2 mutations in different cohorts of lung cancer patients. Mixed results were found with some showing similar mutation rates (An et al., 2012), slightly lower mutation rates ("Comprehensive genomic characterization of squamous cell lung cancers," 2012), or none at all (Sasaki et al., 2012).

4.1.3 Connection of the discoidin domain-containing receptors to the discoipyrroles

Early work by our lab identified the discoipyrroles as mimics of DDR2 knockdown in a prostate cancer cell line (Y. Hu et al., 2013). Additionally, the DPs inhibited the DDR2-associated processes of cell migration. The intriguing suggestion by Hammerman and other

laboratories led us to examine the role of the DPs in DDR2 signaling in NSCLC cells. However, early work in our lab showed that the discoipyrroles did not affect DDR1 or DDR2 activation directly (data not shown). We hypothesized that the discoipyrroles modulated DDR2 related pathways indirectly. Our studies began by examining DPs effect on NSCLC cell lines *in vitro*. Then, we attempted to identify the molecular underpinning of DPs action both *in vivo* and *in vitro* using probe-based chemical biology approaches.

4.3 Exploration of discoipyrroles cytological effects *in vitro*

4.3.1 Toxicity profile of the discoipyrroles in NSCLC cell lines

To study the effect of the DPs in NSCLC, choosing cell lines that were highly informative was key. Therefore, we took advantage of a collaborating laboratory that has characterized over 120 lung cancer cell lines at an in-depth molecular level (oncogene/tumor suppression status, gene expression, intercellular interactions, etc.). Our interest in DDR2 led us to focus on three cell lines: 1) A549, a SCC cell line with a wild-type DDR2 status, 2) HCC366, a SCC cell line with a mutation (L239R) in DDR2, and 3) H2286 a hybrid SCLC /NSCLC line with mutant DDR2 status (I638F). Initial 72 hour cytotoxicity studies of the discoipyrroles demonstrated that the DDR2 mutated cell lines were sensitive to treatment in the nM range, while A549 cell lines were unaffected sub-μM (Figure 55). In preliminary experiments, the DDR2 mutated cells, HCC366 and H2286, were identical in their susceptibility against the discoipyrroles. However, H2286 cells grew more rapidly and consistently and thus were used as the preferable DP-sensitive cell line for growth inhibition assays. HCC366 cells, being of large size, were used preferentially for microscopy.

	HCC366[a]	H2286[a]	A549[b]
DP-A	0.120	0.256	10.2
DP-B	0.190	0.276	8.8
DP-C	0.712	0.796	19.6
DP-D	0.275	0.412	13.4

[a]*ddr2* mutant NSCLC cell line. [b]Wild-type *ddr2* NSCLC cell line.

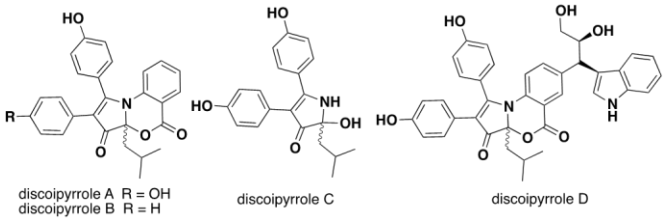


Figure 55. Initial toxicity assays (IC_{50}) and structures of the discoipyrroles.

Due to the propensity for oxidation and hydrophobic nature of the discoipyrroles, a more stable analog was synthesized. A medicinal chemistry program was developed (work unpublished) and yielded various compounds with a diverse range of anti-proliferative activity and pharmacokinetic properties. Particularly, SW212917 and SW208659 were useful as they demonstrated potent and poor growth inhibition, respectively. Compound SW208659 was an invaluable negative control due to its structural similarities to active discoipyrroles. In addition, dasatinib was utilized as a positive control for DDR2 inhibition. Dasatinib is a broad RTK inhibitor whose activity against DDR2 has been demonstrated *in vitro* and *in vivo*, including in the cell lines, HCC366 and H2286 (Terai et al., 2015; C. Xu et al., 2015).

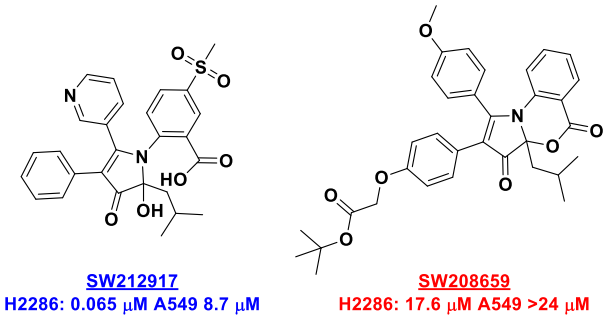


Figure 56. Structures and growth inhibition (IC_{50}) of DP analogs

4.3.2 Cytostatic effect of the DPs

The discoipyrroles were originally hits in a high throughput screen observing a single time point using the CellTiterGlo (Promega) ATP quantification assay. Therefore, we sought to

determine the time course of their cytological effects. Treating the three cell lines with 500 nM of either the active DP analog, inactive DP analog, or dasatinib confirmed the disparate sensitivity between the three. The growth of A549 cells was not altered in response to treatment by any of the compounds. Both HCC366 and H2286 ceased growth after 72 hrs of treatment with both the active DP analog and dasatinib, but not the inactive SW208659 (Figure 57). Interestingly, the growth of sensitive cells was maintained at a basal level. This did not suggest a cytotoxic effect that would result in a decrease of viable cells, but rather a cytostatic effect.

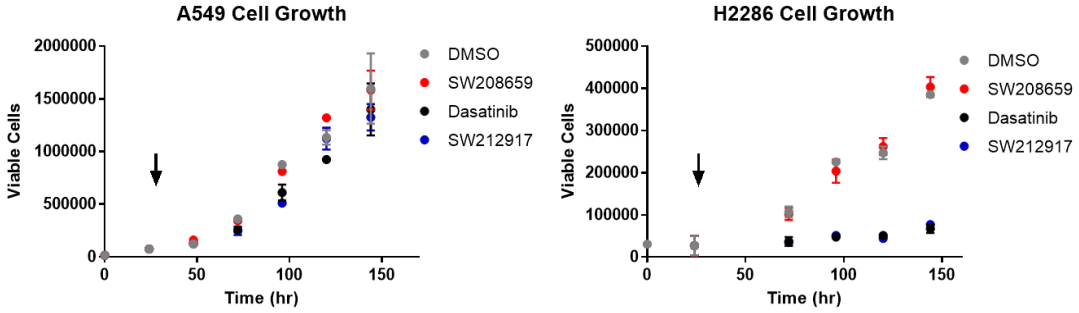


Figure 57. Time course of DP treatment

This cytostatic effect might have been due to the dosing concentration of the DPs. Therefore, a higher dose was tested for the ability to overcome this boundary and result in a cytotoxic effect. However, even when dosed at 5 μM, almost 100X higher concentration than the observed IC_{50}, cell numbers of sensitive NSCLC line did not decrease (Figure 58). To more effectively determine the cytological outcome of these cells, apoptosis was measured indirectly through extracellular DNA quantification. Dosing of SW212917 up to 10 μM did not induce DNA release from cells, a hallmark of apoptosis. This was in contrast to dasatinib and the known mitochondrial inhibitor, rotenone (Figure 58).

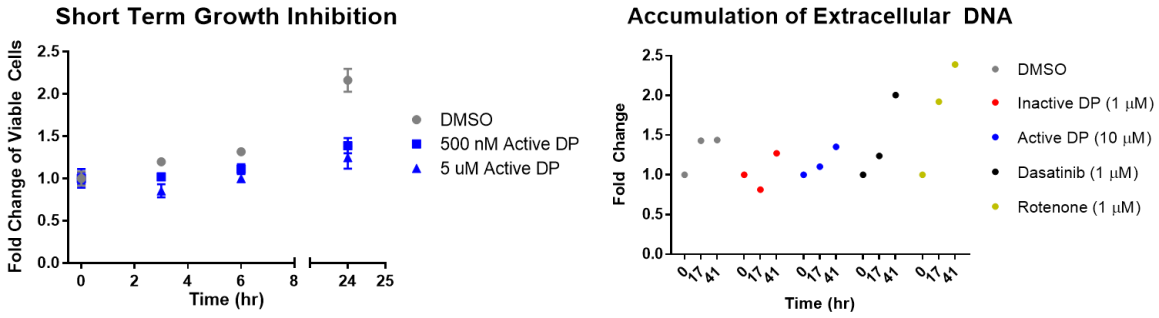


Figure 58. Short term toxicity of DPs

DNA accumulation is commonly used as a first step in determining apoptosis. Therefore, we sought to further characterize cell lines treated with SW212917 using flow cytometry based methods. All three cell lines were cultured for 24, 48, and 72 hrs with SW212917, SW208659, or DMSO. At each time point, sensitive cell lines HCC366 and H2286 showed no evidence of apoptosis by accumulation of cells with DNA content below 2N when treated with SW212917 (Figure 59) or the negative control SW208659. However, the sensitive cell lines, but not the insensitive A549 cell line, demonstrated an accumulation of cells containing 2N DNA content and a decrease in 4N content compared to vehicle. These results suggest that H2286 and HCC366 cells are arrested at G0/G1 in response to SW212917 treatment. A slight accumulation of A549 cells demonstrated a similar G0/G1 arrest phenotype, but to lesser extent than H2286 and HCC366. This validated high throughput screening data that showed the sensitivity of A549 and other resistant cell lines at concentrations of DP in the high uM range.

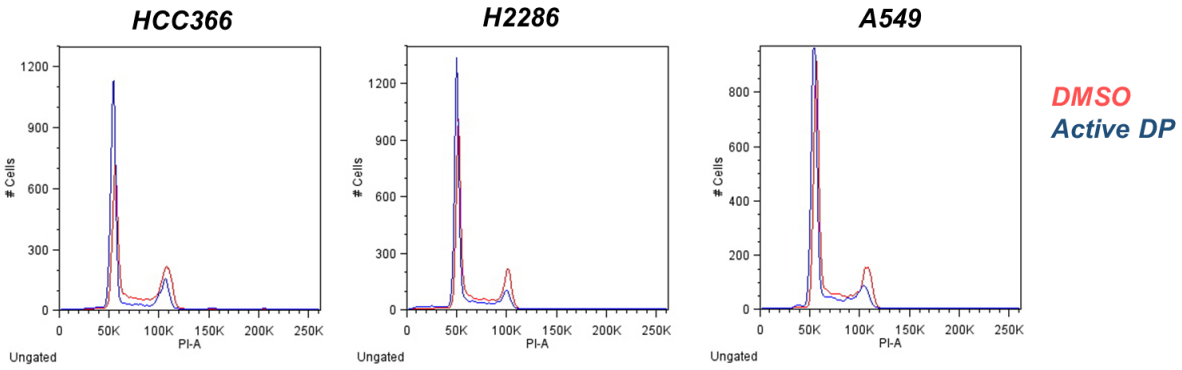


Figure 59. Flow cytometry analysis of DP treated cells

4.3.3 DPs cause irregular cell morphology and affect migration

After determining the cytostatic nature of the discoipyrroles, we sought to characterize the differential effect on these cell lines. During previous toxicity assays, a morphological anomaly was noted in various cell lines. Sensitive cell lines, such as HCC366 and H2286, demonstrated a cell rounding effect that altered the canonical cobblestone-like appearance of epithelial cells. While SW212917 did not inhibit the growth of A549 cells, it induced the irregular phenotype similar to sensitive cell lines (Figure 60). Additionally, the cells seemed to have a disrupted intercellular junction leading to disperse islands of elongated, irregularly packed cells. This information suggested that the discoipyrroles do have a target in the insensitive cell lines, however the binding of the discoipyrroles does not affect the ability for the cells to grow at sub-uM concentrations.

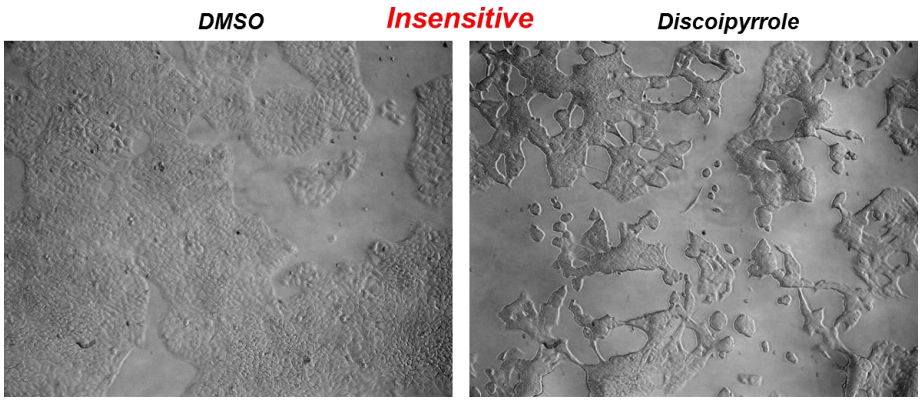


Figure 60. Phenotypic effects of DP treatment on A549

The unoccupied space in a single viewing field as demonstrated in Figure 60 also called into question the previous growth assessment of A549. By eye it seemed that there were fewer A549 cells when treated with SW212917, however cell counting consistently showed no difference. To explain this observation, the insensitive A549 cells would have to be smaller in size and more compacted in arrangement. To test this hypothesis, phalloidin-based actin staining was performed on A549 cells treated with DMSO vehicle and the active DP analog, SW212350. Confocal microscopy of these cells at 48 hours after dosing indeed demonstrated that A549 cells had retracted actin skeletons and appeared to exist in tight groupings of cells that were limited in size, as evident by nuclear and cytoskeletal staining (Figure 61). This suggested that A549 cells were affected by the presence of the discoipyrroles, but were not experiencing growth inhibition. This evidence once again suggested that the target of the discoipyrroles is of varying significance to lung cancer cell growth.

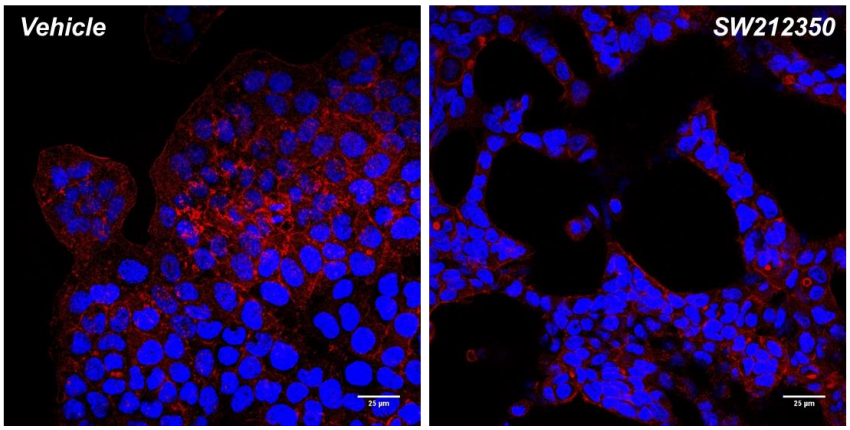


Figure 61. Confocal microscopy of treated A549 cells

4.3.4 Assessment of DPs role in collagen associated pathways

Our observed disparity in growth inhibition was thought to be due to different genetic mutations in DDR2 based on evidence from the literature and our own experiments. The sensitivity of the same three cell lines, HCC366, H2286, and A549, to dasatinib match that of the discoipyrroles. The Hammerman group showed that dasatinib's anti-cancer effect is DDR2 mutation status dependent using mutagenesis models (Terai et al., 2015). In these studies, the sensitivity of A549 and HCC366/H2286 could be switched by modulating DDR2 mutation status. The Hammerman group also demonstrated that dasatinib is cytostatic and not cytotoxic *in vitro* and *in vivo*. Besides phenocopying the sensitivity disparity and cytostatic effect of dasatinib in the same cell lines, the discoipyrroles strongly mapped to knockdown of DDR2 in our original FuSiOn assay (Y. Hu et al., 2013). Follow-up assays in BR5 fibroblasts demonstrated that the discoipyrroles inhibit growth factor mediated migration, a cell process shown to be related to DDR2 function (Gonzalez et al., 2017; Xie et al., 2015; J. Xu et al., 2014). Finally and most importantly, A549 cells were sensitized to discoipyrrole treatment under starving conditions and collagen I stimulation (Figure 62). Thus, when a DDR2-dependent pathway was the primary

source of growth stimulation, the discoipyrroles inhibited the growth of A549 cells (Carafoli & Hohenester, 2013; Vogel et al., 1997; K. Zhang et al., 2013). Morphologically, the discoipyrroles led to a decrease in cytoskeletal branching, a phenotype induced by collagen I (Figure 62). These results demonstrated that the discoipyrroles inhibit collagen signaling pathways and that these pathways dictate the effect on cell growth. With this data in mind, we set out to study the collagen associated pathways of cell migration and adhesion. The cell line A549 provided an excellent model, as the obvious phenotypic effects were decoupled from growth inhibition seen in sensitive cell lines such as HCC366 and H2286.

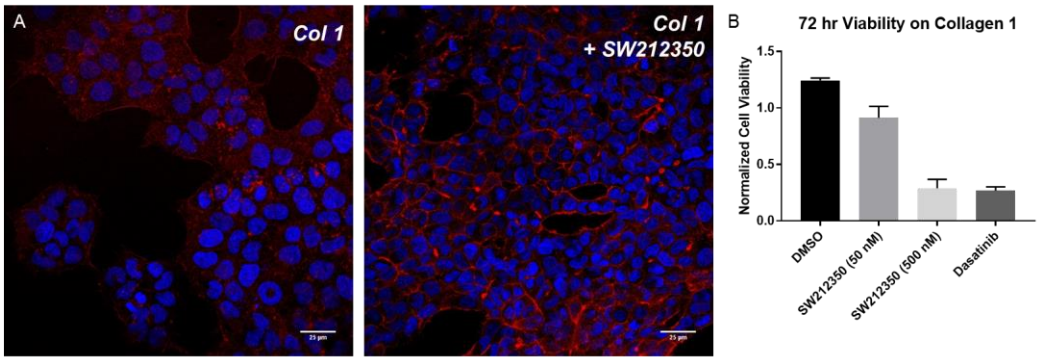


Figure 62. Collagen effects on actin skeleton and DP growth inhibition

A gold standard of migration assessment is movement through membrane filters with pores small enough to prevent inactive movement, but large enough to allow cell passage. Migration assays of A549 showed that the discoipyrroles were capable of inhibiting movement through such membrane in the presence and absence of the known migration stimulating growth factor, TGF-β (Figure 63).

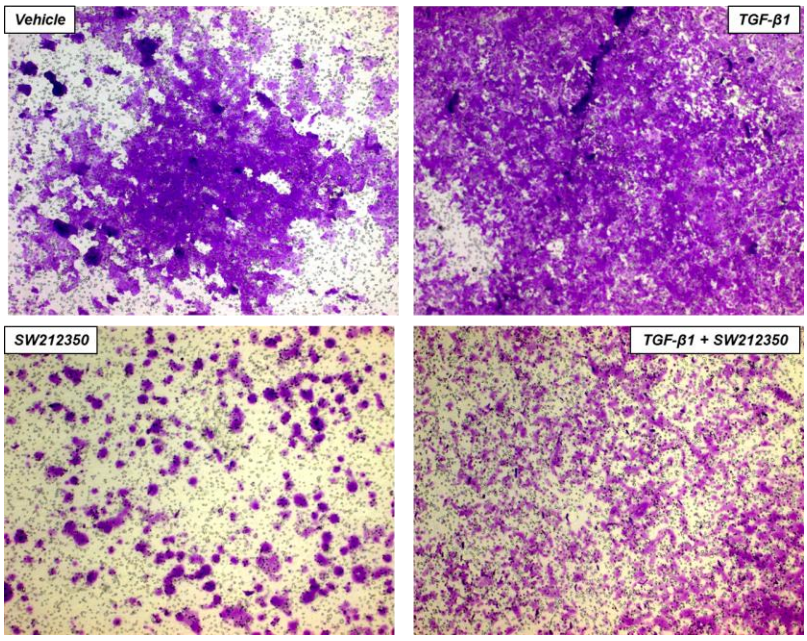


Figure 63. Migration assays of A549 cells

A major cellular process intimately tied to cell migration is epithelial mesenchymal transition (EMT). EMT allows for epithelial cells, which are canonically immobile and tightly packed with neighboring cells, to become mobile and migrate through tight junctions. The DDRs have been implicated in EMT transition (Walsh, Nawshad, & Medici, 2011; J. Xu et al., 2014; K. Zhang et al., 2013). Therefore, we set out to determine if the migration inhibition demonstrated by the discoipyrroles was indicative of EMT inhibition. This was performed by measuring protein markers of an epithelial state (E-cadherin high, vimentin low) and a mesenchymal state (E-cadherin low, vimentin high). Vehicle treated A549 cells underwent EMT readily as did A549 cells treated with the discoipyrroles (Figure 64). These findings were replicated in the sensitive cell line, HCC366 (data not shown), which did not undergo EMT as readily. These results indicated that the inhibition of migration by the discoipyrroles is mediated through a pathway disparate from EMT. The dependence of EMT on DDR2 has been reported in various manners in the literature. In a recent breast cancer study, TGF-β induced EMT could occur normally in a

DDR2 deficient background (K. Zhang et al., 2013). The discoipyrroles inhibitory effect of collagen signaling independent of EMT match the findings of Zhang and co-workers.

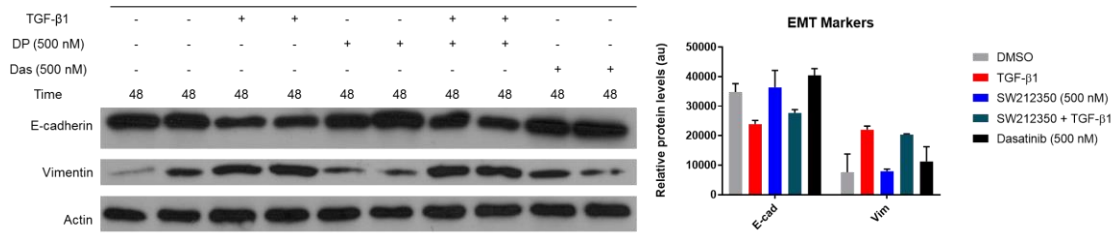


Figure 64. Assessment of EMT in A549 cells treated with DPs

Overall, these results describe the phenotypic effects of the discoipyrroles on different NSCLC cell types. While the discoipyrroles do not inhibit the growth of A549, they seem to inhibit cellular processes associated with collagen stimulation and migration, but not EMT. The same phenotypic effects are seen in the sensitive cell lines H2286 and HCC366, however the increased sensitivity to the DPs obfuscates these assessments. These results are not surprising given that other groups have demonstrated similar effects of dasatinib with the same cell lines. While it is not possible to ascribe these differences to the receptor DDR2 without direct evidence, there is sound reason to pursue this mechanism of the DPs.

4.5 Exploration of the molecular target of the discoipyrroles

4.5.1 Affinity based probes to attempt molecular target identification

The clustered phenotypic effects of the discoipyrroles suggested they were acting on specific cellular pathways. However, we sought to use an unbiased approach to determine what exactly the discoipyrroles were binding to (molecular target) to elicit these phenotypes. A methodology to accomplish such tasks is affinity based enrichment (Figure 65) (Burdine & Kodadek, 2004; Schenone, Dancik, Wagner, & Clemons, 2013). In these experiments, a tag is synthetically incorporated into a structural analog of the molecule of interest. This new tagged

molecule is incubated with live cells or protein lysate at which time it is allowed to bind to its target protein. The incorporated tag can then be used to purify or label the bound protein. Commonly this tag is an alkyne or azide which is amenable to the biorthogonal reaction, copper catalyzed click chemistry (McKay & Finn, 2014). In these reactions alkynes and azides participate in a cycloaddition to generate a heterocyclic ring irreversibly binding the two moieties. The tagged molecule can be conjugated to affinity purification moieties, such as biotin, or to labeling moieties, such as a fluorophore. The diversity of these types of experiments has been explored in many reviews (Abet, Mariani, Truscott, Britton, & Rodriguez, 2014; Lapinsky & Johnson, 2015; Sumranjit & Chung, 2013; P. Yang & Liu, 2015).

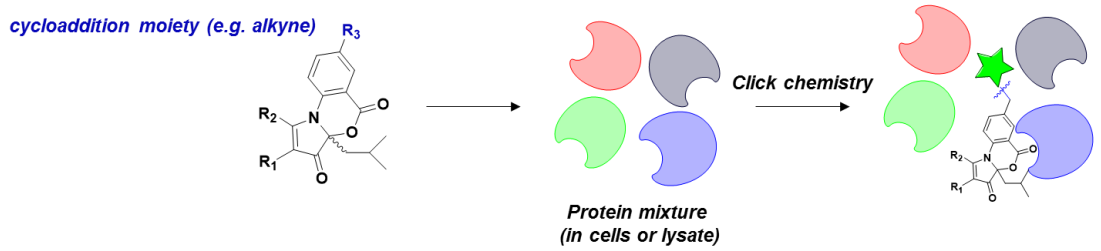


Figure 65. Conceptual overview of affinity based target identification

To adopt this approach we first needed to construct alkyne tagged discoipyrrole analogs. Designing these molecules required consideration for the possible binding mechanisms between the discoipyrroles and their putative protein targets (Figure 66). A compound that binds covalently to its protein target will do so irreversibly and thus present a case where off-rates do not have to be considered for post-binding processing (lysing, cycloaddition reaction, purification, etc.). For compounds that do not bind covalently, such as those binding in hydrophobic pockets via Van der Waals forces, the ability to maintain bonding during processing will be diminished. A common way to combat this is to incorporate functional groups which can be induced by UV light to form covalent bonds promiscuously with proximal proteogenic

functional groups (Mackinnon & Taunton, 2009). A third and more rare scenario could be possible for compounds like the discoipyrroles which can undergo structural rearrangements. The discoipyrroles are known to exist as both their traditional structure and the open-form, where the lactone ring on the eastern side of the molecule is open, presenting a free carboxylic acid and alcohol. When opened, an iminium ion can be formed generating a potent carbon electrophile. This mechanism was suggested by the mechanistic studies of the discoipyrroles presented in Chapter 2. In addition to attack by the carboxylate intramolecularly, one could imagine this iminium being subject to intermolecular nucleophilic addition by amino acids such as lysine or cysteine. However, this interaction could be reversible similar to the intramolecular reaction, which would require a UV-cross linker.

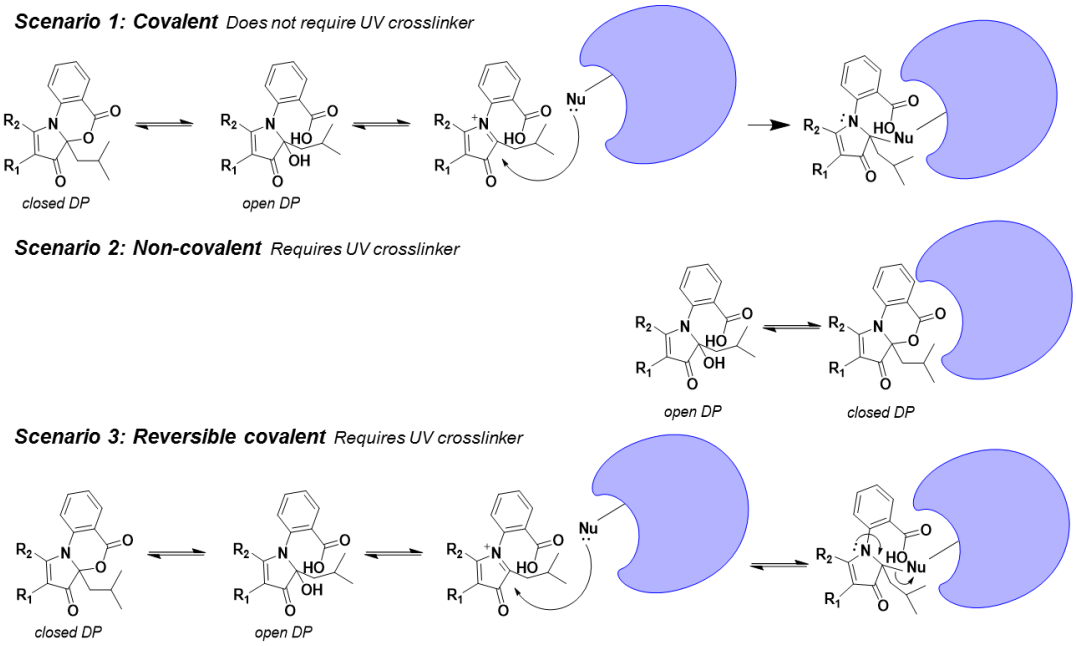


Figure 66. Possible binding modes of DP to a protein target

The exact nature of discoipyrrole binding was not known, however preliminary experiments using an propargylated discoipyrrole analog without a UV cross linker indicated that

these compounds did not bind covalently, eliminating Scenario 1 in Figure 66. To combat weak or reversible binding, a series of alkynylated discoipyrroles were made with UV crosslinking moieties such as aryldiazirines, azides, and benzophenones (Figure 67).

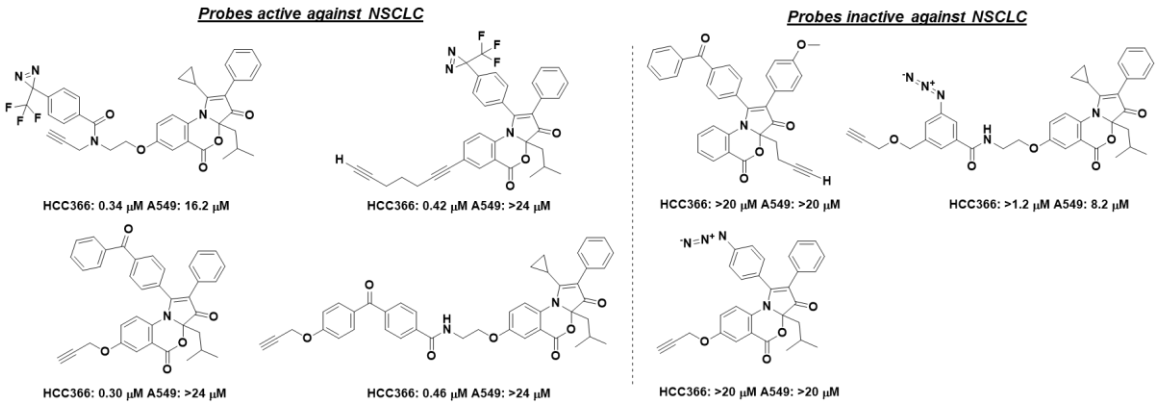


Figure 67. Various alkynylated DP analog probes

4.5.2 Pipeline for preliminary target validation studies

A brief pipeline was designed for assessing the utility of these probes. The first requirement was maintenance of selective activity against NSCLC cell lines, A549 and HCC366. Many of the compounds of diverse alkyne and UV-cross linker placement fulfilled this requirement (Figure 67, left). Active compounds were then tested in a fluorescent labeling experiment to visually confirm that proteins within a cell were being bound specifically by the probe compound (Figure 68). This experiment involved incubation with whole cells for a period of time followed by washing and UV crosslinking at the appropriate wavelength. After crosslinking, these cells were lysed and the resulting lysate used in a cycloaddition reaction with azide-conjugated fluorophores. After quenching and protein precipitation, the protein lysate was electrophoresed using SDS-PAGE and visualized on a fluorescent gel reader. Validation was assessed by fluorescent labeling of bands within the gel that were competed away in a different

treatment condition featuring co- or pre-incubation of the original cells with non-alkynylated analogs.

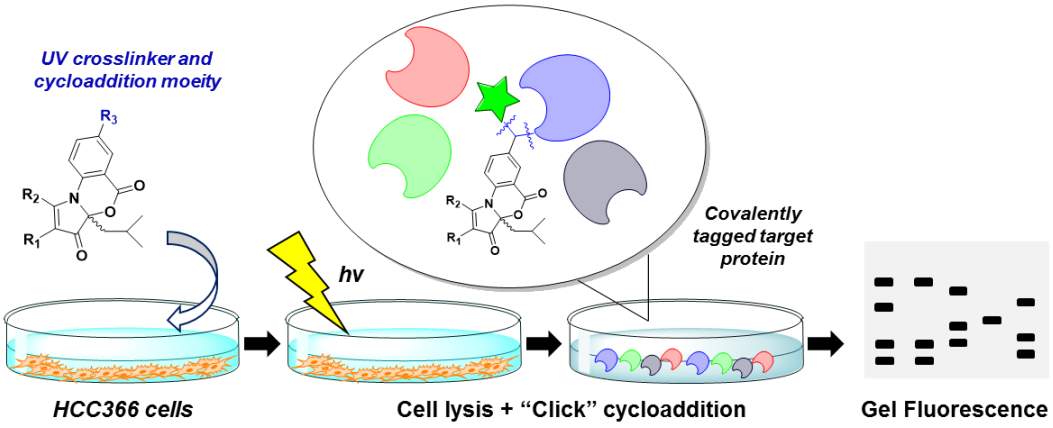


Figure 68. Overview of proteomic visualization

Once this preliminary visualization was confirmed, a probe compound was used in the affinity purification of the protein target. This scenario involved a similar experimental work-flow, albeit the fluorescent probe was replaced by a biotinylated probe. This reaction mixture was then purified using streptavidin-coated beads, eluted, and analyzed via mass spectrometry.

4.5.3 Visualization of DP probes in lysate and fixed cells

With bifunctional probes in hand and a pipeline to validate initial hits, we set out to find the molecular target of the DPs. Incubation of various probes with sensitive and insensitive cells led to no discernible proteome labeling, as demonstrated by a representative fluorescent scan in Figure 69. While bands appear to be enriched in fluorescence, competition with non-labeled DPs was not observed. To address the notion that slow off-rates were responsible for the lack of competition, pre-incubation of various competitors was attempted. However, pre-incubation from 1 hour to 48 hours did not alter the fluorescent patterns. The observation of bands that could not be competed away indicate that the labeled DP probes showed high non-specific

background as a result of promiscuous binding. Decreasing probe concentration, separating subcellular fractions, and alternate click chemistry approaches all did not decrease the non-specific binding of the probe DPs (data not shown).

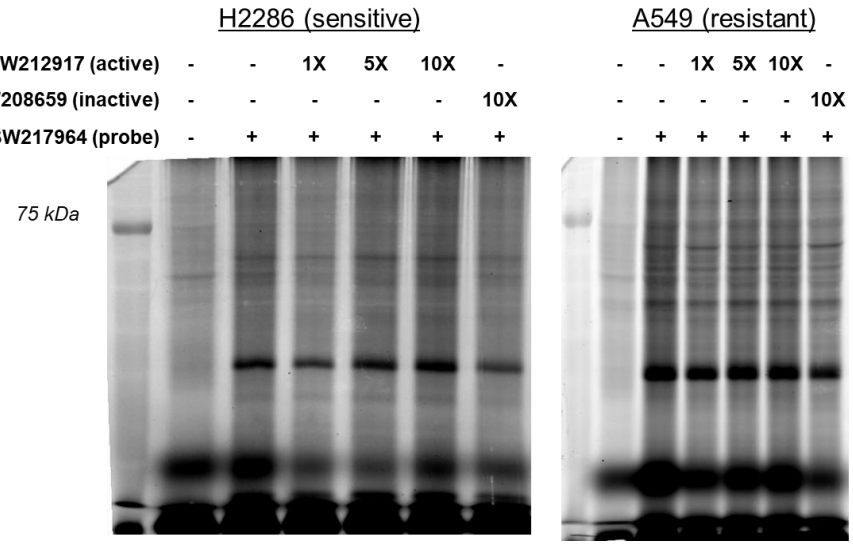


Figure 69. SDS-PAGE used to view proteome labeling

The lack of competitive validation in the proteomic fluorescent labeling suggested that continuing the pipeline of target discovery would be fruitless with the current probes. As biotinylation and streptavidin enrichment required the same protocol as the failed fluorescent proteomic labeling, it was assumed that the non-specific interactions would continue to be problematic. Moreso, it was assumed that these issues would be expanded in mass spectrometry where specificity due to sample contamination is a common problem.

Concurrently, confocal fluorescence microscopy was used to visualize bifunctional probes in the fixed lung cancer cell line, HCC366. Fluorescent staining was seen in probe treated cells compared to vehicle control and cells incubated with 10X non-alkynated probe (Figure 70). The specificity demonstrated by microscopy was in stark contrast to the lack of competition in fluorescent gel electrophoresis. Furthermore, staining of HCC366 was localized to a single

perinuclear location of the cell and occurred within an hour. Three dimensional Z-stacks of the cells indicate that this localization was indeed on the perimeter of the counterstained nucleus.

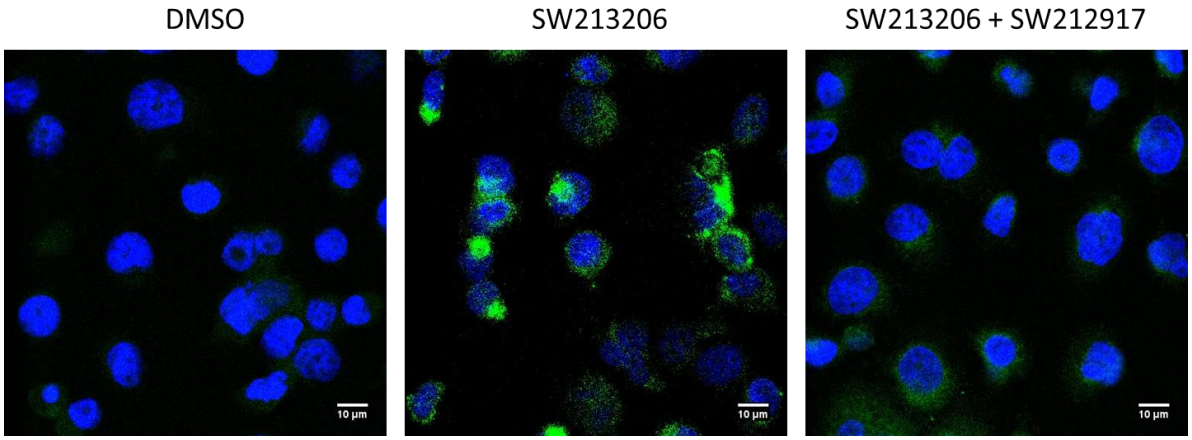


Figure 70. Fluorescence confocal microscopy of probe treated cells

These findings indicate that the discoipyrroles were internalized quickly and specifically to a subcellular locale. The speed of this internalization matched results of unrelated experiments measuring DP cell permeability. At varying time points, media was removed from DP treated cells and the wells were washed thoroughly before fresh media was replaced (Figure 71). At 72 hours, cell growth was measured. Cells treated with DP for 4 hours were inhibited to the same extent as 24 hour treatment. This indicated a rapid internalization and a long-lasting effect of DP binding.

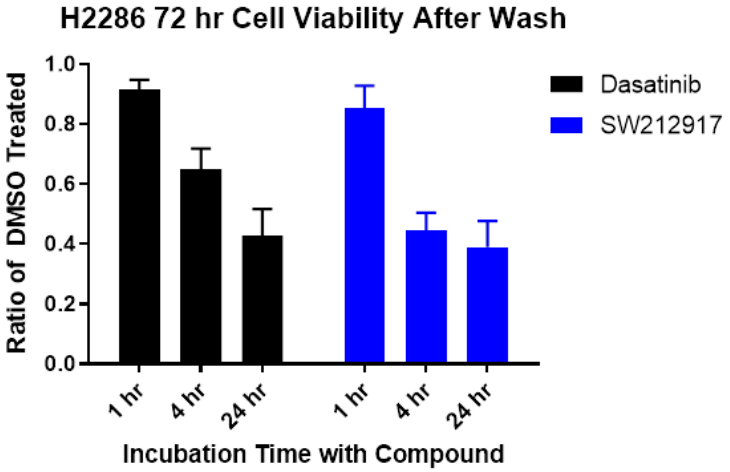


Figure 71. Wash assays of SW212917

The lack of staining in fluorophore treated cells (vehicle) and cells co-incubated with non-alkynylated DP analogs indicate that this localization is specific to the DP scaffold and not the fluorophore (Figure 70). One explanation for this single perinuclear clustering is localization of the DPs to the Golgi apparatus. Typically, the Golgi is visualized as a cluster on one pole of the nucleus (Dejgaard, Dejgaard, & Presley, 2001). Future co-staining experiments are imperative to suggest that the DPs aggregate in the Golgi. However, the wash assays and microscopy suggest that the DPs are not bound for long periods of time on the periphery of the cell as one would expect for receptor binding. Additionally, nuclear staining is not seen with the DPs indicating that their interaction with cells occurs in the cytoplasm or within cytoplasmic organelles.

4.5.4 Discussion of proteome labeling limitations

There are many possible explanations for the inability to identify the molecular target of the discoipyrroles via proteomic labeling approaches. The most obvious of which is the possibility that the discoipyrroles do not bind a protein. While protein binding molecules

dominate the drug discovery world, there are many cases where non-proteogenic molecules are targeted. Particularly of significance are the molecules that intercalate DNA (Rehman, Sarwar, Husain, Ishqi, & Tabish, 2015) and bind lipids (Troeira Henriques & Craik, 2017).

If the discoipyrroles do bind proteins, the target could be transiently expressed or present in low titers due to frequent turnover. Also, binding of discoipyrroles to the protein could elicit degradation. To overcome these obstacles synchronized cell populations, performing affinity labeling in larger batches of cells, or proteasome inhibition could be useful. As fixed cell staining demonstrated specific visualization, the lysis methodology could be troublesome. If the target protein is subject to aggregation or sequestering within insoluble lipids, it might evade solubilization. Relatedly, an insoluble receptor would be a problematic target for these methods.

Finally, the synthesized DP probes could present issues. There are reports detailing non-specific binding of UV cross linkers and the linkers that connect them to the basic bioactive structure (Sakurai, Ozawa, Yamada, Yasui, & Mizuno, 2014; Sumranjit & Chung, 2013). Also, more flexible linkers might be needed to orient the UV cross linker accessible to reactive amino acids. Finally, the DPs could be situated in such a way where UV cross linkers can never be accessed by the bound protein. In the field of target identification, the more diverse the probe set, the higher chance of successful affinity-based identification. Clearly, the current toolbox of DP probes is insufficient to garner successful identification

4.5.5 Gene regulation in response to discoipyrrole treatment

The inability to identify a molecular target via probe based methodologies led to different experimental approaches. One of the alternate approaches was exploration of the transcriptome in the NSCLC cell lines. Microarrays were used to identify genes whose expression was altered

under different discoipyrrole treatment conditions. Cells, both A549 and HCC366, were grown to confluence in 150 cm plates then treated with either DMSO, inactive DP SW208659 at two concentrations, or active DP at the same two concentrations. After 12 hrs, the cells were washed, trypsin digested, and snap-frozen before their RNA was extracted and their transcriptome was analyzed using Illumina HumanHT-12 v4 Expression BeadChip array.

Log_2 fold changes in DP treated cells from vehicle treated cells were calculated resulting in 2,816 significantly altered genes for both cell lines. The cell lines were then separated into two lists and filtered separately. The first filtration was carried out by removing genes that were altered in the presence of the inactive DP analog, SW208659. The resulting list was then filtered down to 236 genes whose expression was upregulated or downregulated in both the 100 nM treatment and 1 uM treatment conditions. Thus, the final list represented genes whose expression was not altered after 24 hour treatment with either vehicle or an inactive DP analog, but was altered after treatment with both concentrations of active DP analog. These lists were then sorted by the relative upregulation or downregulation of these genes and the top hits were input into the pathway analysis program, Reactome, to generate lists of statistically significant enriched pathways (Table 6Table 6. Pathway analysis of DP treated cells) (Croft et al., 2014; Fabregat et al., 2016).

Table 6. Pathway analysis of DP treated cells

A549 Upregulated	p-value	A549 Downregulated	p-value
Alpha-defensins	0.001210547	Nuclear Receptor transcription pathway	0.001798814
Regulation of IFNA signaling	0.001986463	Crosslinking of collagen fibrils	0.015319597
GABA A (rho) receptor activation	0.004047518	Leukotriene receptors	0.034378382
TRAF6 mediated IRF7 activation	0.004386355	Scavenging by Class H Receptors	0.084399485
RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways	0.024189604	Defective SLC34A3 causes Hereditary hypophosphatemic rickets with hypercalciuria (HHRH)	0.091726991
Regulation of gene expression by Hypoxia-inducible Factor	0.028455308	Defective AVP causes neurohypophyseal diabetes insipidus (NDI)	0.120397931
Nef mediated downregulation of CD28 cell surface expression	0.05426421	Neurophilin interactions with VEGF and VEGFR	0.120397931
Defensins	0.067465737	Assembly of collagen fibrils and other multimeric structures	0.128959473
Phase 2 - plateau phase	0.108051458	Synthesis of IP3 and IP4 in the cytosol	0.131390509
Phase 1 - inactivation of fast Na+ channels	0.108051458	Mineralocorticoid biosynthesis	0.135586475

HCC366 Upregulated	p-value	HCC366 Downregulated	p-value
Alpha-defensins	0.015860711	Regulation of IFNA signaling	0.001752061
Defective ABCC8 can cause hypoglycemias and hyperglycemias	0.050504613	TRAF6 mediated IRF7 activation	0.004512192
Defective SLC5A5 causes thyroid dyshormonogenesis 1 (TDH1)	0.050504613	Defective SLC6A2 causes orthostatic intolerance (OI)	0.086532141
Chemokine receptors bind chemokines	0.077045822	Defective SLC6A5 causes hyperekplexia 3 (HKPX3)	0.113683444
Thyroxine biosynthesis	0.080005975	Defective ABCB4 causes progressive familial intrahepatic cholestasis 3, intrahepatic cholestasis of pregnancy 3 and gallbladder disease 1	0.113683444
ATP sensitive Potassium channels	0.098470097	Phase 4 - resting membrane potential	0.12288268
Activation of Matrix Metalloproteinases	0.123210712	RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways	0.125334333
The fatty acid cycling model	0.129099477	Classical Kir channels	0.140029865
GLI proteins bind promoters of Hh responsive genes to promote transcription	0.129099477	Terminal pathway of complement	0.214474317
Ligand-receptor interactions	0.129099477	Phase 3 - rapid repolarisation	0.237832263

Both sensitive and resistant cell lines were altered in pathways associated with immunological responses. The most significantly upregulated pathway in both cell lines were regulation of alpha-defensins which are secreted protective peptides. The exact mechanism of the defensins is not known, but they are thought to function via plasma membrane insertion or polymeric entrapment of microbial pathogens (L. Zhao & Lu, 2014). Additionally, the two most downregulated pathways in HCC366 cells were regulatory pathways of interferon signaling. Interferon receptor signaling has been linked to DDR function and collagen synthesis historically (Granstein, Flotte, & Amento, 1990; Kamohara, Yamashiro, Galligan, & Yoshimura, 2001). Lately, DDR2 knockdown has been shown to negate immunological activation by collagen in human derived monocytes (Poudel, Yoon, Lee, Lee, & Kim, 2012).

These pathways were analyzed in addition to those of manually selected genes of interest from the list of significantly altered genes. However, this information was of little utility without

a handle on the molecular target of the DPs. Many conclusions could be extrapolated from these genes of interest, but validation of over 50 genes via knockdown screening is required to make any significant claims. Future efforts to unveil a molecular target of the DPs will benefit from these collections of genes. Specifically, they could be useful in cross-referencing affinity immunoprecipitations for validation.

4.5.6 GPCR screening of the discoipyrroles

The microarray revealed a number of responses in GPCR proteins to DP treatment (69 genes involved in GPCR signaling altered significantly in active DP treated cells: GPR68, DGKA, RGSL1, RASGRF2, CHRM4, OR5K2, OR4K17, OR52N4, CX3CL1, FGF4, FGF7,CYSLTR2, OR1S1, GRM4, GRM6, CCR8, OR52B2, OR8D2, OR5B3, OR8D1, CCR4, PDGFRB, OR51G2, OR1B1, PRKCB1, NPY1R, TAAR5, OR4P4,GAB2,OR2T8, MMP10, OR52R1, ADRB3, RASA3, LTB4R2, OR4X1,IAPP, OR52E5,GNAZ, RGS16,ADCY4, OR5M10, GALR3, OR8B8, HRH3, OR2W1, NPBWR1, OR5AN1, FYN,S1PR2, PRLH, RGS22, OR7D2, DRD4, OR10K1, P2RY13, OR10G2, WNT3A, GCGR, OR1L6, NMBR, OR10X1, SST, GHRH). The availability of protein binding panels by commercial companies such as DiscoverX made validation studies more obtainable than large scale knockdown screens. The binding of the discoipyrroles to large seven transmembrane receptors would also explain the inability to easily enrich proteins within our affinity labeling protocol. Therefore, we utilized the DiscoverX GPCR panel gpcrMAX to assay the ability of the discoipyrroles to inhibit or activate a variety of GPCRs.

The gpcrMAX assay utilizes recombinant GPCR proteins tagged with a β-galactosidase moiety to easily report activation and inhibition of 168 known GPCRs. The active DP analog,

SW212917, was screened at a concentration of 1 μM against this panel. No major activity (>25% of maximum activity with known ligand/inhibitor) was demonstrated by SW212917 as either an agonist or antagonist (Table 7). However, SW212917 demonstrated modest agonist activity against CXCR4. The GPCR CXCR4 is a chemokine receptor that binds the ligands CXCL12 and MIF to coordinate stem cell homing and cell migration, respectively (Pawig, Klasen, Weber, Bernhagen, & Noels, 2015). This finding is of interest due to the abundance of immunological genes altered in the microarray and CXCR4's unique role as an agonist for migration inhibition signaling. Potentially, the DPs migration inhibition, as evident by the Transwell assays, is tied to its agonism of the CXCR4 or related receptor. There exists many orphan GPCRs not included in this panel. The DPs might be binding a related receptor that controls migration in the lung cancer cell lines tested. This could explain the low agonist activity against CXCR4.

Table 7. GPCR panel results

GPCR	%activity	Activity Type
CXCR4	18	Agonist
ADORA3	13	Antagonist
CALCRL-RAMP1	22	Antagonist
HTR5A	17	Antagonist
MRGPRX1	18	Antagonist
PTGIR	24	Antagonist
RXFP3	16	Antagonist

The low antagonist activity against multiple receptors including CALCRL-RAMP1 and PTGIR is less clear. The former is a complex of two proteins that together form the CGRP receptor. Peptides known as CGRP modulate vasodilation and pain by binding to CALCRL-RAMP1 (Russo, 2015). The signaling pathway of CGRP has been tied to the MAPK pathway as

well as cytokine signaling (Schafers, Svensson, Sommer, & Sorkin, 2003). PTGIR is in the prostaglandin family of receptors and has been shown to bind prostacyclin as its ligand (Midgett, Stitham, Martin, & Hwa, 2011). This receptor-ligand pair has been linked to different functions including vasodilation, inflammation, and immunology (Hata & Breyer, 2004). While these two receptors maintain ties to immunology, a consistent narrative with the functions of the DDRs or the DPs remains less clear.

4.6 Conclusion

Phenotypic screening for bioactive natural products is a powerful approach. The mechanism independent nature of this method allows for diverse targets to be identified and more importantly, for new targets to be discovered. However, pinpointing said targets is an arduous task due to the complicated nature of mammalian cell signaling pathways that are ultimately responsible for the desired phenotype.

In the case of the discoipyrroles, the FuSiOn phenotypic screen suggested that these compounds inhibited the action of the DDR2 receptor in specific lung cancer cell lines to induce cytotoxicity. Preliminary data demonstrated that DDR2 activation was not directly affected by these compounds. However, similarities in the cell specificity to the DDR2 inhibitor dasatinib alongside key DDR2-related phenotypes suggested that the DPs might affect DDR2-signaling.

To interrogate this we took several unbiased approaches to interrogate their action in cells. While valuable information was learned about several phenotypic effects of the DPs, none of these experiments were able to pinpoint their exact molecular target and mechanism of growth inhibition. Probe based affinity approaches to target identification have been an important practice in the field of drug discovery. However, alternative methodologies have been developed

to overcome the short comings of probe based approaches. In the case of the discoipyrroles, these alternative practices may be better suited.

Particularly, a more robust examination of the signaling pathways within the cell will be instrumental. Phosphoproteomic approaches such as SILAC may be able to identify phosphorylation patterns illustrating how the DPs affect migration and growth in cells. These approaches would mitigate the problems encountered in affinity approaches that rely on capturing protein binding events. By examining the cellular environment rather than the compound themselves, the outcome of DP treatment may be more apparent. While the DDR2 signaling pathway has recently been a focus in oncology and other diseases, the novelty of the field inhibits easy validation of potential inhibitors. A SILAC approach would also preclude this type of information deficit from inhibiting mechanism of action studies, as individual protein phosphorylation would be assayed.

In conclusion, the discoipyrroles are a biologically intriguing family of natural products. Their activity in lung cancer cell lines is highlighted by a stark selectivity in growth inhibition, yet global effect on extracellular interactions. Understanding how these compounds affect migration and cell morphology could permit tool development for use in development, differentiation, and mobilization. Clearly, a comprehensive biological approach will be key to discovering more about these potential roles of the discoipyrroles.

4.7 Experimental

Cell culture and cell lines. The cell lines A549, HCC366, and H2286 were graciously donated by the laboratory of John Minna and Bruce Posner. Routine mutation analysis and mycoplasma tests were performed by these labs. Cells under passage number 15 were used. Cells were

cultured in 150 mm tissue culture treated dishes (Corning). Hyclone™ RPMI 1640 (GE) with 5% FBS and Penicillin/Streptomycin (Thermo-Fisher) was used under normal growth conditions in 5% CO_2 37°C incubator. Starve media was constructed with the same components albeit with 0.2% FBS. For all dosing regiments cells were plated, using cell specific seeding concentrations, and allowed to adhere for 16 hrs overnight before media was replaced and compounds were added. All compounds were solubilized in DMSO and the final concentration of DMSO in cultures was 0.1%. Growth inhibition assays. TGF-β (R&D Systems) stock was solubilized in PBS and was dosed from a 1000X stock to a final concentration of 10 ng/mL. Collagen I (Thermo Fisher Scientific) was dosed at a final concentration of 50 μg/mL.

Cell viability assays. Cells were trypsinized and moved into 1.5 mL microcentrifuge tubes. Cell suspensions were spun at 250 xg for 3 minutes. Supernatant was removed and cells were resuspended in 50 uL media by gently flicking the tube. 50 uL of trypan blue (Fisher Scientific) was added and the cells were allowed to sit for 3-5 minutes. Stained cells were loaded onto Bio-Rad TC10 Cell counter and viable cells were enumerated. For all growth inhibition assays, biological triplicates were measured. For the wash assays, at indicated times the cells were washed with 37°C media 3X and their media was replaced.

DNA release assays. Cells were seeded into 96 well plates and allowed to adhere overnight. Compounds were dosed in 0.1% DMSO in triplicate and CellTox Green assay (Promega) was performed by manufacturer specifications.

Flow cytometry. 300,000 cells were plated in a T75. After 24h, cells were treated with 500nM compounds (SW212350), dasatinib or DMSO for 24h. Cells were harvested by trypsinization, combined with their respective media, and 400,000 cells were counted (Beckman-Coulter Z2

particle counter set at >12 micron) and pelleted. Pelleted cells were washed once with ice-cold PBS, resuspended in 1mL of ice-cold PBS, 2.5mL of 100% ethanol was added while gently vortexing, then this solution was stored at -20C for at least overnight. Cells were then pelleted, washed once with ice-cold PBS, and then stained for 1h in a 37C water bath using a solution of .1% Triton X-100 (Sigma), 50ug/mL propidium iodide (ThermoFisher), 2% fetal bovine serum (FBS, Gibco) and 100ug/mL RNase A (Sigma). After 1hr, 5mL of PBS was added and cells were pelleted. Cells were resuspended in PBS+2% FBS, passed through a 70micron strainer, and ran through a BD LSRFortessa flow cytometer. Cells were gated for single cells using PI-A x PI-W, gated for debris using SSC-A x FSC-A and 10,000 cells passing these filters were analyzed. Histograms were generated using FlowJo version 7 and the percent of cells in each phase of the cell cycle was determined by mathematical modeling, using the Dean-Jett and Watson pragmatic algorithm.

Confocal microscopy. Cells were plated on 4 welled Lab-Tek I Chambered Slides and allowed to adhere overnight to 75% confluency. After media replacement, cells were dosed and allowed to incubate for allotted time. Media was removed and each well was washed with PBS 3X. 500 uL of10% paraformaldehyde solution (Fisher Scientific) was added to each well for 20 minutes to allow for fixation. Fixing solution was aspirated and washed 3X with room temperature PBS. 500 uL of 0.2% Triton-X 100 in PBS was then pipetted into each well and allowed to sit for 20 minutes to allow for permeabilization. Permeabilization solution was removed and the wells were washed 3X with room temperature PBS, followed by 2% BSA in PBS. Fresh Click-iT® Cocktail (Click Chemistry Tools) was made according to manufacturer instructions. Alexa-Fluor™ 532 and Alexa-Fluor™ 488 azide (ThermoFisher Scientific) were the fluorophores

utilized in this cocktail. Cells were incubated for 30 minutes protected from light. Cells were washed with 2% BSA in PBS followed by a secondary wash with PBS. After aspiration of PBS, the cell chambers were removed and a drop of Fluoromount-G DAPI (SouthernBiotech) was added to the middle of the chamber. Slides were carefully pressed onto the mounting solution and the chamber was allowed to dry for 30 minutes in the dark. Slides were then visualized using a Leica TCS SP5 confocal microscope. Z-stacks were composed of images taken every 0.33 μm for three-dimensional images. Images were imported with the Bio-Formats extension and ImageJ.

Transwell migration assays. Transwell® inserts (6.5 mm 8.0 μm pore size PET membrane) (Corning) were placed into 12-well plates and wet with 100 uL RPMI for 1 hour in 5% CO_2 37°C incubator. During this time, cells of confluence were trypsinized and counted. After aspirating RPMI from the inserts, cells were seeded at 25k/100 μL/well in serum-free RPMI. 600 μL of RPMI + 10% FBS media was placed in lower chamber with pipette. The lower media was supplemented with growth factors or compounds if indicated. Cells were incubated for variable times from 4 hrs to 48 hrs. Representative images were from 12 hour incubations. At indicated time, Transwells were removed from wells, placed into a new 12-well plate for preparation. Transwells were washed 2X with warmed RPMI and a cotton swab was used to remove cells from the top of the membrane. Cells were then fixed with 70% ethanol in ddH_2O for 10 minutes. A cotton swab was used to remove fixing solution and again scrape non-adherent cells from the top of the membrane. Cells were then stained with 0.2% crystal violet in ddH_2O for 10 minutes. Transwells were then removed with tweezers and rinsed in beakers of ddH_2O thoroughly until all excess staining solution was gone. Cotton swabs were used to remove excess liquid. The

membranes of the Transwells were then removed by cutting the edge of each with a scalpel. The membranes were placed on a cleaned glass slide and a drop of mounting solution was added to the top. Cover slides were pressed gently on the top of the membrane and the slides were allowed to dry in the dark for 15 minutes. A Zeiss AxioObserver epifluorescence microscope was then used to take color images of the membranes. ImageJ was used to process and crop images.

Western blotting. Cells were plated and allowed to adhere overnight. Cells were grown in 12-well dishes until 60-75% confluent. At this time, media was replaced and cells were dosed with indicated amounts of compound or growth factor. After incubation time, media was aspirated and cells were washed 2X with cold PBS. Then lysis buffer (RIPA + ThermoScientific HALT Protease Inhibitor Cocktail, Roche PhosStop Phosphatase Inhibitor, and Sigma Benzonase nuclease) was added to the wells and the plate was rotated at 4°C for 15 minutes. Cell lysates were then moved into 1.5 mL microcentrifuge tubes and spun at 14,000 x g for 15 minutes at 4°C. The supernatant was then moved to new 1.5 mL microcentrifuge tubes and BCA assay (Pierce) was used to measure protein concentrations. Individual lysates were normalized with RIPA buffer and a 4X sample loading buffer was added to each sample to bring the final concentration of SLB to 1X. Samples were loaded onto 4-15% Mini-PROEAN® TGX^{TM} gels (Bio-Rad) and ran for 15 minutes at 60 V, then 90 minutes at 90 V. Gels were removed, washed, and transferred to methanol activated PVDF membranes at 100V for 70 minutes in a refrigerated wet blotting system (Bio-Rad). Membranes were removed and blocked with filtered 5% skim-milk in TBS-T for 1 hr. Membranes were then stained with indicated antibodies overnight at 4°C. Membranes were then washed 5X with TBS-T and stained with secondary antibodies in 5% skim-milk in TBS-T for 1 hr. Membranes were then washed 5X with TBS-T and incubated with

for 5 minutes before being developed. Images were scanned and analyzed with ImageJ and Graphpad Prism.

Table 8. Antibodies List

	Source	*Antigen*	*Dilution*	*Company*
Primary	Rabbit	Vimentin	1:5,000	BD Pharmingen™
	Mouse	E-cadherin	1:10,000	BD Pharmingen™
	Mouse	Actin	1:5,000	Sigma-Aldrich
Secondary	Goat	Mouse	1:5,000	ThermoFisher
	Donkey	Rabbit	1:5,000	Sigma-Aldrich

Direct binding approach. Cells of indicated type were seeded in 150 mm tissue-culture treated dishes and allowed to grow until 75% confluency. Media was replaced and dosing regimen was applied as indicated. Cells were incubated for allotted time then washed with 37°C media several times. Ice cold PBS was added to cells and the plates were irradiated with indicated UV light for each specific UV cross linker (azide: 325 nm, benzophenone: UV-B light, diazirine: 365 nm). Each plate was irradiated on ice to prevent overheating. Cells were kept ~1-1.5 inches from bulb. After irradiation, cells were washed once with ice-cold PBS, then lysis buffer (RIPA + ThermoScientific HALT Protease Inhibitor Cocktail, Roche PhosStop Phosphatase Inhibitor, and Sigma Benzonase nuclease) was added and the plates were rocked at 4°C for 15 minutes. The cell lysates were scraped from the plates and moved into 1.5 mL microcentrifuge tubes. Lysates were spun at 14,000 x g for 15 minutes at 4°C. The supernatant was moved into new microcentrifuge tubes and BCA assay (Pierce) was used to quantify protein levels. Equivalent amounts of protein were used to conduct click chemistry reaction as follows. The total volume of the reaction was 200 µL. 60 µL of protein mixture diluted with PBS, 99 µL PBS, 4 µL 5 mM TBTA pre-mixed in 16 µL t-butanol, 1 µL 12.5 mM AlexaFluor® 532 azide, 10 µL 20 mM

$CuSO_4 \cdot 5\ H_2O$, and10 μL 200 mM sodium ascorbate. All reagents added *in situ* and the reducing agent added last. Reaction mixture was rotated in dark at room temperature for 30 minutes. Then methanol chloroform precipitation was carried out and the pellet was resuspended in 1X sample loading buffer and run on 4%/12% SDS-PAGE for 120 minutes at 90V. The gel was removed and the bottom was removed to eliminate excess probe signal. Gels were visualized on a Typhoon™ scanner (GE) using a channel optimized for the fluorophore AlexaFluor® 532. AlexaFluor® 532 was synthesized as indicated in (Theodoropoulos et al., 2016).

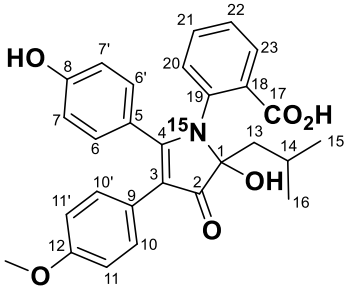
Microarray. Cells were seeded in 150 mm tissue culture treated dishes and allowed to adhere overnight. When cells reached 75% confluency, media was replaced and compounds were dosed at 0.1% DMSO at indicated concentrations. After 12 hrs incubation, media was replaced and cells were trypsinized with 0.025% trypsin solution. After quenching enzyme activity with growth media, cells were spun down at 250 x g and the supernatant was removed. The cell pellets were snap-frozen in liquid nitrogen and submitted to the UT Southwestern Medical Center core facility for Gene Expression and Genomic Sequencing. Data analysis was performed as previously described (Potts et al., 2015).

Appendix 1. Chapter 2

Table 9. NMR table for compound 42

No.	δ_H, mult (*J* in Hz)[a]	δ_C[b]	COSY[a]	HMBC[a]
1	-	92.3	-	13, 14
2	-	199.5	-	13
3	-	110.0	-	10, 10'
4	-	172.2	-	6, 6'
5	-	121.2	-	7, 7'
6, 6'	6.85, d (8.7)	131	7,7'	4, 6, 6', 8
7, 7'	6.45, d (8.7)	114.7	6,6'	5, 6, 6', 8
8	-	158.2	-	6, 6', 7, 7'
9	-	125.5	-	11, 11'
10, 10'	6.92, d (6.8)	129.9	11, 11'	3, 11, 11', 12
11, 11'	6.71, d (6.8)	113.0	10, 10'	9, 10, 10', 12
12	-	156.5	-	10, 10', 11, 11'
13	1.69, d (4.6)	44.5	14	1, 2, 14, 15, 16
14	1.92, m	23.2	13, 15, 16	1, 13, 15, 16
15	1.04, *d* (6.7)	24.4	14	13, 14, 16
16	0.79, *d* (6.7)	23.8	14	13, 14, 15
17	-	169.8	-	20, 23
18	-	135.9	-	20, 21, 22, 23
19	-	143.0	-	20, 22
20	7.24 d (7.7)	130.8	21	17, 22, 135.8, 141.9
21	7.34 t (7.5)	128.4	20, 22	23, 135.8
22	7.17 t (7.5)	126.9	21, 23	21, 23, 20, 135.8, 141.9
23	7.37 d (7.7)	130.3	22	17, 21, 135.8
24	3.67, s	54.8	-	12
OH	9.76, s	-	-	
OH	6.64, s	-	-	
OH	10.71, s	-	-	

Spectra collected in DMSO-d_6 on a) 600 MHz spectrometer or b) 400 MHz spectrometer



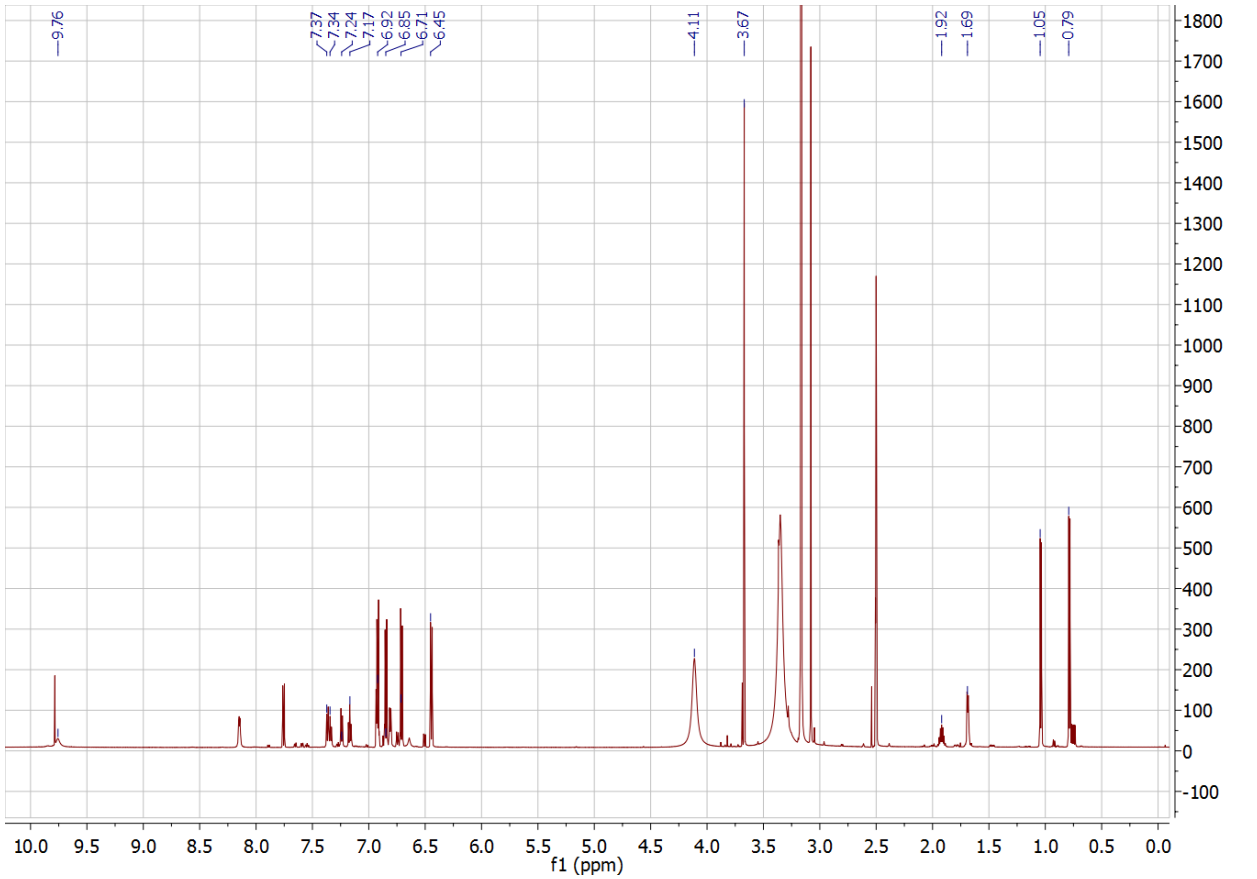


Figure 72. The ^{1}H-NMR spectrum of compound 42

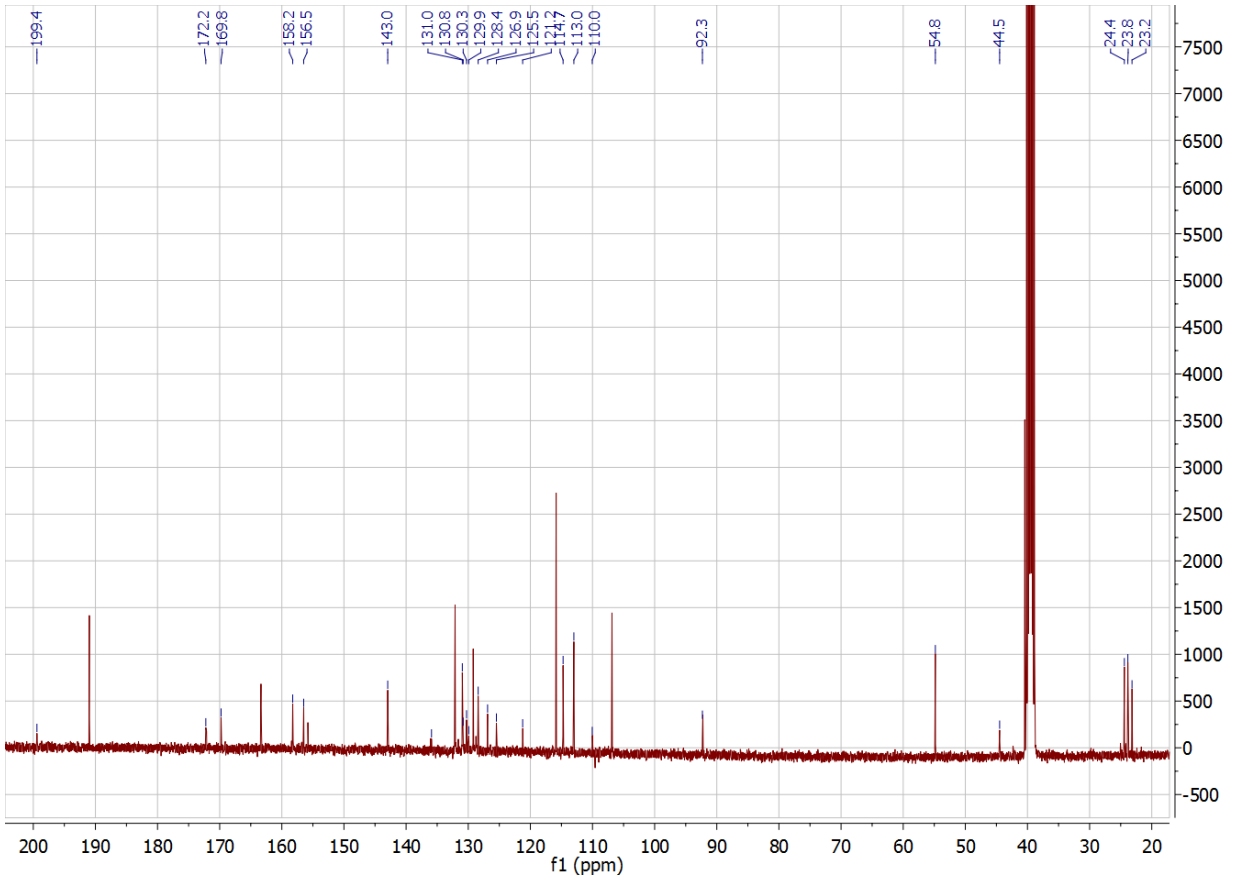


Figure 73. The 13C-NMR spectrum of compound 42

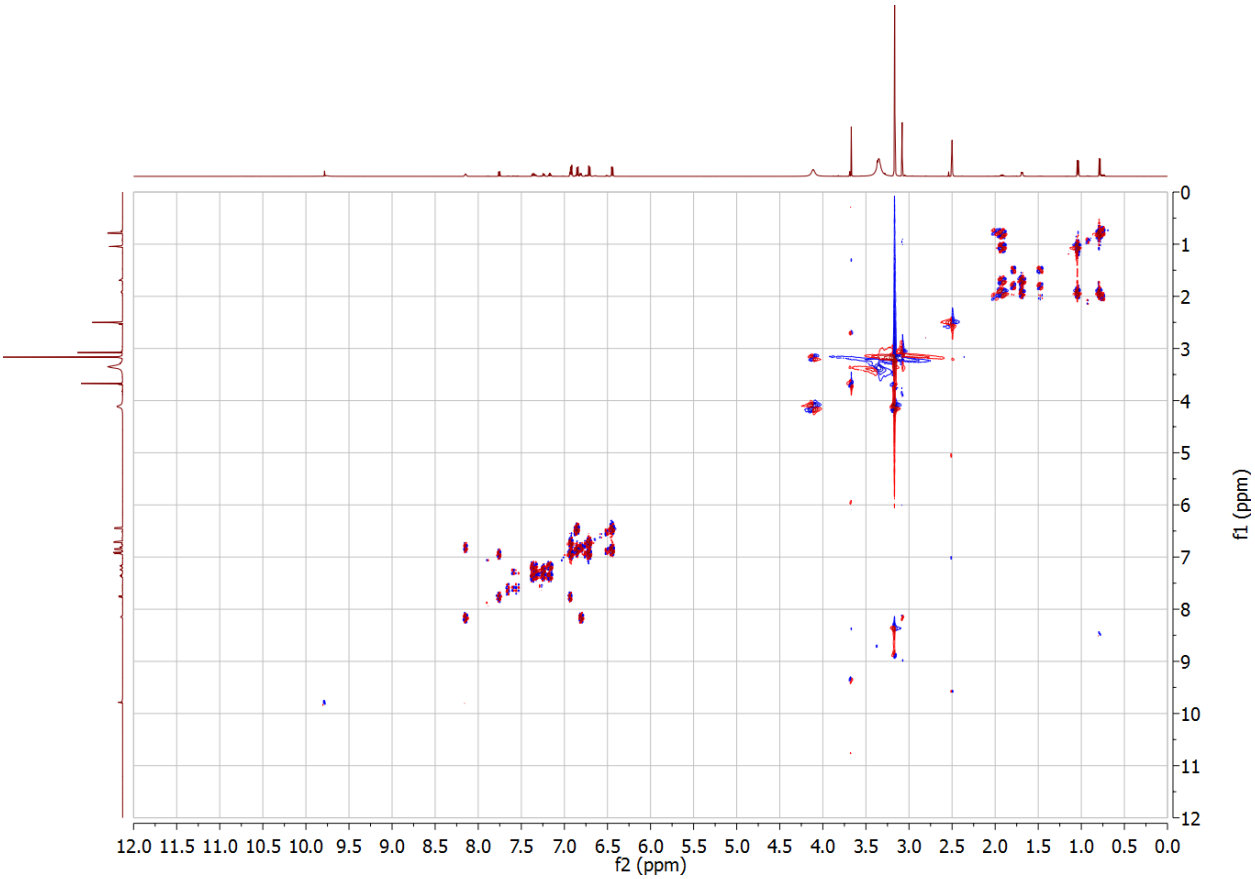


Figure 74. The COSY-NMR spectrum of compound 42

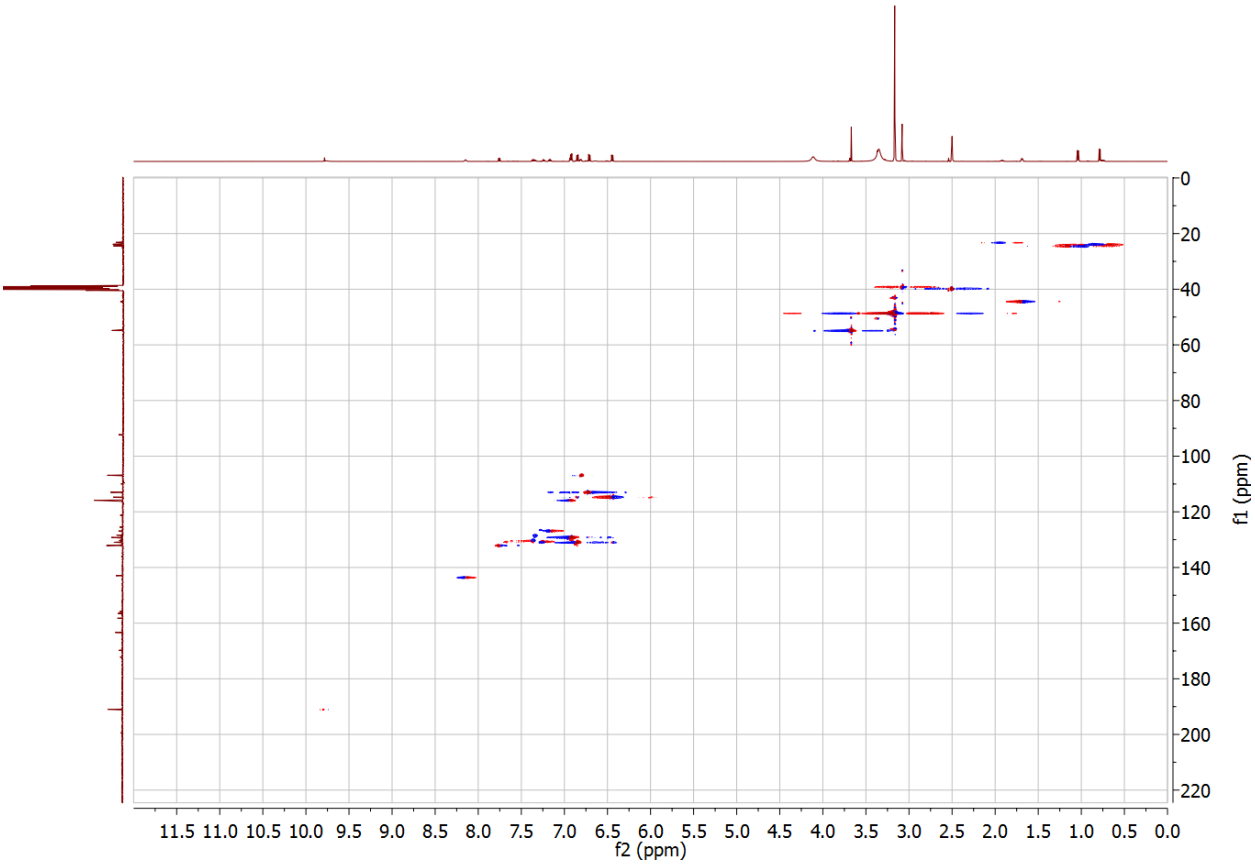


Figure 75. The HSQC-NMR spectrum of compound 42

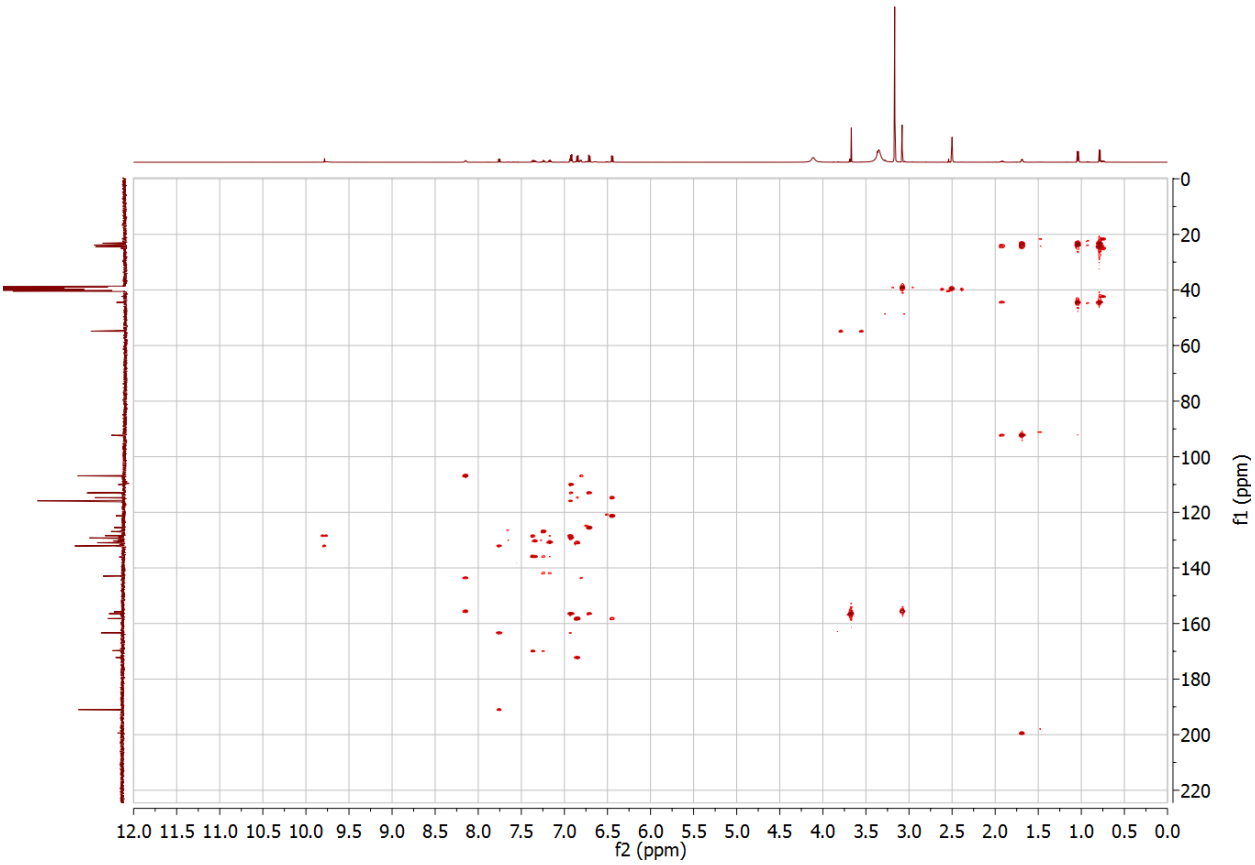


Figure 76. The HMBC-NMR spectrum of compound 42

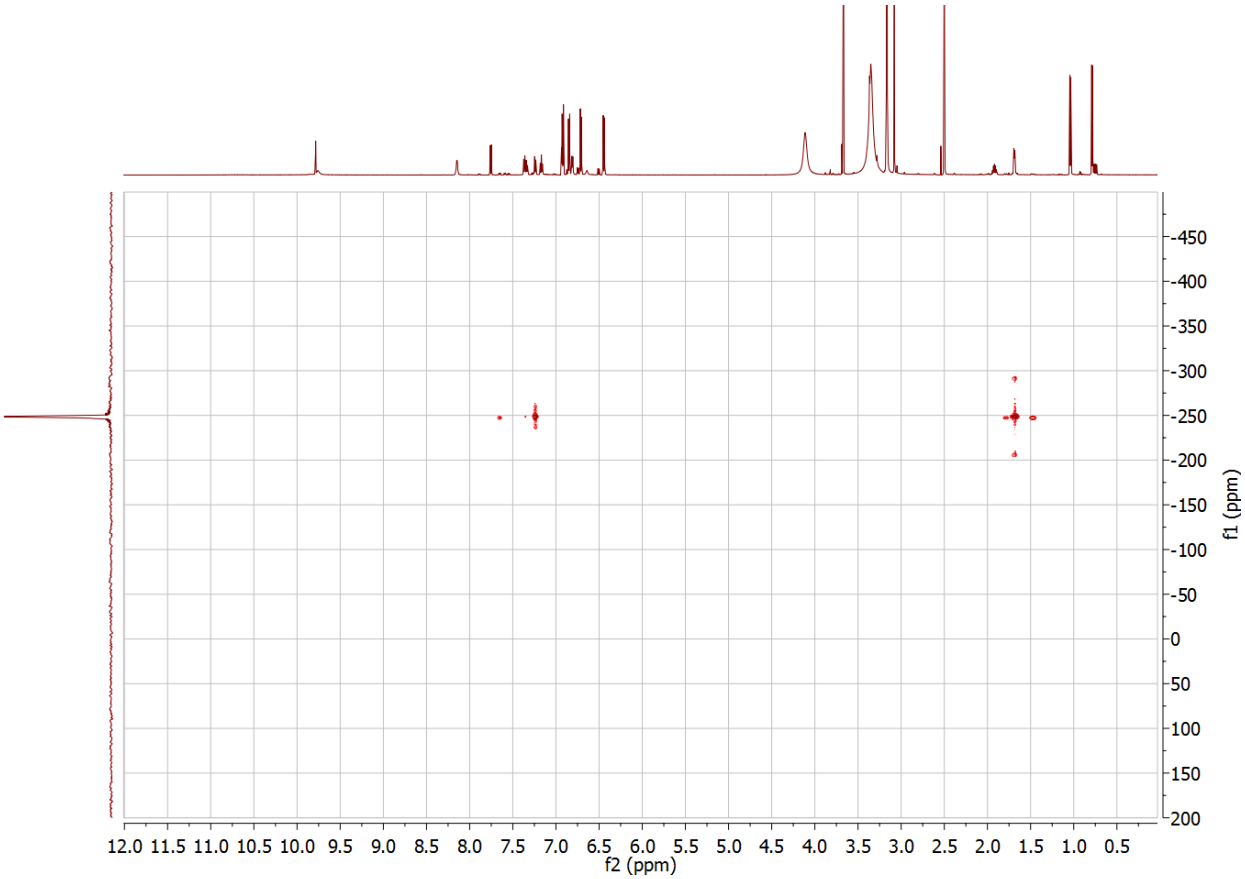


Figure 77. The ^{1}H-^{15}N HMBC NMR spectrum of compound 42

Appendix 2. Chapter 3

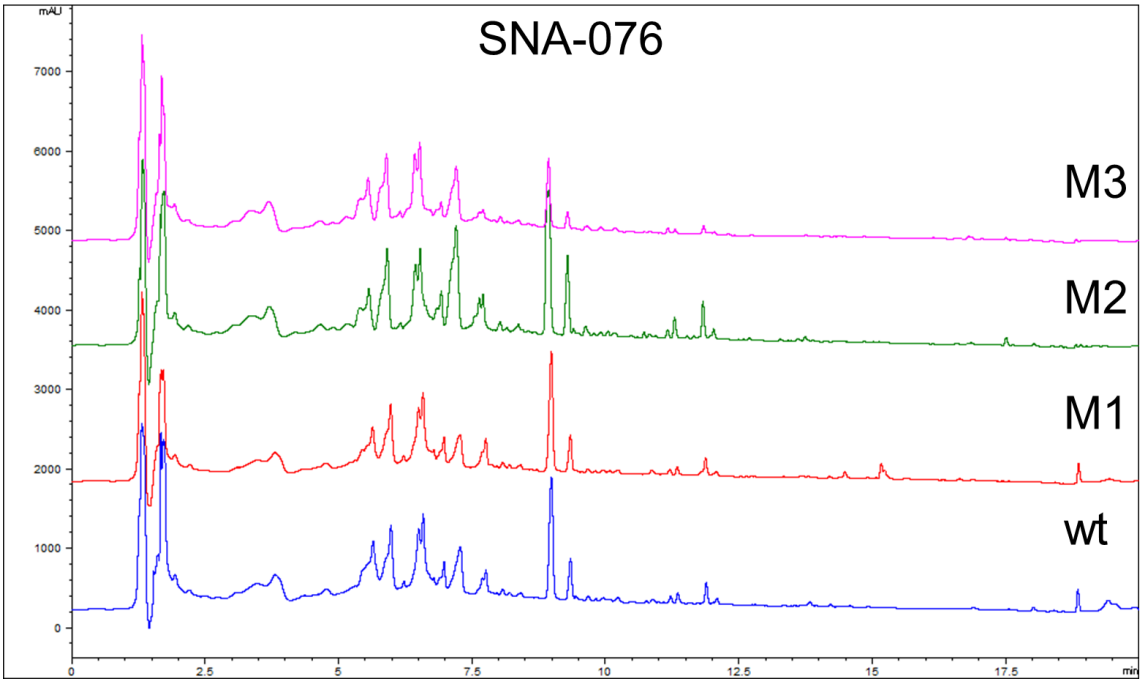


Figure 78. SNA-076 crude extracts (210 nm)

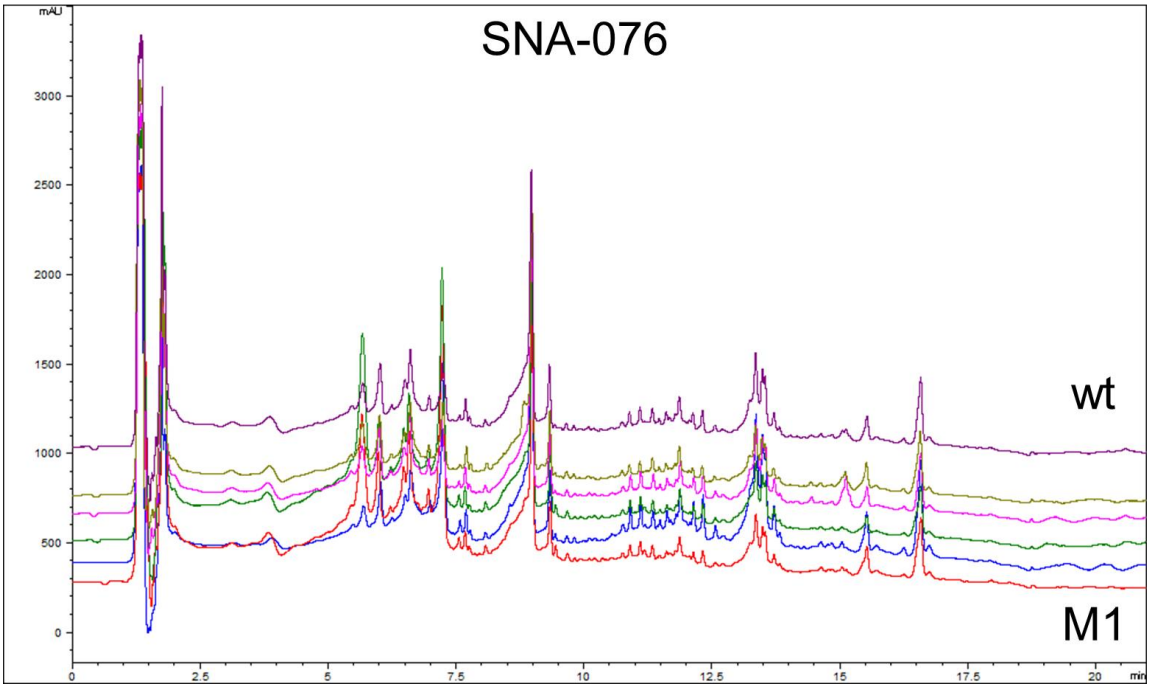


Figure 79. SNA-076 crude extracts (280 nm)

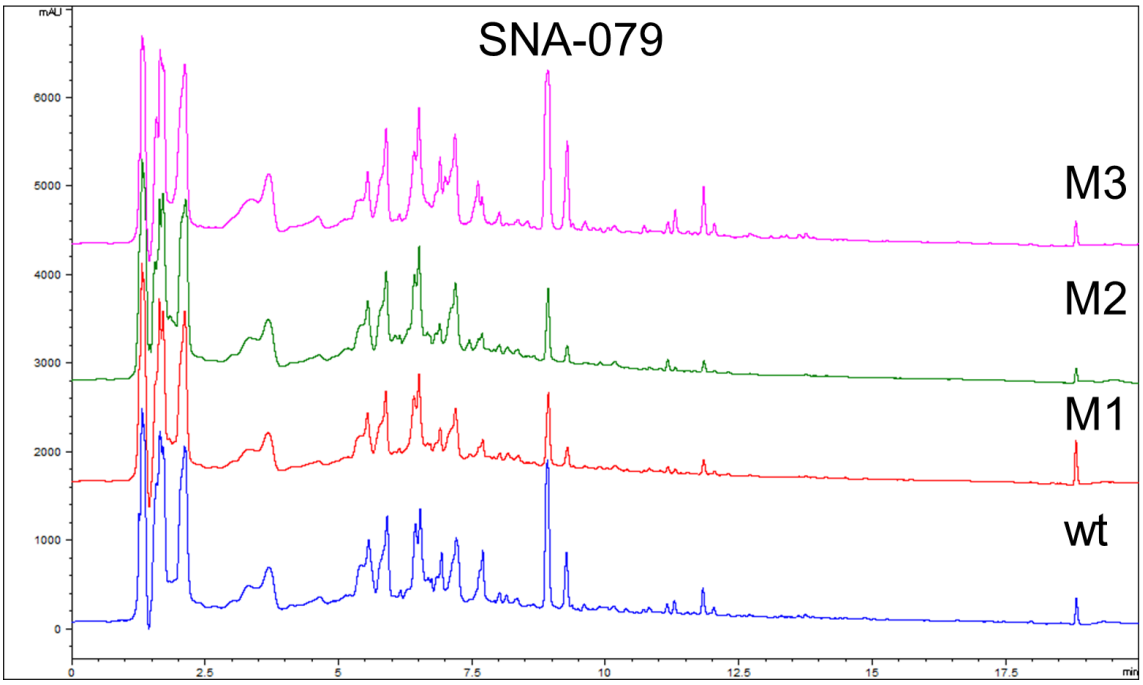


Figure 80. SNA-079 crude extracts (210 nm)

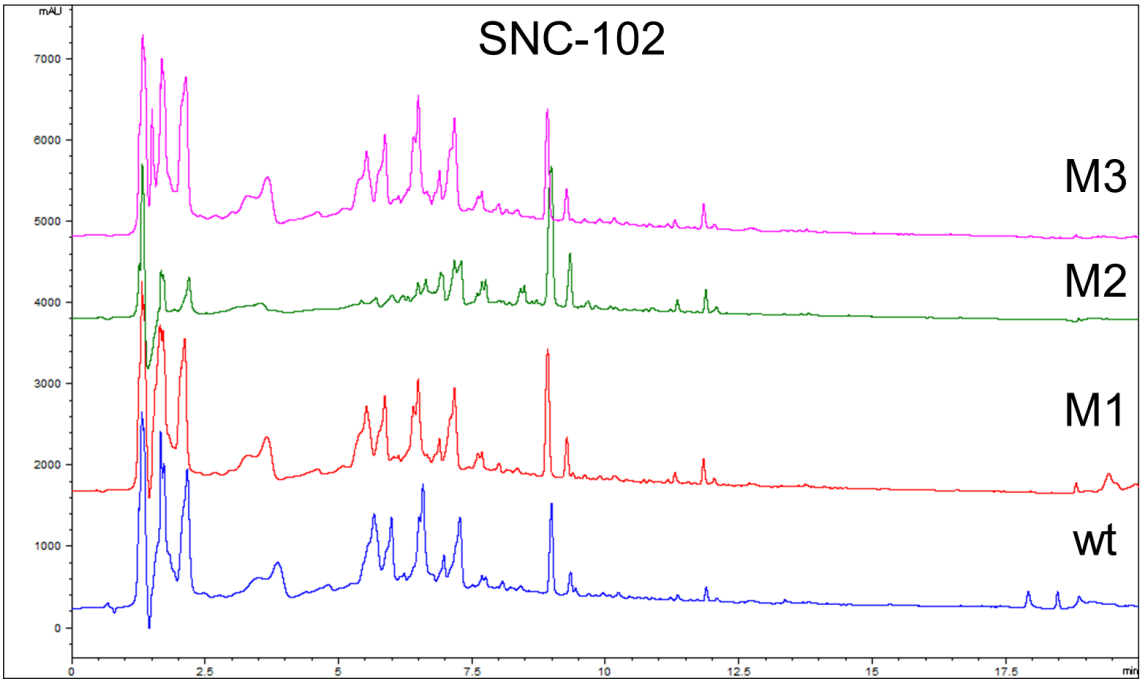


Figure 81. SNC-102 crude extracts (210 nm)

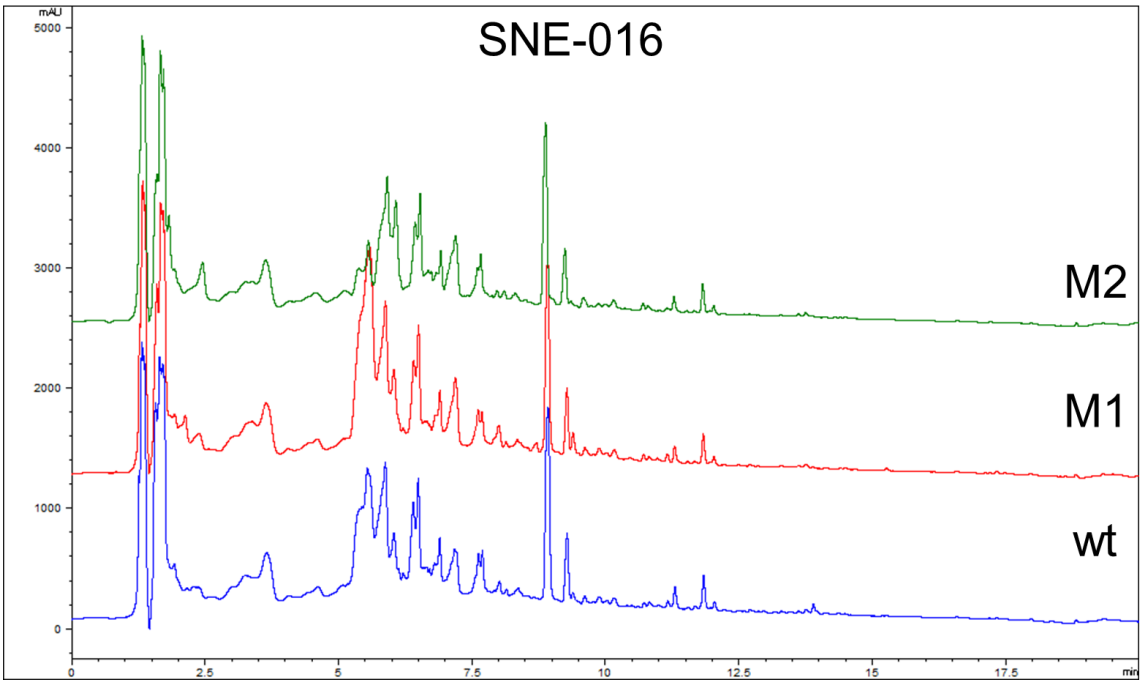


Figure 82. SNE-016 crude extracts (210 nm)

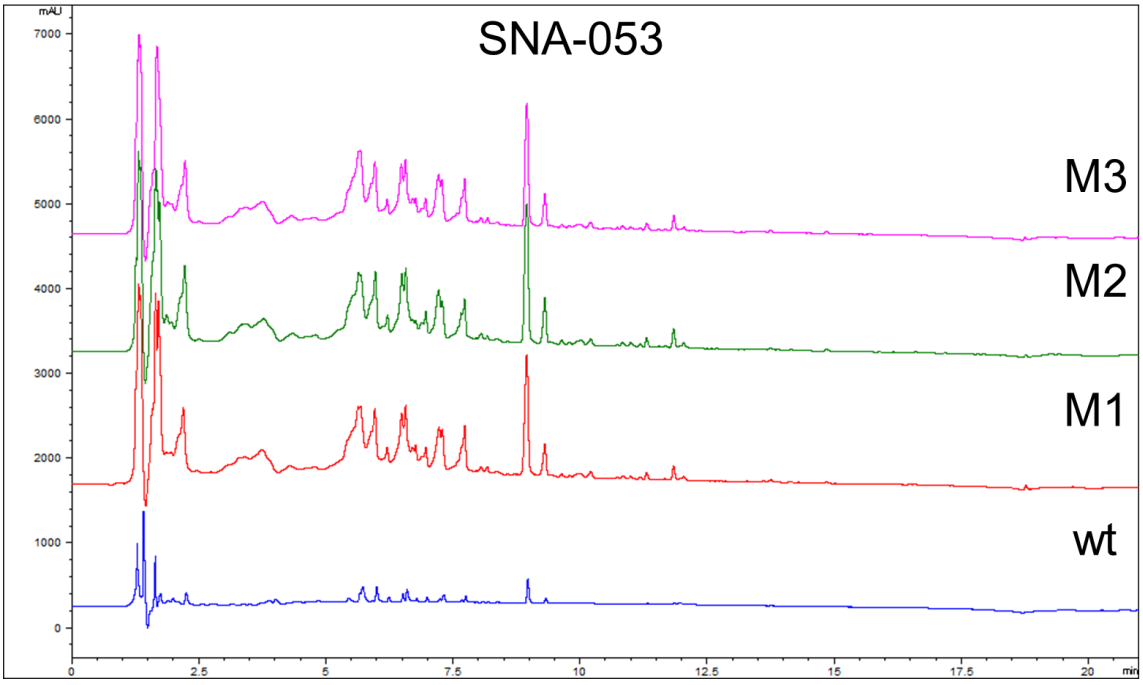


Figure 83. SNA-053 crude extracts (210 nm)

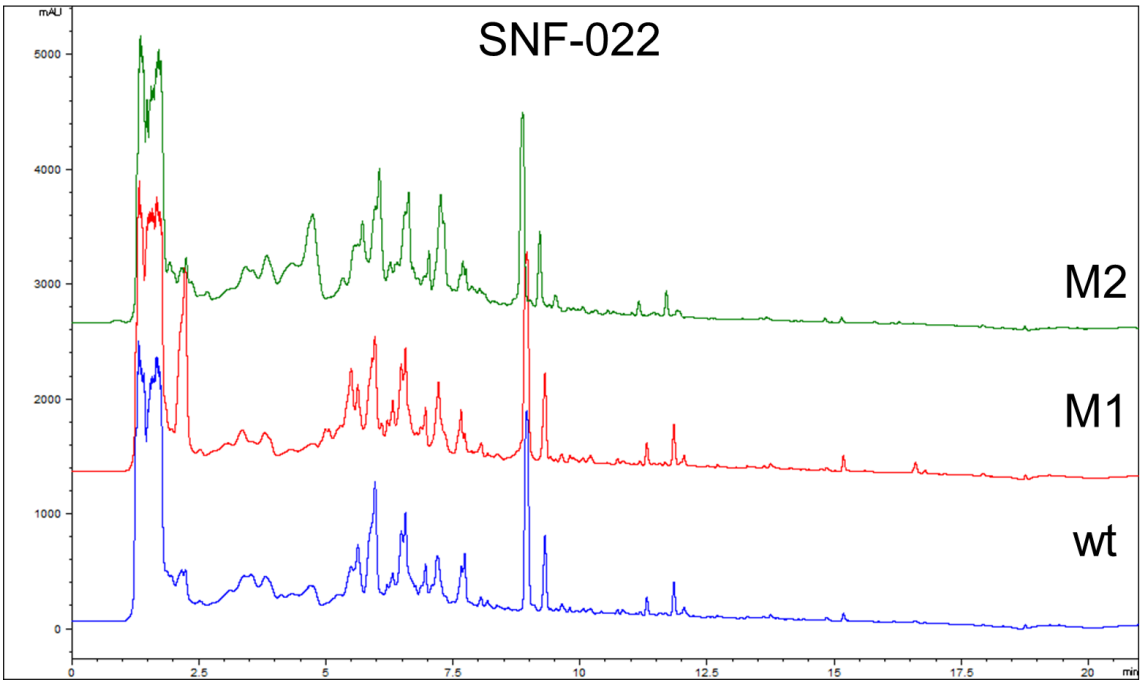


Figure 84. SNF-022 crude extracts (210 nm)

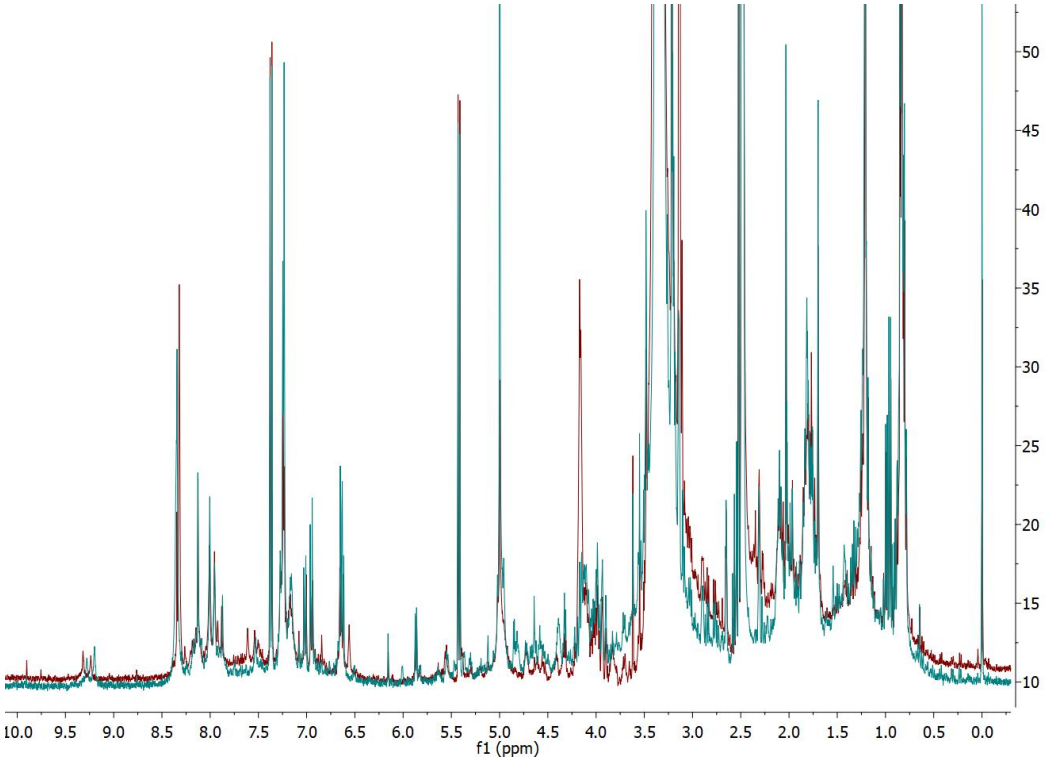


Figure 85. SNA-079 WT (red) and M3 (blue)

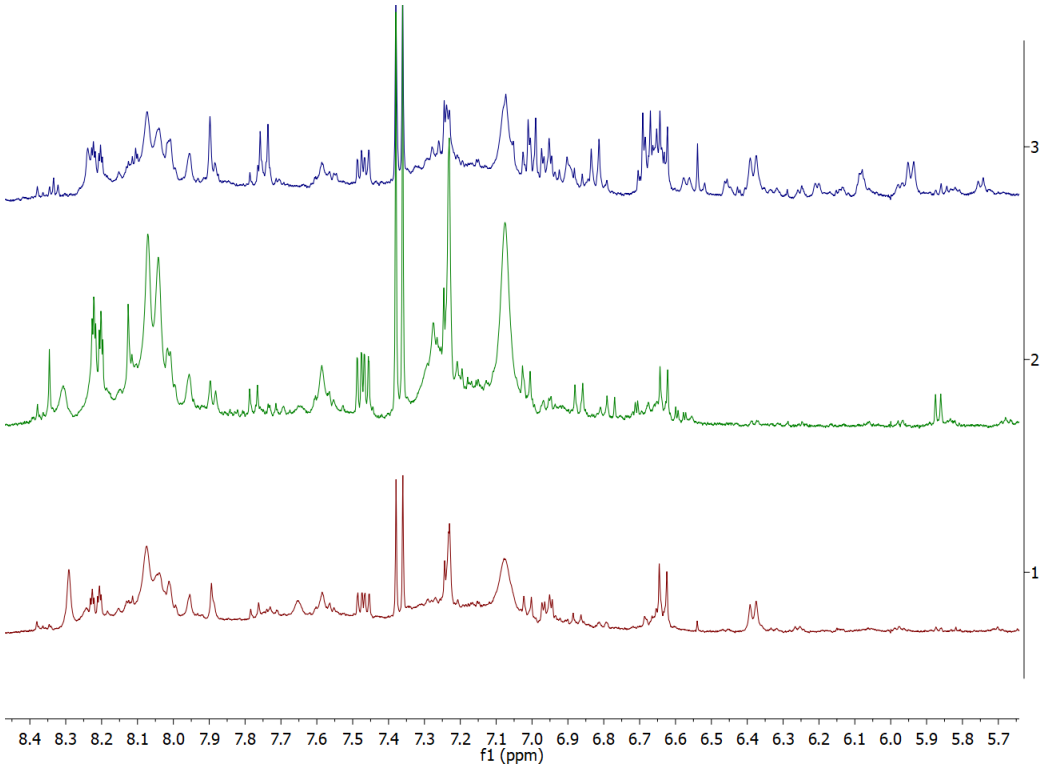


Figure 86. Downfield 1H NMR of SNF-022 WT, M2, M3 (bottom to top)

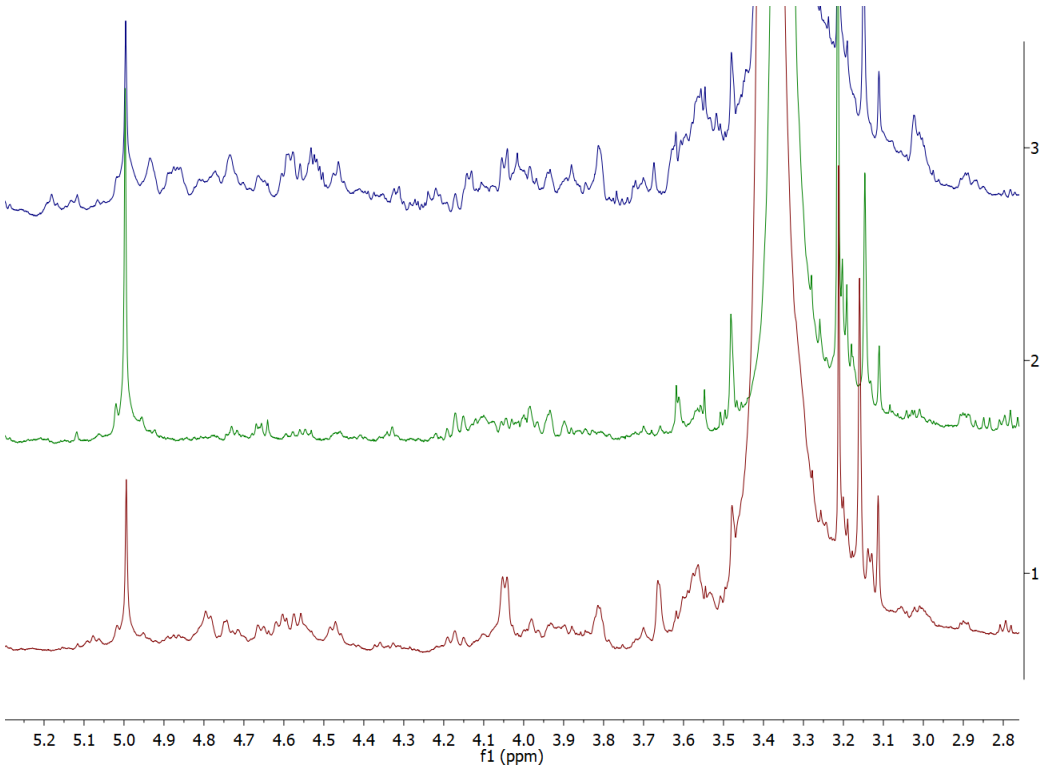


Figure 87. 1H NMR of SNF-022 WT, M2, M3 (bottom to top)

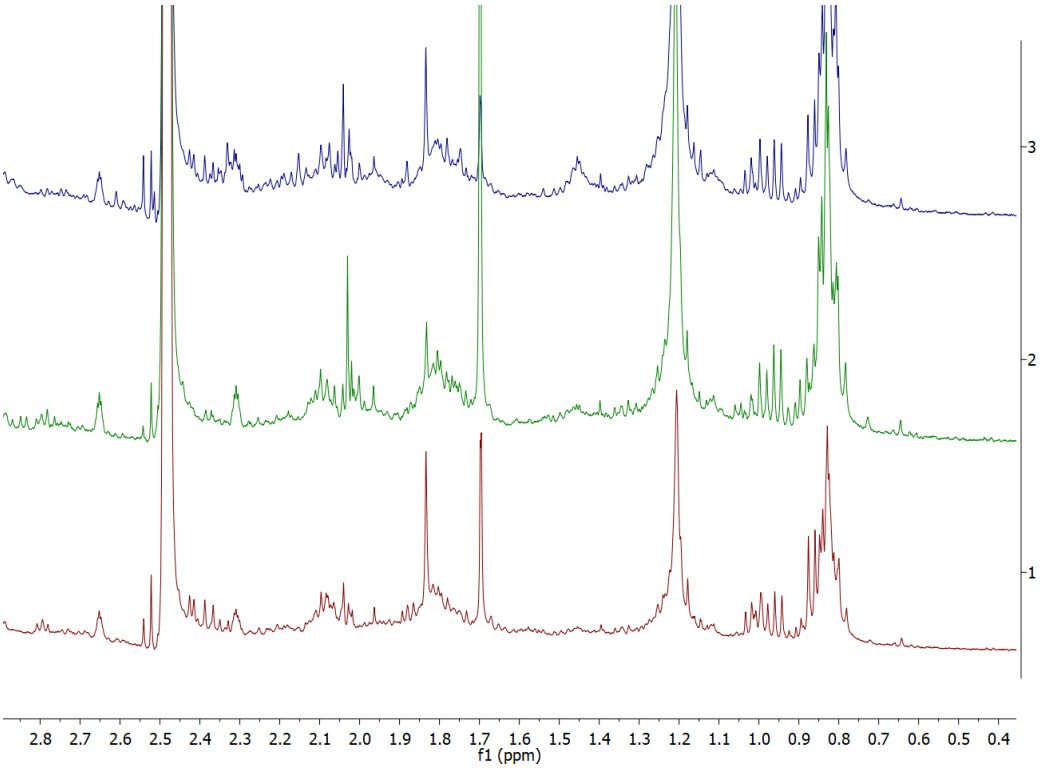


Figure 88. Upfield ^{1}H NMR of SNF-022 WT, M2, M3 (bottom to top)

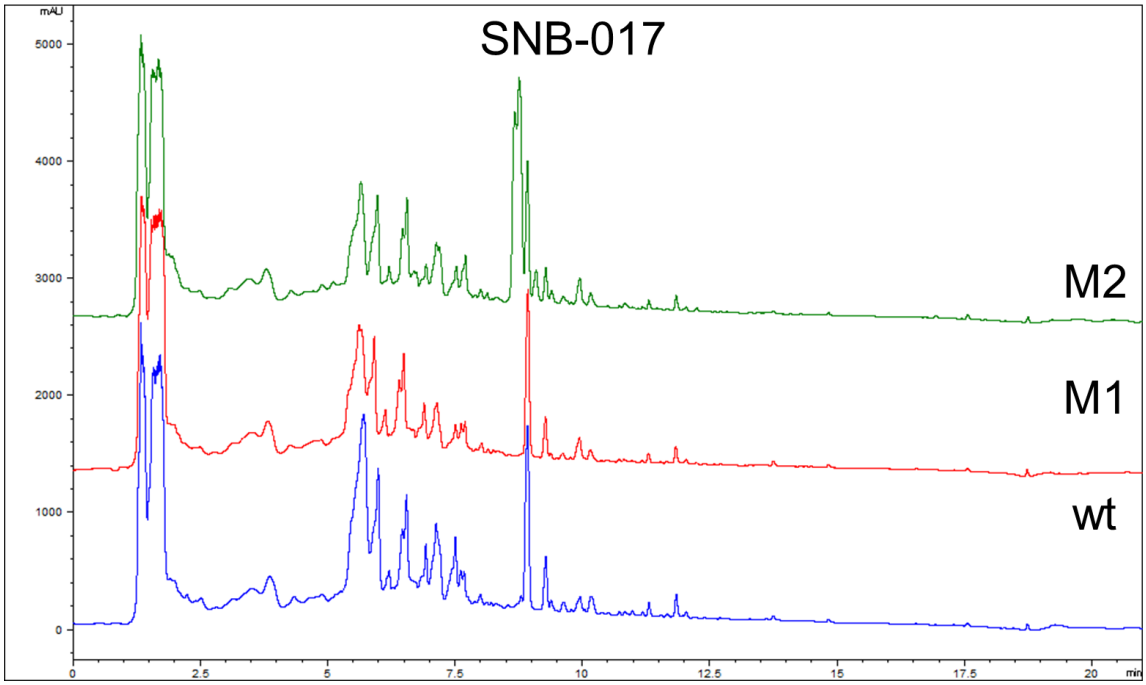


Figure 89. SNB-017 crude extracts (210 nm)

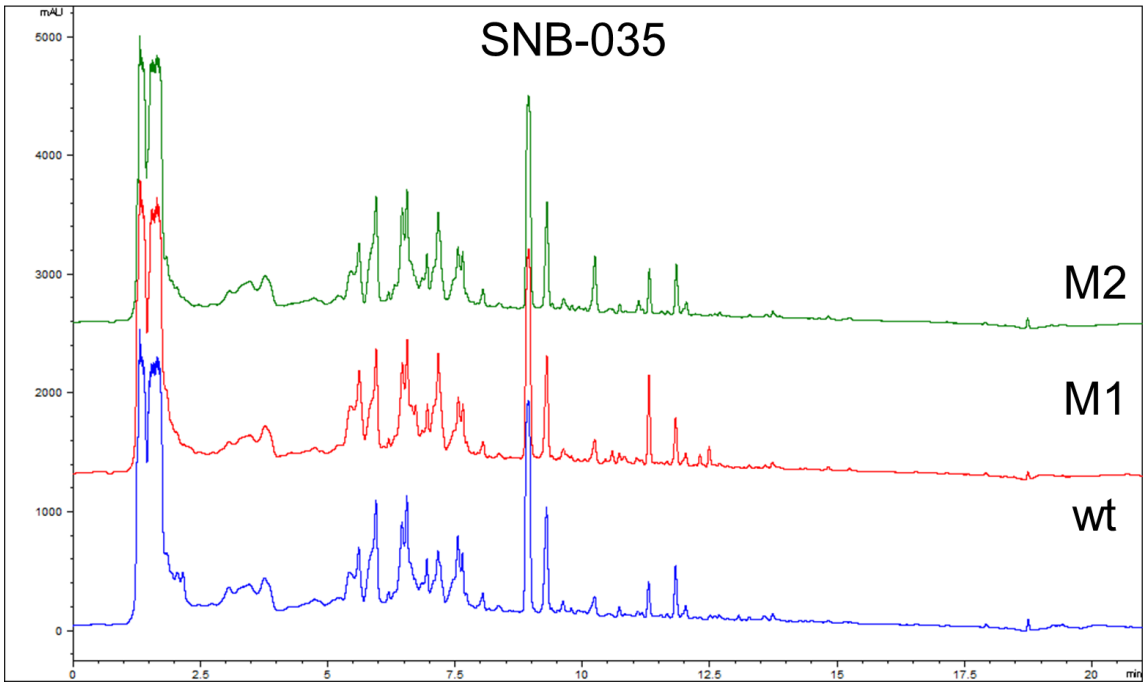


Figure 90. SNB-035 crude extracts (210 nm)

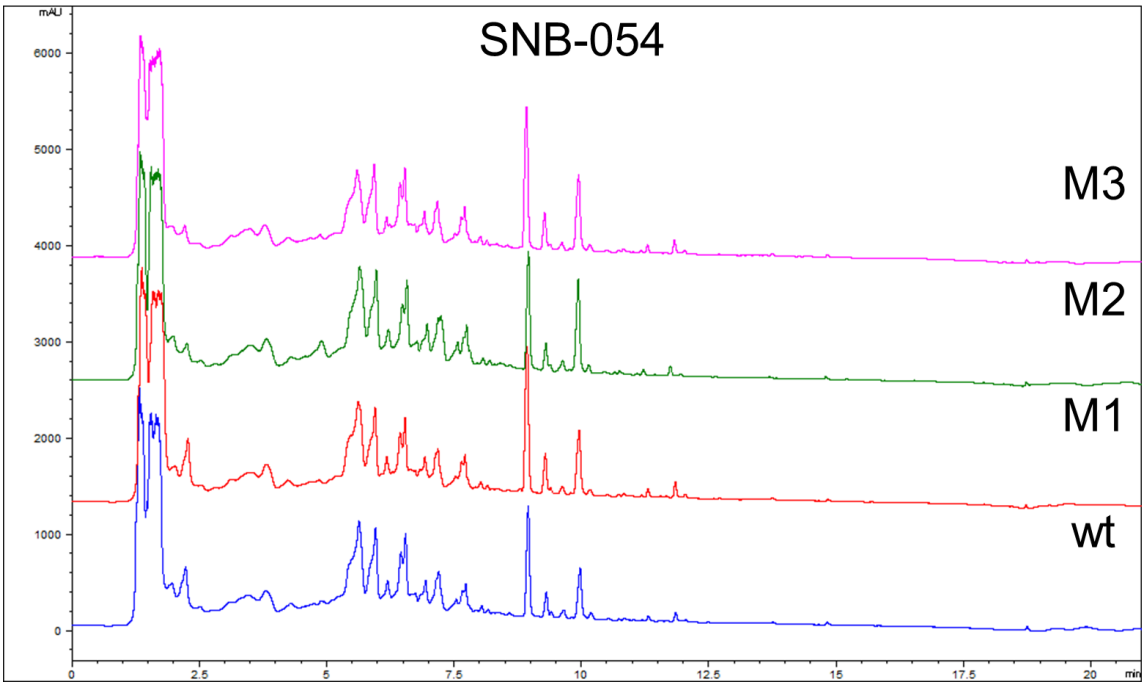


Figure 91. SNB-054 crude extract (210 nm)

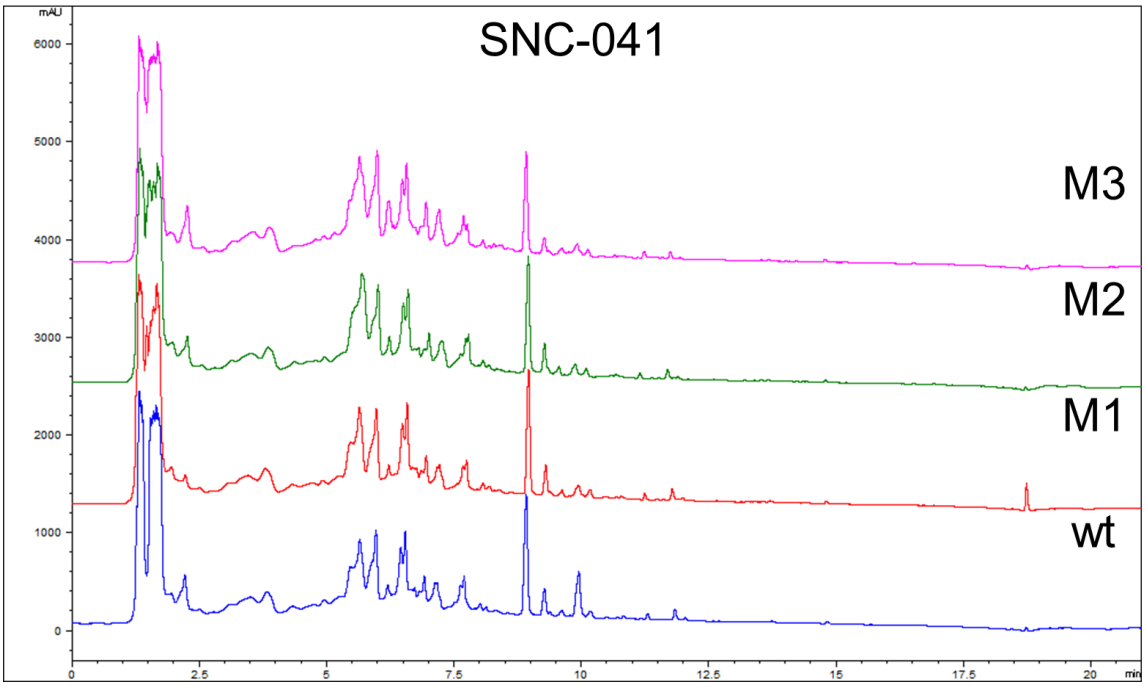


Figure 92. SNC-041 crude extract (210 nm)

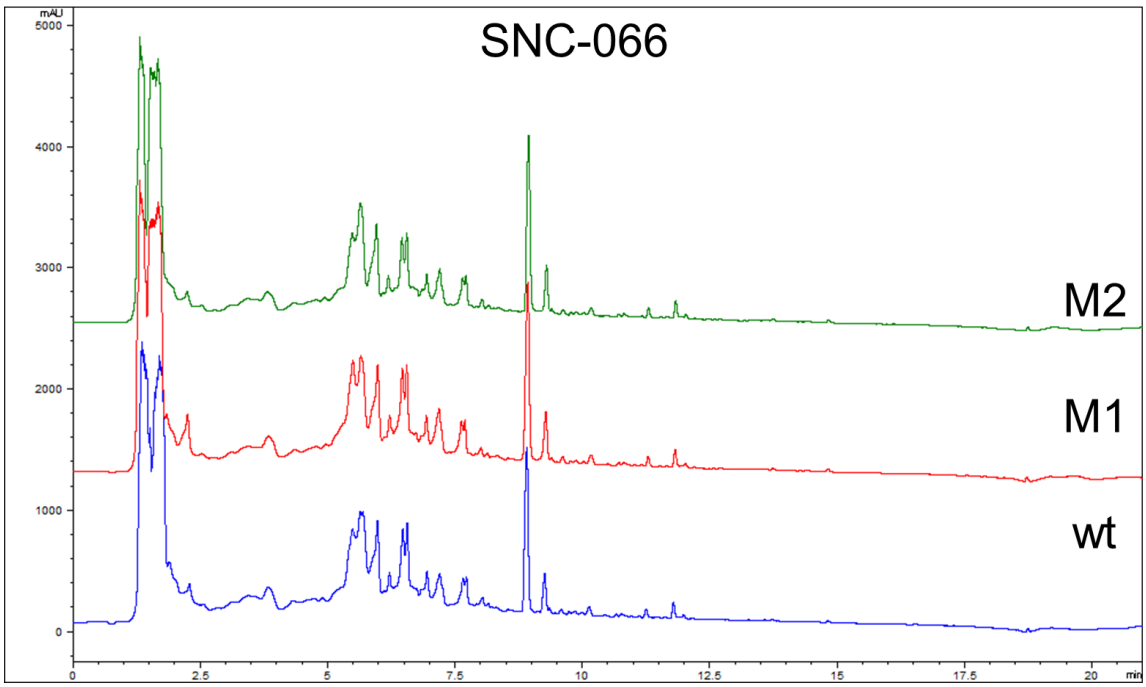


Figure 93. SNC-066 crude extracts (210 nm)

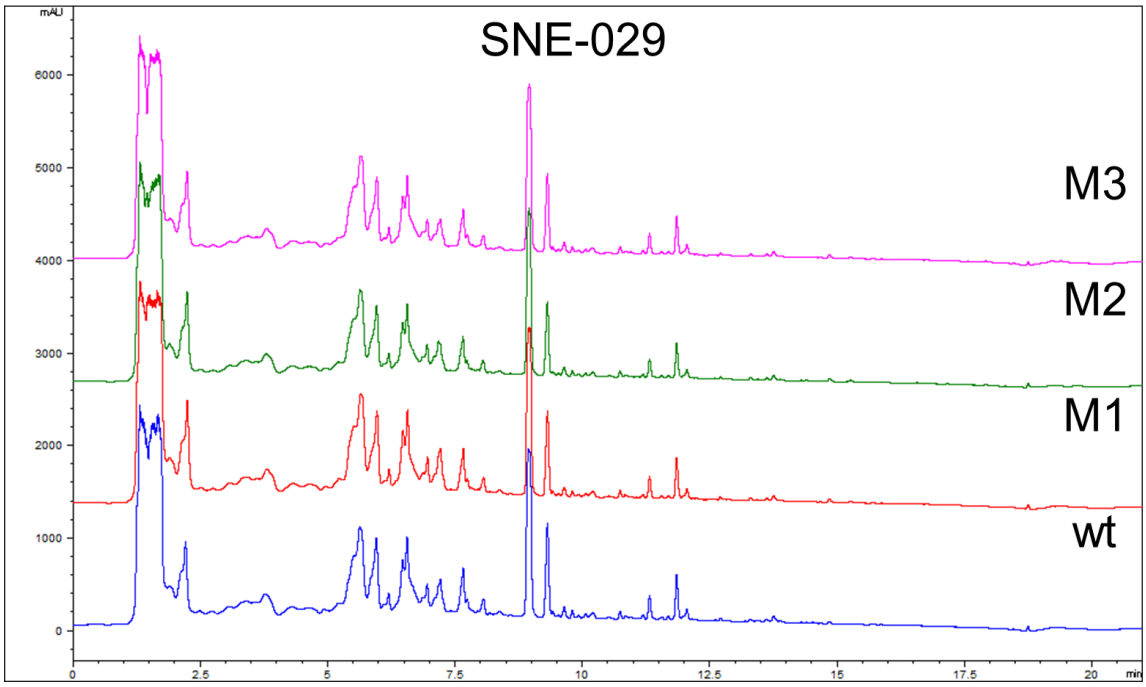


Figure 94. SNE-029 crude extracts (210 nm)

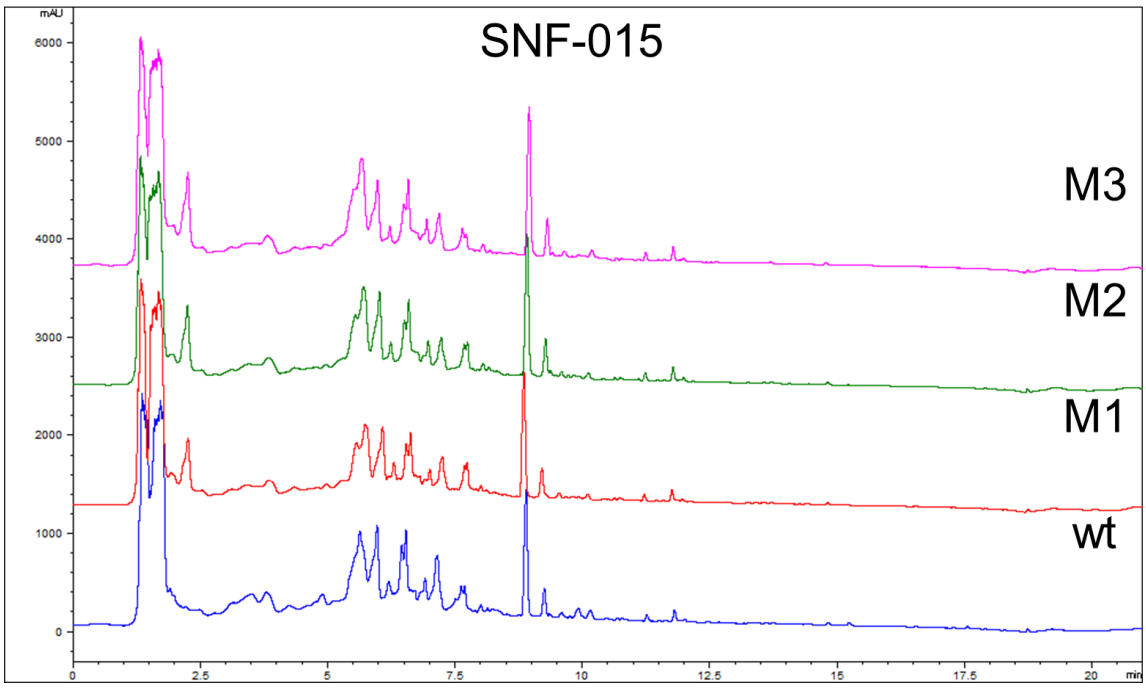


Figure 95. SNF-015 crude extracts (210 nm)

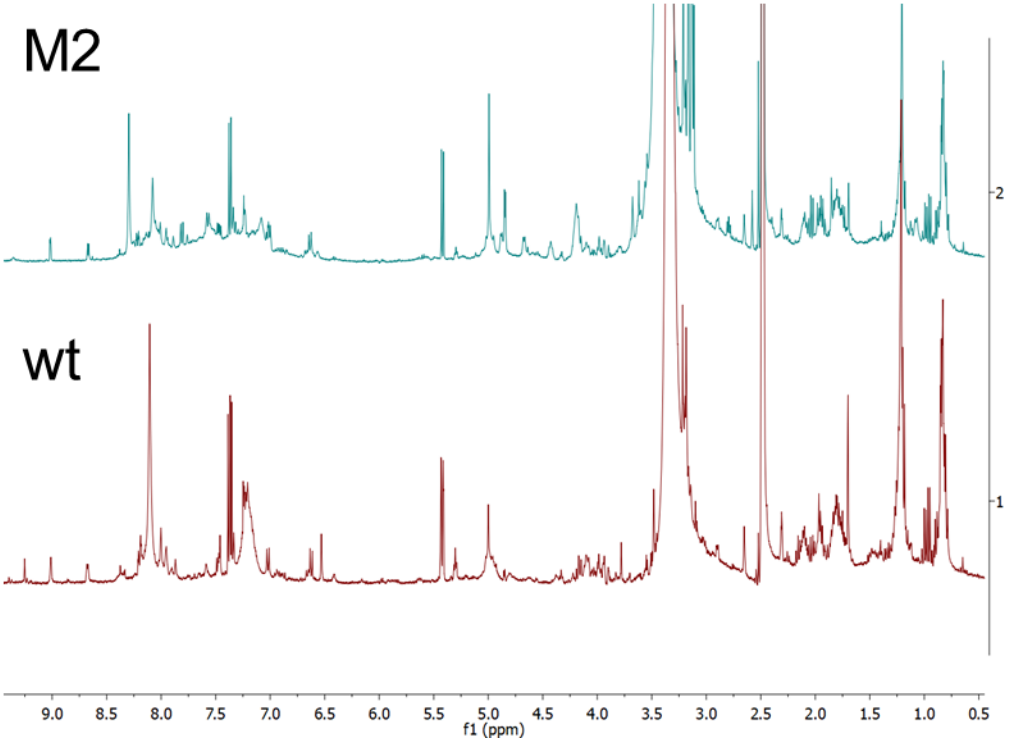


Figure 96. ^{1}H NMR spectra of SNC-119 WT and M2

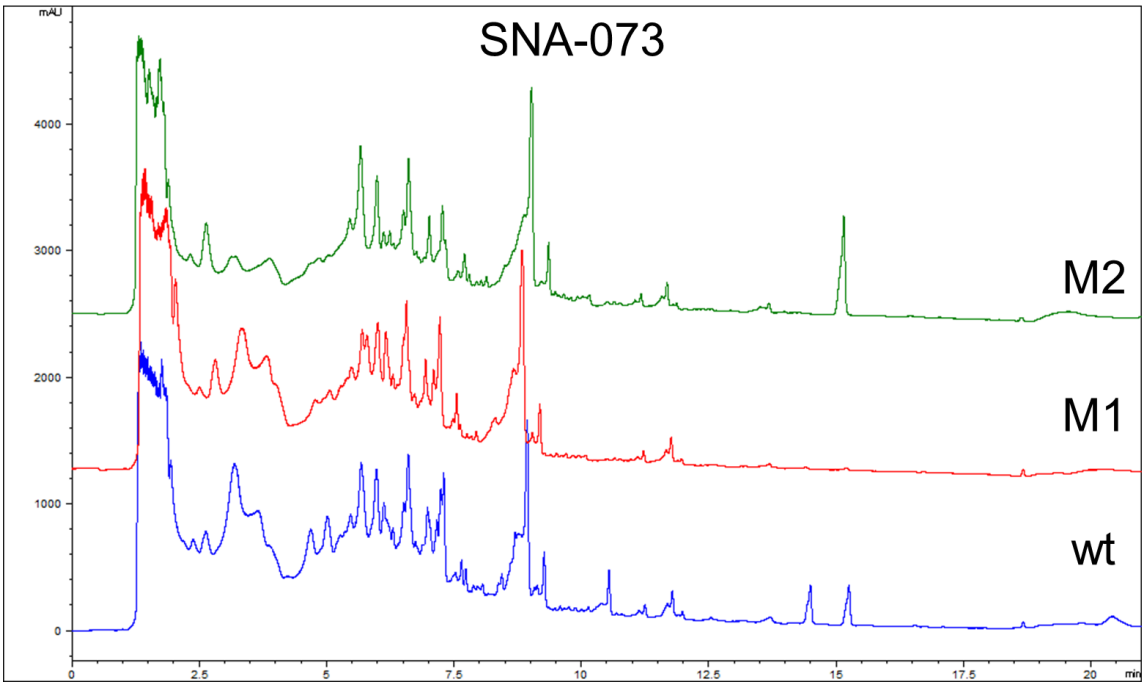


Figure 97. SNA-073 crude extracts (210 nm)

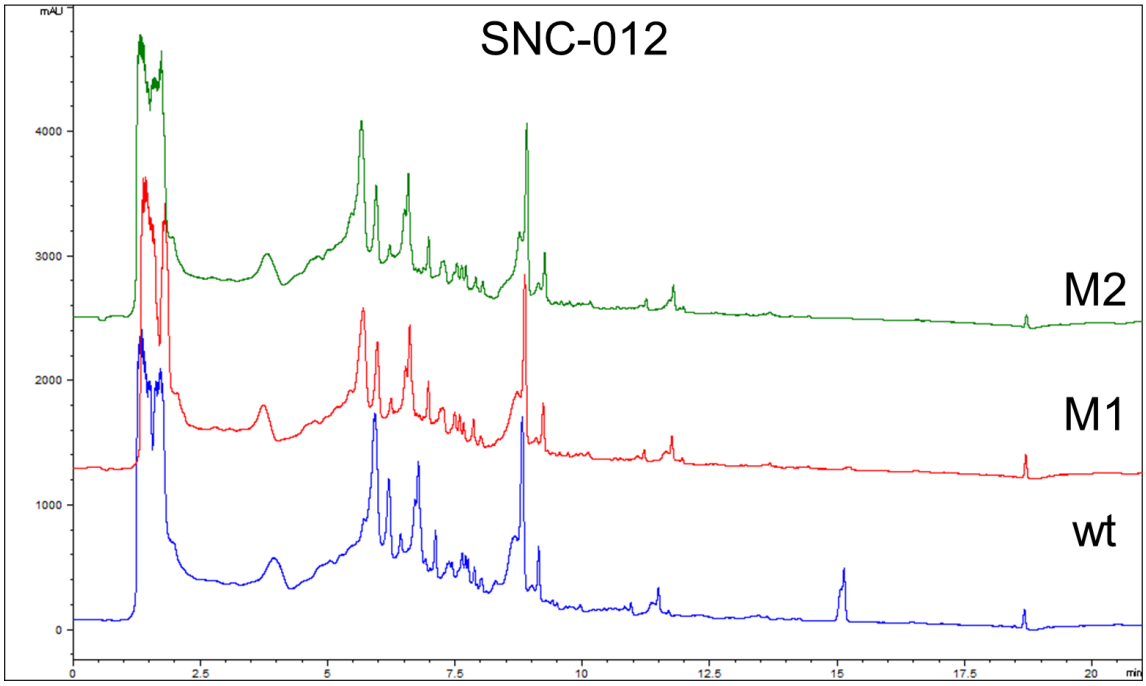


Figure 98. SNC-012 crude extracts (210 nm)

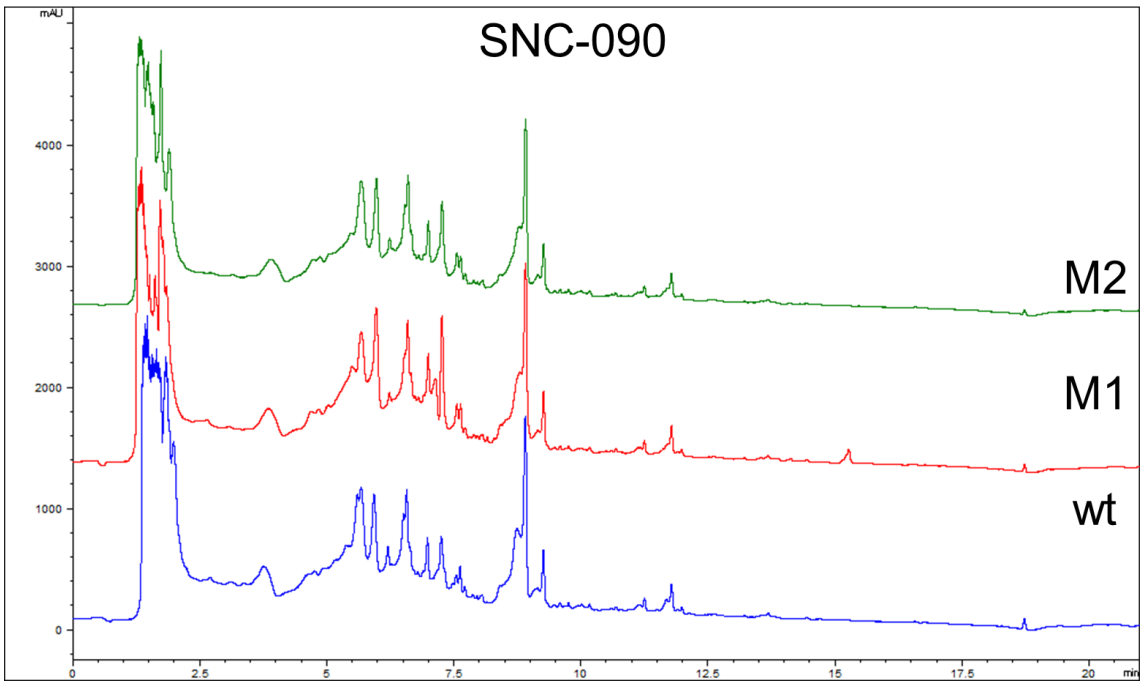


Figure 99. SNC-090 crude extracts (210 nm)

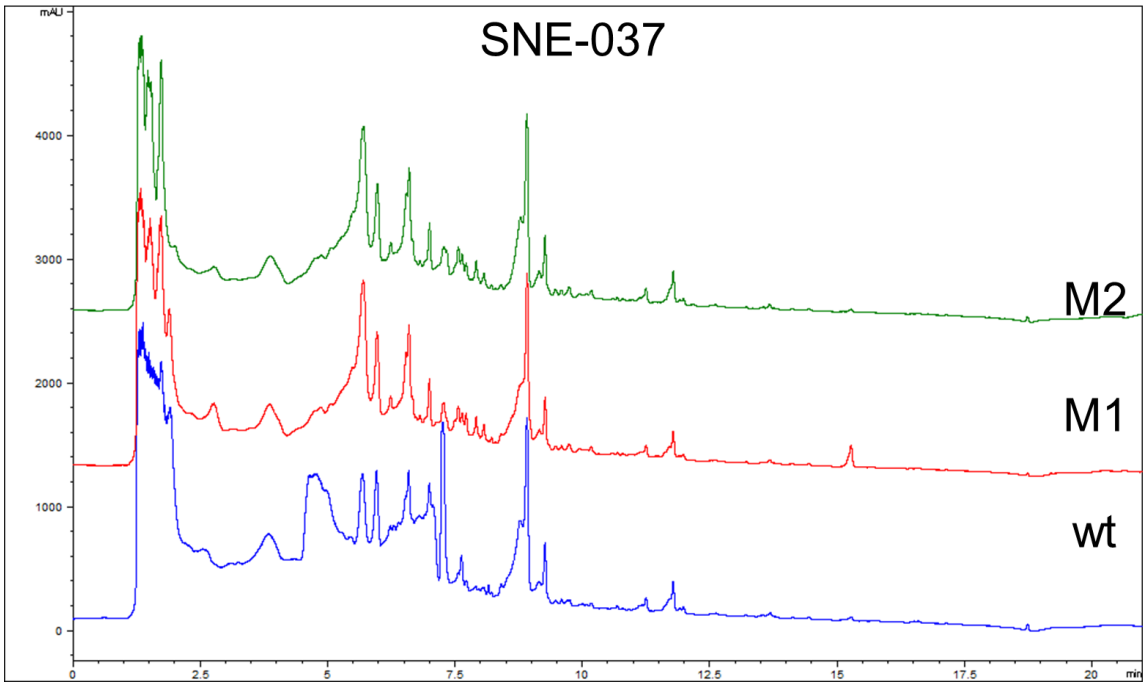


Figure 100. SNE-037 crude extracts (210 nm)

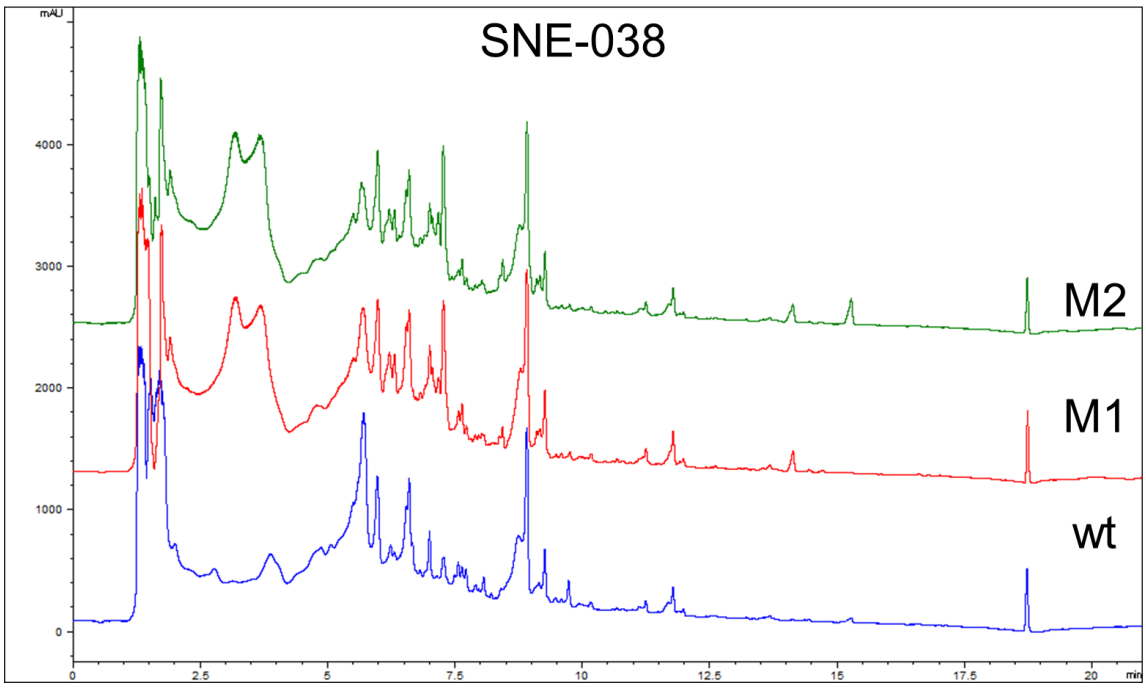


Figure 101. SNE-038 crude extracts (210 nm)

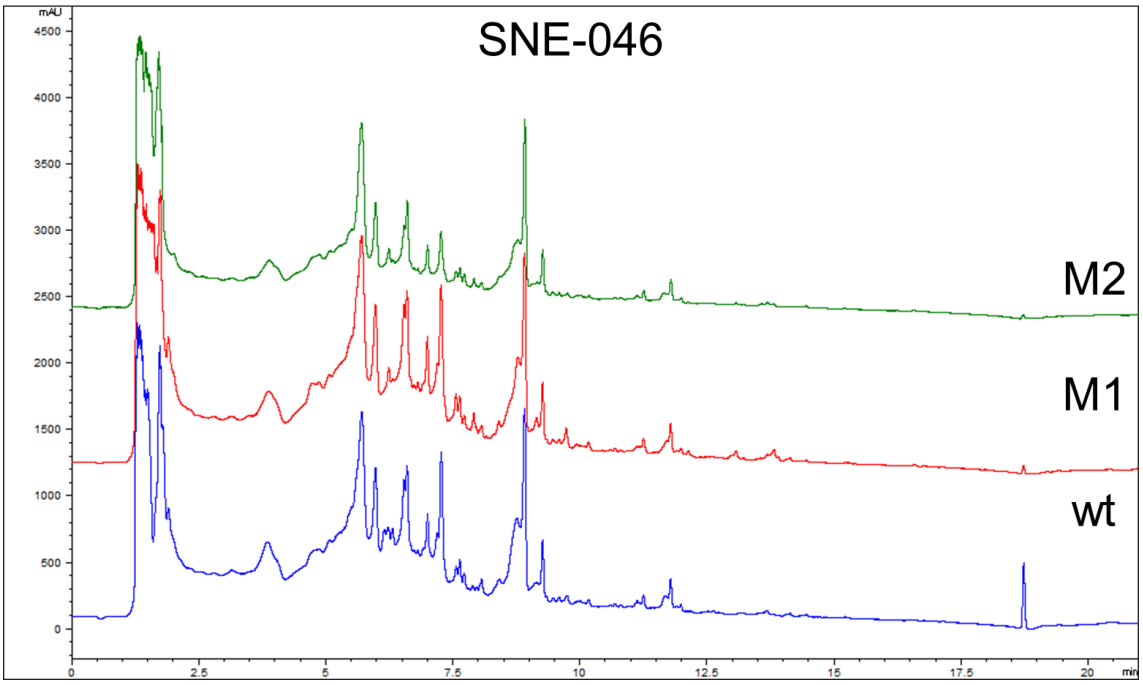


Figure 102. SNE-046 crude extracts (210 nm)

Table 10. Primer list for mutant rpoβ sequencing

SN	Genus	Species	M#	phylum	F-Primer	R-Primer (RC)
SNA-053	Agrococcus-99	terreus-99	3	Actinobacteria	GCTGTCGCAGTTCATGGA	GATSAGRCCGATGTTSGG
SNA-076	Isoptericola-100	variabilis-98	3	Actinobacteria	CCGAACATCGGCCTCATC	GATGAGGCCGATGTTCGG
SNA-079	Isoptericola-99	nanjingensis-99	2	Actinobacteria	CCGAACATCGGCCTCATC	GATGAGGCCGATGTTCGG
SNC-102	Isoptericola-97	variabilis-97	3	Actinobacteria	CCGAACATCGGCCTCATC	GATGAGGCCGATGTTCGG
SNE-016	Promicromonospora-99	sukumoe-99	2	Actinobacteria	TCGCAGCTSTCGCAGTTCA	GATCAGACCGATGTTCGGACC
SNF-022	Euzebyella-99	saccharophila-99	3	Bacteroidetes	AACCAATTATCTCAATTCATGGATCAA	AATCAAACCAATATTTGGCCCTT
SNB-017	Erythrobacter-99	flavus-99	3	Proteobacteria	TCGCAGTTCATGGAYCAG	ATSAGRCCGATRTTCGGG
SNB-035	Erythrobacter-99/Halomonas-99/Bacillus-98/Citromicrobium-98	citreus-99, aquimaris-99	2	Proteobacteria	TCGCAGTTCATGGAYCAG	ATSAGRCCGATRTTCGGG
SNB-054	Erythrobacter-100	citreus-100	3	Proteobacteria	TCGCAGTTCATGGAYCAG	ATSAGRCCGATRTTCGGG
SNC-041	Erythrobacter-100	citreus-100	3	Proteobacteria	TCGCAGTTCATGGAYCAG	ATSAGRCCGATRTTCGGG
SNC-066	Erythrobacter-99	vulgaris-99	2	Proteobacteria	TCGCAGTTCATGGAYCAG	ATSAGRCCGATRTTCGGG
SNC-119	Erythrobacter-100	flavus-100	3	Proteobacteria	TCGCAGTTCATGGAYCAG	ATSAGRCCGATRTTCGGG
SNE-029	Altererythrobacter-99	epoxidivorans-97	3	Proteobacteria	CAGCTDTCGCAGTTCATGGA	GATCAGRCCGATGTTCGG
SNF-015	Erythrobacter-100	citreus-100	3	Proteobacteria	TCGCAGTTCATGGAYCAG	ATSAGRCCGATRTTCGGG
SNA-073	Bacillus-100	endophyticus-99	2	Firmicutes	CTCAGCTTTCACAGTTCATGG	GATAAGCCCAATGTTTGGACC
SNC-012	Bacillus-99	vietnamensis-99	2	Firmicutes	TCTCAATTGTCTCAATTCATGGA	ATCAGTCCGATGTTCGGA
SNC-090	Bacillus-95	indicus-95	2	Firmicutes	TCACAGCTTTCCCAGTTCA	GATCAGACCGATGTTCGGT
SNE-037	Bacillus-99	algicola-99	2	Firmicutes	CTCAGCTTTCTCAATTTATGGACC	AATCAATCCAATGTTCGGTCCC
SNE-038	Bacillus-99	aquimaris-99, veitnamensis-99, alcalophilus-99, subtilis-99,	2	Firmicutes	TCTCAGTTATCTCAATTTATGGATCA	GATAAGTCCAATGTTTGGTCC
					TCACAGCTTTCTCAATTCATGGACCA	GATCAGTCCGATGTTCGGACC
SNE-046	Bacillus-99	baekryungensis-99	2	Firmicutes	CTCAGCTTTCTCAATTTATGGACC	AATCAATCCAATGTTCGGTCCC

BIBLIOGRAPHY

Abe, N., & Hirota, A. (2002). Chemical studies of the radical scavenging mechanism of bisorbicillinol using the 1,1-diphenyl-2-picrylhydrazyl radical. *Chemical Communications*(6), 662-663. doi: 10.1039/B200039N

Abe, N., Murata, T., & Hirota, A. (1998). Novel DPPH Radical Scavengers, Bisorbicillinol and Demethyltrichodimerol, from a Fungus. *Bioscience, Biotechnology, and Biochemistry, 62*(4), 661-666. doi: 10.1271/bbb.62.661

Abe, N., Murata, T., Yamamoto, K., & Hirota, A. (1999). Bisorbibetanone, a novel oxidized sorbicillin dimer, with 1,1-diphenyl-2-picrylhydrazyl radical scavenging activity from a fungus. *Tetrahedron Letters, 40*(28), 5203-5206. doi: http://dx.doi.org/10.1016/S0040-4039(99)00938-7

Abet, V., Mariani, A., Truscott, F. R., Britton, S., & Rodriguez, R. (2014). Biased and unbiased strategies to identify biologically active small molecules. *Bioorg Med Chem, 22*(16), 4474-4489. doi: 10.1016/j.bmc.2014.04.019

Adekambi, T., Drancourt, M., & Raoult, D. (2009). The rpoB gene as a tool for clinical microbiologists. *Trends Microbiol, 17*(1), 37-45. doi: 10.1016/j.tim.2008.09.008

Alifano, P., Palumbo, C., Pasanisi, D., & Tala, A. (2015). Rifampicin-resistance, rpoB polymorphism and RNA polymerase genetic engineering. *J Biotechnol, 202*, 60-77. doi: 10.1016/j.jbiotec.2014.11.024

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410. doi: 10.1016/s0022-2836(05)80360-2

An, S. J., Chen, Z. H., Su, J., Zhang, X. C., Zhong, W. Z., Yang, J. J., . . . Wu, Y. L. (2012). Identification of enriched driver gene alterations in subgroups of non-small cell lung cancer patients based on histology and smoking status. *PLoS One, 7*(6), e40109. doi: 10.1371/journal.pone.0040109

Arisawa, M., Fujita, A., Hayashi, T., Hayashi, K., Ochiai, H., & Morita, N. (1990). Cytotoxic and antiherpetic activity of phloroglucinol derivatives from Mallotus japonicus (Euphorbiaceae). *Chem Pharm Bull (Tokyo), 38*(6), 1624-1626.

Artman, G. D., Grubbs, A. W., & Williams, R. M. (2007). Concise, Asymmetric, Stereocontrolled Total Synthesis of Stephacidins A, B and Notoamide B. *J Am Chem Soc, 129*(19), 6336-6342. doi: 10.1021/ja070259i

Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M. N., Hosaka, T., Ochi, K., . . . Vassylyev, D. G. (2004). Structural basis for transcription regulation by alarmone ppGpp. *Cell, 117*(3), 299-310.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., . . . Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics, 9*, 75. doi: 10.1186/1471-2164-9-75

Bach, T., & Hehn, J. P. (2011). Photochemical reactions as key steps in natural product synthesis. *Angew Chem Int Ed Engl, 50*(5), 1000-1045. doi: 10.1002/anie.201002845

BAEZ, S., SEGURA-AGUILAR, J., WIDERSTEN, M., JOHANSSON, A.-S., & MANNERVIK, B. (1997). Glutathione transferases catalyse the detoxication of oxidized metabolites (o-quinones) of catecholamines and may serve as an antioxidant system preventing degenerative cellular processes. *Biochemical Journal, 324*(1), 25-28. doi: 10.1042/bj3240025

Balskus, E. P., & Walsh, C. T. (2008). Investigating the initial steps in the biosynthesis of cyanobacterial sunscreen scytonemin. *J Am Chem Soc, 130*(46), 15260-15261. doi: 10.1021/ja807192u

Bandaranayake, W. M., Banfield, J. E., & Black, D. S. C. (1980). Postulated electrocyclic reactions leading to endiandric acid and related natural products. *Journal of the Chemical Society, Chemical Communications*(19), 902-903. doi: 10.1039/C39800000902

Bandurraga, M. M., Fenical, W., Donovan, S. F., & Clardy, J. (1982). Pseudopterolide, an irregular diterpenoid with unusual cytotoxic properties from the Caribbean sea whip Pseudopterogorgia acerosa (Pallas) (Gorgonacea). *Journal of the American Chemical Society, 104*(23), 6463-6465. doi: 10.1021/ja00387a059

Baran, P. S., Guerrero, C. A., Hafensteiner, B. D., & Ambhaikar, N. B. (2005). Total Synthesis of Avrainvillamide (CJ-17,665) and Stephacidin B. *Angewandte Chemie International Edition, 44*(25), 3892-3895. doi: 10.1002/anie.200500655

Baran, P. S., Hafensteiner, B. D., Ambhaikar, N. B., Guerrero, C. A., & Gallagher, J. D. (2006). Enantioselective Total Synthesis of Avrainvillamide and the Stephacidins. *J Am Chem Soc, 128*(26), 8678-8693. doi: 10.1021/ja061660s

Beaudry, C. M., Malerich, J. P., & Trauner, D. (2005). Biosynthetic and Biomimetic Electrocyclizations. *Chemical Reviews, 105*(12), 4757-4778. doi: 10.1021/cr0406110

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., . . . Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature, 417*(6885), 141-147. doi: 10.1038/417141a

Bhat, S. G., & Vaidyanathan, C. S. (1976). Involvement of 4-hydroxymandelic acid in the degradation of mandelic acid by Pseudomonas convexa. *J Bacteriol, 127*(3), 1108-1118.

Bode, H. B., Reimer, D., Fuchs, S. W., Kirchner, F., Dauth, C., Kegler, C., . . . Grun, P. (2012). Determination of the absolute configuration of peptide natural products by using stable isotope labeling and mass spectrometry. *Chemistry, 18*(8), 2342-2348. doi: 10.1002/chem.201103479

Brophy, G. C., Mohandas, J., Slaytor, M., Sternhell, S., Watson, T. R., & Wilson, L. A. (1969). Novel lignans from a Cinnamomum sp. from Bougainville. *Tetrahedron Lett*(59), 5159-5162.

Burckle, A. J., Vasilev, V. H., & Burns, N. Z. (2016). A Unified Approach for the Enantioselective Synthesis of the Brominated Chamigrene Sesquiterpenes. *Angewandte Chemie International Edition, 55*(38), 11476-11479. doi: 10.1002/anie.201605722

Burdine, L., & Kodadek, T. (2004). Target identification in chemical genetics: the (often) missing link. *Chem Biol, 11*(5), 593-597. doi: 10.1016/j.chembiol.2004.05.001

Cai, L., & Tu, B. P. (2011). On acetyl-CoA as a gauge of cellular metabolic state. *Cold Spring Harb Symp Quant Biol, 76*, 195-202. doi: 10.1101/sqb.2011.76.010769

Carafoli, F., & Hohenester, E. (2013). Collagen recognition and transmembrane signalling by discoidin domain receptors. *Biochim Biophys Acta, 1834*(10), 2187-2194. doi: 10.1016/j.bbapap.2012.10.014

Chapman, O. L., Engel, M. R., Springer, J. P., & Clardy, J. C. (1971). Total synthesis of carpanone. *Journal of the American Chemical Society, 93*(24), 6696-6698. doi: 10.1021/ja00753a073

Chauthe, S. K., Bharate, S. B., Sabde, S., Mitra, D., Bhutani, K. K., & Singh, I. P. (2010). Biomimetic synthesis and anti-HIV activity of dimeric phloroglucinols. *Bioorganic & Medicinal Chemistry, 18*(5), 2029-2036. doi: https://doi.org/10.1016/j.bmc.2010.01.023

Chen, L., Gong, M.-W., Peng, Z.-F., Zhou, T., Ying, M.-G., Zheng, Q.-H., . . . Zhang, Q.-Q. (2014). The Marine Fungal Metabolite, Dicitrinone B, Induces A375 Cell Apoptosis through the ROS-Related Caspase Pathway. *Marine Drugs, 12*(4), 1939-1958. doi: 10.3390/md12041939

Chen, N. H., Djoko, K. Y., Veyrier, F. J., & McEwan, A. G. (2016). Formaldehyde Stress Responses in Bacterial Pathogens. *Front Microbiol, 7*. doi: 10.3389/fmicb.2016.00257

Chen, X. L., Liu, H. L., Li, J., Xin, G. R., & Guo, Y. W. (2011). Paracaseolide A, first alpha-alkylbutenolide dimer with an unusual tetraquinane oxa-cage bislactone skeleton from Chinese mangrove Sonneratia paracaseolaris. *Org Lett, 13*(19), 5032-5035. doi: 10.1021/ol201809q

Cheng, Q., Xiang, L., Izumikawa, M., Meluzzi, D., & Moore, B. S. (2007). Enzymatic total synthesis of enterocin polyketides. *Nat Chem Biol, 3*(9), 557-558. doi: 10.1038/nchembio.2007.22

Cleland, W. W. (2005). The use of isotope effects to determine enzyme mechanisms. *Arch Biochem Biophys, 433*(1), 2-12. doi: 10.1016/j.abb.2004.08.027

Colosimo, D. A., & MacMillan, J. B. (2016). Detailed Mechanistic Study of the Non-enzymatic Formation of the Discoipyrrole Family of Natural Products. *J Am Chem Soc, 138*(7), 2383-2388. doi: 10.1021/jacs.5b13320

Comprehensive genomic characterization of squamous cell lung cancers. (2012). *Nature, 489*(7417), 519-525. doi: 10.1038/nature11404

Cragg Gordon, M., & Newman David, J. (2005). Biodiversity: A continuing source of novel drug leads *Pure and Applied Chemistry* (Vol. 77, pp. 7).

Craney, A., Ozimok, C., Pimentel-Elardo, S. M., Capretta, A., & Nodwell, J. R. (2012). Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem Biol, 19*(8), 1020-1027. doi: 10.1016/j.chembiol.2012.06.013

Crick, F. (1970). Central dogma of molecular biology. *Nature, 227*(5258), 561-563.

Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., . . . D'Eustachio, P. (2014). The Reactome pathway knowledgebase. *Nucleic Acids Res, 42*(Database issue), D472-477. doi: 10.1093/nar/gkt1102

Czechowski, T., Larson, T. R., Catania, T. M., Harvey, D., Brown, G. D., & Graham, I. A. (2016). Artemisia annua mutant impaired in artemisinin synthesis demonstrates importance of nonenzymatic conversion in terpenoid metabolism. *Proceedings of the National Academy of Sciences of the United States of America, 113*(52), 15150-15155. doi: 10.1073/pnas.1611567113

de Lorenzo, V., Sekowska, A., & Danchin, A. (2015). Chemical reactivity drives spatiotemporal organisation of bacterial metabolism. *FEMS Microbiol Rev, 39*(1), 96-119. doi: 10.1111/1574-6976.12089

Dehn, R., Katsuyama, Y., Weber, A., Gerth, K., Jansen, R., Steinmetz, H., . . . Kirschning, A. (2011). Molecular basis of elansolid biosynthesis: evidence for an unprecedented quinone methide initiated intramolecular Diels-Alder cycloaddition/macrolactonization. *Angew Chem Int Ed Engl, 50*(17), 3882-3887. doi: 10.1002/anie.201006880

Dejgaard, S. Y., Dejgaard, K., & Presley, J. F. (2001). Cell Staining: Fluorescent Labelling of the Golgi Apparatus *eLS*: John Wiley & Sons, Ltd.

Dormeyer, M., Egelkamp, R., Thiele, M. J., Hammer, E., Gunka, K., Stannek, L., . . . Commichau, F. M. (2015). A novel engineering tool in the Bacillus subtilis toolbox: inducer-free activation of gene expression by selection-driven promoter decryptification. *Microbiology, 161*(Pt 2), 354-361. doi: 10.1099/mic.0.000001

Doull, J. L., Ayer, S. W., Singh, A. K., & Thibault, P. (1993). Production of a novel polyketide antibiotic, jadomycin B, by Streptomyces venezuelae following heat shock. *J Antibiot (Tokyo), 46*(5), 869-871.

Doull, J. L., Singh, A. K., Hoare, M., & Ayer, S. W. (1994). Conditions for the production of jadomycin B by Streptomyces venezuelae ISP5230: effects of heat shock, ethanol treatment and phage infection. *J Ind Microbiol, 13*(2), 120-125.

Du, L., Robles, A. J., King, J. B., Powell, D. R., Miller, A. N., Mooberry, S. L., & Cichewicz, R. H. (2014). Crowdsourcing Natural Products Discovery to Access Uncharted Dimensions

of Fungal Metabolite Diversity(). *Angewandte Chemie (International ed. in English), 53*(3), 804-809. doi: 10.1002/anie.201306549

Du, L., You, J., Nicholas, K. M., & Cichewicz, R. H. (2016). Chemoreactive Natural Products that Afford Resistance Against Disparate Antibiotics and Toxins. *55*(13), 4220-4225. doi: 10.1002/anie.201511348

Dupuis, S. N., Veinot, T., Monro, S. M., Douglas, S. E., Syvitski, R. T., Goralski, K. B., . . . Jakeman, D. L. (2011). Jadomycins derived from the assimilation and incorporation of norvaline and norleucine. *J Nat Prod, 74*(11), 2420-2424. doi: 10.1021/np200689w

Eade, S. J., Walter, M. W., Byrne, C., Odell, B., Rodriguez, R., Baldwin, J. E., . . . Moses, J. E. (2008). Biomimetic synthesis of pyrone-derived natural products: exploring chemical pathways from a unique polyketide precursor. *J Org Chem, 73*(13), 4830-4839. doi: 10.1021/jo800220w

Eaton, A. L., Dalal, S., Cassera, M. B., Zhao, S., & Kingston, D. G. (2016). Synthesis and Antimalarial Activity of Mallatojaponin C and Related Compounds. *J Nat Prod, 79*(6), 1679-1683. doi: 10.1021/acs.jnatprod.6b00347

Egan, J. B., Barrett, M. T., Champion, M. D., Middha, S., Lenkiewicz, E., Evers, L., . . . Stewart, A. K. (2014). Whole genome analyses of a well-differentiated liposarcoma reveals novel SYT1 and DDR2 rearrangements. *PLoS One, 9*(2), e87113. doi: 10.1371/journal.pone.0087113

Esquenazi, E., Jones, A. C., Byrum, T., Dorrestein, P. C., & Gerwick, W. H. (2011). Temporal dynamics of natural product biosynthesis in marine cyanobacteria. *Proc Natl Acad Sci U S A, 108*(13), 5226-5231. doi: 10.1073/pnas.1012813108

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., . . . D'Eustachio, P. (2016). The Reactome pathway Knowledgebase. *Nucleic Acids Res, 44*(D1), D481-487. doi: 10.1093/nar/gkv1351

Fedorov, S. N., Radchenko, O. S., Shubina, L. K., Kalinovsky, A. I., Gerasimenko, A. V., Popov, D. Y., & Stonik, V. A. (2001). Aplydactone, a New Sesquiterpenoid with an Unprecedented Carbon Skeleton from the Sea Hare Aplysia dactylomela, and Its Cargill-Like Rearrangement. *Journal of the American Chemical Society, 123*(3), 504-505. doi: 10.1021/ja003254t

Fedorov, S. N., Shubina, L. K., Kalinovsky, A. I., Lyakhova, E. G., & Stonik, V. A. (2000). Structure and absolute configuration of a new rearranged chamigrane-type sesquiterpenoid from the sea hare Aplysia sp. *Tetrahedron Letters, 41*(12), 1979-1982. doi: https://doi.org/10.1016/S0040-4039(00)00077-0

Fenical, W., Jensen, P., Macmillan, J., Hughes, C. C., & Laclair, J. J. (2012). US8133902B2.

Fenical, W., & McConnell, O. (1976). Simple antibiotics from the red seaweed Dasya pedicellata var. Stanfordiana. *Phytochemistry, 15*(3), 435-436. doi: http://dx.doi.org/10.1016/S0031-9422(00)86848-X

Fischer, N., Neumann, P., Konevega, A. L., Bock, L. V., Ficner, R., Rodnina, M. V., & Stark, H. (2015). Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM. *Nature, 520*(7548), 567-570. doi: 10.1038/nature14275

Forget, S. M., Robertson, A. W., Overy, D. P., Kerr, R. G., & Jakeman, D. L. (2017). Furan and Lactam Jadomycin Biosynthetic Congeners Isolated from Streptomyces venezuelae ISP5230 Cultured with Nepsilon-Trifluoroacetyl-l-lysine. *J Nat Prod*. doi: 10.1021/acs.jnatprod.7b00152

Fu, P., Jamison, M., La, S., & MacMillan, J. B. (2014). Inducamides A-C, chlorinated alkaloids from an RNA polymerase mutant strain of Streptomyces sp. *Org Lett, 16*(21), 5656-5659. doi: 10.1021/ol502731p

Fu, P., Legako, A., La, S., & MacMillan, J. B. (2016). Discovery, Characterization, and Analogue Synthesis of Bohemamine Dimers Generated by Non-enzymatic Biosynthesis. *Chemistry, 22*(10), 3491-3495. doi: 10.1002/chem.201600024

Gaca, A. O., Colomer-Winter, C., & Lemos, J. A. (2015). Many means to a common end: the intricacies of (p)ppGpp metabolism and its control of bacterial homeostasis. *J Bacteriol, 197*(7), 1146-1156. doi: 10.1128/jb.02577-14

Gan, L. S., Fan, C. Q., Yang, S. P., Wu, Y., Lin, L. P., Ding, J., & Yue, J. M. (2006). Flueggenines A and B, two novel C,C-linked dimeric indolizidine alkaloids from Flueggea virosa. *Org Lett, 8*(11), 2285-2288. doi: 10.1021/ol060551f

Giera, D. S., & Stark, C. B. W. (2013). Total synthesis of (+/-)-paracaseolide A and initial attempts at a Lewis acid mediated dimerization of its putative biosynthetic precursor. *RSC Advances, 3*(44), 21280-21284. doi: 10.1039/C3RA44590A

Giordano, D., Coppola, D., Russo, R., Denaro, R., Giuliano, L., Lauro, F. M., . . . Verde, C. (2015). Marine Microbial Secondary Metabolites: Pathways, Evolution and Physiological Roles. *Adv Microb Physiol, 66*, 357-428. doi: 10.1016/bs.ampbs.2015.04.001

Gocke, D., Nguyen, C. L., Pohl, M., Stillger, T., Walter, L., & Müller, M. (2007). Branched-Chain Keto Acid Decarboxylase from Lactococcus lactis (KdcA), a Valuable Thiamine Diphosphate-Dependent Enzyme for Asymmetric C□C Bond Formation. *Advanced Synthesis & Catalysis, 349*(8-9), 1425-1435. doi: 10.1002/adsc.200700057

Gomez-Gallego, M., & Sierra, M. A. (2011). Kinetic isotope effects in the study of organometallic reaction mechanisms. *Chem Rev, 111*(8), 4857-4963. doi: 10.1021/cr100436k

Gonzalez, M. E., Martin, E., Anwar, T., Arellano-Garcia, C., Medhora, N., Lama, A., . . . Kleer, C. G. (2017). Mesenchymal stem cell induced DDR2 mediates stromal-breast cancer interactions and metastasis growth. *Cell reports, 18*(5), 1215-1228. doi: 10.1016/j.celrep.2016.12.079

Gould, S. J., Chen, J., Cone, M. C., Gore, M. P., Melville, C. R., & Tamayo, N. (1996). Identification of Prekinamycin in Extracts of Streptomyces murayamaensis. *The Journal of Organic Chemistry, 61*(17), 5720-5721. doi: 10.1021/jo9611024

Gould, S. J., Hong, S. T., & Carney, J. R. (1998). Cloning and heterologous expression of genes from the kinamycin biosynthetic pathway of Streptomyces murayamaensis. *J Antibiot (Tokyo), 51*(1), 50-57.

Gould, S. J., Melville, C. R., Cone, M. C., Chen, J., & Carney, J. R. (1997). Kinamycin Biosynthesis. Synthesis, Isolation, and Incorporation of Stealthin C, an Aminobenzo[b]fluorene. *The Journal of Organic Chemistry, 62*(2), 320-324. doi: 10.1021/jo961486y

Granstein, R. D., Flotte, T. J., & Amento, E. P. (1990). Interferons and collagen production. *J Invest Dermatol, 95*(6 Suppl), 75s-80s.

Gravel, E., & Poupon, E. (2008). Biogenesis and Biomimetic Chemistry: Can Complex Natural Products Be Assembled Spontaneously? *European journal of organic chemistry, 2008*(1), 27-42. doi: 10.1002/ejoc.200700331

Grube, A., Immel, S., Baran, P. S., & Köck, M. (2007). Massadine Chloride: A Biosynthetic Precursor of Massadine and Stylissadine. *Angewandte Chemie International Edition, 46*(43), 8107-8107. doi: 10.1002/anie.200790221

Guo, Y., Yang, T. L., Dong, S. S., Yan, H., Hao, R. H., Chen, X. F., . . . Deng, H. W. (2015). Genetic analysis identifies DDR2 as a novel gene affecting bone mineral density and osteoporotic fractures in Chinese population. *PLoS One, 10*(2), e0117102. doi: 10.1371/journal.pone.0117102

Hallows, W. C., Yu, W., & Denu, J. M. (2012). Regulation of glycolytic enzyme phosphoglycerate mutase-1 by Sirt1 protein-mediated deacetylation. *J Biol Chem, 287*(6), 3850-3858. doi: 10.1074/jbc.M111.317404

Hammerman, P. S., Sos, M. L., Ramos, A. H., Xu, C., Dutt, A., Zhou, W., . . . Meyerson, M. (2011). Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer Discov, 1*(1), 78-89. doi: 10.1158/2159-8274.cd-11-0005

Harned, A. M., & Volp, K. A. (2011). The sorbicillinoid family of natural products: Isolation, biosynthesis, and synthetic studies. *Natural Product Reports, 28*(11), 1790-1810. doi: 10.1039/C1NP00039J

Hata, A. N., & Breyer, R. M. (2004). Pharmacology and signaling of prostaglandin receptors: multiple roles in inflammation and immune modulation. *Pharmacol Ther, 103*(2), 147-166. doi: 10.1016/j.pharmthera.2004.06.003

Herzon, S. B., & Myers, A. G. (2005). Enantioselective Synthesis of Stephacidin B. *J Am Chem Soc, 127*(15), 5342-5344. doi: 10.1021/ja0510616

Hesketh, A., & Ochi, K. (1997). A novel method for improving Streptomyces coelicolor A3(2) for production of actinorhodin by introduction of rpsL (encoding ribosomal protein S12) mutations conferring resistance to streptomycin. *J Antibiot (Tokyo), 50*(6), 532-535.

Hopwood, D. A. (1988). The Leeuwenhoek lecture, 1987. Towards an understanding of gene switching in Streptomyces, the basis of sporulation and antibiotic production. *Proc R Soc Lond B Biol Sci, 235*(1279), 121-138.

Hosaka, T., Ohnishi-Kameyama, M., Muramatsu, H., Murakami, K., Tsurumi, Y., Kodani, S., . . . Ochi, K. (2009). Antibacterial discovery in actinomycetes strains with mutations in

RNA polymerase or ribosomal protein S12. *Nat Biotechnol, 27*(5), 462-464. doi: 10.1038/nbt.1538

Hosaka, T., Xu, J., & Ochi, K. (2006). Increased expression of ribosome recycling factor is responsible for the enhanced protein synthesis during the late growth phase in an antibiotic-overproducing Streptomyces coelicolor ribosomal rpsL mutant. *Mol Microbiol, 61*(4), 883-897. doi: 10.1111/j.1365-2958.2006.05285.x

Hu, H., Zhang, Q., & Ochi, K. (2002). Activation of antibiotic biosynthesis by specified mutations in the rpoB gene (encoding the RNA polymerase beta subunit) of Streptomyces lividans. *J Bacteriol, 184*(14), 3984-3991.

Hu, Y., Legako, A. G., Espindola, A. P., & MacMillan, J. B. (2012). Erythrolic acids A-E, meroterpenoids from a marine-derived Erythrobacter sp. *J Org Chem, 77*(7), 3401-3407. doi: 10.1021/jo300197z

Hu, Y., & MacMillan, J. B. (2011). Erythrazoles A-B, cytotoxic benzothiazoles from a marine-derived Erythrobacter sp. *Org Lett, 13*(24), 6580-6583. doi: 10.1021/ol202944g

Hu, Y., Potts, M. B., Colosimo, D., Herrera-Herrera, M. L., Legako, A. G., Yousufuddin, M., . . . MacMillan, J. B. (2013). Discoipyrroles A-D: isolation, structure determination, and synthesis of potent migration inhibitors from Bacillus hunanensis. *J Am Chem Soc, 135*(36), 13387-13392. doi: 10.1021/ja403412y

Hughes, C. C., MacMillan, J. B., Gaudencio, S. P., Fenical, W., & La Clair, J. J. (2009). Ammosamides A and B target myosin. *Angew. Chem., Int. Ed., 48*(4), 728-732. doi: 10.1002/anie.200804107

Hughes, C. C., MacMillan, J. B., Gaudencio, S. P., Jensen, P. R., & Fenical, W. (2009). The ammosamides: structures of cell cycle modulators from a marine-derived Streptomyces species. *Angew. Chem., Int. Ed., 48*(4), 725-727. doi: 10.1002/anie.200804890

IaA, A., & Turova, A. D. (1956). [Pharmacology of a new alkaloid securinine]. *Farmakol Toksikol, 19*(4), 11-17.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., . . . Omura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nat Biotech, 21*(5), 526-531. doi: http://www.nature.com/nbt/journal/v21/n5/suppinfo/nbt820_S1.html

Ireland, C., Faulkner, D. J., Finer, J. S., & Clardy, J. (1979). Crispatone, a metabolite of the opisthobranch mollusc Tridachia crispata. *Journal of the American Chemical Society, 101*(5), 1275-1276. doi: 10.1021/ja00499a037

Ireland, C., Faulkner, D. J., Solheim, B. A., & Clardy, J. (1978). Tridachione, a propionate-derived metabolite of the opisthobranch mollusc Tridachiella diomedea. *Journal of the American Chemical Society, 100*(3), 1002-1003. doi: 10.1021/ja00471a074

Ireland, C., & Scheuer, P. J. (1979). Photosynthetic Marine Mollusks: In vivo 14C Incorporation into Metabolites of the Sacoglossan Placobranchus ocellatus. *Science, 205*(4409), 922-923. doi: 10.1126/science.205.4409.922

Iwai, L. K., Payne, L. S., Luczynski, M. T., Chang, F., Xu, H., Clinton, R. W., . . . Huang, P. H. (2013). Phosphoproteomics of collagen receptor networks reveals SHP-2 phosphorylation downstream of wild-type DDR2 and its lung cancer mutants. *Biochem J, 454*(3), 501-513. doi: 10.1042/bj20121750

Jakeman, D. L., Bandi, S., Graham, C. L., Reid, T. R., Wentzell, J. R., & Douglas, S. E. (2009). Antimicrobial activities of jadomycin B and structurally related analogues. *Antimicrob Agents Chemother, 53*(3), 1245-1247. doi: 10.1128/aac.00801-08

Jakeman, D. L., Graham, C. L., & Reid, T. R. (2005). Novel and expanded jadomycins incorporating non-proteogenic amino acids. *Bioorg Med Chem Lett, 15*(23), 5280-5283. doi: 10.1016/j.bmcl.2005.08.047

Jamaluddin, M., Rao, P. V., & Vaidyanathan, C. S. (1970). Involvement of the protocatechuate pathway in the metabolism of mandelic acid by Aspergillus niger. *J Bacteriol, 101*(3), 786-793.

Jansen, R., Gerth, K., Steinmetz, H., Reinecke, S., Kessler, W., Kirschning, A., & Muller, R. (2011). Elansolid A3, a unique p-quinone methide antibiotic from Chitinophaga sancti. *Chemistry, 17*(28), 7739-7744. doi: 10.1002/chem.201100457

Jensen, P. R. (2016). Natural Products and the Gene Cluster Revolution.

Jiang, T., Wang, M., Li, L., Si, J., Song, B., Zhou, C., . . . Zou, Z. (2016). Overexpression of the Global Regulator LaeA in Chaetomium globosum Leads to the Biosynthesis of Chaetoglobosin Z. *J Nat Prod, 79*(10), 2487-2494. doi: 10.1021/acs.jnatprod.6b00333

Johnson, J. D., Edman, J. C., & Rutter, W. J. (1993). A receptor tyrosine kinase found in breast carcinoma cells has an extracellular discoidin I-like domain. *Proc Natl Acad Sci U S A, 90*(12), 5677-5681.

Jordan, P. A., & Moore, B. S. (2016). Biosynthetic Pathway Connects Cryptic Ribosomally Synthesized Posttranslationally Modified Peptide Genes with Pyrroloquinoline Alkaloids. *Cell Chem Biol, 23*(12), 1504-1514. doi: 10.1016/j.chembiol.2016.10.009

Juhl, M., & Tanner, D. (2009). Recent applications of intramolecular Diels-Alder reactions to natural product synthesis. *Chem Soc Rev, 38*(11), 2983-2992. doi: 10.1039/b816703f

Kalaitzis, J. A. (2013). Discovery, biosynthesis, and rational engineering of novel enterocin and wailupemycin polyketide analogues. *Methods Mol Biol, 1055*, 171-189. doi: 10.1007/978-1-62703-577-4_13

Kalaitzis, J. A., Cheng, Q., Thomas, P. M., Kelleher, N. L., & Moore, B. S. (2009). In vitro biosynthesis of unnatural enterocin and wailupemycin polyketides. *J Nat Prod, 72*(3), 469-472. doi: 10.1021/np800598t

Kalaitzis, J. A., Izumikawa, M., Xiang, L., Hertweck, C., & Moore, B. S. (2003). Mutasynthesis of enterocin and wailupemycin analogues. *J Am Chem Soc, 125*(31), 9290-9291. doi: 10.1021/ja035973o

Kamohara, H., Yamashiro, S., Galligan, C., & Yoshimura, T. (2001). Discoidin domain receptor 1 isoform-a (DDR1alpha) promotes migration of leukocytes in three-dimensional collagen lattices. *Faseb j, 15*(14), 2724-2726. doi: 10.1096/fj.01-0359fje

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res, 28*(1), 27-30.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res, 42*(Database issue), D199-205. doi: 10.1093/nar/gkt1076

Keller, M. A., Piedrafita, G., & Ralser, M. (2015). The widespread role of non-enzymatic reactions in cellular metabolism. *Curr Opin Biotechnol, 34*, 153-161. doi: 10.1016/j.copbio.2014.12.020

Kelly, K. S., Ochi, K., & Jones, G. H. (1991). Pleiotropic effects of a relC mutation in Streptomyces antibioticus. *J Bacteriol, 173*(7), 2297-2300.

Kenmotsu, H., Serizawa, M., Koh, Y., Isaka, M., Takahashi, T., Taira, T., . . . Yamamoto, N. (2014). Prospective genetic profiling of squamous cell lung cancer and adenosquamous carcinoma in Japanese patients by multitarget assays. *BMC Cancer, 14*, 786. doi: 10.1186/1471-2407-14-786

Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., . . . Zhao, Y. (2006). Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey. *Molecular Cell, 23*(4), 607-618. doi: https://doi.org/10.1016/j.molcel.2006.06.026

Kolm, R. H., Danielson, U. H., Zhang, Y., Talalay, P., & Mannervik, B. (1995). Isothiocyanates as substrates for human glutathione transferases: structure-activity studies. *Biochemical Journal, 311*(2), 453-459.

Kurashige, J., Hasegawa, T., Niida, A., Sugimachi, K., Deng, N., Mima, K., . . . Mimori, K. (2016). Integrated Molecular Profiling of Human Gastric Cancer Identifies DDR2 as a Potential Regulator of Peritoneal Dissemination. *Sci Rep, 6*, 22371. doi: 10.1038/srep22371

Kurita, K. L., Glassey, E., & Linington, R. G. (2015). Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc Natl Acad Sci U S A, 112*(39), 11999-12004. doi: 10.1073/pnas.1507743112

Kurnasov, O., Jablonski, L., Polanuyer, B., Dorrestein, P., Begley, T., & Osterman, A. (2003). Aerobic tryptophan degradation pathway in bacteria: novel kynurenine formamidase. *FEMS Microbiol Lett, 227*(2), 219-227.

Kwon, Y., Park, S., Shin, J., & Oh, D. C. (2014). Application of 13C-labeling and 13C-13C COSY NMR experiments in the structure determination of a microbial natural product. *Arch Pharm Res, 37*(8), 967-971. doi: 10.1007/s12272-013-0254-8

Laiple, K. J., Hartner, T., Fiedler, H.-P., Wohlleben, W., & Weber, T. (2009). The kirromycin gene cluster of Streptomyces collinus Tu 365 codes for an aspartate-[alpha]-decarboxylase, KirD, which is involved in the biosynthesis of the precursor [beta]-alanine. *J Antibiot, 62*(8), 465-468. doi: http://www.nature.com/ja/journal/v62/n8/suppinfo/ja200967s1.html

Lampis, G., Deidda, D., Maullu, C., Madeddu, M. A., Pompei, R., Delle Monachie, F., & Satta, G. (1995). Sattabacins and sattazolins: new biologically active compounds with antiviral properties extracted from a Bacillus sp. *J Antibiot (Tokyo), 48*(9), 967-972.

Lapinsky, D. J., & Johnson, D. S. (2015). Recent developments and applications of clickable photoprobes in medicinal chemistry and chemical biology. *Future Med Chem, 7*(16), 2143-2171. doi: 10.4155/fmc.15.136

Lee, J. C., Strobel, G. A., Lobkovsky, E., & Clardy, J. (1996). Torreyanic Acid: A Selectively Cytotoxic Quinone Dimer from the Endophytic Fungus Pestalotiopsis microspora. *J Org Chem, 61*(10), 3232-3233. doi: 10.1021/jo960471x

Lee, J. C., Yang, X., Schwartz, M., Strobel, G., & Clardy, J. (1995). The relationship between an endangered North American tree and an endophytic fungus. *Chem Biol, 2*(11), 721-727.

Leitinger, B. (2014). Discoidin domain receptor functions in physiological and pathological conditions. *Int Rev Cell Mol Biol, 310*, 39-87. doi: 10.1016/b978-0-12-800180-6.00002-5

Lemmon, M. A., & Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. *Cell, 141*(7), 1117-1134. doi: 10.1016/j.cell.2010.06.011

Li, D., Azoulay, P., & Sampat, B. N. (2017). The applied value of public investments in biomedical research. *Science, 356*(6333), 78-81. doi: 10.1126/science.aal0010

Li, X., Yu, T. K., Kwak, J. H., Son, B. Y., Seo, Y., Zee, O. P., & Ahn, J. W. (2008). Soraphinol C, a new free-radical scavenger from Sorangium cellulosum. *J Microbiol Biotechnol, 18*(3), 520-522.

Lin, Z., Koch, M., Abdel Aziz, M. H., Galindo-Murillo, R., Tianero, M. D., Cheatham, T. E., . . . Schmidt, E. W. (2014). Oxazinin A, a pseudodimeric natural product of mixed biosynthetic origin from a filamentous fungus. *Org Lett, 16*(18), 4774-4777. doi: 10.1021/ol502227x

Lin, Z., Koch, M., Abdel Aziz, M. H., Galindo-Murillo, R., Tianero, M. D., Cheatham, T. E., . . . Schmidt, E. W. (2014). Oxazinin A, a Pseudodimeric Natural Product of Mixed Biosynthetic Origin from a Filamentous Fungus. *Organic Letters, 16*(18), 4774-4777. doi: 10.1021/ol502227x

Lin, Z., Marett, L., Hughen, R. W., Flores, M., Forteza, I., Ammon, M. A., . . . Schmidt, E. W. (2013). Neuroactive diol and acyloin metabolites from cone snail-associated bacteria. *Bioorg Med Chem Lett, 23*(17), 4867-4869. doi: 10.1016/j.bmcl.2013.06.088

Lopez, J. M., Dromerick, A., & Freese, E. (1981). Response of guanosine 5'-triphosphate concentration to nutritional changes and its significance for Bacillus subtilis sporulation. *J Bacteriol, 146*(2), 605-613.

Mackinnon, A. L., & Taunton, J. (2009). Target Identification by Diazirine Photo-Cross-linking and Click Chemistry. *Curr Protoc Chem Biol, 1*, 55-73. doi: 10.1002/9780470559277.ch090167

Magoc, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics, 27*(21), 2957-2963. doi: 10.1093/bioinformatics/btr507

Mahmud, T. (2007). Isotope tracer investigations of natural products biosynthesis: the discovery of novel metabolic pathways. *Journal of Labelled Compounds and Radiopharmaceuticals, 50*(11-12), 1039-1051. doi: 10.1002/jlcr.1391

Martinez-Farina, C. F., & Jakeman, D. L. (2015). Jadomycins, put a bigger ring in it: isolation of seven- to ten-membered ring analogues. *Chem Commun (Camb), 51*(78), 14617-14619. doi: 10.1039/c5cc05571g

Martinez-Farina, C. F., Robertson, A. W., Yin, H., Monro, S., McFarland, S. A., Syvitski, R. T., & Jakeman, D. L. (2015). Isolation and Synthetic Diversification of Jadomycin 4-Amino-l-phenylalanine. *J Nat Prod, 78*(6), 1208-1214. doi: 10.1021/np5009398

Matsuura, B. S., Kolle, P., Trauner, D., de Vivie-Riedle, R., & Meier, R. (2017). Unravelling Photochemical Relationships Among Natural Products from Aplysia dactylomela. *ACS Cent Sci, 3*(1), 39-46. doi: 10.1021/acscentsci.6b00293

McKay, C. S., & Finn, M. G. (2014). Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem Biol, 21*(9), 1075-1101. doi: 10.1016/j.chembiol.2014.09.002

McKerrow, J. H. (2015). Recognition of the role of Natural Products as drugs to treat neglected tropical diseases by the 2015 Nobel prize in physiology or medicine. *Natural Product Reports, 32*(12), 1610-1611. doi: 10.1039/C5NP90043C

Mevers, E., Saurí, J., Liu, Y., Moser, A., Ramadhar, T. R., Varlan, M., . . . Clardy, J. (2016). Homodimericin A: A Complex Hexacyclic Fungal Metabolite. *Journal of the American Chemical Society, 138*(38), 12324-12327. doi: 10.1021/jacs.6b07588

Midgett, C., Stitham, J., Martin, K., & Hwa, J. (2011). Prostacyclin receptor regulation--from transcription to trafficking. *Curr Mol Med, 11*(7), 517-528.

Miller, E. D., Kauffman, C. A., Jensen, P. R., & Fenical, W. (2007). Piperazimycins: cytotoxic hexadepsipeptides from a marine-derived bacterium of the genus Streptomyces. *J Org Chem, 72*(2), 323-330. doi: 10.1021/jo061064g

Minassi, A., Cicione, L., Koeberle, A., Bauer, J., Laufer, S., Werz, O., & Appendino, G. (2012). A Multicomponent Carba-Betti Strategy to Alkylidene Heterodimers – Total Synthesis and Structure–Activity Relationships of Arzanol. *European journal of organic chemistry, 2012*(4), 772-779. doi: 10.1002/ejoc.201101193

Monroe, E. A., Podell, S., Glukhov, E., Allen, E. E., Gerwick, W. H., & Gerwick, L. *Proc Natl Acad Sci U S A.* doi: 10.1073/pnas.1618556114

Musiol, E. M., Greule, A., Hartner, T., Kulik, A., Wohlleben, W., & Weber, T. (2013). The AT(2) domain of KirCI loads malonyl extender units to the ACPs of the kirromycin PKS. *Chembiochem, 14*(11), 1343-1352. doi: 10.1002/cbic.201300211

Newman, D. J., & Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod, 79*(3), 629-661. doi: 10.1021/acs.jnatprod.5b01055

Nicolaou, K. C., Lim, Y. H., Piper, J. L., & Papageorgiou, C. D. (2007). Total Syntheses of 2,2‘-epi-Cytoskyrin A, Rugulosin, and the Alleged Structure of Rugulin. *J Am Chem Soc, 129*(13), 4001-4013. doi: 10.1021/ja0685708

Nicolaou, K. C., Montagnon, T., & Snyder, S. A. (2003). Tandem reactions, cascade sequences, and biomimetic strategies in total synthesis. *Chem Commun (Camb)*(5), 551-564.

Nicolaou, K. C., Papageorgiou, C. D., Piper, J. L., & Chadha, R. K. (2005). The cytoskyrin cascade: a facile entry into cytoskyrin A, deoxyrubroskyrin, rugulin, skyrin, and flavoskyrin model systems. *Angew Chem Int Ed Engl, 44*(36), 5846-5851. doi: 10.1002/anie.200502011

Nicolaou, K. C., Petasis, N. A., & Zipkin, R. E. (1982). The endiandric acid cascade. Electrocyclizations in organic synthesis. 4. Biomimetic approach to endiandric acids A-

G. Total synthesis and thermal studies. *Journal of the American Chemical Society, 104*(20), 5560-5562. doi: 10.1021/ja00384a080

Nicolaou, K. C., Petasis, N. A., Zipkin, R. E., & Uenishi, J. (1982). The endiandric acid cascade. Electrocyclizations in organic synthesis. I. Stepwise, stereocontrolled total synthesis of endiandric acids A and B. *Journal of the American Chemical Society, 104*(20), 5555-5557. doi: 10.1021/ja00384a077

Nicolaou, K. C., Zipkin, R. E., & Petasis, N. A. (1982). The endiandric acid cascade. Electrocyclizations in organic synthesis. 3. "Biomimetic" approach to endiandric acids A-G. Synthesis of precursors. *Journal of the American Chemical Society, 104*(20), 5558-5560. doi: 10.1021/ja00384a079

Nishikawa, H., Miyazaki, T., Nakayama, H., Minematsu, A., Yamauchi, S., Yamashita, K., . . . Mukae, H. (2016). Roles of vacuolar H+-ATPase in the oxidative stress response of Candida glabrata. *FEMS Yeast Res, 16*(5). doi: 10.1093/femsyr/fow054

Noutsias, D., & Vassilikogiannakis, G. (2012). First total synthesis of paracaseolide A. *Org Lett, 14*(14), 3565-3567. doi: 10.1021/ol301481t

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., . . . Pevzner, P. A. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol, 20*(10), 714-737. doi: 10.1089/cmb.2013.0084

Ochi, K. (1986). A decrease in GTP content is associated with aerial mycelium formation in Streptomyces MA406-A-1. *J Gen Microbiol, 132*(2), 299-305. doi: 10.1099/00221287-132-2-299

Ochi, K. (1987). Metabolic initiation of differentiation and secondary metabolism by Streptomyces griseus: significance of the stringent response (ppGpp) and GTP content in relation to A factor. *J Bacteriol, 169*(8), 3608-3616.

Ochi, K., & Hosaka, T. (2013). New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl Microbiol Biotechnol, 97*(1), 87-98. doi: 10.1007/s00253-012-4551-9

Ochi, K., Kandala, J. C., & Freese, E. (1981). Initiation of Bacillus subtilis sporulation by the stringent response to partial amino acid deprivation. *J Biol Chem, 256*(13), 6866-6875.

Ochi, K., & Ohsawa, S. (1984). Initiation of antibiotic production by the stringent response of Bacillus subtilis Marburg. *J Gen Microbiol, 130*(10), 2473-2482. doi: 10.1099/00221287-130-10-2473

Okada, B. K., & Seyedsayamdost, M. R. (2017). Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol Rev, 41*(1), 19-33. doi: 10.1093/femsre/fuw035

Okada, B. K., Wu, Y., Mao, D., Bushin, L. B., & Seyedsayamdost, M. R. (2016). Mapping the Trimethoprim-Induced Secondary Metabolome of Burkholderia thailandensis. *ACS Chemical Biology, 11*(8), 2124-2130. doi: 10.1021/acschembio.6b00447

Okamoto-Hosoya, Y., Hosaka, T., & Ochi, K. (2003). An aberrant protein synthesis activity is linked with antibiotic overproduction in rpsL mutants of Streptomyces coelicolor A3(2). *Microbiology, 149*(Pt 11), 3299-3309. doi: 10.1099/mic.0.26490-0

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., . . . Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res, 33*(17), 5691-5702. doi: 10.1093/nar/gki866

Palleroni, N. J., Reichelt, K. E., Mueller, D., Epps, R., Tabenkin, B., Bull, D. N., . . . Berger, J. (1978). Production of a novel red pigment, rubrolone, by Streptomyces echinoruber sp. nov. I. Taxonomy, fermentation and partial purification. *J Antibiot (Tokyo), 31*(12), 1218-1225.

Palmer, J. M., & Keller, N. P. (2010). Secondary metabolism in fungi: does chromosomal location matter? *Current Opinion in Microbiology, 13*(4), 431-436. doi: 10.1016/j.mib.2010.04.008

Pan, E., Oswald, N. W., Legako, A. G., Life, J. M., Posner, B. A., & Macmillan, J. B. (2013). Precursor-Directed Generation of Amidine Containing Ammosamide Analogs: Ammosamides E-P. *Chem Sci, 4*(1), 482-488. doi: 10.1039/C2SC21442C

Park, J. S., Kagaya, N., Hashimoto, J., Izumikawa, M., Yabe, S., Shin-Ya, K., . . . Kuzuyama, T. (2014). Identification and biosynthesis of new acyloins from the thermophilic bacterium Thermosporothrix hazakensis SK20-1(T). *Chembiochem, 15*(4), 527-532. doi: 10.1002/cbic.201300690

Parmeggiani, A., & Swart, G. W. (1985). Mechanism of action of kirromycin-like antibiotics. *Annu Rev Microbiol, 39*, 557-577. doi: 10.1146/annurev.mi.39.100185.003013

Pawig, L., Klasen, C., Weber, C., Bernhagen, J., & Noels, H. (2015). Diversity and Inter-Connections in the CXCR4 Chemokine Receptor/Ligand Family: Molecular Perspectives. *Front Immunol, 6*, 429. doi: 10.3389/fimmu.2015.00429

Penttila, A., Kapadia, G. J., & Fales, H. M. (1965). The Biosynthesis in Vivo of Methylenebisphloroglucinol Derivatives. *Journal of the American Chemical Society, 87*(19), 4402-4403. doi: 10.1021/ja00947a045

Pettit, R. K. (2011). Small-molecule elicitation of microbial secondary metabolites. *Microbial Biotechnology, 4*(4), 471-478. doi: 10.1111/j.1751-7915.2010.00196.x

Piel, J., Hertweck, C., Shipley, P. R., Hunt, D. M., Newman, M. S., & Moore, B. S. (2000). Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate 'Streptomyces maritimus': evidence for the derailment of an aromatic polyketide synthase. *Chem Biol, 7*(12), 943-955.

Potts, M. B., Kim, H. S., Fisher, K. W., Hu, Y., Carrasco, Y. P., Bulut, G. B., . . . White, M. A. (2013). Using functional signature ontology (FUSION) to identify mechanisms of action for natural products. *Sci Signal, 6*(297), ra90. doi: 10.1126/scisignal.2004657

Potts, M. B., McMillan, E. A., Rosales, T. I., Kim, H. S., Ou, Y. H., Toombs, J. E., . . . White, M. A. (2015). Mode of action and pharmacogenomic biomarkers for exceptional responders to didemnin B. *Nat Chem Biol, 11*(6), 401-408. doi: 10.1038/nchembio.1797

Poudel, B., Yoon, D. S., Lee, J. H., Lee, Y. M., & Kim, D. K. (2012). Collagen I enhances functional activities of human monocyte-derived dendritic cells via discoidin domain receptor 2. *Cell Immunol, 278*(1-2), 95-102. doi: 10.1016/j.cellimm.2012.07.004

Proschak, A., Zhou, Q., Schoner, T., Thanwisai, A., Kresovic, D., Dowling, A., . . . Bode, H. B. (2014). Biosynthesis of the insecticidal xenocyloins in Xenorhabdus bovienii. *Chembiochem, 15*(3), 369-372. doi: 10.1002/cbic.201300694

Raju, R., Piggott, A. M., Conte, M. M., & Capon, R. J. (2010). Heronamides A-C, new polyketide macrolactams from an Australian marine-derived Streptomyces sp. A biosynthetic case for synchronized tandem electrocyclization. *Org Biomol Chem, 8*(20), 4682-4689. doi: 10.1039/c0ob00267d

Rammal, H., Saby, C., Magnien, K., Van-Gulick, L., Garnotel, R., Buache, E., . . . Morjani, H. (2016). Discoidin Domain Receptors: Potential Actors and Targets in Cancer. *Front Pharmacol, 7*, 55. doi: 10.3389/fphar.2016.00055

Razzak, M., & De Brabander, J. K. (2011). Lessons and revelations from biomimetic syntheses. *Nat Chem Biol, 7*(12), 865-875.

Rehman, S. U., Sarwar, T., Husain, M. A., Ishqi, H. M., & Tabish, M. (2015). Studying non-covalent drug-DNA interactions. *Arch Biochem Biophys, 576*, 49-60. doi: 10.1016/j.abb.2015.03.024

Rinkel, J., & Dickschat, J. S. (2015). Recent highlights in biosynthesis research using stable isotopes. *Beilstein Journal of Organic Chemistry, 11*, 2493-2508. doi: 10.3762/bjoc.11.271

Rix, U., Zheng, J., Remsing Rix, L. L., Greenwell, L., Yang, K., & Rohr, J. (2004). The dynamic structure of jadomycin B and the amino acid incorporation step of its biosynthesis. *J Am Chem Soc, 126*(14), 4496-4497. doi: 10.1021/ja031724o

Robertson, A. W., Martinez-Farina, C. F., Smithen, D. A., Yin, H., Monro, S., Thompson, A., . . . Jakeman, D. L. (2015). Eight-membered ring-containing jadomycins: implications for non-enzymatic natural products biosynthesis. *J Am Chem Soc, 137*(9), 3271-3275. doi: 10.1021/ja5114672

Robertson, J., & Stevens, K. (2017). Pyrrolizidine alkaloids: occurrence, biology, and chemical synthesis. *Nat Prod Rep, 34*(1), 62-89. doi: 10.1039/c5np00076a

Robles, A. J., Du, L., Cichewicz, R. H., & Mooberry, S. L. (2016). Maximiscin Induces DNA Damage, Activates DNA Damage Response Pathways, and Has Selective Cytotoxic Activity against a Subtype of Triple-Negative Breast Cancer. *J Nat Prod, 79*(7), 1822-1827. doi: 10.1021/acs.jnatprod.6b00290

Rodriguez, A. D., & Shi, J. G. (1998). The First Cembrane-Pseudopterane Cycloisomerization. *J Org Chem, 63*(3), 420-421.

Roethle, P. A., Hernandez, P. T., & Trauner, D. (2006). Exploring Biosynthetic Relationships among Furanocembranoids: Synthesis of (−)-Bipinnatin J, (+)-Intricarene, (+)-Rubifolide, and (+)-Isoepilophodione B. *Organic Letters, 8*(25), 5901-5904. doi: 10.1021/ol062581o

Russo, A. F. (2015). Calcitonin gene-related peptide (CGRP): a new target for migraine. *Annu Rev Pharmacol Toxicol, 55*, 533-552. doi: 10.1146/annurev-pharmtox-010814-124701

Rutledge, P. J., & Challis, G. L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol, 13*(8), 509-523. doi: 10.1038/nrmicro3496

Sakurai, K., Ozawa, S., Yamada, R., Yasui, T., & Mizuno, S. (2014). Comparison of the reactivity of carbohydrate photoaffinity probes with different photoreactive groups. *Chembiochem, 15*(10), 1399-1403. doi: 10.1002/cbic.201402051

Sasaki, H., Shitara, M., Yokota, K., Okuda, K., Hikosaka, Y., Moriyama, S., . . . Fujii, Y. (2012). DDR2 polymorphisms and mRNA expression in lung cancers of Japanese patients. *Oncol Lett, 4*(1), 33-37. doi: 10.3892/ol.2012.684

Satoh, K. (1995). The high non-enzymatic conjugation rates of some glutathione s-transferase (gst) substrates at high glutathione concentrations. *Carcinogenesis, 16*(4), 869-874. doi: 10.1093/carcin/16.4.869

Schafers, M., Svensson, C. I., Sommer, C., & Sorkin, L. S. (2003). Tumor necrosis factor-alpha induces mechanical allodynia after spinal nerve ligation by activation of p38 MAPK in primary sensory neurons. *J Neurosci, 23*(7), 2517-2521.

Schenone, M., Dancik, V., Wagner, B. K., & Clemons, P. A. (2013). Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol, 9*(4), 232-240. doi: 10.1038/nchembio.1199

Schuep, W., Blount, J. F., Williams, T. H., & Stempel, A. (1978). Production of a novel red pigment, rubrolone, by Streptomyces echinoruber Sp. Nov. II. Chemistry and structure elucidation. *J Antibiot (Tokyo), 31*(12), 1226-1232.

Shao, C. L., Wang, C. Y., Wei, M. Y., Gu, Y. C., She, Z. G., Qian, P. Y., & Lin, Y. C. (2011). Aspergilones A and B, two benzylazaphilones with an unprecedented carbon skeleton from the gorgonian-derived fungus Aspergillus sp. *Bioorg Med Chem Lett, 21*(2), 690-693. doi: 10.1016/j.bmcl.2010.12.005

Sharif, E. U., & O'Doherty, G. A. (2012). Biosynthesis and Total Synthesis Studies on The Jadomycin Family of Natural Products. *European journal of organic chemistry, 2012*(11), 10.1002/ejoc.201101609. doi: 10.1002/ejoc.201101609

Sharma, P., Lygo, B., Lewis, W., & Moses, J. E. (2009). Biomimetic synthesis and structural reassignment of the tridachiahydropyrones. *J Am Chem Soc, 131*(16), 5966-5972. doi: 10.1021/ja900369z

Shentu, X., Liu, N., Tang, G., Tanaka, Y., Ochi, K., Xu, J., & Yu, X. (2016). Improved antibiotic production and silent gene activation in Streptomyces diastatochromogenes by ribosome engineering. *J Antibiot (Tokyo), 69*(5), 406-410. doi: 10.1038/ja.2015.123

Shi, L., & Tu, B. P. (2015). Acetyl-CoA and the Regulation of Metabolism: Mechanisms and Consequences. *Curr Opin Cell Biol, 33*, 125-131. doi: 10.1016/j.ceb.2015.02.003

Shi, X., Wu, H., Lu, J., Duan, H., Liu, X., & Liang, Z. (2016). Screening for major driver oncogene alterations in adenosquamous lung carcinoma using PCR coupled with next-generation and Sanger sequencing methods. *Sci Rep, 6*, 22297. doi: 10.1038/srep22297

Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S., & Ochi, K. (1996). Induction of actinorhodin production by rpsL (encoding ribosomal protein S12) mutations that confer streptomycin resistance in Streptomyces lividans and Streptomyces coelicolor A3(2). *J Bacteriol, 178*(24), 7276-7284.

Shrivastava, A., Radziejewski, C., Campbell, E., Kovac, L., McGlynn, M., Ryan, T. E., . . . Yancopoulos, G. D. (1997). An orphan receptor tyrosine kinase family whose members serve as nonintegrin collagen receptors. *Mol Cell, 1*(1), 25-34.

Smit, B. A., van Hylckama Vlieg, J. E., Engels, W. J., Meijer, L., Wouters, J. T., & Smit, G. (2005). Identification, cloning, and characterization of a Lactococcus lactis branched-chain alpha-keto acid decarboxylase involved in flavor formation. *Appl Environ Microbiol, 71*(1), 303-311. doi: 10.1128/aem.71.1.303-311.2005

Springer, W. R., Cooper, D. N., & Barondes, S. H. (1984). Discoidin I is implicated in cell-substratum attachment and ordered cell migration of Dictyostelium discoideum and resembles fibronectin. *Cell, 39*(3 Pt 2), 557-564.

Stein, J. L., Medland, S. E., Vasquez, A. A., Hibar, D. P., Senstad, R. E., Winkler, A. M., . . . Thompson, P. M. (2012). Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet, 44*(5), 552-561. doi: 10.1038/ng.2250

Steinmetz, H., Gerth, K., Jansen, R., Schlager, N., Dehn, R., Reinecke, S., . . . Muller, R. (2011). Elansolid A, a unique macrolide antibiotic from Chitinophaga sancti isolated as two

stable atropisomers. *Angew Chem Int Ed Engl, 50*(2), 532-536. doi: 10.1002/anie.201005226

Steinmetz, H., Gerth, K., Jansen, R., Schläger, N., Dehn, R., Reinecke, S., . . . Müller, R. (2011). Elansolid A, a Unique Macrolide Antibiotic from Chitinophaga sancti Isolated as Two Stable Atropisomers. *Angewandte Chemie International Edition, 50*(2), 532-536. doi: 10.1002/anie.201005226

Steinmetz, H., Zander, W., Shushni, M. A. M., Jansen, R., Gerth, K., Dehn, R., . . . Müller, R. (2012). Precursor-Directed Syntheses and Biological Evaluation of New Elansolid Derivatives. *ChemBioChem, 13*(12), 1813-1817. doi: 10.1002/cbic.201200228

Stichnoth, D., Kolle, P., Kimbrough, T. J., Riedle, E., de Vivie-Riedle, R., & Trauner, D. (2014). Photochemical formation of intricarene. *Nat Commun, 5*, 5597. doi: 10.1038/ncomms6597

Stocking, E. M., & Williams, R. M. (2003). Chemistry and biology of biosynthetic Diels-Alder reactions. *Angew Chem Int Ed Engl, 42*(27), 3078-3115. doi: 10.1002/anie.200200534

Strauch, E., Takano, E., Baylis, H. A., & Bibb, M. J. (1991). The stringent response in Streptomyces coelicolor A3(2). *Mol Microbiol, 5*(2), 289-298.

Streuli, C. H., & Akhtar, N. (2009). Signal co-operation between integrins and other receptor systems. *Biochem J, 418*(3), 491-506. doi: 10.1042/bj20081948

Sumranjit, J., & Chung, S. J. (2013). Recent advances in target characterization and identification by photoaffinity probes. *Molecules, 18*(9), 10425-10451. doi: 10.3390/molecules180910425

Takeuchi, K., & Ito, F. (2011). Receptor tyrosine kinases and targeted cancer therapeutics. *Biol Pharm Bull, 34*(12), 1774-1780.

Tanaka, K., Sasaki, A., Cao, H.-Q., Yamada, T., Igarashi, M., Komine, I., . . . Kiyota, H. (2011). Synthesis and Biotransformation of Plausible Biosynthetic Intermediates of Salicylaldehyde-Type Phytotoxins of Rice Blast Fungus, Magnaporthe grisea. *European journal of organic chemistry, 2011*(31), 6276-6280. doi: 10.1002/ejoc.201100771

Tang, B., Bray, C. D., & Pattenden, G. (2006). A biomimetic total synthesis of (+)-intricarene. *Tetrahedron Letters, 47*(36), 6401-6404. doi: https://doi.org/10.1016/j.tetlet.2006.06.150

Tarhonskaya, H., Rydzik, A. M., Leung, I. K. H., Loik, N. D., Chan, M. C., Kawamura, A., . . . Schofield, C. J. (2014). Non-enzymatic chemistry enables 2-hydroxyglutarate-mediated activation of 2-oxoglutarate oxygenases. *Nat Commun, 5*, 3423. doi: 10.1038/ncomms4423

Teles, H. L., Sordi, R., Silva, G. H., Castro-Gamboa, I., Bolzani Vda, S., Pfenning, L. H., . . . Araujo, A. R. (2006). Aromatic compounds produced by Periconia atropurpurea, an endophytic fungus associated with Xylopia aromatica. *Phytochemistry, 67*(24), 2686-2690. doi: 10.1016/j.phytochem.2006.09.005

Terai, H., Tan, L., Beauchamp, E. M., Hatcher, J. M., Liu, Q., Meyerson, M., . . . Hammerman, P. S. (2015). Characterization of DDR2 Inhibitors for the Treatment of DDR2 Mutated Nonsmall Cell Lung Cancer. *ACS Chem Biol, 10*(12), 2687-2696. doi: 10.1021/acschembio.5b00655

Terashima, M., Togashi, Y., Sato, K., Mizuuchi, H., Sakai, K., Suda, K., . . . Nishio, K. (2016). Functional Analyses of Mutations in Receptor Tyrosine Kinase Genes in Non-Small Cell Lung Cancer: Double-Edged Sword of DDR2. *Clin Cancer Res, 22*(14), 3663-3671. doi: 10.1158/1078-0432.ccr-15-2093

Thanapipatsiri, A., Gomez-Escribano, J. P., Song, L., Bibb, M. J., Al-Bassam, M., Chandra, G., . . . Bibb, M. J. (2016). Discovery of Unusual Biaryl Polyketides by Activation of a Silent Streptomyces venezuelae Biosynthetic Gene Cluster. *Chembiochem, 17*(22), 2189-2198. doi: 10.1002/cbic.201600396

Theodoropoulos, P. C., Gonzales, S. S., Winterton, S. E., Rodriguez-Navas, C., McKnight, J. S., Morlock, L. K., . . . Nijhawan, D. (2016). Discovery of tumor-specific irreversible inhibitors of stearoyl CoA desaturase. *Nat Chem Biol, 12*(4), 218-225. doi: 10.1038/nchembio.2016

http://www.nature.com/nchembio/journal/v12/n4/abs/nchembio.2016.html#supplementary-information

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res, 22*(22), 4673-4680.

Tojo, S., Tanaka, Y., & Ochi, K. (2015). Activation of Antibiotic Production in Bacillus spp. by Cumulative Drug Resistance Mutations. *Antimicrob Agents Chemother, 59*(12), 7799-7804. doi: 10.1128/aac.01932-15

Trifonov, L. S., Dreiding, A. S., Hoesch, L., & Rast, D. M. (1981). Isolation of Four Hexaketides from Verticillium intertextum. Preliminary communication. *Helvetica Chimica Acta, 64*(6), 1843-1846. doi: 10.1002/hlca.19810640616

Troeira Henriques, S., & Craik, D. J. (2017). Cyclotide Structure and Function: The Role of Membrane Binding and Permeation. *56*(5), 669-682. doi: 10.1021/acs.biochem.6b01212

Uchida, R., Shiomi, K., Inokoshi, J., Masuma, R., Kawakubo, T., Tanaka, H., . . . Omura, S. (1996). Kurasoins A and B, new protein farnesyltransferase inhibitors produced by Paecilomyces sp. FO-3684. I. Producing strain, fermentation, isolation, and biological activities. *J Antibiot (Tokyo), 49*(9), 932-934.

Uchida, R., Shiomi, K., Sunazuka, T., Inokoshi, J., Nishizawa, A., Hirose, T., . . . Omura, S. (1996). Kurasoins A and B, new protein farnesyltransferase inhibitors produced by Paecilomyces sp. FO-3684. II. Structure elucidation and total synthesis. *J Antibiot (Tokyo), 49*(9), 886-889.

Vasamsetty, L., Khan, F. A., & Mehta, G. (2013). Total synthesis of a novel oxa-bowl natural product paracaseolide A via a ‘putative’ biomimetic pathway. *Tetrahedron Letters, 54*(27), 3522-3525. doi: https://doi.org/10.1016/j.tetlet.2013.04.097

Vasamsetty, L., Sahu, D., Ganguly, B., Khan, F. A., & Mehta, G. (2014). Total synthesis of novel bioactive natural product paracaseolide A and analogues: computational evaluation of a ‘proposed’biomimetic Diels–Alder reaction. *Tetrahedron, 70*(45), 8488-8497.

Vizcaino, M. I., & Crawford, J. M. (2015). The colibactin warhead crosslinks DNA. *Nat Chem, 7*(5), 411-417. doi: 10.1038/nchem.2221

Vogel, W., Gish, G. D., Alves, F., & Pawson, T. (1997). The discoidin domain receptor tyrosine kinases are activated by collagen. *Mol Cell, 1*(1), 13-23.

Wagner, Gregory R., & Hirschey, Matthew D. (2014). Nonenzymatic Protein Acylation as a Carbon Stress Regulated by Sirtuin Deacylases. *Molecular Cell, 54*(1), 5-16. doi: https://doi.org/10.1016/j.molcel.2014.03.027

Wallace, S., Schultz, E. E., & Balskus, E. P. (2015). Using Non-Enzymatic Chemistry to Influence Microbial Metabolism. *Current opinion in chemical biology, 25*, 71-79. doi: 10.1016/j.cbpa.2014.12.024

Walsh, L. A., Nawshad, A., & Medici, D. (2011). Discoidin domain receptor 2 is a critical regulator of epithelial-mesenchymal transition. *Matrix biology : journal of the International Society for Matrix Biology, 30*(4), 243-247. doi: 10.1016/j.matbio.2011.03.007

Wang, L. L., Candito, D., Drager, G., Herrmann, J., Muller, R., & Kirschning, A. (2017). Harnessing a p-Quinone Methide Intermediate in the Biomimetic Total Synthesis of the Highly Active Antibiotic 20-Deoxy-Elansolid B1. *Chemistry, 23*(22), 5291-5298. doi: 10.1002/chem.201605884

Wang, P., Hong, G. J., Wilson, M. R., & Balskus, E. P. (2017). Production of Stealthin C Involves an S–N-Type Smiles Rearrangement. *Journal of the American Chemical Society, 139*(8), 2864-2867. doi: 10.1021/jacs.6b10586

Wang, T., & Hoye, T. R. (2015). Diels-Alderase-free, bis-pericyclic, [4+2] dimerization in the biosynthesis of (+/-)-paracaseolide A. *Nat Chem, 7*(8), 641-645. doi: 10.1038/nchem.2281

Weber, T., Laiple, K. J., Pross, E. K., Textor, A., Grond, S., Welzel, K., . . . Wohlleben, W. (2008). Molecular analysis of the kirromycin biosynthetic gene cluster revealed beta-alanine as precursor of the pyridone moiety. *Chem Biol, 15*(2), 175-188. doi: 10.1016/j.chembiol.2007.12.009

Webster, R., Gaspar, B., Mayer, P., & Trauner, D. (2013). Studies toward the biomimetic total synthesis of (-)-PF-1018. *Org Lett, 15*(8), 1866-1869. doi: 10.1021/ol400508s

Wei, H., Qiao, C., Liu, G., Yang, Z., & Li, C.-c. (2013). Stereoselective Total Syntheses of (−)-Flueggine A and (+)-Virosaine B. *Angewandte Chemie International Edition, 52*(2), 620-624. doi: 10.1002/anie.201208261

Weinreb, S. M. (2009). Total synthesis of the Securinega alkaloids. *Natural Product Reports, 26*(6), 758-775. doi: 10.1039/B902265A

Weiss, M. (2014). Crowding, diffusion, and biochemical reactions. *Int Rev Cell Mol Biol, 307*, 383-417. doi: 10.1016/b978-0-12-800046-5.00011-4

Wolf, H., & Zahner, H. (1972). [Metabolic products of microorganisms. 99. Kirromycin]. *Arch Mikrobiol, 83*(2), 147-154.

Xie, B., Lin, W., Ye, J., Wang, X., Zhang, B., Xiong, S., . . . Tan, G. (2015). DDR2 facilitates hepatocellular carcinoma invasion and metastasis via activating ERK signaling and stabilizing SNAIL1. *Journal of Experimental & Clinical Cancer Research : CR, 34*(1), 101. doi: 10.1186/s13046-015-0218-6

Xu, C., Buczkowski, K. A., Zhang, Y., Asahina, H., Beauchamp, E. M., Terai, H., . . . Hammerman, P. S. (2015). NSCLC Driven by DDR2 Mutation Is Sensitive to Dasatinib and JQ1 Combination Therapy. *Mol Cancer Ther, 14*(10), 2382-2389. doi: 10.1158/1535-7163.mct-15-0077

Xu, J., Lu, W., Zhang, S., Zhu, C., Ren, T., Zhu, T., . . . Su, J. (2014). Overexpression of DDR2 contributes to cell invasion and migration in head and neck squamous cell carcinoma. *Cancer Biology & Therapy, 15*(5), 612-622. doi: 10.4161/cbt.28181

Yan, Y., Yang, J., Yu, Z., Yu, M., Ma, Y. T., Wang, L., . . . Huang, S. X. (2016). Non-enzymatic pyridine ring formation in the biosynthesis of the rubrolone tropolone alkaloids. *Nat Commun, 7*, 13083. doi: 10.1038/ncomms13083

Yang, J., Liang, Q., Wang, M., Jeffries, C., Smithson, D., Tu, Y., . . . Li, X. C. (2014). UPLC-MS-ELSD-PDA as a powerful dereplication tool to facilitate compound identification from small-molecule natural product libraries. *J Nat Prod, 77*(4), 902-909. doi: 10.1021/np4009706

Yang, P., & Liu, K. (2015). Activity-based protein profiling: recent advances in probe development and applications. *Chembiochem, 16*(5), 712-724. doi: 10.1002/cbic.201402582

Yokomakura, A., Hong, J., Ohuchi, K., Oh, S. E., Lee, J. Y., Mano, N., . . . Naganuma, A. (2012). Increased production of reactive oxygen species by the vacuolar-type (H(+))-ATPase inhibitors bafilomycin A1 and concanamycin A in RAW 264 cells. *J Toxicol Sci, 37*(5), 1045-1048.

Zarins-Tutt, J. S., Barberi, T. T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D. J., & Goss, R. J. (2016). Prospecting for new bacterial metabolites: a glossary of approaches

for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat Prod Rep, 33*(1), 54-72. doi: 10.1039/c5np00111k

Zhang, H., Zhang, C. R., Zhu, K. K., Gao, A. H., Luo, C., Li, J., & Yue, J. M. (2013). Fluevirosines A-C: a biogenesis inspired example in the discovery of new bioactive scaffolds from Flueggea virosa. *Org Lett, 15*(1), 120-123. doi: 10.1021/ol303146a

Zhang, J., Qian, Z., Wu, X., Ding, Y., Li, J., Lu, C., & Shen, Y. (2014). Juanlimycins A and B, ansamycin macrodilactams from Streptomyces sp. *Org Lett, 16*(10), 2752-2755. doi: 10.1021/ol501072t

Zhang, K., Corsa, C. A., Ponik, S. M., Prior, J. L., Piwnica-Worms, D., Eliceiri, K. W., . . . Longmore, G. D. (2013). The Collagen Receptor Discoidin Domain Receptor 2 Stabilizes Snail1 Protein to Facilitate Breast Cancer Metastasis. *Nature cell biology, 15*(6), 677-687. doi: 10.1038/ncb2743

Zhao, B.-X., Wang, Y., Li, C., Wang, G.-C., Huang, X.-J., Fan, C.-L., . . . Ye, W.-C. (2013). Flueggedine, a novel axisymmetric indolizidine alkaloid dimer from Flueggea virosa. *Tetrahedron Letters, 54*(35), 4708-4711. doi: https://doi.org/10.1016/j.tetlet.2013.06.097

Zhao, B. X., Wang, Y., Zhang, D. M., Jiang, R. W., Wang, G. C., Shi, J. M., . . . Ye, W. C. (2011). Flueggines A and B, two new dimeric indolizidine alkaloids from Flueggea virosa. *Org Lett, 13*(15), 3888-3891. doi: 10.1021/ol201410z

Zhao, L., & Lu, W. (2014). Defensins in innate immunity. *Curr Opin Hematol, 21*(1), 37-42. doi: 10.1097/moh.0000000000000005